A Non-Parametric Asymptotic Analysis of Inventory Planning with Censored Demand

Woonghee Tim Huh
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
email: huh@ieor.columbia.edu `http://www.columbia.edu/~th2113/`

Paat Rusmevichientong
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853
email: paatrus@cornell.edu `http://legacy.orie.cornell.edu/~paatrus/`

We study stochastic inventory planning with lost sales and instantaneous replenishment, where contrary to the classical inventory theory, the knowledge of the demand distribution is not available. Furthermore, we observe only the sales quantity in each period, and lost sales are unobservable, that is, demand data are censored. The manager must make an ordering decision in each period based only on historical sales data. Excess inventory is either perishable or carried over to the next period. In this setting, we propose non-parametric adaptive policies that generate ordering decisions over time. We show that the T-period average expected cost of our policy differs from the benchmark newsvendor cost – the minimum expected cost that would have incurred if the manager had known the underlying demand distribution – by at most $O(1/\sqrt{T})$.

Key words: Newsvendor Problem; Nonparametric Demand; Censored Demand; Online Optimization

MSC2000 Subject Classification: Primary: 90B30, 62G07; Secondary: 68T05

OR/MS subject classification: Primary:Inventory/Production:Uncertainty:Stochastic; Secondary: Analysis of Algorithms, Statistics:Nonparametric

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 often 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 when excess demand is lost, the information available to the inventory manager is further limited since she does not observe the realized demand but only observes 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.

In our model, time periods are indexed forward by $t = 1, \ldots, T$, and random demands in each period $D_1, D_2, \ldots$ are independent and identically distributed. We denote by D a generic random variable with the same common distribution. We develop an adaptive inventory policy $\phi = (y_t \mid t \geq 1)$, where the decision y_t represents 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*, neither assuming any prior knowledge of the underlying demand distribution nor observing any lost sales quantity. We assume that the inventory decision is made at the beginning of each period and the replenishment lead time is instantaneous. We consider two separate models in which excess inventory at the end of a period either perishes or is carried over to the next period.

For any order-up-to level y, let $Q(y)$ denote the expected overage and underage cost in a period, where the overage cost and the underage cost are linear and the expectation is taken with respect to the underlying (yet unknown) demand distribution. Had we known the underlying demand distribution, it is a well-known result that the minimum expected cost corresponds to the newsvendor cost, that is, $\min_{y \geq 0} Q(y) = Q\left(y^{NV}\right)$, where y^{NV} denotes the newsvendor quantity.

To assess the quality of an inventory policy $\phi = (y_t : t \geq 1)$, we use the newsvendor cost $Q\left(y^{NV}\right)$ as the benchmark, and compare it to the average expected cost over time under ϕ, that is, we consider the

T-period average expected regret

$$\Delta_T(\phi) \quad = \quad E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q\left(y^{NV}\right) \ ,$$

where $T \geq 1$. Note that Δ_T is nonnegative by the definition of $Q\left(y^{NV}\right)$. A major result of this paper is to propose an adaptive inventory policy ϕ whose average expected cost converges to the newvendor cost benchmark at the rate of $1/\sqrt{T}$, that is , $\Delta_T(\phi) = O\left(1/\sqrt{T}\right)$. Our convergence results hold for both perishable and non-perishable inventory settings. To our knowledge, this represents the first algorithm with a provable rate of convergence for this problem.

1.1 Algorithm Overview We briefly outline the ideas behind our algorithm for perishable products. The problem described above is difficult to solve optimally because it involves multiple periods; the current period's decision affects the censoring of the current demand data, which in turn affects future inventory decisions. However, the newsvendor benchmark $Q\left(y^{NV}\right)$ corresponds to the minimum of a single-period convex cost function Q. Given the order-up-to level y_t in period t, it turns out that we can compute an unbiased estimate of a subgradient of Q at y_t *using only the sales (censored demand) data at time* t. This result enables us to leverage the online convex optimization method for minimizing the convex function Q, by adjusting the order-up-to level in the next period based on the subgradient of Q evaluated at the order-up-to level of the current period. Under our proposed algorithm, the sequence of order-up-to levels $(y_t : t \geq 1)$ is generated as follows: for any $t \geq 1$,

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

where ϵ_t denotes the step size in period t and $H_t(y_t)$ denotes an unbiased estimate of the subgradient of Q at y_t, which is computed solely from the sales data in period t. 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 newsvendor quantity. (Later in the paper, we extend the result to the case where the knowledge of $\bar{y}$ is not available. See Section 3.1 for more details.) By choosing $\epsilon_t = O\left(1/\sqrt{t}\right)$, we show that the average expected cost converges to the newsvendor benchmark $Q\left(y^{NV}\right)$ at the rate of $O\left(1/\sqrt{T}\right)$ (Theorem 2 in Section 2.2). We also show how we can obtain better convergence rates of $O(\log T/T)$ by imposing additional assumptions on the problem (in Section 3.5). Our analysis makes use of recent results in the online convex optimization literature.

In the case of non-perishable inventory, the major difficulty in applying existing online convex optimization is the dependency of decisions from one period to another; the order-up-to level decision in each period is constrained by the current on-hand inventory because left-over inventory is carried over to the next period. Thus, the target order-up-to level computed using the above stochastic gradient descent method may not be feasible because it may be less than the on-hand inventory. We circumvent this difficulty by establishing a relationship between the amount of inventory in excess of the target level and the waiting time process in a GI/D/1 queue. By controlling the step size of the gradient descent method, we prove that, over T periods, the average expected inventory in excess of our target order-up-to levels is also at most $O\left(1/\sqrt{T}\right)$. Thus, the average incremental holding cost is at most $O\left(1/\sqrt{T}\right)$, still giving us the desired convergence result.

1.2 Literature Review and Our Contributions *Classical Inventory Theory:* 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 when 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 corresponds to a fractile – a ratio involving per-unit overage and underage costs – of the underlying demand distribution. Whether excess inventory is perishable or not, the newsvendor-based base-stock policy is optimal. (See, for example, Karlin and Scarf [21].) 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.

Bayesian Approaches: 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 the underlying demand belongs, but she is uncertain about its parameters. She has an initial prior belief regarding the uncertainty of the parameter values, and this belief is continually updated based on historical realized demands by computing posterior distributions. Early papers such as Scarf [36, 35], Karlin [20] and Iglehart [17] 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 [30], Chang and Fyffe [6], and Azoury [2]. Lovejoy [27] shows that a simple myopic inventory policy based on a critical fractile is optimal or near-optimal. In all of the 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, however, 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 [9], who shows the effect of censoring in estimating the parameter of the Poisson demand distribution. In the Bayesian 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. The main result, here, 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. [14] and Ding et al. [10]. A recent paper by Lu et al. [28] 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. In this case, the optimal inventory level may be higher or lower than the myopic solution. Lariviere and Porteus [24] study this case with a particular distribution called the "newsvendor distribution" (Braden and Freimer [4]), and provide sufficient conditions for the stock-higher result to hold. Using a sample-path argument, Lu et al. [29] prove that, in general, the stock-higher result does not hold. Chen and Plambeck [7] also consider the Bayesian learning of product substitution.

In the case where the manager knows the distribution family to which demand belongs, but does not know either its parameters or its priors, Liyanage and Shanthikumar [26] 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. [8] 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 sometimes difficult to parsimoniously update the prior distribution as pointed out by Nahmias [31]. Additionally, in many applications, it is unclear which particular prior distribution one should be using.

We point out subtle but important differences in the formulation of the objective function between the Bayesian approach and our non-parametric method. In the Bayesian framework, the expected cost in each period is accounted for using the *Bayesian estimate* of the demand distribution in that period. This estimate is computed based on the assumed family of distributions that the underlying demand belongs to as well as historical observations. In other words, the expected cost in a period depends not on the underlying demand distribution, but *on the manager's belief of the demand distribution in that period.* (See, for example, Ding et al. [10]). The inventory planning problem under the Bayesian framework can thus be formulated using a Markov Decision Process whose state at time t corresponds to the posterior distribution ϕ_t of the underlying demand distribution based on observations up to time t. We then have the following dynamic programming recursion relating the cost-to-go functions V_t's:

$$V_t(\phi_t) = \inf_{y_t \geq 0} \left\{ Q_{\phi_t}(y_t) + E\left[V_{t+1}\left(\phi_{t+1}(\phi_t, y_t) \right) \right] \right\},$$

where the random variable $\phi_{t+1}(\phi_t, y_t)$ corresponds to the posterior distribution of the demand in period $t+1$, which depends on the posterior distribution ϕ_t in period t and the order-up-to decision y_t. Note that the expected cost $Q_{\phi_t}(y)$ incurred in period t is *computed based on the posterior distribution* ϕ_t *of the demand in period* t, representing the manager's belief about the demand in period t.

In contrast, our method assumes that there exists a **unique true underlying demand distribution** (even though the manager does not know it *a priori*). The expected cost in each period, $Q(\cdot)$, is *always computed with respect to this unique true underlying demand distribution.* Our analysis uses the newsvendor cost $Q(y^{NV})$ as the benchmark (corresponding to the minimum expected cost that the manager would incur had she known the true underlying demand distribution), and examines the rate at which the average expected cost $\sum_{t=1}^{T} Q(y_t)/T$ converges to this newsvendor benchmark cost. Our method does not use any prior distribution.

Non-Parametric Approaches. In this paper, we take a non-parametric approach, where the inventory manager knows neither the demand distribution nor the distribution family to which the demand belongs. The manager must make an ordering decision in each period based only on historical sales (censored demand) data. There is a limited number of non-parametric approaches dealing with censored demand data. One such approach is based on a variant of a stochastic approximation algorithm that finds the critical fractile of the demand distribution using censored demand samples. Using this approach, Burnetas and Smith [5] develop an adaptive algorithm for ordering and pricing when inventory is perishable. They show that the average profit converges to the optimal, but they do not establish the rate of convergence. 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 the case of non-perishable inventory.

Another non-parametric approach that utilizes censored data to estimate the newsvendor cost function by recognizing its convexity is the Concave, Adaptive Value Estimation (CAVE) algorithm. This algorithm successively approximates the cost function with a sequence of piecewise linear functions. When inventory is perishable, Godfrey and Powell [13] show that the CAVE algorithm has good numerical performance, but does not prove any convergence result. Powell et al. [33] extend this line of research and propose a modified algorithm that produces an asymptotically optimal solution. (In both of the above papers, the speed of convergence was addressed experimentally.)

While these methods mentioned above have been used only for the perishable inventory case, our algorithms apply to both perishable and non-perishable inventories. Furthermore, we provide the convergence rates of our algorithms in both cases; to our knowledge, our rates of convergence represent the first such results for these problems.

We mention other non-parametric approaches in the inventory literature. Recently, Levi et al. [25] study a multi-period inventory system without any knowledge of the demand distribution, when *uncensored* samples from the demand distributions are available. They compute the sample size required to achieve a certain level of accuracy with high probability. Also with uncensored demand data, Chang et al. [**?**] propose an adaptive algorithm using results from multi-armed bandit problems (see Lai and Robbins [23] and Auer et al. [1] for more details). Another approach with uncensored demand data is the bootstrap method, as shown in Bookbinder and Lordahl [3], to estimate the fractile of the demand distribution. Yet another approach is applicable when the manager has limited access to the demand distribution (such as mean and standard deviation). The objective is to compute the optimal stocking quantity that will provide the maximum expected profit against the worst possible demand for that stocking quantity. See Scarf [34], Jagannathan [18], and Gallego and Moon [12]. Perakis and Roels [32] present an algorithm for minimizing regrets from not ordering the optimal quantity.

Online Convex Optimization. The analysis of the algorithms developed in this paper is based on recent developments in computer science. The aim of online convex optimization, as in regular convex optimization, is to minimize a convex function defined over a convex compact 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 [37] has shown that the average T-period cost converges to the optimal cost at the rate of $O\left(1/\sqrt{T}\right)$. This result was extended by Flaxman et al. [11] to the case where the optimizer instead obtains an unbiased estimator of the gradient. Under additional technical assumptions on the shape of the convex function, a modified algorithm by Hazan et al. [15] achieves a faster convergence rate of $O(\log(T))$. The case where the available information is an unbiased estimator of the objective value, not its derivative, has been studied by Flaxman et al. [11] and Kleinberg [22].

Mathematical Contributions: Our paper offers the following contributions to the mathematical inventory theory.

- Motivated by realistic constraints faced by an inventory manager, we offer non-parametric adaptive inventory policies that do not require any prior knowledge of the underlying demand distribution and make the ordering decision in each period based only on historical sales (censored demand) data. We also establish the first rate of convergence guarantee for this class of inventory problems.
- While our proof technique for the perishable inventory case relies on existing results on online convex optimization, existing analysis however no longer applies if inventory is not perishable; the order-up-to level decision in each period is constrained by the decisions made in the earlier periods. In this case, we introduce a new proof technique by establishing a connection between the application of the stochastic gradient method and the waiting time process in a single server $GI/D/1$ queue, whose service time parameter is related to the step size of the gradient descent method (see Theorem 6 in Section 2.3.2). We believe that this new insight is of independent interest, and may be applicable to other online optimization problems where the decision in each period may be constrained by past decisions.
- The existing analysis of online optimization methods requires the compactness of the feasible set, which, in the inventory model, corresponds to the assumption the manager knows an upper bound on the optimal order-up-to level (the newsvendor quantity) *a priori*. In many applications, however, this information on the upper bound might not be available. We introduce a new variation of the stochastic gradient descent method that does not require any knowledge of the upper bound on the optimal order-up-to level (Section 3.1). We show that for any $\delta > 0$, there is an adaptive algorithm whose average expected cost converges to the newsvendor cost benchmark at the rate of $O\left(\left(1/T^{0.5-\delta}\right) + \left(A^{1/\delta}/T\right)\right)$, where the constant A is independent of δ (Theorem 7). Our technique is applicable to a general online convex optimization problem in removing the assumption that the compact feasible set is known *a priori.*

1.3 Organization This paper is organized as follows. In Section 2, we describe the problem in detail, and propose an adaptive policy. We establish that the T-period average expected cost of the policy converges to the newsvendor cost benchmark at a rate of $O\left(1/\sqrt{T}\right)$ in both perishable and non-perishable inventory settings. In Section 3, we consider several generalizations and extensions, including the case where an upper bound on the newsvendor quantity is not available *a priori.* We conclude in Section 4.

2. Adaptive Inventory Control

In this section, we develop an adaptive inventory policy and prove its convergence. We present our problem formulation in Section 2.1, and state the algorithm and our main result (Theorem 2) in Section 2.2. After establishing a connection between the application of the stochastic gradient descent method and the waiting time process in a GI/D/1 queue in Section 2.3, we prove Theorem 2 in Section 2.4.

2.1 Problem Formulation We consider a multi-period inventory system with stationary demand, where any demand that cannot be satisfied immediately is lost. Excess inventory in each period is either scrapped entirely (perishable) or carried over to the next period (non-perishable). (In Section 3.3, we discuss the case of partially perishable inventory – where only a fraction of the inventory is scrapped.) Both overage and underage costs are linear. Let $D_1, D_2, \ldots$ denote the sequence of nonnegative demand random variables, where D_t denotes the demand in period t. While the manager knows that the demand is independent and identically distributed in each period, we assume that she does not know its distribution *a priori.* She observes only the sales quantity in each period, corresponding to the minimum of the demand and the stocking quantity; she does not observe lost sales.

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

(i) 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 $x_1 = 0$. In the case of perishable inventory, we have $x_t = 0$ for each t.

(ii) 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.

(iii) The demand D_t in period t is realized and we denote its realized value by d_t. The manager does **not** observe d_t, instead she observes the sales quantity $\min\{d_t, y_t\}$.

(iv) The overage and underage cost associated with this period is $h \cdot [y_t - d_t]^+ + b \cdot [d_t - y_t]^+$. While the manager does not observe the quantity of lost sales, we assume that she incurs the goodwill loss of b per unit. The inventory at the beginning of the next period is given by $x_{t+1} = 0$ in the perishable case, and $x_{t+1} = [y_t - d_t]^+$ in the non-perishable case.

Note that our cost-minimization formulation is equivalent to the following profit-maximization version. Let $\bar{c}$ be the purchase cost and let $\bar{p}$ be the selling price per unit, where $\bar{p} \geq \bar{c} \geq 0$. Let $\bar{h}$ be the per-unit holding cost in the case of excess inventory, and let $\bar{b}$ be the goodwill lost in the case of unsatisfied demand. Then, the T-period profit is

$$\sum_{t=1}^{T} \left(\bar{c} \cdot u_t + \bar{p} \cdot \min\{d_t, y_t\} - \bar{h} \cdot [y_t - d_t]^+ - \bar{b} \cdot [d_t - y_t]^+ \right) .$$

Since $u_t = y_t - x_t$ and $x_t = y_{t-1} - \min\{d_{t-1}, y_{t-1}\}$, it equals

$$\bar{c} \cdot (x_{T+1} - x_1) + (\bar{p} - \bar{c}) \cdot \sum_{t=1}^{T} d_t - \sum_{t=1}^{T} \left(\bar{h} \cdot [y_t - d_t]^+ + (\bar{b} + \bar{p} - \bar{c}) \cdot [d_t - y_t]^+ \right) .$$

Under any reasonably policy, the first term is finite, and does not affect the long-run average cost. The second term is a constant independent of the decisions. The third term represents the overage and underage cost, where $h = \bar{h}$ and $b = \bar{b} + \bar{p} - \bar{c}$. Thus, in this paper, we suppose $c = 0$, and under the long-run average cost criterion, this assumption is without loss of any generality by appropriately modifying h and b parameters. Interested readers are referred to Veinott and Wagner [**?**] and Janakiraman and Muckstadt [19] for more details.

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

$$Q(y) \quad = \quad h \cdot E[y - D]^+ + b \cdot E[D - y]^+ \ , \tag{1}$$

where D denotes the demand random variable, having the same distributions as $D_1, D_2, \ldots$. This single-period cost function is also known as the newsvendor cost function. It is well-known that $Q(\cdot)$ is convex since its left-derivative is

$$\lim_{\varepsilon \downarrow 0} \frac{Q(y + \varepsilon) - Q(y)}{\varepsilon} \quad = \quad h \cdot P[y \geq D] - b \cdot P[y < D] \ , \tag{2}$$

and that Q achieves its minimum at the newsvendor quantity given by

$$y^{NV} = \inf\left\{ y \geq 0 \mid F(y) \geq b/(b+h) \right\} ,$$

where F denotes the distribution function of the demand D. (See, for example, Zipkin [**?**] or Porteus [**?**].)

Since the manager does not know the demand distribution, she does not know the function Q. In both perishable and non-perishable inventory settings, we aim to find a sequence of inventory levels $(y_t : t \geq 1)$ whose average expected cost $E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right]$ converges to the newsvendor cost $Q\left(y^{NV}\right)$. We require that *the inventory level* y_t *depends only on the sales quantities observed by the manager during the previous* $t-1$ *periods.*

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 single-period newsvendor model, and ordering up to y^{NV} in each period is also optimal for this problem; in such case, the expected cost incurred in each period is $Q\left(y^{NV}\right)$. Under this myopic 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, in the case of non-perishable inventory, her decision in each period may be *tightly* constrained by the carry-over inventory from the previous period.

2.2 AIM Algorithm In this section, we define the Adaptive Inventory Management (AIM) algorithm that generates an asymptotically optimal sequence of inventory levels. To facilitate our discussion and analysis, let us introduce the following assumption that will be used throughout Section 2.

Assumption 1 *The manager knows an upper bound $\bar{y}$ on the newsvendor quantity y^{NV}, that is, $y^{NV} \leq \bar{y}$. Furthermore, in the non-perishable inventory case, she also knows a lower bound $\rho > 0$ on the expected demand, that is,* $0 < \rho < E[D_1]$.

The above assumption is introduced primarily to simplify the description and analysis of our AIM algorithm. *We emphasize that even when the above assumption fails, we can still develop variations of the AIM algorithm that yield an asymptotically optimal sequence of inventory levels with similar convergence rates. These variations and extensions are considered in Sections 3.1 and 3.2.*

The AIM algorithm 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, with $y_t \geq \hat{y}_t$ for all t. The two sequences are recursively defined as follows. Set $y_1 = \hat{y}_1$ to any value in $[0, \bar{y}]$. For $t \geq 1$, let

$$\begin{aligned} \hat{y}_{t+1} &= P_{[0,\bar{y}]}\left(\hat{y}_t - \epsilon_t H_t(\hat{y}_t)\right) \qquad \text{and} \\ y_{t+1} &= \max\left\{\hat{y}_{t+1},\ x_{t+1}\right\}\ , \end{aligned} \tag{3}$$

where the function $P_{[0,\bar{y}]}(\cdot)$ denotes 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., $P_{[0,\bar{y}]}(z) = \max\{\min\{z, \bar{y}\},\ 0\}$. The step size ϵ_t is given by

$$\epsilon_t = \frac{\gamma \bar{y}}{\max\{b, h\}\sqrt{t}} \qquad \text{for some } \gamma > 0\ , \tag{4}$$

and the random variable $H_t(\hat{y}_t)$ is defined as

$$H_t(\hat{y}_t) = \begin{cases} h, & \text{if } D_t < \hat{y}_t\ , \\ -b, & \text{if } D_t \geq \hat{y}_t\ . \end{cases} \tag{5}$$

Use of Historical Sales Data in the AIM Algorithm: We emphasize that the random variable $H_t(\hat{y}_t)$ appearing in the update equation for the AIM algorithm (see Equation (3)) can be computed based on the sales (censored demand) data observed by the manager in period t. In the perishable inventory case where $\hat{y}_t = y_t$, the event $D_t \geq \hat{y}_t$ corresponds to zero ending inventory (that is, sales equal inventory), and the event $D_t < \hat{y}_t$ corresponds to strictly positive ending inventory. These events are observable by the manager, who sees the inventory level and the sales quantity in each period. In the non-perishable inventory case where $y_t \geq \hat{y}_t$, the event $D_t \geq \hat{y}_t$ is equivalent to the case where the ending inventory in period t is at most $y_t - \hat{y}_t$; thus, this event is also observable. In both cases, we can compute $H_t(\hat{y}_t)$ based on the observed sales quantity and inventory level y_t in period t.

The main result of this section is given in Theorem 2, which states that the expected running average cost of the AIM algorithm converges to the newsvendor benchmark cost $Q\left(y^{NV}\right)$ at the rate of $O(1/\sqrt{T})$. The proof of Theorem 2 appears in Section 2.4. Furthermore, we provide an example in Section 2.5 showing the convergence rate of $\Theta(1/\sqrt{T})$.

Theorem 2 *Under Assumption 1, the sequence of order-up-to levels $(y_t : t \geq 1)$ generated by the AIM algorithm has the following properties.*

- *Perishable Inventory Case: For any $T \geq 1$,*

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

- *Non-Perishable Inventory Case: Suppose $E[D_1^6] < \infty$ and $\gamma \leq (\rho \max\{b, h\})/(h\bar{y})$. There is a constant C such that for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \leq \frac{C}{\sqrt{T}}\ .$$

The explicit formula for the constant C above is given in the proof of Theorem 2 in Section 2.4 (Equation 9). According to Theorem 2, when excess inventory is perishable, the average expected cost under the AIM algorithm converges to the newsvendor benchmark for any choice of the scaling parameter γ used in the definition of the step size. The algorithm and performance analysis use the knowledge of $\bar{y}$, but not ρ. When excess inventory is non-perishable, however, the AIM algorithm uses both $\bar{y}$ and ρ, requiring the scaling parameter γ be sufficiently small relative to ρ. As we mentioned earlier, we will generalize the AIM algorithm to the settings when $\bar{y}$ and ρ are not known *a priori* later in Section 3.1 and 3.2.

2.3 Preliminaries In this section, we present and prove properties of online convex programming (Section 2.3.1) and establish a connection between the gradient descent method and a queueing process (Section 2.3.2). We use these results in the proof of Theorem 2.

2.3.1 Online Convex Programming 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 [37] 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. [11]. Lemma 3 below is a minor adaption of this result to the case where the objective function may not be differentiable, and this lemma is used to establish Theorem 2. The proof of Lemma 3 appears in Appendix A, and it will be modified later in Section 3.1 to address a case where the domain of the function may *not* be bounded.

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

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

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 $z \in S$.

Lemma 3 *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., $g(z) \in \bigtriangledown\Phi(z)$. For any $z \in S$, let $H(z)$ be an n-dimensional random vector defined on S such that $E\left[H(z) \mid z\right] \;=\; g(z)$. Suppose that there exists $\bar{B}$ such $\|H(z)\| \leq \bar{B}$ with probability one for all $z \in S$. Let w_1 be any point in S. For any $t \geq 1$, recursively define*

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

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

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

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

2.3.2 Connection Between the Stochastic Gradient Descent Method and the Waiting Time Process in a GI/D/1 Queue A $GI/D/1$ queue denotes a single server queue with general identical and independently distributed (IID) inter-arrival times and deterministic service times. For any $\theta > 0$, consider the stochastic process $(W_t(\theta) \mid t \geq 0)$ defined by the following Lindley's equation: $W_0(\theta) = 0$ and

$$W_{t+1}(\theta) \quad = \quad \left[W_t(\theta) + \theta - D_t\right]^+ \;, \tag{6}$$

where $D_1, D_2, \ldots$ are independent and identically distributed demand random variables. For any $i \geq 1$, define a random variable $\tau_i(\theta)$ by $\tau_i(\theta) = \inf\left\{t > \tau_{i-1}(\theta) \mid W_t(\theta) = 0\right\}$, where $\tau_0(\theta) = 0$. Let $J_i(\theta) = \left\{s \mid \tau_{i-1}(\theta) < s \leq \tau_i(\theta)\right\}$. The random variable $W_t(\theta)$ can be interpreted as 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 θ. Then, $|J_i(\theta)|$ corresponds to the length of the i^{th} busy period. Proposition 4 below establishes an upper bound on the second moment of $|J_i(\theta)|$ for any $\theta > 0$. The proof is based on the Markov's Inequality and the Hoeffding Inequality, and is given in Appendix B.

Proposition 4 *Suppose* $0 < \theta < E\left[D_1\right]$.

(i) If $E\left[D_1^6\right] < \infty$, *then* $E\left[\left|J_1(\theta)\right|^2\right] \leq 7\pi^2 E\left[\left(D_1 - E[D_1]\right)^6\right] \Big/ \left(E\left[D_1\right] - \theta\right)^6$.

(ii) If there exists $\bar{D} > 0$ *such that* $D_1 \leq \bar{D}$ *with probability one, then* $E\left[\left|J_1(\theta)\right|^2\right] \leq 2\alpha/(1-\alpha)^2$, *where* $\alpha = \exp\left\{-2 \cdot \left(E\left[D_1\right] - \theta\right)^2 / \bar{D}^2\right\}$.

Note that the condition $E\left[D_1^6\right] < \infty$ is satisfied for a large class of demand distributions, including Gaussian, Poisson, Geometric, Negative Binomial, light-tail distributions, and any distribution with a finite support.

For any $\theta > 0$, we define an auxiliary stochastic process $(Z_t(\theta) \mid t \geq 0)$ where

$$Z_{t+1}(\theta) \quad = \quad \left[Z_t(\theta) + \frac{\theta}{\sqrt{t}} - D_t\right]^+ , \tag{7}$$

and $Z_0(\theta) = 0$. The process $(Z_t(\theta) \mid t \geq 0)$ is closely related to the original waiting time process $(W_t(\theta) \mid t \geq 0)$ as shown in the following lemma.

Lemma 5 *For any* $\theta > 0$ *and* $T \geq 1$, $E\left[\sum_{t=1}^T Z_t(\theta)\right] \leq 2\theta E\left[\left|J_1(\theta)\right|^2\right] \sqrt{T}$.

The main idea behind the proof of Lemma 5 is to circumvent the difficulty of working with the non-stationary Z_t process by using the process W_t given in (6), which is stationary and dominates the original process along each sample path. The proof of Lemma 5 appears in Appendix C.

The main result of this section is stated in the following theorem, which shows the connection between the inventory $y_t - \hat{y}_t$ in excess of our target level $\hat{y}_t$ under the stochastic gradient descent method and the waiting time of the t^{th} customer in the queueing process $Z_t(\rho)$. Our proof approach is novel to the best of our knowledge, and it is instrumental in upper-bounding the impact of the period-to-period dependency of the ordering decisions (in the non-perishable inventory setting).

Theorem 6 *Under Assumption 1, suppose that excess inventory is non-perishable. If* $\gamma \leq \left(\rho \max\{b, h\}\right)/(h\bar{y})$ *holds, then for any* t, $y_t - \hat{y}_t \leq Z_t(\rho)$ *with probability one.*

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

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

where d_t denotes the realized demand in period t. If $x_{t+1} \leq \hat{y}_{t+1}$, then by definition of the AIM algorithm, 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 algorithm starting at $x_1 = 0$ does not let any target inventory level $\hat{y}_t$ exceed $\bar{y}$, we have that $\bar{y} \geq x_{t+1} > \hat{y}_{t+1}$, which implies that

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

where the last inequality follows from the fact that $\hat{y}_{t+1} < \bar{y}$, and thus, $\hat{y}_t - \epsilon_t H_t(y_t) \leq P_{[0,\bar{y}]}\left(\hat{y}_t - \epsilon_t H_t(y_t)\right)$. From $y_{t+1} = y_t - d_t$, it follows

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

where the last inequality follows from the fact that $H_t(y_t) \leq h$. Thus, we prove the claim.

Now, consider the stochastic process $(Z_t(\rho) \mid t \geq 1)$ defined in Equation (7) by

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

where $Z_1(\rho) = 0$. Since $\gamma \leq \left(\rho \max\{b, h\}\right)/(h\bar{y})$, it follows from the definition of ϵ_t (Equation (4)) that for any $t \geq 1$,

$$h\epsilon_t \quad = \quad \frac{h\gamma\bar{y}}{\max\{b, h\}\sqrt{t}} \;\leq\; \frac{\rho}{\sqrt{t}}.$$

Since $y_{t+1} - \hat{y}_{t+1} \leq \left[(y_t - \hat{y}_t) + h\epsilon_t - d_t\right]^+$ for all t from the above claim and $y_1 - \hat{y}_1 = 0$, it follows, from the recursive definition of Z_t process, that $y_t - \hat{y}_t \leq Z_t(\rho)$ holds with probability one. □

2.4 Proof of the Rate of Convergence for the AIM Algorithm (Theorem 2) In this section, we prove Theorem 2 for both the perishable and non-perishable inventory settings simultaneously. The proof relies on the results established in Section 2.3. We express

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \quad = \quad \Lambda_1(T) \ + \ \Lambda_2(T) \ ,$$

where

$$\Lambda_1(T) \ = \ E\left[\frac{1}{T}\sum_{t=1}^{T} Q(\hat{y}_t)\right] - Q(y^{NV}) \quad \text{and} \quad \Lambda_2(T) \ = \ E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t) - Q(\hat{y}_t)\right] \ . \tag{8}$$

We will first show that

$$\Lambda_1(T) \quad \le \quad \left(\gamma + \frac{1}{\gamma}\right)\frac{\bar{y}\ \max\{b,h\}}{\sqrt{T}}.$$

This result follows from the fact that the expected single-period cost Q, given in Equation (1), is convex, and has a left-derivative given by $h \cdot \mathcal{P}\{D_1 < y\} - b \cdot \mathcal{P}\{D_1 \ge y\}$. It follows from the definition of $H_t(\hat{y}_t)$ (see Equation (5)) that $E\left[H_t\left(\hat{y}_t\right) \middle| \hat{y}_t\right]$ is an unbiased estimate of the left-derivative of Q. Moreover, it is easy to verify that $|H(\cdot)| \le \max\{b,h\}$. Let $S = [0, \bar{y}]$ and $\bar{B} = \max\{b,h\}$. The desired result follows directly from Lemma 3.

In the perishable inventory case, we have $\hat{y}_t = y_t$, which implies that $\Lambda_2(T) = 0$. Thus, the result of Theorem 2 for perishable inventory follows directly from the bound on $\Lambda_1(T)$. For the remainder of this section, we focus on the non-perishable case. Since $y_t = \max\{\hat{y}_t, x_t\}$ for each t, it follows

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

It follows from Theorem 6 that $y_t - \hat{y}_t \le Z_t(\rho)$ with probability one, and therefore, for any $T \ge 1$,

$$\begin{aligned} \Lambda_2(T) \quad &= \quad E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t) - Q(\hat{y}_t)\right] \ \le \ h \cdot E\left[\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)\right] \\ &\le \quad h \cdot E\left[\frac{1}{T}\sum_{t=1}^{T} Z_t(\rho)\right] \quad \le \quad \frac{2h\rho E[|J_1(\rho)|^2]}{\sqrt{T}} \ , \end{aligned}$$

where the last inequality follows from Lemma 5. To complete the proof, note that if $E\left[D_1^6\right] < \infty$, it follows from Proposition 4 that

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \quad &= \quad \Lambda_1(T) \ + \ \Lambda_2(T) \\ &\le \quad \left(\gamma + \frac{1}{\gamma}\right)\frac{\bar{y}\ \max\{b,h\}}{\sqrt{T}} + \frac{2h\rho E\left[|J_1(\rho)|^2\right]}{\sqrt{T}} \ \le \ \frac{C}{\sqrt{T}}, \end{aligned}$$

where

$$C \quad = \quad \left(\gamma + \frac{1}{\gamma}\right)\bar{y}\ \max\{b,h\} + \frac{14\pi^2 h\rho E[(D_1 - E[D_1])^6]}{(E[D_1] - \rho)^6}, \tag{9}$$

completing the proof of the theorem.

2.5 Example: $\mathbf{\Theta(1/\sqrt{T})}$ Convergence Rate for Theorem 2 In Theorem 2, we have shown that the expected running average cost of the AIM algorithm converges to the newsvendor benchmark at the rate of $O(1/\sqrt{T})$. In this section, we show by example that this rate can indeed be $\Theta(1/\sqrt{T})$. Hazan et al. [15] have established such a lower bound on the convergence rate in an adversarial setting, but not for the stochastic non-adversarial setting.

Suppose that each D_t assumes 0, 1 or 2 with the equal probability. Let $b = h = 1$. Then,

$$Q(y) \quad = \quad \begin{cases} 1 - y/3 & \text{for } y \in [0,1] \\ 1/3 + y/3 & \text{for } y \in [1,2] \ , \end{cases}$$

which is minimized at $y^{NV} = 1$ with $Q(y^{NV}) = 2/3$. We consider the perishable inventory only. Then,

$$y_{t+1} \quad = \quad \begin{cases} y_t - 2\gamma/\sqrt{t} & \text{if } D_t < y_t \\ y_t + 2\gamma/\sqrt{t} & \text{if } D_t \geq y_t. \end{cases}$$

Observe that $y_t \geq 1$ implies $P[y_{t+1} \geq 1 + 2\gamma/\sqrt{t}] \geq 1/3$, and that $y_t \leq 1$ implies $P[y_{t+1} \leq 1 - 2\gamma/\sqrt{t}] \geq 1/3$. Thus,

$$E[Q(y_{t+1})] - Q(y^{NV}) \quad \geq \quad P\left[y_{t+1} \geq 1 + 2\gamma/\sqrt{t} \text{ or } y_{t+1} \leq 1 - 2\gamma/\sqrt{t}\right] \cdot \frac{2\gamma}{3} \cdot \frac{1}{\sqrt{t}} \quad \geq \quad \frac{2\gamma}{9} \cdot \frac{1}{\sqrt{t}} \ ,$$

implying that $\sum_{t=1}^{T} E[Q(y_t)]/T - Q(y^{NV}) = \Omega(1/\sqrt{T})$. Combining this result with Theorem 2, we obtain that $\sum_{t=1}^{T} E[Q(y_t)]/T - Q(y^{NV}) = \Theta(1/\sqrt{T})$.

3. Generalizations and Remarks on the AIM Algorithm

In the previous section, we describe and analyze the AIM algorithm under the assumption that the manager has prior knowledge on the upper bound $\bar{y}$ of the newsvendor quantity y^{NV} and she knows a lower bound ρ on the expected demand (Assumption 1). We develop generalizations of the AIM algorithm that remain asymptotically optimal with similar convergence rates *even when these assumptions fail.* In Section 3.1 and 3.2, we discuss the case when the manager does not have prior knowledge of $\bar{y}$ and ρ, respectively. Then, in Section 3.3, we show that the original AIM algorithm from Section 2 continues to work even when only a fraction of excess inventory perishes. Section 3.4 considers the case of discrete demand and discrete ordering quantities. In Section 3.5, we discuss how to improve the convergence rate to $O\left(\log(T)/T\right)$ under a minor technical condition.

3.1 Without Prior Knowledge of $\bar{y}$

The description of the AIM algorithm in Section 2 depends on $\bar{y}$, an upper bound on the newsvendor stocking quantity y^{NV}. In this section, we consider the case when $\bar{y}$ is not known *a priori.* We will show that for any $\delta > 0$, there is a modified AIM algorithm whose average expected cost converges to the newsvendor cost benchmark at the rate of $O\left(\left(1/T^{0.5-\delta}\right) + \left(A^{1/\delta}/T\right)\right)$, where the constant A is independent of δ. Thus, even when $\bar{y}$ is not known *a priori*, we can still obtain an asymptotically optimal sequence of order-up-to levels with a convergence rate that is comparable to the case when the upper bound is known in advance. For ease of exposition, we focus on the perishable inventory case only; a similar result can be shown for the non-perishable case as well.

Let $\delta > 0$ be given. The modified AIM algorithm generates a sequence of order-up-to levels $\left(\hat{y}_t^\delta : t \geq 1\right)$ whose definition is similar to the original AIM algorithm described in Section 2.2, with the changes listed below. The key idea is to iteratively expand the domain of the projection operator in order to achieve a target convergence rate.

Highlights of the modified AIM algorithm:

- Redefine the step size ϵ_t in Equation (4) so that it does not depend on $\bar{y}$, i.e., for any $t \geq 1$,

$$\epsilon_t \quad = \quad \frac{1}{\max\{b,h\} \cdot \sqrt{t}} \ .$$

- Modify the updating of order-up-to levels in Equation (3) so that the projection operation does not depend on $\bar{y}$. We define

$$\hat{y}_{t+1}^\delta = P_{[0,t^{\delta/2}]}\left(\hat{y}_t^\delta - \epsilon_t H_t(\hat{y}_t^\delta)\right) \ .$$

Note that the range of the projection operation $\left[0, t^{\delta/2}\right]$ increases with t and depends on the parameter δ. Theorem 7 shows that the modified AIM algorithm has a T-period average expected regret of $O\left(\left(1/T^{0.5-\delta}\right) + \left(A^{1/\delta}/T\right)\right)$.

Theorem 7 *Consider the perishable inventory case. Suppose the manager does not know any upper bound on y^{NV} a priori. For any $\delta \in (0, 1/2)$, the sequence of order-up-to levels $\left(\hat{y}_t^\delta : t \geq 1\right)$ generated by the modified AIM algorithm has the following property: for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q\left(\hat{y}_t^\delta\right)\right] - Q\left(y^{NV}\right) \quad \leq \quad \max\{b,h\} \cdot \left[\frac{1}{T^{0.5-\delta}} + \frac{1}{T^{0.5}} + \frac{\left(y^{NV}\right)^{1+\frac{2}{\delta}}}{T}\right] \ .$$

Proof. Let $\delta > 0$ be given and let $\bar{B} = \max\{b, h\}$. For any $t \geq 1$, let $\alpha_t = t^{\delta/2}$, and let y_t^* be the minimizer of Q in the restricted domain $[0, \alpha_t]$, i.e., $y_t^* = \min\{y^{NV},\ \alpha_t\}$. Then, for any $T \geq 1$,

$$E\left[\sum_{t=1}^{T} Q(\hat{y}_t^\delta) - Q(y^{NV})\right] \;=\; E\left[\sum_{t<(y^{NV})^{2/\delta}} \left\{Q(\hat{y}_t^\delta) - Q(y^{NV})\right\}\right] \;+\; E\left[\sum_{t\geq(y^{NV})^{2/\delta}} \left\{Q(\hat{y}_t^\delta) - Q(y^{NV})\right\}\right] .$$

To establish the result of Theorem 7, it suffices to prove the following two inequalities.

$$E\left[\sum_{t<(y^{NV})^{2/\delta}} \left\{Q(\hat{y}_t^\delta) - Q(y^{NV})\right\}\right] \;\leq\; \bar{B}\left(y^{NV}\right)^{1+\frac{2}{\delta}} , \qquad \text{and} \tag{10}$$

$$E\left[\sum_{t\geq(y^{NV})^{2/\delta}} \left\{Q(\hat{y}_t^\delta) - Q(y^{NV})\right\}\right] \;\leq\; \bar{B}\cdot T^{0.5+\delta} \;+\; \bar{B}\cdot T^{0.5} . \tag{11}$$

To establish the inequality in Equation (10), note that $t < (y^{NV})^{2/\delta}$ implies $\alpha_t = t^{\delta/2} < y^{NV}$. Since $\hat{y}_t^\delta \in [0, \alpha_t]$, it follows $0 \leq y^{NV} - \hat{y}_t^\delta < y^{NV}$. Thus, $Q(\hat{y}_t^\delta) - Q(y^{NV})$ is bounded above by $\bar{B} \cdot y^{NV}$. Thus, the left-hand-side of (10) is bounded above by $\bar{B} \cdot y^{NV}$ multiplied by $(y^{NV})^{2/\delta}$, obtaining the bound in the right-side of (10).

Now we prove the inequality in Equation (11). Let $t_\circ = \lceil (y^{NV})^{2/\delta} \rceil$. Observe that most of the arguments in the proofs of Lemma 3 remain valid in this case. In particular, Equation (15) in the proof of Lemma 3 in Appendix A holds. Thus,

$$\begin{aligned}\sum_{t=t_\circ}^{T} E\left[Q(\hat{y}_t^\delta) - Q(y_t^*)\right] \;&\leq\; \sum_{t=t_\circ}^{T} \left\{ \frac{E\left\|\hat{y}_t^\delta - y_t^*\right\|^2}{2\epsilon_t} - \frac{E\left\|\hat{y}_{t+1}^\delta - y_t^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(\hat{y}_t^\delta)\right\|^2 \right\} \\ &\leq\; \sum_{t=t_\circ}^{T} \left\{ \frac{E\left\|\hat{y}_t^\delta - y^{NV}\right\|^2}{2\epsilon_t} - \frac{E\left\|\hat{y}_{t+1}^\delta - y^{NV}\right\|^2}{2\epsilon_t} \right\} \;+\; \frac{\bar{B}^2}{2} \cdot \sum_{t=t_\circ}^{T} \epsilon_t ,\end{aligned}$$

where the last inequality follows from the fact that $y_t^* = \min\{y^{NV},\ \alpha_t\} = y^{NV}$ for $t \geq t_\circ$. Since $\epsilon_t = 1/(\bar{B}\sqrt{t})$, the second sum above is bounded above by $\bar{B} \cdot T^{0.5}$. Thus, it remains to show that the first term above is bounded by $\bar{B} \cdot T^{0.5+\delta}$.

Consider

$$\begin{aligned}&\sum_{t=t_\circ}^{T} \left\{ \frac{E\left\|\hat{y}_t^\delta - y^{NV}\right\|^2}{2\epsilon_t} - \frac{E\left\|\hat{y}_{t+1}^\delta - y^{NV}\right\|^2}{2\epsilon_t} \right\} \\ &\quad\leq\; \frac{E\left\|\hat{y}_{t_\circ}^\delta - y^{NV}\right\|^2}{2\epsilon_{t_\circ}} \;+\; \frac{1}{2}\sum_{t=t_\circ}^{T} \left[\frac{1}{\epsilon_{t+1}} - \frac{1}{\epsilon_t}\right] E\left\|\hat{y}_{t+1}^\delta - y^{NV}\right\|^2 \\ &\quad\leq\; \frac{\alpha_{t_\circ}^2}{2\epsilon_{t_\circ}} \;+\; \frac{1}{2}\sum_{t=t_\circ}^{T} \left[\frac{1}{\epsilon_{t+1}} - \frac{1}{\epsilon_t}\right] \alpha_{t+1}^2 ,\end{aligned}$$

where the last inequality follows from $\left\|\hat{y}_t^\delta - y^{NV}\right\| \leq \alpha_t$ for all $t \geq t_\circ$. Then,

$$\frac{\alpha_{t_\circ}^2}{2\epsilon_{t_\circ}} \;+\; \frac{1}{2}\sum_{t=t_\circ}^{T} \left[\frac{1}{\epsilon_{t+1}} - \frac{1}{\epsilon_t}\right] \alpha_{t+1}^2 \;=\; \frac{1}{2}\left[\frac{\alpha_{T+1}^2}{\epsilon_{T+1}} \;+\; \sum_{t=t_\circ}^{T} \frac{1}{\epsilon_t}\left\{\alpha_t^2 - \alpha_{t+1}^2\right\}\right] \;\leq\; \frac{\alpha_{T+1}^2}{2\epsilon_{T+1}},$$

where the last inequality follows from $\alpha_t \leq \alpha_{t+1}$. Using the definition of α_{T+1} and ϵ_{T+1} and the fact that $(T+1)^{0.5+\delta} \leq 2T^{0.5+\delta}$ for $\delta \in (0, 0.5)$, we have

$$\frac{\alpha_{T+1}^2}{2\epsilon_{T+1}} \;=\; \frac{\bar{B} \cdot (T+1)^\delta \cdot \sqrt{T+1}}{2} \;=\; \frac{\bar{B}}{2}(T+1)^{0.5+\delta} \;\leq\; \bar{B} \cdot T^{0.5+\delta},$$

which is the desired result. □

3.2 Without Prior Knowledge of ρ In the non-perishable inventory case of Theorem 2, we have required that the scaling parameter γ used in the step size of the AIM algorithm to be small relative to ρ, a lower bound on the expected demand. This requirement assumes that the manager knows ρ *a priori.* In this section, using a more refined analysis, we show that the AIM algorithm remains asymptotically optimal with the same convergence rate *for any choice of the parameter* γ. We assume, for ease of exposition, that the manager knows $\bar{y}$ *a priori.* The intuition underlying our analysis follows from the fact that the step size $\epsilon_t = O(1/\sqrt{t})$ given in Equation (4) decreases in t; thus, even when γ is large, its impact on the running average cost decreases over time.

For any θ and z, consider a random walk starting at z with the IID increment of $\{\theta - D_s \mid s \geq 1\}$. Let $K(\theta, z)$ be the first hitting time of $(-\infty, 0]$ by the random walk, i.e.,

$$K(\theta, z) = \min \left\{ t \geq 0 \;\middle|\; z + \sum_{s=1}^{t} (\theta - D_s) \leq 0 \right\}. \tag{12}$$

Also, recall that for any θ, $J_i(\theta)$ denotes the i^{th} renewal period associated with the stochastic process $W_t(\theta)$ defined in Equation (6). The following theorem establishes an error bound for the AIM algorithm under any arbitrary scaling parameter γ. We note that by Equation (8) in Section 2.4, it suffices to establish a bound on $\Lambda_2(T)$ since the bound for $\Lambda_1(T)$ is already given in the proof of Theorem 2.

Theorem 8 *Let $m = (h\gamma\bar{y}) / \max\{b, h\}$. Consider the AIM Algorithm for the non-perishable inventory case, and suppose that the manager knows an upper bound $\bar{y}$ on y^{NV}. For any choice of the parameter γ, we have, for any $T \geq 1$ and $\rho' < \min\{m, E[D_1]\}$,*

$$\Lambda_2(T) \quad \leq \quad \frac{\max\{b,h\}\bar{y}}{T} \left\{ \left\lceil (m/\rho')^2 \right\rceil + E\left[K\left(\rho', \frac{2\sqrt{2}\, m^2}{\rho'} \right) \right] \right\} \quad + \quad \frac{2hmE[|J_1(\rho')|^2]}{\sqrt{T}} \; .$$

Theorem 8 provides an upper bound on $\Lambda_2(T)$ *for any choice of scaling parameter* γ. The second term in the above error bound is similar to the bound that appears in the proof of Theorem 2 (Section 2.4), reflecting the average excess inventory above the target levels $\hat{y}_t$'s. The first term in the error bound reflects a bound on the amount of time that is required for the impact of $\gamma/\sqrt{t}$ to become small and the stochastic process $W_t(\theta)$ to have a renewal at 0. The bound in Theorem 8 is $O\left(1/\sqrt{T}\right)$ provided that both $|J_1(\cdot)|^2$ and $K(\cdot, \cdot)$ have finite expectations. The bound on $E\left[|J_1(\cdot)|^2\right]$ is finite and given in Proposition 4. Before we proceed to the proof of Theorem 8, Lemma 9 below provides an explicit bound on $E[K(\cdot, \cdot)]$ that depends only on b, h, γ, and the moment of the demand.

Lemma 9 *If $\theta < E[D_1]$ and $E\left[D_1^4\right] < \infty$, then for any $z > 0$,*

$$E[K(\theta, z)] \quad \leq \quad \frac{2z}{E[D_1] - \theta} \quad + \quad \frac{8\pi^2 E[D_1 - E[D_1]]^4}{(E[D_1] - \theta)^4}.$$

The proof of Lemma 9, based on Markov's Inequality and the definition of $K(\theta, z)$, appears in Appendix D. Below is the proof of Theorem 8.

Proof. [Proof of Theorem 8] Fix any ρ' satisfying $\rho' < \min\{m, E[D_1]\}$. Using a similar argument as in the proof of Theorem 6, we can show that for any $t \geq 1$,

$$y_{t+1} - \hat{y}_{t+1} \leq [(y_t - \hat{y}_t) + h\epsilon_t - D_t]^+ = \left[(y_t - \hat{y}_t) + \frac{m}{\sqrt{t}} - D_t \right]^+ ,$$

where the equality follows from the definition of m and ϵ_t. As in the proof of Theorem 6, we can show that the above inequality implies that $y_t - \hat{y}_t \leq Z_t(m)$ with probability one for all t. Since $Q(y_t) - Q(\hat{y}_t) \leq h(y_t - \hat{y}_t)$, we have that

$$\Lambda_2(T) \;=\; \frac{1}{T} \sum_{t=1}^{T} E[Q(y_t) - Q(\hat{y}_t)] \;\leq\; \frac{1}{T} \sum_{t=1}^{\tilde{K}} E[Q(y_t) - Q(\hat{y}_t)] + \frac{h}{T} \sum_{t=\tilde{K}+1}^{T} E[Z_t(m)], \tag{13}$$

where

$$\tilde{T} = \min\left\{ t \in \mathbb{Z}_+ \;\middle|\; \frac{m}{\sqrt{t}} \leq \rho' \right\} \quad \text{and} \quad \tilde{K} = \min\left\{ t > \tilde{T} \;\middle|\; Z_t(m) = 0 \right\}.$$

We will now bound each of the terms in the above sum separately. We will provide an upper bound on the first term in the right-most expression of (13). By definition of $\tilde{T}$, we have that $\tilde{T} = \left\lceil (m/\rho')^2 \right\rceil \leq 1 + (m/\rho')^2$, which implies that, with probability one,

$$Z_{\tilde{T}}(m) \;\leq\; \sum_{t=1}^{\tilde{T}} \frac{m}{\sqrt{t}} \;\leq\; 2m\sqrt{\tilde{T}} \;\leq\; 2m\sqrt{1+(m/\rho')^2} \;\leq\; 2m\sqrt{2\,(m/\rho')^2} \;=\; \frac{2\sqrt{2}\; m^2}{\rho'}\ .$$

Moreover, for any $\tilde{T} < t \leq \tilde{K}$, the definition of $\tilde{T}$ implies

$$Z_t(m) \;=\; \left[Z_{t-1}(m) + \frac{m}{\sqrt{t-1}} - D_{t-1}\right]^+ \;\leq\; \left[Z_{t-1}(m) + \rho' - D_{t-1}\right]^+ ,$$

and since $Z_{\tilde{T}}(m) \leq \left(2\sqrt{2}\; m^2\right)/\rho'$, it follows from the definition of $K(\cdot\ ,\ \cdot)$ that

$$E\left[\tilde{K}\right] \;=\; E\left[\tilde{T} + (\tilde{K} - \tilde{T})\right] \;\leq\; \left\lceil (m/\rho')^2 \right\rceil + E\left[K\left(\rho', \frac{2\sqrt{2}\; m^2}{\rho'}\right)\right].$$

Therefore, since $\left|Q\left(y^1\right) - Q\left(y^2\right)\right| \leq \max\{b,h\}\bar{y}$ for any $y^1, y^2 \in [0,\bar{y}]$, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{\tilde{K}} E\left[Q(y_t) - Q(\hat{y}_t)\right] \;&\leq\; \frac{\max\{b,h\}\bar{y}}{T} E\left[\tilde{K}\right] \\ &\leq\; \frac{\max\{b,h\}\bar{y}}{T}\left\{\left\lceil (m/\rho')^2 \right\rceil + E\left[K\left(\rho', \frac{2\sqrt{2}\; m^2}{\rho'}\right)\right]\right\}\ ,\end{aligned}$$

which is the first term of the bound given in the statement of Theorem 8.

We now consider the second term in Equation (13), corresponding to time periods $\tilde{K}+1$ through T. From the above definitions of $\tilde{T}$ and $\tilde{K}$, a modification of arguments used in the proofs of Lemma 5 and Theorem 6 yields

$$\frac{h}{T}\sum_{t=\tilde{K}+1}^{T} E\left[Z_t(m)\right] \;\leq\; \frac{2hmE[|J_1(\rho')|^2]}{\sqrt{T}}. \tag{14}$$

Please see Appendix E for details. This completes the proof of Theorem 8. □

3.3 Partial Perishability of Excess Inventory In Section 2, we have assumed that excess inventory is either perishable or non-perishable, i.e., $x_{t+1} = 0$ or $x_{t+1} = [y_t - d_t]^+$. We consider the case when only some of the excess inventory is perishable. Let σ_t represent all events up to the end of period t. In particular, recall that $[y_t - d_t]^+$ is the excess inventory at the end of period t. Now, suppose that the inventory level at the beginning of period $t+1$ is given by

$$x_{t+1} \;=\; \Upsilon([y_t - d_t]^+,\ \sigma_t)\ ,$$

where $\Upsilon(\cdot,\cdot)$ is a random function whose value lies between 0 and $[y_t - d_t]^+$ with probability one. This model of partial perishability is quite general, and can handle, for example, age-dependent perishability. Note that $\Upsilon([y_t - d_t]^+,\ \sigma_t) = 0$ corresponds to the perishable inventory case, and $\Upsilon([y_t - d_t]^+,\ \sigma_t) = [y_t - d_t]^+$ corresponds to the non-perishable inventory case.

With the partially perishable inventories, the AIM algorithm described in Section 2.2 is still applicable, and it can be shown that the convergence rate in Theorem 2 (for the non-perishable inventory case) remains valid as indicated in the following corollary.

Corollary 10 *Under Assumption 1, suppose that excess inventory is partially perishable with $E[D_1^6] < \infty$ and $\gamma \leq (\rho\max\{b,h\})/(h\bar{y})$. Then, the sequence of order-up-to levels $(y_t : t \geq 1)$ generated by the AIM algorithm has the following properties: there exist a constant C such that for any $T \geq 1$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} Q(y_t)\right] - Q(y^{NV}) \;\leq\; \frac{C}{\sqrt{T}}\ ,$$

where the constant C is the same as the one given in Theorem 2 for the case of non-perishable inventory.

The proof of the above corollary parallels the proof of Theorem 2 for non-perishable inventory, and we omit the details. The key observation is that the excess inventory level in each period in the partially perishable case does not exceed the corresponding quantity in the non-perishable case.

3.4 Discrete Demand and Discrete Ordering Quantities In Section 2, we assume that the order quantity in each period can be any nonnegative real number, while the demand distribution is either continuous or discrete. In practice, the set of possible ordering quantities may be constrained to a discrete set. In this section, we study adaptive inventory control when both the demand and ordering quantities are nonnegative integers. For simplicity, we consider the perishable inventory case only under Assumption 1.

In the AIM Algorithm of Section 2, the update in the inventory decision is given by (3)-(5). The value of $H_t(y_t)$ represents an estimate of the left-derivative of Q at $\hat{y}_t$, which depends on $\mathbf{1}\left[D_t < \hat{y}_t\right]$, corresponding to whether or not there exists excess inventory. All of the analysis of Section 2 remains valid if we have instead an estimator for the right-derivative of Q. The estimator of the right-derivative depends on $\mathbf{1}\left[D_t \leq \hat{y}_t\right]$, corresponding to whether or not there exists any lost sales. For this estimator, it is not sufficient to observe historical inventory quantities and sales quantities; we also need an indicator for whether lost sales has occurred. In this section, we assume that this lost sales indicator is available, and thus can obtain an estimator for the right-derivative of Q.

To address the integrality constraint, we introduce the following variant of the AIM algorithm, which we will call AIM-Discrete. The AIM-Discrete 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

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

where $\epsilon_t = \bar{y}/\left(\max\{b,h\}\sqrt{t}\right)$ as defined previously, and $\widehat{H}_t\left(z_t\right) = -b + (b+h)\cdot \mathbf{1}\left[D_t \leq z_t\right]$. 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 $Q\left(y_t\right)$.

Assuming that lost sales indicator is available, we now argue that we can compute the estimator $\widehat{H}_t\left(z_t\right)$, which represents an unbiased estimator of the right-derivative of the newsvendor cost function $Q(\cdot)$ at z_t. If z_t happens to be an integer, then $z_t = y_t$ and the result follows from the lost sales indicator assumption. Suppose that z_t is not an integer. It follow from the definition that Q is a piece-wise linear function with integer breakpoints (see Equation (1)). Thus, for any non-integer $z_t \in \Re_+$, the slope of Q at z_t is the same as the right-derivative of Q at $\lfloor z_t \rfloor$ (or equivalently the left-derivative of Q at $\lceil z_t \rceil$). Thus, we have

$$\widehat{H}_t\left(z_t\right) \quad = \quad \begin{cases} -b \ + \ (h+b)\cdot \mathbf{1}[D_t \leq y_t] \ , & \text{if } y_t = \lfloor z_t \rfloor \\ -b \ + \ (h+b)\cdot \mathbf{1}[D_t \leq y_t - 1] \ , & \text{if } \lfloor z_t \rfloor < y_t = \lceil z_t \rceil. \end{cases}$$

The above events are observable: $\mathbf{1}[D_t \leq y_t]$ can be determined from the lost sales indicator, and $\mathbf{1}[D_t \leq y_t - 1]$ can be computed from the sales quantity. It is straightforward to show that the expected value of $\widehat{H}_t\left(z_t\right)$ is a subgradient of Q at z_t. The reason we require the right-derivative is due to the fact that the AIM-Discrete Algorithm involves probabilistic rounding; we need to find a gradient when z_t is rounded down. We remark that without the lost sales indicator, the above approach does not work; in case of $y_t = \lfloor z_t \rfloor$, we cannot obtain an estimate of the slope of Q at z_t.

Since Q is piecewise linear, it can be shown that $Q(z_t) = E\left[Q\left(y_t\right) \mid z_t\right]$. Thus, we obtain the following performance guarantee of AIM-Discrete. The proof of Theorem 11 is based on the observations of this section and Theorem 2, and is therefore omitted.

Theorem 11 *Under Assumption 1, consider the case where excess inventory is perishable. Suppose that the lost sales indicator is available for observation. Then, the AIM-Discrete Algorithm satisfies, for any* $T \geq 1$,

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

3.5 Improving the Rate of Convergence If we impose additional assumptions on the problem, we can improve the convergence rate of the AIM algorithm to $O(\log T/T)$. Suppose that the demand is continuous random variable with a continuous density function f such that $\inf_{x\in[0,\bar{y}]} f(x) \geq \alpha > 0$. In this case, we can show that the one-period expected overage and underage cost function $Q(\cdot)$ is strictly convex because for any $y \in \mathbb{R}$, it can be shown from (2) that Q is differentiable and

$$Q'(y) \quad = \quad h \cdot P[y \geq D] - b \cdot P[y < D] \;=\; (b+h)F(y) - b \;.$$

Thus,

$$Q''(y) \quad = \quad (b+h)f(y) \;\geq\; \alpha \cdot (b+h).$$

If we modify the step size of the original AIM algorithm so that for $\epsilon_t = 1/\left(\alpha(b+h)t\right)$ for all t, then it follows from Theorem 1 in Hazan et al. [15] that, for the perishable inventory case,

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

4. Conclusion Motivated by the constraints faced by inventory managers, we present a non-parametric asymptotic analysis of the inventory planning problem with censored demand. Building upon the recent results on online convex optimization, we propose an adaptive inventory policy for both perishable and non-perishable products, and establish the rate of convergence. By doing so, we extend the applicability of the existing online convex optimization literature to the cases where the feasible set may not be bounded and where the decisions may be dependent from one period to another. Our work offers many interesting directions for future research. We have established a relationship between the gradient descent method and the waiting time process in a single-server queue. It would be interesting to explore if similar relationships exist in more general settings. We can also consider extensions of our results to the case where the replenishment lead-time is positive, or the inventory system consists of multiple products or multiple echelons.

Appendix A. Proof of Lemma 3 Proof. [Proof of Lemma 3] The proof of Lemma 3 is a minor adaptation of Flaxman et al. [11]. We first claim

$$E\left[\Phi(w_t) - \Phi(w^*)\right] \quad \leq \quad \frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2 \;. \tag{15}$$

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 $E\left[\langle H(w_t),\; w_t - w^*\rangle\right]$ on the value of w_t, and taking an expectation,

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

Therefore, it follows that $E\left[\langle g(w_t),\; w_t - w^*\rangle\right]$ is bounded above by the right-hand side of (15). 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, we complete the proof of the claim (15).

Now, it remains to prove that the summation of the right-hand sides of (15) over $t = 1,\ldots,T$ is bounded above by $(\gamma + 1/\gamma)\; diam(S)\; \bar{B}\; \sqrt{T}$. Observe

$$\sum_{t=1}^{T}\left\{\frac{E\left\|w_t - w^*\right\|^2}{2\epsilon_t} - \frac{E\left\|w_{t+1} - w^*\right\|^2}{2\epsilon_t} + \frac{\epsilon_t}{2} E\left\|H(w_t)\right\|^2\right\}$$

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

where

$$\begin{aligned}
\frac{diam(S)^2}{2\epsilon_{T+1}} &= \frac{diam(S)\ \bar{B}}{2\gamma}\sqrt{T+1} \ \le\ \frac{diam(S)\ \bar{B}}{\gamma}\sqrt{T} \\
\frac{\bar{B}^2}{2}\sum_{t=1}^{T}\epsilon_t &= \frac{diam(S)\ \bar{B}\ \gamma}{2}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \ \le\ \frac{diam(S)\ \bar{B}\ \gamma}{2}\int_0^T t^{-1/2}dt \ =\ diam(S)\ \bar{B}\ \gamma\sqrt{T} \ .
\end{aligned}$$

Thus, we complete the required proof. □

Appendix B. Proof of Proposition 4 The proof of Proposition 4 makes use of the following lemma.

Lemma 12 *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables such that $EX_1 = 0$. Then, for any $n \ge 1$,*

$$E\left[\sum_{i=1}^{n} X_i\right]^4 \le 3n^2 EX_1^4 \quad \text{if } EX_1^4 < \infty, \quad \text{and} \qquad E\left[\sum_{i=1}^{n} X_i\right]^6 \le 21n^3 EX_1^6 \quad \text{if } EX_1^6 < \infty.$$

Proof. It follows from the multinomial theorem and the fact that $EX_1 = 0$ that

$$\begin{aligned}
E\left[\sum_{i=1}^{n} X_i\right]^4 &= \binom{n}{2}\frac{4!}{2!\ 2!}EX_1^2 EX_2^2 + \binom{n}{1}EX_1^4 = 6\binom{n}{2}EX_1^2EX_2^2 + \binom{n}{1}EX_1^4 \\
E\left[\sum_{i=1}^{n} X_i\right]^6 &= \binom{n}{3}\frac{6!}{2!\ 2!\ 2!}EX_1^2EX_2^2EX_3^2 \ +\ \binom{n}{2}\frac{6!}{3!\ 3!}EX_1^3EX_2^3 \\
&\quad +\ 2\binom{n}{2}\frac{6!}{4!\ 2!}EX_1^4EX_2^2 \ +\ \binom{n}{1}EX_1^6 \\
&= 90\binom{n}{3}EX_1^2EX_2^2EX_3^2 \ +\ 20\binom{n}{2}EX_1^3EX_2^3 \ +\ 30\binom{n}{2}EX_1^4EX_2^2 \ +\ \binom{n}{1}EX_1^6 \ .
\end{aligned}$$

By Jensen's Inequality, $EX_1^2EX_2^2 \le EX_1^4$, and each of $(EX_1^2)^3$, $(EX_1^3)^2$ and $EX_1^4EX_2^2$ is bounded above by EX_1^6. We have

$$\begin{aligned}
E\left[\sum_{i=1}^{n} X_i\right]^4 &\le \left(6\binom{n}{2} + \binom{n}{1}\right)EX_1^4 \\
&= (3n(n-1) + n)\,EX_1^4 \\
&= \left(3n^2 - 2n\right)EX_1^4 \\
&\le 3n^2 EX_1^4 \\
E\left[\sum_{i=1}^{n} X_i\right]^6 &\le \left[90\binom{n}{3} + 50\binom{n}{2} + \binom{n}{1}\right]\cdot EX_1^6 \\
&= (15n(n-1)(n-2) + 25n(n-1) + n)\,EX_1^6 \\
&= \left(15n^3 - 20n^2 + 6n\right)EX_1^6 \\
&\le 21n^3 EX_1^6,
\end{aligned}$$

which is the desired result. □

Here is the proof of Proposition 4.

Proof. [Proof of Proposition 4] Observe

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

We need to establish an upper bound on $\mathcal{P}\{|J_1(\theta)| \geq \ell\}$. The event $|J_1(\theta)| \geq \ell$ occurs if and only if the cumulative sum $\sum_{s=1}^{\ell'}(\theta - D_s)$ remains non-negative for all $\ell' = 1, \ldots, \ell$ remains positive. Thus, $|J_1(\theta)| \geq l$ implies $\sum_{s=1}^{\ell}(\theta - D_s) \geq 0$. Therefore,

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

(i) Consider the first case where $E\left[D_1^6\right] < \infty$. Since $ED_1 > \theta$, it follows from Markov's Inequality that

$$\begin{aligned} \mathcal{P}\left\{\sum_{s=1}^{\ell}(E[D_s] - D_s) \geq \ell\cdot(E[D_1]-\theta)\right\} &\leq \mathcal{P}\left\{\left[\sum_{s=1}^{\ell}(E[D_s] - D_s)\right]^6 \geq \ell^6\cdot(E[D_1]-\theta)^6\right\} \\ &\leq \frac{E\left[\sum_{s=1}^{\ell}(E[D_s]-D_s)\right]^6}{\ell^6\cdot(E[D_1]-\theta)^6} \\ &\leq \frac{E\left[(D_1 - ED_1)^6\right]}{(E[D_1]-\theta)^6}\cdot\frac{21\ \ell^3}{\ell^6} , \end{aligned}$$

where the last inequality follows from Lemma 12. Therefore,

$$E[|J_1(\theta)|^2] \leq 2\sum_{\ell=1}^{\infty}\ell\cdot\mathcal{P}\left\{|J_1(\theta)|\geq\ell\right\} \leq 2\sum_{\ell=1}^{\infty}\frac{21}{\ell^2}\cdot\frac{E\left[(D_1-ED_1)^6\right]}{(E[D_1]-\theta)^6} \leq 7\pi^2\frac{E\left[(D_1-ED_1)^6\right]}{(E[D_1]-\theta)^6} ,$$

where the last inequality follows from $\sum_{l=1}^{\infty} 1/l^2 = \pi^2/6$. Thus, we obtain the desired result.

(ii) We will now consider the second case when $D_1 \leq \bar{D}$ with probability one. Recall the following inequality due to Hoeffding [16]. 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

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

Therefore, since $\alpha \in (0,1)$, it follows

$$E[|J_1(\theta)|^2] \leq 2\sum_{\ell=1}^{\infty}\ell\cdot\mathcal{P}\left\{|J_1(\theta)|\geq\ell\right\} \leq 2\sum_{\ell=1}^{\infty}\ell\alpha^{\ell} = \frac{2\alpha}{(1-\alpha)^2} ,$$

where the last equality follows from $\frac{d}{d\alpha}\sum_{l=0}^{\infty}\alpha^l = \frac{d}{d\alpha}1/(1-\alpha)$. □

Appendix C. Proof of Lemma 5 For any $i \geq 1$, recall that the random variable τ_i denotes the i^{th} renewal time of the stochastic process $(W_t(\theta) \mid t \geq 1)$ defined in (6), and $|J_i(\theta)| = |\{s \mid \tau_{i-1}(\theta) < s \leq \tau_i(\theta)\}|$ denotes the length of the i^{th} renewal period. For each $t \geq 1$, let the random variable $i(\theta, t)$ denote the index k such that $J_k(\theta)$ contains t. The following lemma establishes an upper bound on the expected value of $\left|J_{i(\theta,t)}(\theta)\right|$.

Lemma 13 *For any $t \geq 1$, $E\left[\left|J_{i(\theta,t)}(\theta)\right|\right] \leq E[|J_1(\theta)|^2]$.*

Proof. Since the collection of $J_i(\theta)$'s are disjoint and partition the natural numbers, $i(\cdot)$ is well-defined. Consider the following recursive equation defined by conditioning on the time of the first renewal:

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

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

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

By iteratively applying the above recursion, we have

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

completing the proof of the proposition. □

Here is the proof of Lemma 5.

Proof. [Proof of Lemma 5] From the definition of $(W_t(\theta) \mid t \geq 1)$, we observe that $Z_t(\theta) \leq W_t(\theta)$ with probability one. We introduce another stochastic process $(V_t(\theta) \mid t \geq 1)$, where $V_t(\theta)$ corresponds to the cumulative inflow in the renewal cycle containing t, without accounting for its outflow, i.e.,

$$V_t(\theta) \ = \ \sum_{t'=1}^{t} \frac{\theta}{\sqrt{t'}} \cdot \mathbf{1}\left[t' \in J_{i(\theta,t)}(\theta)\right] \ = \ \sum_{t'} \frac{\theta}{\sqrt{t'}} \cdot \mathbf{1}\left[\tau_{i(\theta,t)-1}(\theta) < t' \leq t\right] \ . \tag{16}$$

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

$$\begin{aligned} Z_t(\theta) &\leq \sum_{t'} \frac{\theta}{\sqrt{t'}} \cdot \mathbf{1}\left[\tau_{i-1}(\theta) < t' \leq t\right] \\ &= \sum_{t'} \frac{\theta}{\sqrt{t'}} \cdot \mathbf{1}\left[\tau_{i(\theta,t)-1}(\theta) < t' \leq t\right] \ = \ V_t(\theta), \end{aligned} \tag{17}$$

which is the desired claim.

Thus, to complete the proof of Lemma 5, it suffices to establish

$$\frac{1}{\theta} \cdot \sum_{t=1}^{T} E\left[V_t(\theta)\right] \ \leq \ 2E[|J_1(\theta)|^2]\sqrt{T} \ .$$

For any fixed T,

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

Therefore,

$$\begin{aligned}
\frac{1}{\theta}\cdot\sum_{t=1}^{T} E\left[V_t(\theta)\right] &\leq \sum_{s=1}^{T}\frac{1}{\sqrt{s}}\, E\left[\left|J_{i(\theta,s)}(\theta)\right|\right]\\
&\leq \sum_{s=1}^{T}\frac{1}{\sqrt{s}}\, E\left[\left|J_1(\theta)\right|^2\right]\\
&\leq 2E\left[\left|J_1(\theta)\right|^2\right]\sqrt{T}\ ,
\end{aligned}$$

where the second inequality follows from Lemma 13, and the final inequality follows from the fact that $\sum_{s=1}^{T} 1/\sqrt{s} \leq 2\sqrt{T}$. □

Appendix D. Proof of Lemma 9

Proof. Let $\theta < E\left[D_1\right]$ be given. For any $z > 0$, we have

$$\begin{aligned}
E\left[K(\theta,z)\right] &= \sum_{s=0}^{\infty}\mathcal{P}\left\{K(\theta,z) > s\right\}\\
&\leq \frac{2z}{E\left[D_1\right]-\theta} + \sum_{s\geq\frac{2z}{E\left[D_1\right]-\theta}}\mathcal{P}\left\{K(\theta,z) > s\right\}.
\end{aligned}$$

For any $s \geq 2z\big/\left(E\left[D_1\right]-\theta\right)$, it follows from the definition of $K(\theta,z)$ in Equation (12) that

$$\begin{aligned}
\mathcal{P}\left\{K(\theta,z) > s\right\} &\leq \mathcal{P}\left\{z + \sum_{i=1}^{s}(\theta - D_i) > 0\right\}\\
&= \mathcal{P}\left\{\sum_{i=1}^{s}\left(E\left[D_i\right]-D_i\right) > s\left(E\left[D_1\right]-\theta-\frac{z}{s}\right)\right\}\\
&\leq \mathcal{P}\left\{\sum_{i=1}^{s}\left(E\left[D_i\right]-D_i\right) > s\cdot\frac{E\left[D_1\right]-\theta}{2}\right\}\\
&\leq \frac{16E\left[\sum_{i=1}^{s}\left(E\left[D_i\right]-D_i\right)\right]^4}{s^4\left(E\left[D_1\right]-\theta\right)^4}\\
&\leq \frac{16\cdot 3\cdot s^2 E\left[D_1 - E\left[D_1\right]\right]^4}{s^4\left(E\left[D_1\right]-\theta\right)^4},
\end{aligned}$$

where the second inequality follows from $s \geq 2z\big/\left(E\left[D_1\right]-\theta\right)$, the third inequality follows from Markov's inequality, and the final inequality follows from Lemma 12. Therefore,

$$\begin{aligned}
E\left[K(\theta,z)\right) &\leq \frac{2z}{E\left[D_1\right]-\theta} + \sum_{s\geq\frac{2z}{E\left[D_1\right]-\theta}}\frac{48E\left[D_1-E\left[D_1\right]\right]^4}{s^2\left(E\left[D_1\right]-\theta\right)^4}\\
&\leq \frac{2z}{E\left[D_1\right]-\theta} + \frac{48E\left[D_1-E\left[D_1\right]\right]^4}{\left(E\left[D_1\right]-\theta\right)^4}\sum_{s\geq 1}\frac{1}{s^2}\\
&= \frac{2z}{E\left[D_1\right]-\theta} + \frac{8\pi^2 E\left[D_1-E\left[D_1\right]\right]^4}{\left(E\left[D_1\right]-\theta\right)^4},
\end{aligned}$$

where the last equality follows from the fact that $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$. □

Appendix E. Proof of Claim (14) in the Proof of Theorem 8 We will now establish an upper bound on $Z_t(m)$ for $t > \tilde{K}$, using arguments similar to the proof of Lemma 5 given in Appendix C. It follows from the definition of $\tilde{T}$ and $\tilde{K}$ that $Z_{\tilde{K}}(m) = 0$ and $t > \tilde{K}$ implies that $m/\sqrt{t-1} \leq \rho'$. Thus, for any $t > \tilde{K}$,

$$Z_t(m) \;=\; \left[Z_{t-1}(m) + \frac{m}{\sqrt{t-1}} - D_{t-1}\right]^+ \;\leq\; \left[Z_{t-1}(m) + \rho' - D_{t-1}\right]^+ \;\leq\; W_t\left(\rho'\right).$$

Thus, when $Z_t(m) > 0$, then $\tau_{i-1}(\rho') < t < \tau_i(\rho')$ for some i. Using the same argument as in Equation (17), we conclude that for $t > \tilde{K}$,

$$Z_t(m) \quad \leq \quad \sum_{t'} \frac{m}{\sqrt{t'}} \mathbf{1}\left[\tau_{i-1}(\rho') < t' \leq t\right] \quad = \quad \frac{m}{\rho'} \sum_{t'} \frac{\rho'}{\sqrt{t'}} \mathbf{1}\left[\tau_{i-1}(\rho') < t' \leq t\right] \quad = \quad \frac{m}{\rho'} V_t(\rho'),$$

where the last inequality follows from the fact that $\tau_{i-1}(\rho') < t < \tau_i(\rho')$. Therefore,

$$\sum_{t=\tilde{K}+1}^{T} E[Z_t(m)] \quad \leq \quad \frac{m}{\rho'} \sum_{t=\tilde{K}+1}^{T} E[V_t(\rho')] \quad \leq \quad \frac{m}{\rho'} \cdot \sum_{t=1}^{T} E[V_t(\rho')] \quad \leq \quad \frac{m}{\rho'} \cdot 2\rho' \cdot E[|J_1(\rho')|^2] \sqrt{T}\ ,$$

where the last inequality follows as in the proof Lemma 5 since $\rho' < E\left[D_1\right]$. We thus obtain

$$\frac{h}{T} \sum_{t=\tilde{K}+1}^{T} E\left[Z_t(m)\right] \quad \leq \quad \frac{2hmE[|J_1(\rho')|^2]}{\sqrt{T}}\ ,$$

proving Claim (14).

Acknowledgments. We thank Jack Muckstadt for discussions on the stochastic inventory, and Karl Sigman and Gennady Samorodnitsky for discussions on stochastic processes. We also thank Retsef Levi for his feedback on an earlier version of this paper. Comments provided by the referee board have improved the content and exposition of this paper. Research is supported partially by NSF grants DMS-0732169 and DMS-0732196.

References

[1] P. Auer, N. Cesa-Bianchi, and P. Fisher. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.

[2] K. S. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31(9):1150–1160, 1985.

[3] J. H. Bookbinder and A. E. Lordahl. Estimation of inventory reorder level using the bootstrap statistical procedure. *IEE Transactions*, 21:302–312, 1989.

[4] D. J. Braden and M. Freimer. Information dynamics of censored observations. *Management Science*, 37:1390–1404, 1991.

[5] A. N. Burnetas and C. E. Smith. Adaptive ordering and pricing for perishable products. *Operations Research*, 48(3):436–443, 2000.

[6] S. H. Chang and D. E. Fyffe. Estimation of forecast errors for seasonal style-goods sales. *Management Science*, 18(2):B89–B96, 1971.

[7] L. Chen and E. Plambeck. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing and Service Operations Management*, 10(2):236–256, 2008.

[8] L. Y. Chu, J. G. Shanthikumar, and Z.-J. M. Shen. Solving operational statistics via a bayesian analysis. *Operations Research Letters*, 36(1):110–116, 2008.

[9] S. A. Conrad. Sales data and the estimation of demand. *Operations Research Quarterly*, 27(1):123–127, 1976.

[10] X. Ding, M. L. Puterman, and A. Bisi. The censored newsvendor and the optimal acquisition of information. *Operations Research*, 50(3):517–527, 2002.

[11] A. D. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. *Working Paper*, 2004.

[12] G. Gallego and I. Moon. The distribution free newboy problem: Review and extensions. *Journal of the Operations Research Society*, 44(8):825–834, 1993.

[13] G. A. Godfrey and W. B. Powell. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science*, 47:1101–1112, 2001.

[14] G. Harpaz, W. Y. Lee, and R. L. Winkler. Optimal output decisions of a competitive firm. *Management Science*, 28:589–602, 1982.

[15] E. Hazan, A. Kalai, S. Kale, and A. Agarwal. Logarithmic regret algorithms for online convex optimization. In *Proceedings of the 19th Annual Conference on Learning Theory*, 2006.

[16] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of American Statistical Association*, 58:13–30, 1963.

[17] D. L. Iglehart. The dynamic inventory problem with unknown demand distribution. *Management Science*, 10(3):429–440, 1964.

[18] R. Jagannathan. Minimax procedure for a class of linear programs under uncertainty. *Operations Research*, 25(1):173–177, 1977.

[19] G. Janakiraman and J. A. Muckstadt. Inventory control in directed networks: A note on linear costs. *Operations Research*, 52(3):491–495, 2004.

[20] S. Karlin. Dynamic inventory policy with varying stochastic demands. *Management Science*, 6(3):231–258, 1960.

[21] S. Karlin and H. Scarf. Inventory models of the arrow-harris-marschak type with time lag. In K. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, 1958.

[22] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 2004.

[23] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[24] M. A. Lariviere and E. L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45(3):346–363, 1999.

[25] R. Levi, R. Roundy, and D. B. Shmoys. Provably near-optimal sampling-based algorithms for stochastic inventory control models. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, 2006.

[26] L. H. Liyanage and J. G. Shanthikumar. A practical inventory control policy using operational statistics. *Operations Research Letters*, 33:341–348, 2005.

[27] W. S. Lovejoy. Myopic policies for some inventory models with uncertain demand distributions. *Management Science*, 36(6):724–738, 1990.

[28] X. Lu, J.-S. Song, and K. Zhu. Dynamic inventory planning for perishable products with censored demand data. *Working Paper*, 2004.

[29] X. Lu, J.-S. Song, and K. Zhu. Inventory control with unobservable lost sales and bayesian updates. *Working Paper*, 2005.

[30] G. R. Murray and E. A. Silver. A bayesian analysis of the style goods inventory problem. *Management Science*, 12(11):785–797, 1966.

[31] S. Nahmias. Demand estimation in lost sales inventory systems. *Naval Research Logistics*, 41:739–757, 1994.

[32] G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

[33] W. Powell, A. Ruszczynski, and H. Topaloglu. Learning algorithms for separable approximations of discrete stochastic optimization problems. *Mathematics of Operations Research*, 29(4):814–836, 2004.

[34] H. Scarf. A min-max solution of an inventory problem. In k. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.

[35] H. Scarf. Some remarks on bayes solutions to the inventory problem. *Naval Research Logistics Quarterly*, 7:591–596, 1960.

[36] H. E. Scarf. Bayes solution to the statistical inventory problem. *Annals of Mathematical Statistics*, 30(2):490–508, 1959.

[37] Martin Zinkevich. Online convex programming and generalizaed infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, DC, 2003.