Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models

Retsef Levi
Sloan School of Management, MIT, Cambridge, MA, 02139, USA
email: retsef@mit.edu

Robin O. Roundy
School of ORIE, Cornell University, Ithaca, NY 14853, USA
email: robin@orie.cornell.edu

David B. Shmoys
School of ORIE and Dept. of Computer Science, Cornell University, Ithaca, NY 14853, USA
email: shmoys@cs.cornell.edu http://www.orie.cornell.edu/~shmoys

In this paper, we consider two fundamental inventory models, the single-period newsvendor problem and its multi-period extension, but under the assumption that the explicit demand distributions are not known and that the only information available is a set of independent samples drawn from the true distributions. Under the assumption that the demand distributions are given explicitly, these models are well-studied and relatively straightforward to solve. However, in most real-life scenarios, the true demand distributions are not available or they are too complex to work with. Thus, a sampling-driven algorithmic framework is very attractive, both in practice and in theory.

We shall describe how to compute sampling-based policies, that is, policies that are computed based only on observed samples of the demands without any access to, or assumptions on, the true demand distributions. Moreover, we establish bounds on the number of samples required to guarantee that with high probability, the expected cost of the sampling-based policies is arbitrarily close (i.e., with arbitrarily small relative error) compared to the expected cost of the optimal policies which have full access to the demand distributions. The bounds that we develop are general, easy to compute and do not depend at all on the specific demand distributions.

Key words: Inventory, Approximation ; Sampling ;Algorithms ; Nonparametric

MSC2000 Subject Classification: Primary: 90B05 , ; Secondary: 62G99 ,

OR/MS subject classification: Primary: inventory/production , approximation/heuristics ; Secondary: production/scheduling , approximation/heuristics, learning

---

1. Introduction In this paper, we address two fundamental models in stochastic inventory theory, the *single-period newsvendor model* and its multiperiod extension, under the assumption that the explicit demand distributions are not known and that the only information available is a set of independent samples drawn from the true distributions. Under the assumption that the demand distributions are specified explicitly, these models are well-studied and usually straightforward to solve. However, in most real-life scenarios, the true demand distributions are not available or they are too complex to work with. Usually, the information that is available comes from historical data, from a simulation model, and from forecasting and market analysis of future trends in the demands. Thus, we believe that a *sampling-driven* algorithmic framework is very attractive, both in practice and in theory. In this paper, we shall describe how to compute *sampling-based policies*, that is, policies that are computed based only on observed samples of the demands without *any* access to and assumptions on the true demand distributions. This is usually called a *non-parametric approach*. Moreover, we shall prove that the quality (expected cost) of these policies is very close to that of the optimal policies that are defined with respect to the true underlying demand distributions.

In the single-period newsvendor model, a random demand D for a single commodity occurs in a single period. At the beginning of the period, *before* the actual demand is observed, we decide how many units of the commodity to order, and this quantity is denoted by y. Next, the actual demand d (the realization of D) is observed and is satisfied to the maximum extent possible from the units that were ordered. At the end of the period, a per-unit *holding cost* $h \geq 0$ is incurred for each unused unit of the commodity, and a per-unit *lost-sales* penalty cost $b \geq 0$ is incurred for each unmet unit of demand. The goal is to minimize the total expected cost. This model is usually easy to solve *if* the demand distribution is specified explicitly by means of a cumulative distribution function (CDF). However, we are not aware of any optimization algorithm with analytical error bounds in the case where only samples are available

and no other parametric assumption is taken.

For the newsvendor model, we take one of the most common approaches to stochastic optimization models that is also used in practice, and solve the *sample average approximation* (SAA) counterpart [39]. The original objective function is the expectation of some random function taken with respect to the true underlying probability distributions. Instead, in the SAA counterpart the objective function is the average value over finitely many independent random samples that are drawn from the probability distributions either by means of *Monte Carlo* sampling or based on available historical data (see [39] for details). In the newsvendor model the samples will be drawn from the (true) demand distribution and the objective value of each order level will be computed as the average of its cost with respect to each one of the samples of demand. The SAA counterpart of the newsvendor problem is extremely easy to solve.

We also provide a novel analysis regarding the number of samples required to guarantee that, with a specified *confidence probability*, the expected cost of an optimal solution to the SAA counterpart has a small specified relative error. Here, small relative error means that the ratio between the expected cost of the optimal solution to the SAA, with respect to the *original* objective function, and the optimal expected cost (of the original model) is very close to 1. The upper bounds that we establish on the number of samples required are general, easy to compute and apply to *any* demand distribution with finite mean. In particular, neither the algorithm nor its analysis require *any* other assumption on the demand distribution. The bounds depend on the specified confidence probability and the relative error mentioned above, as well as on the ratio between the per-unit holding and lost-sales penalty costs. However, they *do not depend* on the specific demand distribution. Conversely, our results indicate what kind of guarantees one can hope for, given historical data with fixed size. The analysis has two novel aspects. First, instead of approximating the objective function and its value, we use first-order information, and stochastically estimate one-sided derivatives. This is motivated by the fact that the newsvendor cost function is convex and hence, optimal solutions can be characterized in a compact way through first-order information. The second novel aspect of the analysis is that we establish a connection between first-order information and bounds on the relative error of the objective value. Moreover, the one-sided derivatives of the newsvendor cost function are nicely bounded and are expressed through the CDF of D. This implies that they can be estimated accurately with a bounded number of samples [18, 44, 11].

In the multiperiod extension, there is a *sequence* of independent (not necessarily identically distributed) random demands for a single commodity, which need to be satisfied over a discrete planning horizon of a finite number of periods. At the beginning of each period we can place an order for any number of units. This order is assumed to arrive after a (fixed) *lead time* of several periods. Only then do we observe the actual demand in the period. Excess inventory at the end of a period is carried to the next period incurring a per-unit holding cost. Symmetrically, each unit of unsatisfied demand is carried to the next period incurring a per-unit *backlogging penalty* cost. The goal is to find an ordering policy with minimum total expected cost. The multiperiod model can be formulated as a tractable dynamic program, where at each stage we minimize a single-variable convex function. Thus, the optimal policies can be efficiently computed, *if* the demand distributions are specified explicitly (see [47] for details).

As was pointed out in [43], solving and analyzing the SAA counterparts for multistage stochastic models seem to be very hard in general. Instead of solving the SAA counterpart of the multiperiod model, we propose a dynamic programming framework that departs from previous sampling-based algorithms. The approximate policy is computed in stages backward in time via a dynamic programming approach. The main challenge here arises from the fact that in a backward dynamic programming framework, the optimal solution in each stage heavily depends on the solutions already computed in the previous stages of the algorithm. Therefore, the algorithm maintains a *shadow dynamic program* that 'imitates' the *exact dynamic program* that would have been used to compute the exact optimal policy, if the explicit demand distributions were known. That is, in each stage, we consider a subproblem that is similar to the corresponding subproblem in the exact dynamic program that is defined with respect to the optimal solutions. However, this subproblem is defined with respect to the *approximate solutions* for the subsequent periods already computed by the algorithm in the previous stages. The algorithm is carefully designed to maintain (with high probability) the convexity of each one of the subproblems that are being solved throughout the execution of the algorithm. Thus, in each stage there is a single-variable convex minimization problem that is solved approximately. As in the newsvendor case, first-order information is used to approximately solve the subproblem in each stage of the algorithm. To do so, we use some general

structural properties of these functions to establish a central lemma (Lemma 3.3) that relates first-order information of these functions to relative errors with respect to their optimal objective value. We believe that this lemma will have additional applications in approximating other classes of stochastic dynamic programs. As was true for the newsvendor cost function, the one-sided derivatives of these functions are nicely bounded. Thus, the Hoeffding inequality implies that they can be estimated using only a bounded number of samples. The analysis indicates that the relative error of the approximation procedure in each stage of the algorithm is carefully controlled, which leads to policies that, with high probability, have small relative error. The upper bounds on the number of samples required are easy to compute and do not depend on the specific demand distributions. In particular, they grow as a polynomial in the number of periods. To the best of our knowledge, this is the first result of its kind for multistage stochastic models and for stochastic dynamic programs. In particular, the existing approaches to approximating stochastic dynamic programs do not admit constant worst-case guarantees of the kind discussed in this work (see [45]).

We believe that this work sets the foundations for additional sampling-based algorithms for stochastic inventory models and stochastic dynamic programs with analyzed performance guarantees. In particular, it seems very likely that the same algorithms and analysis described in this paper will be applicable to a (minimization) multiperiod model with Markov modulated demand process.

We next relate our work to the existing literature. There has been a lot of work to study the newsvendor model with only partial information on the underlying demand distribution. (This is sometimes called the *distribution-free newsvendor model*.)

The most popular parametric approach is the *Bayesian* framework. Under this approach, we assume to know a parametric family of distributions to which the true distribution belongs, but we are uncertain about the specific values of the parameters. Our belief regarding the uncertainty of the parameter values is updated through *prior* and *posterior* distributions based on observations that we collect over time. However, in many applications it is hard to parsimoniously update the prior distributions [30]. This approach has been applied to the newsvendor model and several other inventory models (see, for example, [2, 20, 23, 29, 38, 37]). In particular, the Bayesian approach has been applied to the newsvendor model and its multiperiod extension, but with *censored demands*. By censored demands we mean that only sales are observable, that is, in each period where the demand exceeds the available inventory, we do not observe the exact demand (see, for example, [10, 12, 25, 27, 28]). In recent work Liyanage and Shantikumar [26] have introduced a new approach that is called *operational statistics*. In this approach the optimization and estimation are done simultaneously.

The sample average approximation method has been analyzed in several recent papers. Kleywegt, Shapiro and Homem-De-Mello [24], Shapiro and Nemirovski [43] and Shapiro [40] have considered the SAA in a general setting of two-stage discrete stochastic optimization models (see [35] for discussion on two-stage stochastic models). They have shown that the optimal value of the SAA problem converges to the optimal value of the original problem with probability 1 as the number of samples grows to infinity. They have also used large-deviation results to show that the *additive* error of an optimal solution to the SAA model (i.e., the difference between its objective value and the optimal objective value of the original problem) converges to zero with probability 1 as the number of samples grows to infinity. Moreover, they have developed bounds on the number of samples required to guarantee a certain *confidence probability* that an optimal solution to the SAA model provides a certain additive error. Their bounds on the number of samples depend on the variability and other properties of the objective function as well as on the diameter of the feasible region. Hence, some of these bounds might be hard to compute in scenarios in which nothing is known about the demand distributions. Shapiro, Homem-De-Mello and Kim [42, 41] have also focused on two-stage stochastic models and considered the probability of the event that an optimal solution to the SAA model is in fact an optimal solution to the original problem. Under the assumption that the probability distributions have finite support and the original problem has a unique optimal solution, they have used large-deviation results to show that this probability converges to 1 exponentially fast as the number of samples grows to infinity. In contrast, our focus is on *relative errors* and our analysis is significantly different.

In addition, Swamy and Shmoys [46], Charikar, Chekuri and Pál [9] and Nemirovski and Shapiro [31] have analyzed the SAA counterparts of a class of two-stage stochastic linear and integer programs and established bounds on the number of samples required to guarantee that, with specified high confidence

probability, the optimal solution to the corresponding SAA model has a small specified relative error. Like ours, these bounds are easy to compute and do not depend on the underlying probability distributions. However, these results do not seem to capture the models we consider in this work. Moreover, for multistage stochastic linear programs Swamy and Shmoys [46] have shown that the SAA model is still effective in providing a good solution to the original problem, but the bounds on the number of samples and the running time of the algorithms grow exponentially with the number of stages.

In subsequent work, Huh and Rusmevichientong [19] have applied a non-parametric approach to the newsvendor model and the multiperiod model with censored demands. For these models they have shown that a stochastic variant of the classical *gradient descent* method has convergence rate proportional to the square root of the number of periods. That is, the average running cost converges in expectation to the optimal expected cost as the number of periods considered goes to infinity.

The *robust* or the *min-max* optimization approach is yet another way to address the uncertainty regarding the exact demand distributions in supply chain models including the maximization variant of the newsvendor problem; see for example [36, 13, 14, 3, 32, 1, 4]. (This approach has been applied to many other stochastic optimization models.) These method is attractive in scenarios where there is no information about the demand distributions. However, the resulting solution can be very conservative.

Other approaches have been applied to this type of inventory models. *Infinitesimal perturbation analysis* is a sampling-based stochastic gradient estimation technique that has been extensively explored in the context of solving stochastic supply chain models (see [15], [16] and [22] for several examples). The *concave adaptive value estimation* (CAVE) procedure successively approximates the objective cost function with a sequence of piecewise linear functions [17, 33]. The *bootstrap method* [7] is a non-parametric approach that aims to estimate the newsvendor quantile of the demand distribution. Another non-parametric approach is based on a stochastic approximation algorithm that approximates the newsvendor quantile of the demand distribution directly, based on censored demand samples [8]. However, to the best of our knowledge, except from asymptotic convergence results, there is no theoretical analysis of bounds on the number of samples required to guarantee a solution with small relative (or additive) error, with a high confidence probability.

The rest of the paper is organized as follows. In Section 2 we discuss the single-period newsvendor model, and in Section 3 we proceed to discuss the multiperiod model. In Section 4 we consider the case of approximating myopic policies. Finally, in Section 5 we provide a proof for a general multidimensional version of Lemma 3.3.

2. Newsvendor Problem In this section, we consider the minimization variant of the classical single-period newsvendor problem. Our goal is to find an ordering level y that minimizes the cost function $C(y) = E[h(y-D)^+ + b(D-y)^+]$, where h is the per-unit holding cost, b is the per-unit lost-sales penalty, $x^+ = \max(x, 0)$ and the expectation is taken with respect to the random demand D.

The newsvendor problem is a well-studied model and much is known about the properties of its objective function C and its optimal solutions [47]. It is well-known that $C(y)$ is convex in y. Moreover, it is easy to derive explicit expressions for the right-hand and left-hand derivatives of C, denoted by $C^r(y)$ and $C^l(y)$, respectively. Using a standard dominated convergence theorem (see [5]), the order of integration (expectation) and the limit (derivatives) can be interchanged, and the one-sided derivatives of C can be expressed explicitly. We get $C^r(y) = -b + (b+h)F(y)$, where $F(y) := Pr(D \le y)$ is the CDF of D, and $C^l(y) = -b + (b+h)Pr(D < y)$. The right-hand and the left-hand derivatives are equal at all continuity points of F. In particular, if F is continuous, then C is continuously differentiable with $C'(y) = -b + (b+h)F(y)$.

Using the explicit expressions of the derivatives, one can characterize the optimal solution y^*. Specifically, $y^* = \inf\{y : F(y) \ge \frac{b}{b+h}\}$. That is, y^* is the $\frac{b}{b+h}$ quantile of the distribution of D. It is easy to check that if F is continuous we have $C'(y^*) = 0$, i.e., not surprisingly, y^* zeros the derivative. In the more general case, we get $C^r(y^*) \ge 0$ and $C^l(y^*) \le 0$, which implies that 0 is a subgradient at y^*, and that the optimality conditions for $C(y)$ are satisfied (see [34] for details). Moreover, if the distribution of the demand D is given explicitly, then it is usually easy to compute an optimal solution y^*.

Finally, we note that all of the above is valid for any demand distribution D with $E[|D|] < \infty$, including cases when negative demand is allowed. It is clear that in the case where $E[|D|] = \infty$, the problem is not

well-defined, because any ordering policy will incur infinite expected cost.

2.1 Sample Average Approximation In most real-life scenarios, the demand distribution is not known and the only information available is data from past periods. Consider a model where instead of an explicitly specified demand distribution there is a black box that generates independent samples of the demand drawn from the true distribution of D. Assuming that the demands in all periods are independent and identically distributed (i.i.d) random variables, distributed according to D, this will correspond to available data from past periods or to samples coming from a simulation procedure or from a marketing experiment that can be replicated. Note that there is no assumption on the actual demand distribution. In particular, there is no parametric assumption, and there are no assumptions on the existence of higher moments (beyond the necessary assumption that $E[|D|] < \infty$). A natural question that arises is how many demand samples from the black box or, equivalently, how many historical obsevations are required to be able to find a provably good solution to the original newsvendor problem. By a provably good solution, we mean a solution with expected cost at most $(1+\epsilon)C(y^*)$ for a specified $\epsilon > 0$, where $C(y^*)$ is the optimal expected cost that is defined with respect to the true demand distribution D.

Our approach is based on the natural and common idea of solving the *sample average approximation (SAA)* counterpart of the problem. Suppose that we have N independent samples of the demand D, denoted by $d^1, \ldots, d^N$. The SAA counterpart is defined in the following way. Instead of using the demand distribution of D, which is not available, we assume that each one of the samples of D occurs with a probability of $\frac{1}{N}$. Now define the newsvendor problem with respect to this induced *empirical* distribution. In other words, the problem is defined as

$$\min_{y \geq 0} \hat{C}(y) := \frac{1}{N} \sum_{i=1}^{N} (h(y - d^i)^+ + b(d^i - y)^+).$$

Throughout the paper we use the symbol *hat* to denote quantities and objects that are computed with respect to the *random samples* drawn from the true demand distributions. For example, we distinguish between deterministic functions such as C above, that are defined by taking expectations with respect to the underlying demand distributions, and their SAA counterparts (denoted by $\hat{C}$), which are random variables because they are functions of the random samples which are drawn from the demand distributions. In addition, all expectations are taken with respect to the true underlying demand distributions, unless stated otherwise.

Let $\hat{Y} = \hat{Y}(N)$ denote the optimal solution to the SAA counterpart. Note again that $\hat{Y}$ is a random variable that is dependent on the specific N (independent) samples of D. Clearly, for each given N samples of the demand D, $\hat{y}$ (the realization of $\hat{Y}$) is defined to be the $\frac{b}{b+h}$ quantile of the samples, i.e., $\hat{y} = \inf\{y : \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(d^i \leq y) \geq \frac{b}{b+h}\}$ (where $\mathbb{1}(d^i \leq y)$ is the indicator function which is equal to 1 exactly when $d^i \leq y$). It follows immediately that $\hat{y} = \min_{1 \leq j \leq N}\{d^j : \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(d^i \leq d^j) \geq \frac{b}{b+h}\}$. Hence, given the demand samples $d^1, \ldots, d^N$, the optimal solution to the SAA counterpart, $\hat{y}$, can be computed very efficiently by finding the $\frac{b}{b+h}$ quantile of the samples. This makes the SAA counterpart very attractive to solve.

Next we address the natural question of how the SAA counterpart is related to the original problem as a function of the number of samples N. Consider any specified *accuracy level* $\epsilon > 0$ and a *confidence level* $1 - \delta$ (where $0 < \delta < 1$). We will show that there exists a number of samples $N = N(\epsilon, \delta, h, b)$ such that, with probability at least $1 - \delta$, the optimal solution to the SAA counterpart defined on N samples, has an expected cost $C(\hat{Y})$ that is at most $(1+\epsilon)C(y^*)$. Note that we compare the expected cost of $\hat{y}$ (the realization of $\hat{Y}$) to the optimal expected cost that is defined with respect to the true distribution of D. As we will show, the number N of required samples is polynomial in $\frac{1}{\epsilon}$ and $\log(\frac{1}{\delta})$, and is also dependent on the minimum of the values $\frac{b}{b+h}$ and $\frac{h}{b+h}$ (that define the optimal solution y^* above).

In the first step of the analysis we shall establish a connection between first-order information and bounds on the relative error of the objective value. To do so we introduce a notion of 'closeness' between an approximate solution $\hat{y}$ and the optimal solution y^*. Here 'close' does not mean that $|y^* - \hat{y}|$ is small, but that $F(\hat{y}) = Pr(D \leq \hat{y})$ is 'close' to $F(y^*)$. Recall, that $F(y) := Pr(D \leq y)$ (for each $y \in \mathbb{R}$), and let $\bar{F}(y) := Pr(D \geq y) = 1 - F(y) + Pr(D = y)$ (here we depart from traditional notation). Observe that by the definition of y^* as the $\frac{b}{b+h}$ quantile of D, $F(y^*) \geq \frac{b}{b+h}$ and $\bar{F}(y^*) \geq \frac{h}{b+h}$. The following definition provides a precise notion of what we mean by 'close' above.

Definition 2.1 *Let $\hat{y}$ be some realization of $\hat{Y}$ and let $\alpha > 0$. We will say that $\hat{y}$ is* α-accurate *if* $F(\hat{y}) \geq \frac{b}{b+h} - \alpha$ ***and*** $\bar{F}(\hat{y}) \geq \frac{h}{b+h} - \alpha$.

This definition can be translated to bounds on the right-hand and left-hand derivatives of C at $\hat{y}$. Observe that $Pr(D < y) = 1 - \bar{F}(y)$. It is straightforward to verify that we could equivalently define $\hat{y}$ to be $\alpha-$accurate exactly when $C^r(\hat{y}) \geq -\alpha(b+h)$ and $C^l(\hat{y}) \leq \alpha(b+h)$. This implies that there exists a subgradient $r \in \partial C(\hat{y})$ such that $|r| \leq \alpha(b+h)$. Intuitively, this implies that, for α sufficiently small, 0 is 'almost' a subgradient at $\hat{y}$, and hence $\hat{y}$ is 'close' to being optimal.

LEMMA 2.1 *Let $\alpha > 0$ and assume that $\hat{y}$ is $\alpha-$accurate. Then:*

(i) $C(\hat{y}) - C(y^*) \leq \alpha(b+h)|\hat{y} - y^*|$.

(ii) $C(y^*) \geq (\frac{hb}{b+h} - \alpha \max(b,h))|\hat{y} - y^*|$.

PROOF. Suppose $\hat{y}$ is $\alpha-$accurate. Clearly, either $\hat{y} \geq y^*$ or $\hat{y} < y^*$. Suppose first that $\hat{y} \geq y^*$. We will obtain an upper bound on the difference $C(\hat{y}) - C(y^*)$. Clearly, if the realized demand d is within $(-\infty, \hat{y})$, then the difference between the costs incurred by $\hat{y}$ and y^* is at most $h(\hat{y} - y^*)$. On the other hand, if d falls within $[\hat{y}, \infty)$, then y^* has higher cost than $\hat{y}$, by exactly $b(\hat{y} - y^*)$. Now since $\hat{y}$ is assumed to be $\alpha-$accurate, we know that

$$Pr([D \in [\hat{y}, \infty)]) = Pr(D \geq \hat{y}) = \bar{F}(\hat{y}) \geq \frac{h}{b+h} - \alpha.$$

We also know that

$$Pr([D \in [0, \hat{y})]) = Pr(D < \hat{y}) = 1 - \bar{F}(\hat{y}) \leq 1 - (\frac{h}{b+h} - \alpha) = \frac{b}{b+h} + \alpha.$$

This implies that

$$C(\hat{y}) - C(y^*) \leq h(\frac{b}{b+h} + \alpha)(\hat{y} - y^*) - b(\frac{h}{b+h} - \alpha)(\hat{y} - y^*) = \alpha(b+h)(\hat{y} - y^*).$$

Similarly, if $\hat{y} < y^*$, then for each realization $d \in (\hat{y}, \infty)$ the difference between the costs of $\hat{y}$ and y^*, respectively, is at most $b(y^* - \hat{y})$, and if $d \in (-\infty, \hat{y}]$, then the cost of y^* exceeds the cost of $\hat{y}$ by exactly $h(y^* - \hat{y})$. Since $\hat{y}$ is assumed to be $\alpha-$accurate, we know that

$$Pr(D \leq \hat{y}) = F(\hat{y}) \geq \frac{b}{b+h} - \alpha,$$

which also implies that

$$Pr(D > \hat{y}) = 1 - F(\hat{y}) \leq \frac{h}{b+h} + \alpha.$$

We conclude that

$$C(\hat{y}) - C(y^*) \leq b(\frac{h}{b+h} + \alpha)(y^* - \hat{y}) - h(\frac{b}{b+h} - \alpha)(y^* - \hat{y}) = \alpha(b+h)(y^* - \hat{y}).$$

The proof of part (i) then follows.

The above arguments also imply that if $\hat{y} \geq y^*$ then $C(y^*) \geq E[\mathbb{1}(D \geq \hat{y})b(\hat{y} - y^*)] = b\bar{F}(\hat{y})(\hat{y} - y^*)$. We conclude that $C(y^*)$ is at least $b(\frac{h}{b+h} - \alpha)(\hat{y} - y^*)$. Similarly, in the case $\hat{y} < y^*$, we conclude that $C(y^*)$ is at least $E[\mathbb{1}(D \leq \hat{y})h(y^* - \hat{y})] \geq h(\frac{b}{b+h} - \alpha)(y^* - \hat{y})$. In other words, $C(y^*) \geq (\frac{hb}{b+h} - \alpha \max(b,h))|\hat{y} - y^*|$. This concludes the proof of the lemma. □

We note that there are examples in which the two inequalities in Lemma 2.1 above are simultaneously tight. Next we show that for a given accuracy level ϵ, if α is suitably chosen, then the cost of the approximate solution $\hat{y}$ is at most $(1+\epsilon)$ times the optimal cost, i.e., $C(\hat{y}) \leq (1+\epsilon)C(y^*)$.

COROLLARY 2.1 *For a given accuracy level $\epsilon \in (0, \leq 1]$, if $\hat{y}$ is α-accurate for $\alpha = \frac{\epsilon}{3}\frac{\min(b,h)}{b+h}$, then the cost of $\hat{y}$ is at most $(1+\epsilon)$ times the optimal cost, i.e., $C(\hat{y}) \leq (1+\epsilon)C(y^*)$.*

PROOF. Let $\alpha = \frac{\epsilon}{3}\frac{\min(b,h)}{b+h}$. By Lemma 2.1, we know that in this case $C(\hat{y}) - C(y^*) \leq \alpha(b+h)|\hat{y} - y^*|$ and that $C(y^*) \geq (\frac{hb}{b+h} - \alpha\max(b,h))|\hat{y} - y^*|$. It is then sufficient to show that $\alpha(b+h) \leq \epsilon(\frac{hb}{b+h} - \alpha\max(b,h))$. Indeed,

$$\begin{aligned}\alpha(b+h) &\leq (2+\epsilon)\alpha\max(b,h) - \epsilon\alpha\max(b,h) \\ &= \frac{(2+\epsilon)\epsilon}{3}\frac{\max(b,h)\min(b,h)}{b+h} - \epsilon\alpha\max(b,h) \leq \epsilon(\frac{hb}{b+h} - \alpha\max(b,h)).\end{aligned}$$

In the first equality we just substitute $\alpha = \frac{\epsilon}{3}\frac{\min(b,h)}{b+h}$. The second inequality follows from the assumption that $\epsilon \leq 1$. We conclude that $C(\hat{Y}) - C(y^*) \leq \epsilon C(y^*)$, from which the corollary follows. □

To complete the analysis we shall next establish upper bounds on the number of samples N required in order to guarantee that $\hat{y}$, the realization of $\hat{Y}$, is $\alpha-$accurate with high probability (for each specified $\alpha > 0$ and confidence probability $1-\delta$). Since $\hat{Y}$ is the sample $\frac{b}{b+h}$ quantile and y^* is the true $\frac{b}{b+h}$ quantile, we can use known results regarding the convergence of sample quantiles to the true quantiles or more generally, the convergence of the empirical CDF $F_N(y)$ to the true CDF $F(y)$. (For N independent random samples all distributed according to D, we define $F_N(y) := \frac{1}{N}\sum_{i=1}^{N} X^i$, where for each $i = 1, \ldots, N$, $X_i = \mathbb{1}(D^i \leq D)$, and $D^1, \ldots, D^N$ are i.i.d. according to D.)

Lemma 2.2 below is a direct consequence of the fact that the empirical CDF converges uniformly and exponentially fast to the true CDF. This can be proven as a special case of several well-known known results in probability and statistics, such as the Hoeffding Inequality [18] and Vapnik-Chervonenkis theory [44, 11].

LEMMA 2.2 *For each $\alpha > 0$ and $0 < \delta < 1$, if the number of samples is $N \geq N(\alpha,\delta) = \frac{1}{2}\frac{1}{\alpha^2}\log(\frac{2}{\delta})$, then $\hat{Y}$, the $\frac{b}{b+h}$ quantile of the sample, is $\alpha-$accurate with probability at least $1-\delta$.*

Combining Lemma 2.1, Corollary 2.1 and Lemma 2.2 above, we can obtain the following theorem.

THEOREM 2.2 *Consider a newsvendor problem specified by a per-unit holding cost $h > 0$, a per-unit backlogging penalty $b > 0$ and a demand distribution D with $E[D] < \infty$. Let $0 < \epsilon \leq 1$ be a specified accuracy level and $1-\delta$ (for $0 < \delta < 1$) be a specified confidence level. Suppose that $N \geq \frac{9}{2\epsilon^2}(\frac{\min(b,h)}{b+h})^{-2}\log(\frac{2}{\delta})$ and the SAA counterpart is solved with respect to N i.i.d samples of D. Let $\hat{Y}$ be the optimal solution to the SAA counterpart and $\hat{y}$ denote its realization. Then, with probability at least $1-\delta$, the expected cost of $\hat{Y}$ is at most $1+\epsilon$ times the expected cost of an optimal solution y^* to the newsvendor problem. In other words, $C(\hat{Y}) \leq (1+\epsilon)C(y^*)$ with probability at least $1-\delta$.*

We note that the required number of samples does not depend on the demand distribution D. On the other hand, N depends on the square of the reciprocal of $\frac{\min(b,h)}{b+h}$. This implies that the required number might be large when $\frac{b}{b+h}$ is very close to either 0 or 1. Since the optimal solution y^* is the $\frac{b}{b+h}$ quantile of D, this is consistent with the well-known fact that in order to approximate an extreme quantile one needs many samples. The intuitive explanation is that if, for example, $\frac{b}{b+h}$ is close to 1, it will take many samples before we see the event $[D > y^*]$. We also note that the bound above is insensitive to scaling of the parameters h and b. It is important to keep in mind that these are worst-case upper bounds on the number of samples required, and it is likely that in many cases a significantly fewer number of samples will suffice. Moreover, with additional assumptions on the demand distribution it might be possible to get improved bounds.

Finally, the above result holds for newsvendor models with positive per-unit ordering cost as long as $E[D] \geq 0$. Suppose that the per-unit ordering cost is some $c > 0$ (i.e., if y units are ordered a cost of cy is incurred). Without loss of generality, we can assume that $c < b$ since otherwise the optimal solution is to order nothing. Consider now a modified newsvendor problem with holding cost and penalty cost parameters $\bar{h} = h + c > 0$ and $\bar{b} = b - c > 0$, respectively. It is readily verified that the modified cost function $\bar{C}(y) = E[\bar{h}(y-D)^+ + \bar{b}(D-y)^+]$ is such that $C(y) = \bar{C}(y) + cE[D]$ and hence the two problems are equivalent. Moreover, if $E[D] \geq 0$ and if the solution $\hat{y}$ guarantees a $1+\epsilon$ accuracy level for the modified problem, then it does so also with respect to the original problem, since the cost of each feasible solution is increased by the same positive constant $cE[D]$. Observe that our analysis allows negative demand.

3. Multiperiod Model In this section, we consider the multi-period extension of the newsvendor problem. The goal now is to satisfy a *sequence* of random demands for a single commodity over a planning horizon of T discrete periods (indexed by $t = 1, \ldots, T$) with minimum expected cost. The random demand in period t is denoted by D_t. We assume that $D_1, \ldots, D_T$ are independent but not necessarily identically distributed.

Each feasible policy P makes decisions in T stages, one decision at the beginning of each period, specifying the number of units to be ordered in that period. Let $Q_t \geq 0$ denote the size of the order in period t. This order is assumed to arrive instantaneously and only then is the demand in period t observed (d_t will denote the realization of D_t). At the end of this section, we discuss the extension to the case where there is a positive *lead time* of several periods until the order arrives. For each period $t = 1, \ldots, T$, let X_t be the *net inventory* at the beginning of the period. If the net inventory X_t is positive, it corresponds to physical inventory that is left from pervious periods (i.e., from periods $1, \ldots, t-1$), and if the net inventory is negative it corresponds to unsatisfied units of demand from previous periods. The dynamics of the model are captured through the equation $X_t = X_{t-1} + Q_{t-1} - D_{t-1}$ (for each $t = 2, \ldots, T$). Costs are incurred in the following way. At the end of period t, consider the net inventory x_{t+1} (the realization of X_{t+1}). If $x_{t+1} > 0$, i.e., there are excess units in inventory, then a per-unit holding cost $h_t > 0$ is incurred for each unit in inventory, leading to a total cost of $h_t x_{t+1}$ (the parameter h_t is the per unit cost for carrying one unit of inventory from period t to $t+1$). If, on the other hand, $x_{t+1} < 0$, i.e., there are units of unsatisfied demand, then a per-unit *backlogging penalty cost* $b_t > 0$ is incurred for each unit of unsatisfied demand, and the total cost is $-b_t x_{t+1}$. In particular, all of the unsatisfied units of demand will stay in the system until they are satisfied. That is, b_t plays a role symmetric to that of h_t and can be viewed as the per-unit cost for carrying one unit of shortage from period t to $t+1$. We assume that the per-unit ordering cost in each period is equal to 0. At the end of this section, we shall relax this assumption. The goal is to find an ordering policy that minimizes the overall expected holding and backlogging cost.

The decision of how many units to order in period t can be equivalently described as the level $Y_t \geq X_t$ to which the net inventory is raised (where clearly $Q_t = Y_t - X_t \geq 0$). Thus, the multi-period model can be viewed as consisting of a sequence of *constrained* newsvendor problems, one in each period. The newsvendor problem in period t is defined with respect to D_t, h_t and b_t, under the constraint that $y_t \geq x_t$ (where x_t and y_t are the respective realizations of X_t and Y_t). However, these newsvendor problems are linked together. More specifically, the decision in period t may constrain the decision made in future periods since it may impact the net inventory in these periods. Thus, myopically minimizing the expected newsvendor cost in period t is, in general, not optimal with respect to the total cost over the entire horizon. This makes the multi-period model significantly more complicated. Nevertheless, if we know the explicit (independent) demand distributions $D_1, \ldots, D_t$, this model can be solved to optimality by means of dynamic programming. The multi-period model is well-studied. We present a summary of the main known results regarding the structure of optimal policies (see [47] for details), emphasizing those facts that will be essential for our results. This serves as background for the subsequent discussion about the sampling-based algorithm and its analysis.

3.1 Optimal Policies It is a well-known fact that in the multi-period model described above, the class of *base-stock policies* is optimal. A base-stock policy is characterized by a set of target inventory (base-stock) levels associated with each period t and each possible state of the system in period t. At the beginning of each period t, a base-stock policy aims to keep the inventory level as close as possible to the target level. Thus, if the inventory level at the beginning of the period is below the target level, then the base-stock policy will order up to the target level. If, on the other hand, the inventory level at the beginning of the period is higher than the target, then no order is placed.

An optimal base-stock policy has two important properties. First, the optimal base-stock level in period t does not depend on any decision made (i.e., orders placed) prior to period t. In particular, it is independent of X_t. Second, its optimality is conditioned on the execution of an optimal base-stock policy in the future periods $t+1, \ldots, T$. As a result, optimal base-stock policies can be computed using dynamic programming, where the optimal base-stock levels are computed by a backward recursion from period T to period 1. The main problem is that the state space in each period might be very large, which makes the relevant dynamic program computationally intractable. However, the demands in different periods are assumed to be independent in the model discussed here, and the corresponding dynamic program is

therefore usually easy to solve, *if* we know the demand distributions explicitly. In particular, an optimal base-stock policy in this model consists of T base-stock levels, one for each period.

Next, we present a dynamic programming formulation of the model discussed above and highlight the most relevant aspects. In the following subsection, we shall show how to use a similar dynamic programming framework to construct a sampling-based policy that approximates an optimal base-stock policy.

Let $C_t(y_t)$ be the newsvendor cost associated with period t (for $t = 1, \ldots, T$) as a function of the inventory level y_t after ordering, i.e.,

$$C_t(y_t) = E[h_t(y_t - D_t)^+ + b_t(D_t - y_t)^+].$$

For each $t = 1, \ldots, T$, let $V_t(x_t)$ be the optimal (minimum) expected cost over the interval $[t, T]$ assuming that the inventory level at the beginning of period t is x_t and that optimal decisions are going to be made over the entire horizon $(t, T]$. Also let $U_t(y_t)$ be the expected cost over the horizon $[t, T]$ given that the inventory level in period t was raised to y_t (after the order in period t was placed) and that an optimal policy is followed over the interval $(t, T]$. Clearly, $U_T(y_T) = C_T(y_T)$ and $V_T(x_T) = \min_{y_T \geq x_T} C_T(y_T)$. Now for each $t = 1, \ldots, T-1$,

$$U_t(y_t) = C_t(y_t) + E[V_{t+1}(y_t - D_t)]. \tag{1}$$

We can now write, for each $t = 1, \ldots, T$,

$$V_t(x_t) = \min_{y_t \geq x_t} U_t(y_t). \tag{2}$$

Observe that the optimal expected cost V_t has two parts, the newsvendor (or the period) cost, C_t and the expected future cost, $E[V_{t+1}(y_t - D_t)]$ (where the expectation is taken with respect to D_t). The decision in period t affects the future cost since it affects the inventory level at the beginning of the next period.

The above dynamic program provides a correct formulation of the model discussed above (see [47] for a detailed discussion). The goal is to compute $V_1(x_1)$, where x_1 is the inventory level at the beginning of the horizon, which is given as an input. The following fact provides insight with regard to why this formulation is indeed correct and to why base-stock policies are optimal.

Fact 3.1 *Let $f : \mathbb{R} \mapsto \mathbb{R}$, be a real-valued convex function with a minimizer r (i.e., $f(r) \leq f(y)$ for each $y \in \mathbb{R}$). Then the following holds:*

(i) The function $w(x) = \min_{y \geq x} f(y)$ is convex in x.

(ii) For each $x \leq r$, we have $w(x) = f(r)$, and for each $x > r$, we have $w(x) = f(x)$.

Using Fact 3.1 above, it is straightforward to show that, for each $t = 1, \ldots, T$, the function $U_t(y_t)$ is convex and attains a minimum, and that the function $V_t(x_t)$ is convex. The proof is done by induction over the periods, as follows. The claim is clearly true for $t = T$ since U_T is just a newsvendor cost function and $V_T(x_T) = \min_{y_T \geq x_T} U_T(y_T)$. Suppose now that the claim is true for $t+1, \ldots, T$ (for some $t < T$). From (1), it is readily verified that U_t is convex since it is a sum of two convex functions. It attains a minimum because $\lim_{y_t \to \infty} U_t(y_t) = \infty$ and $\lim_{y_t \to -\infty} U_t(y_t) = \infty$. The convexity of V_t follows from Fact 3.1 above. This also implies that base-stock policies are indeed optimal. Moreover, if the demand distributions are explicitly specified, it is usually straightforward to recursively compute optimal base-stock levels $R_1, \ldots, R_T$, since they are simply minimizers of the functions $U_1, \ldots, U_T$, respectively. More specifically, if the demand distributions are known explicitly, we can compute R_T, which is a minimizer of a newsvendor cost function, then recursively define U_{T-1} and solve for its minimizer R_{T-1} and so on. In particular, if the minimizers $R_{t+1}, \ldots, R_T$ were already computed, then $U_t(y_t)$ is a convex function of a single variable and hence it is relatively easy to compute its minimizer. Throughout the paper we assume, without loss of generality, that for each $t = 1, \ldots, T$, the optimal base-stock level in period t is denoted by R_t and that this is the *smallest* minimizer of U_t (in case it has more than one minimizer). The minimizer R_t of U_t can then be viewed as the best policy in period t conditioning on the fact that the optimal base-stock policy $R_{t+1}, \ldots, R_T$ will be executed over $[t+1, T]$.

By applying Fact 3.1 above to V_{t+1} and U_{t+1}, we see that the function U_t can be expressed as,

$$U_t(y_t) = C_t(y_t) + E[\mathbb{1}(y_t - D_t \le R_{t+1})U_{t+1}(R_{t+1}) + \mathbb{1}(y_t - D_t > R_{t+1})U_{t+1}(y_t - D_t)]. \quad (3)$$

Clearly this is a continuous function of y_t. As in the newsvendor model, one can derive explicit expressions for the right-hand and left-hand derivatives of the functions $U_1, \ldots, U_T$, as follows. Assume first that all the demand distributions are continuous. This implies that the functions $U_1, \ldots, U_T$ are all continuously differentiable. The derivative of $U_T(y_T)$ is $U'_T(y_T) = C'_T = -b_T + (h_T + b_T)F_T(y_T)$, where F_T is the CDF of D_T. Now consider the function $U_t(y_t)$ for some $t < T$. Using the dominated convergence theorem, one can change the order of expectation and integration to get

$$U'_t(y_t) = C'_t(y_t) + E[V'_{t+1}(y_t - D_t)]. \quad (4)$$

However, by Fact 3.1 and (3) above, the derivative $V'_{t+1}(x_{t+1})$ is equal to 0 for each $x_{t+1} \le R_{t+1}$ and is equal $U'_{t+1}(x_{t+1})$ for each $x_{t+1} > R_{t+1}$ (where R_{t+1} is the minimal minimizer of U_{t+1}). This implies that

$$E[V'_{t+1}(y_t - D_t)] = E[\mathbb{1}(y_t - D_t > R_{t+1})U'_{t+1}(y_t - D_t)]. \quad (5)$$

Applying this argument recursively, we obtain

$$U'_t(y_t) = C'_t(y_t) + E[\sum_{j=t+1}^{T} \mathbb{1}(A_{jt}(y_t))C'_j(y_t - D_{[t,j)})], \quad (6)$$

where $D_{[t,j)}$ is the accumulated demand over the interval $[t, j)$ (i.e., $D_{[t,j)} = \sum_{k=t}^{j-1} D_k$), and $A_{jt}(y_t)$ is the event that for each $k \in (t, j]$ the inequality $y_t - D_{[t,k)} > R_k$ holds. Observe that $y_t - D_{[t,k)}$ is the inventory level at the beginning of period k, assuming that we order up to y_t in period t and do not order in any of the periods $t+1, \ldots, k-1$. If $y_t - D_{[t,k)} \le R_k$, then the optimal base-stock level in period k is reachable, and the decision made in period t does not have any impact on the future cost over the interval $[k, T]$. However, if $y_t - D_{[t,s)} > R_s$ for each $s = t+1, \ldots, k$, then the optimal base-stock level in period k is not reachable due to the decision made in period t, and the derivative $C'_k(y_t - D_{[t,k)})$ accounts for the corresponding impact on the cost in period k. The derivative of U_t consists of a sum of derivatives of newsvendor cost functions multiplied by the respective indicator functions.

For general (independent) demand distributions, the functions $U_1, \ldots, U_t$ might not be differentiable, but similar arguments can be used to derive explicit expressions for the right-hand and left-hand derivatives of U_t, denoted by U^r_t and U^l_t, respectively. This is done by replacing C'_j by C^r_j and C^l_j (see Section 2 above), respectively, in the above expression of U'_t (for each $j = t, \ldots, T$). In addition, in the right-hand derivative each of the events $A_{jt}(y_t)$ is defined with respect to weak inequalities $y_t - D - [t, k) \ge R_k$, $\forall k \in (t, j]$. This also provides an optimality criterion for finding a minimizer R_t of U_t, namely, $U^r_t(R_t) \ge 0$ and $U^l_t(R_t) \le 0$. If the demand distributions are given explicitly, it is usually easy to evaluate the one-sided derivatives of U_t. This suggests the following approach for solving the dynamic program presented above. In each stage, compute R_t such that $0 \in \partial U_t(R_t)$, by considering the respective one-sided derivatives of U_t. In the next subsection, we shall use a similar algorithmic approach, but with respect to an approximate base-stock policy and under the assumption that the only information about the demand distributions is available through a black box.

3.2 Approximate Base-Stock Levels To solve the dynamic program described above requires knowing the explicit demand distributions. However, as mentioned before, in most real-life scenarios these distributions are either not available or are too complicated to work with directly. Instead we shall consider this model under the assumption that the only access to the true demand distribution is through a black box that can generate independent sample-paths from the true demand distributions $D_1, \ldots, D_T$. As in the newsvendor model discussed in Section 2, the goal is to find a policy with expected cost close to the expected cost of an optimal policy that is assumed to have full access to the demand distributions. In particular, we shall describe a sampling-based algorithm that, for each specified accuracy level ϵ and confidence level δ, computes a base-stock policy such that with probability at least $1 - \delta$, the expected cost of the policy is at most $1 + \epsilon$ times the expected cost of an optimal policy. Throughout the paper, we use $R_1, \ldots, R_T$ to denote the *minimal optimal base-stock-level*, i.e., the optimal base-stock policy. That is, for each $t = 1, \ldots, T$, the base-stock level R_t is the smallest minimizer of U_t defined above. Next we provide an overview of the algorithm and its analysis.

An overview of the algorithm and its analysis. First note that our approach departs from the SAA method or the IPA methods discussed in Sections 1 and 2. Instead, it is based on a dynamic programming framework. That is, the base-stock levels of the policy are computed using a backward recursion. In particular, the approximate base-stock level in period t, denoted by $\tilde{R}_t$, is computed based on the previously computed approximate base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$. If $T = 1$, then this reduces to solving the SAA of the single-period newsvendor model, already discussed in Section 2. However, if $T > 1$ and the base-stock levels are approximated recursively, then the issue of convexity needs to be carefully addressed. It is no longer clear whether each subproblem is still convex, and whether base-stock policies are still optimal. More specifically, assume that some (approximate) base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t+1, T]$, not necessarily an optimal one, was already computed in previous stages of the algorithm. Now let $\tilde{U}_t(y_t)$ be the expected cost over $[t, T]$ of a policy that orders up to y_t in period t and then follows the base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over $[t+1, T]$ (as before, expectations are taken with respect to the underlying demand distributions $D_1, \ldots, D_T$). Let $\tilde{V}_t(x_t)$ be the minimum expected cost over $[t, T]$ over all ordering policies in period t, given that the inventory level at the beginning of the period is x_t and that the policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is followed over $[t+1, T]$. Clearly, $\tilde{V}_t(x_t) = \min_{y_t \geq x_t} \tilde{U}_t(y_t)$. The functions $\tilde{U}_t$ and $\tilde{V}_t$ play analogous roles to those of U_t and V_t, respectively, but are defined with respect to $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ instead of $R_{t+1}, \ldots, R_T$. The functions U_t and V_t define a *shadow dynamic program* to the one described above that is based on the functions U_t and V_t. From now on, we will distinguish functions and objects that are defined with respect to the approximate policy $\tilde{R}_1, \ldots, \tilde{R}_T$ by adding the *tilde* sign above them. The convexity of U_t and V_t and the optimality of base-stock policies are heavily based on the optimality of $R_{t+1}, \ldots, R_T$ (using Fact 3.1 above). Since the approximate policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is not necessarily optimal, the functions $\tilde{U}_t$ and $\tilde{V}_t$ might not be convex. Hence, it is possible that no base-stock policy in period t is optimal. In order to keep the subproblem (i.e., the function $\tilde{U}_t$) in each stage tractable, the algorithm is going to maintain (with high probability) an invariant under which the convexity of $\tilde{U}_t$ and $\tilde{V}_t$ and the optimality of base-stock policies are preserved (see Definition 3.2 and Lemma 3.1, where we establish the resulting convexity of the functions $\tilde{U}_t$ and $\tilde{V}_t$). Assuming that $\tilde{U}_t$ and $\tilde{V}_t$ are indeed convex, it would be natural to compute the smallest minimizer of $\tilde{U}_t$, denoted by $\bar{R}_t$. However, this also requires full access to the explicit demand distributions. Instead, the algorithm takes the following approach. In each stage $t = T, \ldots, 1$, the algorithm uses a sampling-based procedure to compute a base-stock level $\tilde{R}_t$ that, with high probability, has two properties. First, the base-stock level $\tilde{R}_t$ is a good approximation of the minimizer $\bar{R}_t$, in that $\tilde{U}_t(\tilde{R}_t)$ is close to the minimum value $\tilde{U}_t(\bar{R}_t)$, i.e., it has a small relative error. Second, $\tilde{R}_t$ is greater or equal than $\bar{R}_t$. It is this latter property that preserves the invariant of the algorithm, and in particular, preserves the convexity of $\tilde{U}_{t-1}$ and $\tilde{V}_{t-1}$ in the next stage.

The justification for this approach is given in Lemma 3.2, where it is shown that the properties of $\tilde{R}_t, \ldots, \tilde{R}_T$ also guarantee that small errors relative to $\tilde{U}_t(\bar{R}_t), \ldots, \tilde{U}_T(\bar{R}_T)$, respectively, accumulate but have impact only on the expected cost over $[t, T]$ and do not propagate to the interval $[1, t)$. Thus, applying this approach recursively leads to a base-stock policy for the entire horizon with expected cost close to the optimal expected cost. Analogous to the newsvendor cost function, the functions $\tilde{U}_1, \ldots, \tilde{U}_T$ also have similar explicit expressions for the one-sided derivatives that are also bounded, and hence can be estimated accurately with samples. However, in order to compute such an $\tilde{R}_t$ in each stage, it is essential to establish an explicit connection between first order information, i.e., information about the value of the one-sided derivatives of $\tilde{U}_t$ at a certain point, and the bounded relative error that this guarantees relative to $\tilde{U}_t(\tilde{R}_t)$. This is done in Lemma 3.3 below, which plays a similar central role to Lemma 2.1 in the previous section. Finally, in Lemma 3.4, Corollaries 3.1 and 3.2, and Lemma 3.5, it is shown how the one-sided derivatives of $\tilde{U}_t$ can be estimated using samples in order to compute an $\tilde{R}_t$ that maintains the two required properties, with high probability.

Next we discuss the invariant of the algorithm that preserves the convexity of the functions $\tilde{U}_t$ and $\tilde{V}_t$ above and the optimality of a base-stock policy in period t. In the case where there exists an optimal ordering policy in period t which is a base-stock policy (i.e., $\tilde{U}_t$ is convex), let $\bar{R}_t = \bar{R}_t | \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ be the smallest minimizer of $\tilde{U}_t$, i.e., the smallest optimal base-stock level in period t, given that the policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is followed in periods $t+1, \ldots, T$. If the optimal ordering policy in period t given $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is not a base-stock policy, we say that $\bar{R}_t$ does not exist. The invariant of the algorithm is given in the next definition.

Definition 3.2 *A base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ for the interval $[t+1, T]$ is called an* upper base-stock

policy *if, for each* $j = t+1, \ldots, T$, $\bar{R}_j$ *exists, and the inequality* $\tilde{R}_j \geq \bar{R}_j$ *holds.*

The algorithm is going to preserve this invariant by computing in each stage $t = T, \ldots, 1$ an $\tilde{R}_t$ such that with high probability, $\tilde{R}_t \geq \bar{R}_t$. In the next two lemmas we shall show several important structural properties of upper base-stock policies. In the first of these lemmas, for each $j = 1, \ldots, T$, we shall show that if an upper base-stock policy $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$ is followed over $[j+1, T]$, then the convexity of the functions $\tilde{U}_j$ and $\tilde{V}_j$ is preserved and there exists an optimal ordering policy in period j which is a base-stock policy.

LEMMA 3.1 *Let* $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ *be an upper base-stock policy over* $[t+1, T]$. *For each* $j = t, \ldots, T$, *let* $\tilde{U}_j(y_j)$ *be the expected cost over* $[j, T]$ *of a policy that orders up to* y_j *in period* j, *and then follows* $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$. *Similarly, let* $\tilde{V}_j(x_j)$ *be the minimum expected cost over* $[j, T]$ *given that at the beginning of period* j *the inventory level is* x_j *and that over* $[j+1, T]$ *the base-stock policy* $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$ *is followed. (Thus,* $\tilde{V}_j(x_j) = \min_{y_j \geq x_j} \tilde{U}_j(y_j)$.*) Then, for each period* $j = t, \ldots, T$,

(i) The functions $\tilde{U}_j$ *and* $\tilde{V}_j$ *are convex and* $\tilde{U}_j$ *attains a minimum.*

(ii) Given that over $[j+1, T]$ *we follow the base-stock policy* $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$, *there exists an optimal ordering policy in period* j *which is a base-stock policy, i.e.,* $\bar{R}_j$ *does exist.*

PROOF. For each $k = t+1, \ldots, T$, let $\tilde{G}_k(x_k)$ be the expected cost of the base-stock policy $\tilde{R}_k, \ldots, \tilde{R}_T$ over the interval $[k, T]$ given that there are x_k units in inventory at the beginning of period k. Thus, for each $t < T$, $\tilde{U}_t(y_t) = C_t(y_t) + E[\tilde{G}_{t+1}(y_t - D_t)]$.

The proof follows by induction on $j = T, \ldots, t$. For $j = T$, observe that $\tilde{U}_T = U_T = C_T$ and $\tilde{V}_T(x_T) = V_T(x_T) = \min_{y_T \geq x_T} \tilde{U}_T(y_T)$, which implies that both $\tilde{U}_T$ and $\tilde{V}_T$ are convex, $\tilde{U}_T$ attains a minimum and $\bar{R}_T = R_T$ is indeed an optimal base-stock policy in period T. In particular, $\bar{R}_T$ is the smallest minimizer of $\tilde{U}_T$. Now assume that the claim is true for $j > t$, i.e., for each of the periods $j, \ldots, T$. In particular, the functions $\tilde{U}_j, \ldots, \tilde{U}_T$ are convex, $\bar{R}_j, \ldots, \bar{R}_T$ are their respective smallest minimizers, and the functions $\tilde{V}_j, \ldots, \tilde{V}_T$ are convex. Consider now the function $\tilde{U}_{j-1}(y_{j-1}) = C_{j-1}(y_{j-1}) + E[\tilde{G}_j(y_{j-1} - D_{j-1})]$. Since C_{j-1} is convex, it is sufficient to show that $\tilde{G}_j$ is convex. By induction, $\bar{R}_j$ is a minimizer of $\tilde{U}_j$ and $\tilde{R}_j \geq \bar{R}_j$. Hence, the function $B_j(x_j)$ can be expressed as $B_j(x_j) = \max\{\tilde{V}_j(x_j), \tilde{U}_j(\tilde{R}_j)\}$, which implies that it is indeed convex, since it is the maximum of two convex functions. It is straightforward to see that $\tilde{U}_{j-1}$ is convex and has a minimizer, where its smallest minimizer is denoted by $\bar{R}_{j-1}$. By Fact 3.1, we conclude that $\tilde{V}_{j-1}(x_{j-1})$ is also convex and that there exists an optimal base-stock policy in period $j-1$ (assuming that the policy $\tilde{R}_j, \ldots \tilde{R}_T$ is followed over $[j, T]$). □

In the next lemma we consider the case where an upper base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t+1, T]$ is known to provide a good solution for that interval. More specifically, for each $j = t+1, \ldots, T$, the expected cost of the base-stock policy $\tilde{R}_j, \ldots, \tilde{R}_T$ is assumed to be close to optimal over the interval $[j+1, T]$, i.e., $\tilde{U}_j(\tilde{R}_j) \leq \alpha_j U_j(R_j)$ for some $\alpha_j \geq 1$. We shall show that this gives rise to a good policy over the entire horizon $[1, T]$.

LEMMA 3.2 *For some* $t < T$, *let* $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ *be an upper base-stock policy for the interval* $[t+1, T]$ *and let* $R_t, \ldots, R_T$ *be the minimal optimal base-stock policy over* $[t, T]$. *Consider the function* $\tilde{U}_t(y_t)$ *and its smallest minimizer* $\bar{R}_t$. *Furthermore, assume that for each* $j = t+1, \ldots, T$, *the cost of the base-stock policy* $\tilde{R}_j, \ldots, \tilde{R}_T$ *is at most* β_j *times the optimal expected cost over that interval (where* $\beta_j \geq 1$*), i.e.,* $\tilde{U}_j(\tilde{R}_j) \leq \beta_j U_j(R_j)$. *(Recall that* U_j *is defined with respect to the optimal base-stock policy.) Let* $\beta = \max_j \beta_j$. *Then the expected cost of the approximate (upper) base-stock policy* $\bar{R}_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ *over the interval* $[t, T]$ *is at most* β *times the expected cost of an optimal base-stock policy over that interval.*

PROOF. Recall that for each $j = t+1, \ldots, T$ the function $U_j(y_j)$ is defined with respect to the optimal base-stock levels $R_{j+1}, \ldots, R_T$, and the function $\tilde{U}_j(y_j)$ is defined with respect to the base-stock levels $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$.

Suppose first that under the assumptions in the lemma, the following *structural claim* is true. For each $j > t$, consider the interval $[j, T]$. Then the expected cost of the policy $R_j, \tilde{R}_{j+1}, \ldots, \tilde{R}_T$ over that interval is at most β times the respective (optimal) expected cost of the policy $R_j, R_{j+1}, \ldots, R_T$.

Now consider the *modified* base-stock policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$. This policy consists of the optimal base-stock level R_t in period t, followed by $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the rest of the interval. We shall first show that the expected cost of this policy over the interval $[t, T]$ is at most β times the optimal expected cost over that interval. However, given that the policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is followed over $[t, T]$ the optimal base-stock level in period t is $\bar{R}_t$. This implies that the policy $\bar{R}_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is even better than $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$, from which the proof of the lemma follows.

Focusing on the policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$, we note that in period t this policy is identical to the optimal base-stock policy $R_t, \ldots, R_T$. It is clear that the two policies incur the same cost in period t (for each possible realization of D_t). Moreover, at the beginning of period $t+1$ both policies will have the same inventory level X_{t+1}. Let $\tilde{t} \geq t+1$ be the first period after t in which either the modified or the optimal policy placed an order. Observe that $\tilde{t}$ is a random variable. Over the (possibly empty) interval $[t+1, \tilde{t})$ neither the modified policy nor the optimal policy ordered. This can happen only if, for each $j \in [t+1, \tilde{t})$ the inventory level of the two policies is higher than the respective base-stock levels R_j and $\tilde{R}_j$. Moreover, this implies that over that interval the two policies have exactly the same inventory level, and therefore, they incur exactly the same cost. Now in the period $\tilde{t}$ exactly one of two cases applies. If the modified policy places an order, then by our assumption, it is clear that the expected cost that the modified policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ incurs over the interval $[\tilde{t}, T]$ is at most β times the expected cost of the optimal base-stock policy $R_t, \ldots, R_T$ over that interval. Now consider the case in which the optimal policy places an order in $\tilde{t}$ and the modified policy does not. Recall that at the beginning of period $\tilde{t}$, the inventory level of both policies is the same and is equal to $X_{t+1} - D_{[t+1,\tilde{t})}$. Thus, this event can occur only if

$$\tilde{R}_{\tilde{t}} \leq X_{t+1} - D_{[t+1,\tilde{t})} < R_{\tilde{t}}.$$

(Indeed the optimal policy will order up to $R_{\tilde{t}}$ and the modified policy will not). Let $\bar{R}_{\tilde{t}}$ be the smallest minimizer of $\tilde{U}_{\tilde{t}}$, i.e., the best order up-to level in $\tilde{t}$ given that the policy $\tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$ is followed over $[\tilde{t}+1, T]$. By the assumption that $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is an upper base-stock policy it follows that $\tilde{R}_{\tilde{t}} \geq \bar{R}_{\tilde{t}}$. (Since $\tilde{R}_T \geq \bar{R}_T = R_T$ we conclude that if the above case applies, then, with probability 1, $\tilde{t} < T$.) In particular, over the interval $[\tilde{t}, T]$, the optimal policy executed the policy $R_{\tilde{t}}, R_{\tilde{t}+1}, \ldots, R_T$. By the structural claim above, we also know that the expected cost of policy $R_{\tilde{t}}, \tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$ over $[\tilde{t}, T]$ is at most β times the expected cost of the optimal policy $R_{\tilde{t}}, R_{\tilde{t}+1}, \ldots, R_T$ over that interval. Observe that $X_{t+1} - D_{[t+1,\tilde{t})}$ falls between $\bar{R}_{\tilde{t}}$ and $R_{\tilde{t}}$ (see the inequality above and the fact that $\tilde{R}_{\tilde{t}} \geq \bar{R}_{\tilde{t}}$). Because $\tilde{U}_{\tilde{t}}$ is convex, this implies that the expected cost of the modified policy over the interval $[\tilde{t}, T]$ is no greater than the expected cost of the policy $R_{\tilde{t}}, \tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$, i.e., at most β times the optimal expected cost over that interval. (Observe that the modified policy executed $X_{t+1} - D_{[t+1,\tilde{t})}, \tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$.)

It is left to show that the structural claim is indeed valid. It is readily verified that the claim is true for $j = T$ and $j = T-1$. The proof of the induction step is done by arguments identical to the arguments used above. The proof then follows. □

Consider now an upper base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ such that for each $t = 1, \ldots, T$, the base-stock level $\tilde{R}_t$ is a good approximation of $\bar{R}_t = \bar{R}_t | \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (by Lemma 3.1 above we know that $\bar{R}_t$ is well-defined). More specifically, for each $t = 1, \ldots, T$, we have $\tilde{R}_t \geq \bar{R}_t$ and $\tilde{U}_t(\tilde{R}_t) \leq (1+\epsilon_t)\tilde{U}_t(\bar{R}_t)$ (for some specified $0 \leq \epsilon_t$), where $\tilde{U}_t(y_t)$ is defined with respect to $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (see above) and $\bar{R}_t$ is its smallest minimizer. Using Lemmas 3.1 and 3.2, it is straightforward to verify (by backward induction on the periods) that, for each $s = 1, \ldots, T$, the expected cost of the base-stock policy $\tilde{R}_s, \ldots, \tilde{R}_T$ is at most $\prod_{j=s}^{T}(1+\epsilon_j)$ times the optimal expected cost over the interval $[s, T]$.

For $s = T$, the claim is trivially true. Now assume it is true for some $s > 1$. Applying Lemma 3.2 above with $t = s-1$ and $\beta = \prod_{j=s}^{T}(1+\epsilon_j)$ we conclude that the policy $\bar{R}_{s-1}, \tilde{R}_s, \ldots, \tilde{R}_T$ has expected cost at most $\beta = \prod_{j=s}^{T}(1+\epsilon_j)$ times the optimal expected cost over the interval $[s-1, T]$. Now by the definition of $\tilde{R}_{s-1}$, we conclude that the policy $\tilde{R}_{s-1}, \tilde{R}_s, \ldots, \tilde{R}_T$ has expected cost at most $(1+\epsilon_{s-1})\beta$ times the optimal expected cost over $[s-1, T]$, from which the claim follows. In particular, this implies that the expected cost of the base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ over the entire horizon $[1, T]$ is at most $\prod_{t=1}^{T}(1+\epsilon_t)$ times the optimal expected cost. In other words, the properties of an upper base-stock policy guarantee that errors over the interval $[t+1, T]$ do not propagate to the interval $[1, t]$.

In order to construct such an upper base-stock policy we need to compute a base-stock level $\tilde{R}_t$, in each stage $t = T, \ldots, 1$, with the following two properties. To preserve the invariant of an upper base-

stock policy, it is required that $\tilde{R}_t \geq \bar{R}_t$. In addition, $\tilde{R}_t$ is required to have a small relative error with respect to $\tilde{U}_t(y_t)$ and its minimizer $\bar{R}_t$, i.e., $\tilde{U}_t(\tilde{R}_t) \leq (1+\epsilon_t)\tilde{U}_t(\bar{R}_t)$. Recall that, if the invariant of an upper base-stock policy is preserved, the function $\tilde{U}_t$ is convex with (one-sided) derivatives as given in (7) above. This suggests the same approach as before, i.e., use first-order information in order to find a point with objective value close to optimal. However, unlike the newsvendor cost function, the minimizer of $\tilde{U}_t$ is not a well-defined quantile of the distribution of D_t and it is less obvious how to establish a connection between the value of the one-sided derivatives of $\tilde{U}_t$ at some point y and the relative error of that point with respect to the minimum expected cost $\tilde{U}_t(\bar{R}_t)$. This is established in the next lemma which has an analogous central role to that of Lemma 2.1 above. Observe that, for each $t = 1, \ldots, T$, the function $\tilde{U}_t$ is bounded from below by the newsvendor cost C_t, i.e., for each y, we have $\tilde{U}_t(y) \geq C_t(y)$. Now $C_t(y) = E[h_t(y-D_t)^+ + b_t(D_t-y)^+]$ and for each fixed y the function $h_t(y-D_t)^+ + b_t(D_t-y)^+$ is convex in D_t. Applying Jensen's inequality we conclude that the inequality $C_t(y) \geq h_t(y-E[D_t])^+ + b_t(E[D_t]-y)^+$ holds for each y. For each $t = 1, \ldots, T$, the function $h_t(y - E[D_t])^+ + b_t(E[D_t] - y)^+$ is piecewise linear and convex with a minimum attained at $y = E[D_t]$ and equal to zero. Moreover, it provides a lower bound on $\tilde{U}_t(y)$. This structural property of the functions $\tilde{U}_1, \ldots, \tilde{U}_T$ can be used to establish an explicit connection between first order information and relative errors. The next lemma is specialized to the specific setting of the functions $\tilde{U}_1, \ldots, \tilde{U}_T$. In Section 5, we present a general version of this lemma that is valid in the multi-dimensional case. We believe that this structural lemma will have additional applications in different settings.

Before we state and prove the lemma, we introduce the following definition.

Definition 3.3 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be convex and finite. A point $y \in \mathbb{R}^m$ is called an $\alpha-$point of f if there exists a subgradient r of the function f at y with Euclidean norm less than α, i.e., there exists $r \in \partial f(y)$ with $||r||_2 \leq \alpha$.*

LEMMA 3.3 *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be convex and finite with a minimum at y^*, i.e., $f(y^*) \leq f(y)$ for each $y \in \mathbb{R}$. Suppose that $\bar{f}(y) = h(y-d)^+ + b(d-y)^+$ (where $h,\ b > 0$) is a convex piecewise linear function with minimum equal to 0 at d, such that $f(y) \geq \bar{f}(y)$ for each $y \in \mathbb{R}$. Let $0 < \epsilon \leq 1$ be the specified accuracy level, and let $\alpha = \frac{\epsilon}{3}\min(b,h)$. If $\hat{y}$ is an $\alpha-$point of f, then $f(\hat{y}) \leq (1+\epsilon)f(y^*)$.*

PROOF. Let $\lambda = \min(b,h)$. Since $\hat{y}$ is an $\alpha-$point of f, we conclude that there exists some $r \in \partial f(\hat{y})$ with $|r| \leq \alpha$, and by the definition of a subgradient, $f(\hat{y}) - f(y^*) \leq \alpha|\hat{y} - y^*|$. Now let $d_1 \leq d \leq d_2$ be the two points where $\bar{f}$ takes the minimum value of f, i.e., $\bar{f}(d_1) = \bar{f}(d_2) = f(y^*)$. Let $L_1 = d - d_1$ and $L_2 = d_2 - d$. Clearly, $f(y^*) = bL_1 = hL_2$, which implies that $f(y^*) \geq \frac{\lambda}{2}(L_1 + L_2)$. Moreover, since $f(y) \geq \bar{f}(y)$ for each y, we conclude that $y^* \in [d_1, d_2]$.

Now consider the point $\hat{y}$. Suppose first that $\hat{y} \in [d_1, d_2]$. This implies that $|\hat{y} - y^*| \leq L_1 + L_2$. We now get that

$$f(\hat{y}) - f(y^*) \leq \alpha|\hat{y} - y^*| \leq \alpha(L_1 + L_2) = \frac{\epsilon}{3}\lambda(L_1 + L_2) \leq \frac{2}{3}\epsilon f(y^*).$$

The equality is a substitution of $\alpha = \frac{\epsilon}{3}\lambda$. The claim then follows.

Now assume that $\hat{y} \notin [d_1, d_2]$. Without loss of generality, assume $\hat{y} > d_2$ (a symmetric proof applies if $\hat{y} < d_1$). Let $x = \hat{y} - d_2$. Since f is convex, it is clear that d_2 is also an $\alpha-$point of f. By the same arguments used above we conclude that $f(d_2) - f(y^*) \leq \frac{2}{3}\epsilon f(y^*)$ and that $f(\hat{y}) - f(d_2) \leq \alpha|\hat{y} - d_2| = \alpha x$. This implies that

$$f(\hat{y}) - f(y^*) = f(\hat{y}) - f(d_2) + f(d_2) - f(y^*) \leq \alpha x + \frac{2}{3}\epsilon f(y^*).$$

It is then sufficient to show that $\alpha x \leq \frac{1}{3}\epsilon f(y^*)$. We first bound x from above. Now $\bar{f}(\hat{y}) = f(y^*) + hx \geq f(y^*) + \lambda x$. In addition, $f(\hat{y}) \leq f(d_2) + \alpha x \leq (1 + \frac{2}{3}\epsilon)f(y^*) + \alpha x$. However, $\bar{f}(\hat{y}) \leq f(\hat{y})$. We conclude that

$$x \leq \frac{2}{3}\frac{\epsilon f(y^*)}{\lambda - \alpha} = \frac{2}{\lambda}\frac{\epsilon f(y^*)}{(3-\epsilon)} \leq \frac{\epsilon f(y^*)}{\lambda}.$$

The equality is the substitution of $\alpha = \frac{\epsilon}{3}\lambda$. The last inequality is because $\epsilon \leq 1$. However, this implies that $\alpha x \leq \frac{\epsilon f(y^*)}{3}$, from which the claim follows. □

Lemma 3.3 above establishes an explicit connection between $\alpha-$points of the functions $\tilde{U}_t$ (for $t = 1,\ldots,T$) and the relative error they guarantee. We note that slightly tighter bounds can be proven using somewhat more involved algebra.

Since the demand distributions are not available, it is left to show how the one-sided derivatives of these functions can be estimated with high accuracy and high confidence probability using random samples.

We next derive expressions of the one-sided derivatives of $\tilde{U}_{j-1}$, similar to (6) above. Note that

$$\tilde{U}_{j-1}(y_{j-1}) = C_{j-1}(y_{j-1}) + E[\mathbb{1}(y_{j-1} - D_{j-1} \le \tilde{R}_j)\tilde{U}_j(\tilde{R}_j) + \mathbb{1}(y_{j-1} - D_{j-1} > \tilde{R}_j)\tilde{U}_j(y_{j-1} - D_{j-1})].$$

It is readily verified that $\tilde{U}_{j-1}$ is a continuous function of y_{j-1}. If we take the right-hand derivative and apply this process recursively (similar to (4)-(6) above), we get that the right-hand derivative of $\tilde{U}_{j-1}$ is

$$\tilde{U}^r_{j-1}(y_{j-1}) = C^r_{j-1}(y_{j-1}) + E[\sum_{k=j}^{T} \mathbb{1}(\tilde{A}_{k,j-1}(y_{j-1}))C^r_k(y_{j-1} - D_{[j-1,k)})]. \tag{7}$$

The events $\tilde{A}_{k,j-1}(y_{j-1})$ are defined with respect to $\tilde{R}_j,\ldots,\tilde{R}_T$ instead of $R_j,\ldots,R_T$ (see (6) above). We get a similar expression for the left-hand derivative of $\tilde{U}_{j-1}$ by replacing C^r_k by C^l_k for each $k = j-1,\ldots,T$. As in the case of U^r_t, the events $\tilde{A}_{jt}$ in $\tilde{U}^r_t$ are defined with respect to weak inequalities. It is easy to verify that the right-hand and left-hand derivatives of the function $\tilde{U}_t$ are bounded within the interval $[-\sum_{j=t}^{T} b_t, \sum_{j=t}^{T} h_t]$. The next lemma shows that for each y, there exist explicit computable bounded random variables with expectations equal to $\tilde{U}^r_t(y)$ and $\tilde{U}^l_t(y)$, respectively. This implies that the right-hand and the left-hand derivatives of the function $\tilde{U}_t$ can be evaluated stochastically with high accuracy and high probability (using the Hoeffding inequality).

LEMMA 3.4 *For each $t = 1,\ldots,T-1$, $j > t$ and y_t consider the random variable $\tilde{M}^r_{tj}(y_t) = \mathbb{1}(\tilde{A}_{jt}(y_t))(-b_j + (h_j+b_j)\mathbb{1}(D_j \le y_t - D_{[t,j)}))$. Then $E[\tilde{M}^r_{tj}(y)] = E[\mathbb{1}(\tilde{A}_{jt}(y_t))C^r_j(y_t - D_{[t,j)})]$.*

PROOF. First note that the expectations above are taken with respect to the underlying demand distributions $D_t,\ldots,D_T$. In particular, the base-stock levels $\tilde{R}_{t+1},\ldots,\tilde{R}_T$ that define the events $\tilde{A}_{jt}$ are assumed to be known deterministically. Using conditional expectations we can write

$$\begin{aligned}
E[\tilde{M}^r_{tj}(y)] &= E[E[\tilde{M}^r_{tj}(y)|D_{[t,j)}]] \\
&= E[E[\mathbb{1}(\tilde{A}_{jt}(y_t))(-b_j + (h_j+b_j)\mathbb{1}(D_j \le y_t - D_{[t,j)}))|D_{[t,j)}]] \\
&= E[\mathbb{1}(\tilde{A}_{jt}(y_t))E[-b_j + (h_j+b_j)\mathbb{1}(D_j \le y_t - D_{[t,j)})|D_{[t,j)}]] \\
&= E[\mathbb{1}(\tilde{A}_{jt}(y_t))C^r_j(y_t - D_{[t,j)})].
\end{aligned}$$

We condition on $D_{[t,j)}$, so the indicator $\mathbb{1}(\tilde{A}_{jt}(y_t))$ is known deterministically. In the last equality we use the definition of C^r_j and uncondition. The claim then follows. □

In a similar way, we define the random variable $\tilde{M}^l_{tj}(y_t)$ by replacing the indicator $\mathbb{1}(D_j \le y_t - D_{[t,j)})$ in the definition of $\tilde{M}^r_{tj}$ by the indicator $\mathbb{1}(D_j < y_t - D_{[t,j)})$, for each $1 \le t < j \le T$ and y_t.

Considering (7), we immediately get the following corollary.

COROLLARY 3.1 *For each $t = 1,\ldots,T$ and y_t, the right-hand derivative of $\tilde{U}_t$ is given by $\tilde{U}^r_t(y_t) = E[-b_t + (h_t+b_t)\mathbb{1}(D_t \le y_t) + \sum_{j=t+1}^{T} \tilde{M}^r_{tj}(y_t)]$.*

COROLLARY 3.2 *For each $t = 1,\ldots,T$ and y_t, the left-hand derivative of $\tilde{U}_t$ is given by $\tilde{U}^l_t(y_t) = E[-b_t + (h_t+b_t)\mathbb{1}(D_t < y_t) + \sum_{j=t+1}^{T} \tilde{M}^l_{tj}(y_t)]$.*

Lemma 3.4 and Corollaries 3.1 and 3.2 imply that we can estimate the right-hand and left-hand derivatives of the functions $\tilde{U}_1,\ldots,\tilde{U}_T$ with a bounded number of samples. For each $t = 1,\ldots,T$, take N_t samples from the demand distributions $D_t,\ldots,D_T$ that are *independent* of $\tilde{R}_{t+1},\ldots,\tilde{R}_T$, i.e., independent of the samples taken in previous stages. Let $\hat{\tilde{U}}^r_t$ and $\hat{\tilde{U}}^l_t$ be the respective right-hand and left-hand sampling-based estimators of the one-sided derivatives of $\tilde{U}_t$. Note that the base-stock levels $\tilde{R}_{t+1},\ldots,\tilde{R}_T$ have already been computed based on independent samples from the previous stages. Thus,

at stage t the function $\tilde{U}_t$ and its one-sided derivatives are considered to be deterministic. On the other hand, $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ are random and determined by the specific N_t samples that define them.

To simplify the notation, we will assume from now on that $b_t = b > 0$ and $h_t = h > 0$, for each $t = 1, \ldots, T$. To evaluate the right-hand and left-hand derivatives of $\tilde{U}_t$ at a certain point y_t, consider each sample path $d_t^i, \ldots, d_T^i$ (for $i = 1, \ldots, N_t$), evaluate the random variables $-b + (h+b)\mathbb{1}(D_t \leq y_t) + \sum_{j=t+1}^T \tilde{M}_{tj}^r(y_t)$ and $-b + (h+b)\mathbb{1}(D_t < y_t) + \sum_{j=t+1}^T \tilde{M}_{tj}^l(y_t)$, respectively, and then average over the N_t samples. Note that each of the variables $\tilde{M}_{tj}^r$ (respectively, $\tilde{M}_{tj}^l$) can take values only within $[-b, h]$, so $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ are also bounded. By arguments similar to Lemma 2.2 (using the Hoeffding inequality), it easy to compute the number of samples N_t required to guarantee that the minimizer is an α_t-point of the function $\tilde{U}_t$ with probability at least $1 - \delta_t$ (for specified accuracy and confidence levels). However, as we have already seen, in the multiperiod setting it is also essential to preserve the invariant $\tilde{R}_t \geq \bar{R}_t$ to ensure that the problem in the next stage is still convex. That is, we wish to find an α_t-point $\tilde{R}_t$ of $\tilde{U}_t$ but with the additional property that $\tilde{R}_t \geq \bar{R}_t$, where $\bar{R}_t$ is the smallest minimizer of $\tilde{U}_t$.

We compute $\tilde{R}_t$ in the following way. Given N_t samples, let $\tilde{R}_t$ be the minimal point with sample right-hand derivative at least $\frac{\alpha_t}{2}$ (i.e., $\hat{\tilde{U}}_t^r(\tilde{R}_t) \geq \frac{\alpha_t}{2}$). That is, $\tilde{R}_t := \inf\{y : \hat{\tilde{U}}_t^r(y) \geq \frac{\alpha_t}{2}\}$. First observe that $\tilde{R}_t$ is well-defined for each $0 < \alpha_t \leq 2h(T-t+1)$, since the one-sided derivatives of the function $\tilde{U}_t$ are bounded within the interval $[-b(T-t+1), h(T-t+1)]$. By the definition of $\tilde{R}_t$, it is also clear that $\hat{\tilde{U}}_t^l(\tilde{R}_t) < \frac{\alpha_t}{2}$. Lemma 3.5 analyzes the required number of samples N_t to guarantee that, with probability at least $1-\delta_t$, the point $\tilde{R}_t$ is both an α_t-point of $\tilde{U}_t$ and satisfies $\tilde{R}_t \geq \bar{R}_t$. For the proof of this Lemma we use the Hoeffding inequality; below we bring a variant of this well-known result, which is relevant for the proof.

THEOREM 3.4 *(Hoeffding Inequality [18]). Let $X^1, \ldots, X^N$ be i.i.d. random variables such that $X^1 \in [\beta_1, \beta_2]$ (i.e., $Pr(X^1 \in [\beta_1, \beta_2]) = 1$) for some $\beta_1 < \beta_2$. Then, for each $\alpha > 0$, we have*

(i) $Pr(\frac{1}{N}\sum_{i=1}^N X^i - E[X^1] \geq \alpha) \leq e^{-2\alpha^2 N/(\beta_2-\beta_1)^2}$.

(ii) $Pr(|\frac{1}{N}\sum_{i=1}^N X^i - E[X^1]| \geq \alpha) \leq 2\exp^{-2\alpha^2 N/(\beta_2-\beta_1)^2}$.

LEMMA 3.5 *Consider some stage t during the execution of the algorithm and let $0 < \alpha_t$ and $0 < \delta_t < 1$. Suppose we generate $N_t \geq 2((b+h)(T-t+1))^2 \frac{1}{\alpha_t^2}\log(\frac{2}{\delta_t})$ independent samples of the demands $D_t, \ldots, D_T$ that are also independent of the base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$, and use them to compute $\tilde{R}_t = \inf\{y : \hat{U}_t^r(y) \geq \frac{\alpha_t}{2}\}$. Then, with probability at least $1-\delta_t$, the base-stock level $\tilde{R}_t$ is an α_t-point of $\tilde{U}_t$, and $\tilde{R}_t \geq \bar{R}_t$.*

PROOF. First note that $\tilde{R}_t$ is a random variable, a function of the specific N_t samples, and that $\bar{R}_t$ is a deterministic quantity (the smallest minimizer of the function $\tilde{U}_t$ that is assumed to be known deterministically). Observe that the event $[\tilde{R}_t \geq \bar{R}_t] \cap [\tilde{R}_t$ is an α_t- point$]$ contains the event $[\tilde{U}_t^r(\tilde{R}_t) \geq 0] \cap [\tilde{U}_t^l(\tilde{R}_t) \leq \alpha_t]$. It is then sufficient to show that each of the events $[U_t^r(\tilde{R}_t) < 0]$ and $[U_t^l(\tilde{R}_t) > \alpha_t]$ has probability at most $\frac{\delta_t}{2}$.

By the optimality and minimality of $\bar{R}_t$, we know that $\bar{R}_t = \inf\{y : \tilde{U}_t^r(y) \geq 0\}$ and that $\tilde{U}_t^r(\bar{R}_t) \geq 0$. This implies that the event $[\tilde{U}_t^r(\tilde{R}_t) < 0]$ is equivalent to the event $[\tilde{R}_t \in (-\infty, \bar{R}_t)]$. For a decreasing nonnegative sequence of numbers $\tau_k \downarrow 0$ define the monotone increasing sequence of events $B_k = [\hat{\tilde{U}}_t^r(\bar{R}_t - \tau_k) \geq \frac{\alpha_t}{2}]$ (note that $\hat{\tilde{U}}_t^r(y)$ is a random variable dependent on the specific N_t samples). By the monotonicity of $\hat{\tilde{U}}_t^r$, it is readily verified that $B_k \uparrow \bar{B}$ (where $\bar{B}$ is the limit set), and $[\tilde{R}_t \in (-\infty, \bar{R}_t)] \subseteq \bar{B}$. However, by the Hoeffding inequality (applied to the random samples of $-b + (h+b)\mathbb{1}(D_t \leq \bar{R}_t - \tau_k) + \sum_{j=t+1}^T \tilde{M}_{tj}^r(\bar{R}_t - \tau_k)$ defined above) and the specific choice of N_t, we conclude that, for each k, we have $Pr(B_k) \leq \frac{\delta_t}{2}$, which implies that $Pr([\tilde{R}_t \in (-\infty, \bar{R}_t)]) \leq Pr(\bar{B}) \leq \frac{\delta_t}{2}$.

Now consider the function $\tilde{U}_t$, and let q be the maximal point with left-hand derivative at most α_t, i.e., $q = \sup\{y : \tilde{U}_t^l(y) \leq \alpha_t\}$. Since $\tilde{U}_t^l$ is left continuous, we conclude that $\tilde{U}^l(q) \leq \alpha_t$. This implies that the event $[\tilde{U}_t^l(\tilde{R}_t) > \alpha_t]$ is equivalent to the event $[\tilde{R}_t \in (q, \infty)]$. Define the monotone increasing sequence

of events $L_k = [\hat{\tilde{U}}_t^l(q+\tau_k) \leq \frac{\alpha_t}{2}]$. By the monotonicity of $\hat{\tilde{U}}_t^l$ it is clear that $L_k \uparrow \bar{L}$ (where $\bar{L}$ is the limit set), and that $[\tilde{U}_t^l(\tilde{R}_t) > \alpha_t] \subseteq \bar{L}$. Using the Hoeffding inequality (now applied to the random samples of $-b+(h+b)\mathbb{1}(D_t \leq \bar{R}_t - \tau_k) + \sum_{j=t+1}^T \tilde{M}_{tj}^l(\bar{R}_t - \tau_k)$) and the choice of N_t, we conclude that for each k, $Pr(L_k) \leq \frac{\delta_t}{2}$. This implies that $Pr([\tilde{U}_t^l(\tilde{R}_t) > \alpha_t] \leq Pr(\bar{L}) \leq \frac{\delta_t}{2}$. It is now clear that $\tilde{R}_t \geq \bar{R}_t$ and that $\tilde{R}_t$ is and α_t-point with probability at least $1-\delta_t$. □

Note that it is relatively easy to compute $\tilde{R}_t$ above. In particular, it is readily verified that the functions $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ change values in at most $(2(T-t)+1)N_t$ distinct points y_t. This and other properties enable us to compute $\tilde{R}_t$ in relatively efficient ways.

3.3 An Algorithm Next we shall provide a detailed description of the algorithm and a complete analysis of the number of samples required. For ease of notation we will assume that $h_t = h > 0$ and $b_t = b > 0$ for each $t = 1, \ldots, T$.

For a specified accuracy level ϵ (where $0 < \epsilon \leq 1$) and confidence level δ (where $0 < \delta < 1$), let $\epsilon_t = \frac{\epsilon}{2T}$ and $\delta_t = \frac{\delta}{T}$. For each $t = 1, \ldots, T$, let $\alpha_t = \frac{\epsilon_t}{3}\min(b,h)$ and $\delta_t = \delta_t$. The algorithm computes a base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ in the following way. In stage $t = T, \ldots, 1$, consider the function $\tilde{U}_t$ defined above with respect to the previously computed base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (where $\tilde{U}_T = C_T$). Use the black box to generate $N_t = 2((b+h)(T-t+1))^2 \frac{1}{\alpha_t^2}\log(\frac{2}{\delta_t})$ independent samples of the demands $D_t, \ldots, D_T$ and compute $\tilde{R}_t = \inf\{y : \hat{\tilde{U}}_t^r(y) \geq \frac{\alpha_t}{2}\}$. Note that in each stage, the algorithm is using an *additional* N_t samples that are independent of the samples used in previous stages. In the next theorem we show that the algorithm computes a base-stock policy that satisfies the required accuracy and confidence levels.

THEOREM 3.5 *For each specified accuracy level ϵ (where $0 < \epsilon \leq 1$) and confidence level δ (where $0 < \delta < 1$), the algorithm computes a base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ such that with probability at least $1-\delta$, the expected cost of the policy is at most $1+\epsilon$ times the optimal expected cost.*

PROOF. For each $t = 1, \ldots, T$, let I_t be the event that $\tilde{R}_t, \ldots, \tilde{R}_T$ is an upper base-stock policy, and that for each $j = t, \ldots, T$, $\tilde{R}_j$ is α_j-point of $\tilde{U}_j$. In particular, by the choice of α_j and Lemma 3.3, $\tilde{U}_j(\tilde{R}_j) \leq (1+\epsilon_j)\tilde{U}_j(\bar{R}_j)$, where $\bar{R}_j$ is the smallest minimizer of $\tilde{U}_j$.

Lemma 3.2 implies that for $\epsilon \leq \frac{1}{2}$, if I_1 occurs then $\tilde{R}_1, \ldots, \tilde{R}_T$ is an upper base-stock policy with expected cost at most $\prod_{t=1}^T(1+\epsilon_t) = (1+\frac{\epsilon}{2T})^T \leq 1+\epsilon$ times the optimal expected cost. It is then sufficient to show that I_1 occurs with probability at least $1-\delta$.

Clearly, $I_T \supseteq I_{T-1} \supseteq \ldots \supseteq I_1$. This implies that $Pr(I_1) = Pr(\cap_{t=1}^T I_t)$. It is then sufficient to show that $Pr([\cap_{t=1}^T I_t]^C) \leq \delta$.

We first show that for each $t = 1, \ldots, T$ the event I_t occurs with positive probability. The proof is done by induction on $t = T, \ldots, 1$. For $t = T$ the claim follows trivially. Now assume that the claim is true for some $1 < t \leq T$ and consider the event I_{t-1}. We then have

$$Pr(I_{t-1}) = Pr(I_t \cap I_{t-1}) = Pr(I_t)Pr(I_{t-1}|I_t).$$

By induction we know that $Pr(I_t) > 0$, so conditioning on I_t is well-defined. Since the samples in each stage are independent of each other, Lemma 3.5 and the choice of N_t implies that $Pr(I_{t-1}|I_t) \geq 1-\delta_t > 0$. The claim then follows.

Now observe that we can write $(\cap_{t=1}^T I_t)^C$ as

$$[\cap_{t=1}^T I_t]^C = I_T^C \cup [I_T \cap I_{T-1}^C] \cup [I_{T-1} \cap I_{T-2}^C] \cup \ldots \cup [I_2 \cap I_1^C].$$

However, given our choice of N_t and the fact that the samples in each stage are independent of each other, for each $t = 1, \ldots, T$, Lemma 3.5 implies that $Pr(I_T^C) \leq \delta_T$ and $Pr(I_t \cap I_{t-1}^C) = Pr(I_t)Pr(I_{t-1}^C|I_t) \leq \delta_{t-1}$. We conclude that $Pr([\cap_{t=1}^T I_t]^C) \leq \sum_{t=1}^T \delta_t$, which implies that $Pr(I_1) \geq 1 - \sum_{t=1}^T \delta_t = 1-\delta$ as required. The proof then follows. □

The next corollary provides upper bounds on the total number of samples needed from each of the random variables $D_1, \ldots, D_T$, denoted by $\mathcal{N}_t$.

COROLLARY 3.3 *For each $t = 1, \ldots, T$, specified accuracy level $0 < \epsilon \leq \frac{1}{2}$ and confidence level $0 < \delta < 1$, the algorithm requires at most $\mathcal{N}_t$ independent samples of D_t, where*

$$\mathcal{N}_t \geq 72\frac{T^2}{\epsilon^2}(\frac{\min(b,h)}{h+b})^{-2}\log(\frac{2T}{\delta})\sum_{j=1}^{t}(T-j+1)^2.$$

Observe that the number of samples required is increasing in the periods. In particular, it is of order $O(T^4)$ for the first period and increasing to order of $O(T^5)$ in the last period. The bounds do not depend on the specific demand distributions but do depend on the square of the reciprocal of $\frac{\min(b,h)}{b+h}$.

We note that the algorithm can be applied in the presence of a positive lead time, per-unit ordering costs and a discount factor over time. The exact dynamic program described above can be extended in a straightforward way to capture these features in a way that preserves the convexity of the problem (see [47] for details). Similarly, the shadow dynamic program can still be used to construct an approximate base-stock policy with the same properties described above. Moreover, the bounds on the number of required samples stay very similar to the bounds established above.

4. Approximating Myopic Policies In many cases, finding optimal policies can be computationally demanding, regardless of whether we have access to the demand distributions or not. As a result, researchers and practitioners have paid attention to *myopic policies*. In a myopic policy we aim, in each period $t = 1, \ldots, T$, to minimize the expected cost (the newsvendor cost) in that period, ignoring the future costs. This provides what is called a *myopic base-stock policy*. As we have already mentioned, myopic policies may not be optimal in general. However, in many cases the myopic policy performs well, and in some cases it is even optimal. In this section, we shall describe a simple and very efficient sampling-based procedure that computes a policy that, with high specified confidence probability, has expected cost very close to the expected cost of the myopic policy. In particular, if a myopic policy is optimal then the expected cost of the approximate policy is close to optimal. We let $R_1^m, \ldots, R_T^m$ denote the *minimal myopic policy*, where for each $t = 1, \ldots, T$, the base-stock level R_t^m is the smallest minimizer of $C_t(y)$ the newsvendor cost in that period.

The sampling-based procedure is based on solving the newsvendor problems in each one of the periods independently. Consider each of the functions $C_1, \ldots, C_T$ and find, for each one of them, an approximate minimizer by means of solving the SAA counterpart. Let $\tilde{R}_t^m$ be the approximate solution in period t. In order to guarantee that the approximate policy has expected cost close the the expected myopic cost, it might not be sufficient to simply take $\tilde{R}_t^m$ to be the minimizer of the corresponding SAA model as discussed in Section 2. The problem is that if we approximate the exact myopic base-stock level from above, i.e., if we get $\tilde{R}_t^m \geq R_t^m$, we might impact the inventory level in the next period in a way that will incur high costs. In turn, we will approximate R_t^m from below. More precisely, we will compute $\tilde{R}_t^m$, for each period t, that is an $\epsilon-$point with respect to C_t and is no greater than R_t^m, with high probability.

The procedure is symmetric to the computation of $\tilde{R}_t$ in Section 3 above. Given N_t samples, consider the SAA counterpart and focus on the one-sided derivatives $\hat{C}_t^r$ and $\hat{C}_t^l$. We will compute $\tilde{R}_t^m$ as the maximum point with sample left-hand derivative value at most $-\frac{\epsilon_t}{2}$ (where $0 < \epsilon_t \leq b_t$, i.e., $\tilde{R}_t^m = \sup\{y : \hat{C}_t^l(y) \leq -\frac{\epsilon_t}{2}\}$). By a proof symmetric to the one of Lemma 3.5 above, we get the following lemma.

LEMMA 4.1 *For each $t = 1, \ldots, T$, consider specified $0 < \epsilon_t \leq b_t$ and $0 < \delta_t < 1$. Further consider the SAA counterpart of C_t defined for $N_t \geq 2(b_t + h_t)^2\frac{1}{\epsilon_t^2}\log(\frac{2}{\delta_t})$ samples. Compute $\tilde{R}_t^m = \sup\{y : \hat{C}_t^l(y) \leq -\frac{\epsilon_t}{2}\}$. Then, with probability at least $1 - \delta_t$, the base-stock level $\tilde{R}_t^m$ is an ϵ_t-point of C_t and $\tilde{R}_t^m \leq R_t^m$.*

Moreover, for each specified accuracy level $0 < \epsilon' \leq 1$ and confidence level $0 < \delta' < 1$, let $\epsilon_t' = \epsilon'$, $\epsilon_t = \frac{\min(b_t,h_t)}{3}\epsilon_t'$ and $\delta_t' = \delta_t = \frac{\delta'}{T}$. Now apply the above procedure for each of the periods $t = 1, \ldots, T$ for number of samples N_t as specified in Lemma 4.1 above, and compute an approximate policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$. We claim that, with probability at least $1 - \delta'$, this policy has expected cost at most $1 + \epsilon'$ times the expected cost of the myopic policy.

THEOREM 4.1 *Consider the policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$ computed above. Then, with probability at least $1 - \delta'$, it has expected cost at most $1 + \epsilon'$ times the expected cost of the myopic policy.*

PROOF. Let I be the event that, for each $t = 1, \ldots, T$, the base-stock level $\tilde{R}_t^m$ is ϵ_t-point with respect to C_t and $\tilde{R}_t^m \leq R_t^m$ (the corresponding myopic base-stock level). By the choice of N_t, it is readily verified that $Pr(I) \geq 1 - \delta'$. We claim that under the event I, the expected cost of the policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$ is at most $1 + \epsilon'$ times the expected cost of the myopic policy.

For each $t = 1, \ldots, T$, let $\tilde{X}_t$ and X_t be the respective inventory levels of the approximate policy and of the myopic policy at the beginning of period t (note that these are two random variables). Then $E[C_t(\tilde{R}_t^m \vee \tilde{X}_t)]$ and $E[C_t(R_t^m \vee X_t)]$ are the respective expected costs of the approximate policy and of the myopic policy in period t. It is sufficient to show that for each $t = 1, \ldots, T$, we have

$$E[C_t(\tilde{R}_t^m \vee \tilde{X}_t)] \leq (1 + \epsilon')E[C_t(R_t^m \vee X_t)].$$

Condition now on any realization of the demands $D_1, \ldots, D_{t-1}$ which results respective inventory levels $\tilde{x}_t$ and x_t (where these are the respective realizations of $\tilde{X}_t$ and X_t). Then one of the following cases applies:

Case 1. The base-stock level $\tilde{R}_t$ is reachable, i.e., $\tilde{x}_t \leq \tilde{R}_t^m$. The inequality then immediately follows by the fact that $\tilde{R}_t^m$ is an ϵ_t-point of C_t.

Case 2. The inventory level of the approximated policy, $\tilde{x}_t$, is between $\tilde{R}_t^m$ and R_t^m, i.e., $\tilde{R}_t^m \leq \tilde{x}_t \leq R_t^m$. It is readily verified that $\tilde{x}_t$ is also an ϵ_t-point of C_t which implies that the inequality still holds.

Case 3. Finally, consider the case where the inventory level of the approximate policy is above the myopic base-stock, i.e., $\tilde{x}_t > R_t^m$. Observe that under the event I, we know that $\tilde{x}_t \leq x_t$, with probability 1. In particular, this implies that $C_t(\tilde{x}_t) \leq C_t(x_t)$. This concludes the proof. □

Observe that in this case the number of samples required from each demand distribution is significantly smaller. It is of order $O(\log(T))$ instead of $O(T^5)$.

5. General Structural Lemma In this section, we provide a proof of a multi-dimensional version of the key structural Lemma 3.3 above. In many stochastic dynamic programs, one of the main challenges is to evaluate the future expected cost that results from the decision being made in the current stage. This cost function is often very complex to evaluate. However, there are cases where there are analytical expressions for subgradients, which can then be estimated accurately using sampling-based methods in a way similar to the one described above. In such cases, one of the natural issues to address is how to relate first-order information to relative errors. We believe that the following lemma provides effective tools to establish such relations for certain convex objective functions. This can lead to algorithms with rigorous analysis of their worst-case performance guarantees. More specifically, there are cases where we can derive piecewise linear functions that provide a lower bound on the real objective function value (see for example [21]). Piecewise linear approximations are also used in different heuristics for two-stage stochastic models (see, for example, [6]).

The following lemma indicates that in convex minimization models with the property that there exists a nonnegative piecewise linear function that lower bounds the objective function value, there exists an explicit relation between first-order information and relative errors. Thus, we believe that this lemma will have applications in analyzing the worst-case performance of approximation algorithms for stochastic dynamic programs and stochastic two-stage models.

LEMMA 5.1 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be convex and finite with a global minimizer denoted by y^*. Further assume that there exists a function $\bar{f} : \mathbb{R}^m \mapsto \mathbb{R}$ convex, nonnegative and piecewise linear, such that $\bar{f}(u) \leq f(u)$ for each $u \in \mathbb{R}^m$. Without loss of generality assume that $\bar{f}(u) = \lambda||u||_2$ for each $u \in \mathbb{R}^m$ and some $\lambda > 0$. Let $0 < \epsilon' \leq 1$ and $\epsilon = \frac{\epsilon'\lambda}{3}$. Then if $\hat{y} \in \mathbb{R}^m$ is an $\epsilon-$point (see definition 3.3 above) of f, its objective value $f(\hat{y})$ is at most $1 + \epsilon'$ times the optimal objective value $f(y^*)$, i.e., $f(\hat{y}) \leq (1 + \epsilon')f(y^*)$.*

PROOF. The proof follows along the lines of Lemma 3.3 above. Let $L = \frac{f(y^*)}{\lambda}$ and $\mathcal{B}(0, L) \subseteq \mathbb{R}^m$ be a ball of radius L around the origin. It is straightforward to verify that $y^* \in \mathcal{B}(0, L)$. Now by definition we have $f(y^*) = \lambda L$ and by the fact that $\hat{y}$ is an $\epsilon-$point of f, we know that $f(\hat{y}) - f(y^*) \leq \epsilon||\hat{y} - y^*||_2$.

First consider the case where $\hat{y} \in \mathcal{B}(0, L)$. Then it is readily verified that $||\hat{y} - y^*||_2 \leq 2L$. However, this implies that

$$f(\hat{y}) - f(y^*) \leq \epsilon||\hat{y} - y^*||_2 \leq 2\epsilon L = \frac{2}{3}\epsilon'\lambda L = \frac{2}{3}\epsilon' f(y^*).$$

Now consider the case where $\hat{y} \notin \mathcal{B}(0, L)$, in particular, $||\hat{y}||_2 > L$. Let $d \in \mathcal{B}(0, L)$ be the point on the line between $\hat{y}$ and the origin with $||d||_2 = L$. Since f is convex it is readily verified that the point d is also an $\epsilon-$point of f. Then clearly, the point d satisfies $f(d) - f(y^*) \leq \frac{2}{3}\epsilon' f(y^*)$. Let $x = ||\hat{y} - d||_2$. The rest of the proof proceeds exactly like the proof of Lemma 3.3. □

Acknowledgments. This research was conducted while the first author was a PhD student in the ORIE department at Cornell University.
The research of the first author was supported partially by a grant from Motorola and NSF grants CCR-9912422&CCR-0430682.
The research of the second author was supported partially by a grant from Motorola, NSF grants DMI-0075627 & DMI-0500263.
The research of the third author was supported partially by NSF grants CCR-0430682 & DMI-0500263.
We thank Shane Henderson for stimulating discussions and helpful suggestions that made the paper significantly better.
The first author thanks Shane Henderson and Adrian Lewis for stimulating discussions and helpful suggestions.
We thank the Associate Editor and the two anonymous referees for their comments that helped improving the exposition of this paper

References

[1] S. Ahmed, U. Çakmak, and A. Shapiro. Coherent risk measures in inventory problems. Unpublished manuscript, 2005.

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

[3] D. Bertsimas and A. Thiele. A robust optimization approach to supply chain management. In *Proceedings of 14th IPCO*, pages 86–100, 2004.

[4] D. Bienstock and N. Özbay. Computing robust base-stock levels. Technical Report CORC Report, TR-2005-09, IEOR Department, Columbia University, 2006.

[5] P. Billingsley. *Probability and Measure*. John Wiley & Sons, 1995. Third edition.

[6] J. Birge and S. W. Wallace. A separable piecewise linear upper bound for stochastic linear programs. *SIAM Journal on Control and Optimization*, 26:1–14, 1988.

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

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

[9] M. Charikar, C. Chekuri, and M. Pál. Sampling bounds for stochastic optimization. In *Proceedings of APPROX-RANDOM 2005*, pages 257–269, 2005.

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

[11] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistics theory of pattern rocognition*. Springer, 1996. Chapter 12, pages 196-198.

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

[13] G. Gallego and I. Moon. A min-max distribution newsboy problem: Review and extensions. *Journal of Operational Research Society*, 44:825–834, 1993.

[14] G. Gallego, J. K. Ryan, and D. Simchi-Levi. Minimax analysis for discrete finite horizon inventory models. *IIE Transactions*, pages 861–874, 2001.

[15] P. Glasserman and Y. C. Ho. *Gradient estimation via pertubation analysis*. Kluwer Academic Publishers, 1991.

[16] P. Glasserman and S. Tayur. Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Science*, 41:263–282, 1995.

[17] 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.

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

[19] W. T. Huh and P. Rusmevichientong. A non-parametric approach to stochastic inventory planning with lost-sales and censored demand. Technical Report 1427, School of OR&IE, Cornell University, 2006. Submitted to *Operations Research*.

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

[21] T. Iida and P. Zipkin. Approximate solutions of a dynamic forecast-inventory model. Working paper, 2001.

[22] R. Kapuscinski and S. Tayur. Optimal policies and simulation based optimization for capacitated production inventory systems. In *Quantative models for supply chain management*, chapter 2. Kluwer Academic Publisher, 1998.

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

[24] A. J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479–502, 2001.

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

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

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

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

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

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

[31] A. Nemirovski and A. Shapiro. On complexity of Shmoys - Swamy class of two-stage linear stochastic programming problems. Eprint: www.optimization-online.org, 2006.

[32] G. Perakis and G. Roels. The distribution-free newsvendor: Inventory management with limited demand information. Unpublished manuscript, 2005.

[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] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1972.

[35] A. Ruszczyński and A. Shapiro. Stochastic programming models. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter 6. Elsevier, 2003.

[36] H. Scarf. A min-max solution to an inventory problem. In K. J. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the mathematical theory of inventory and production*, chapter 12, pages 201–209. Stanford University Press, 1958.

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

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

[39] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter 6. Elsevier, 2003.

[40] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathemeatical Programming*, 2006. Forthcoming.

[41] A. Shapiro and T. Homem-De-Mello. On the rate of convergence of Monte Carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11:70–86, 2000.

[42] A. Shapiro, T. Homem-De-Mello, and J. Kim. Conditioning of convex piecewise linear stochastic programs. *Mathematical Programming*, 94:1–19, 2002.

[43] A. Shapiro and A. Nemirovski. On the complexity of stochastic programming problems. Eprint: www.optimization-online.org, 2005.

[44] G. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Wiley, New York, 1986.

[45] J. Si, A. G. Barto, W. B. Powell, and D. Wunch II. *Handbook of learning and approximate dynamic programming*. Wiley & Sons, Inc. Publications, 2004.

[46] C. Swamy and D. B. Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. In *Proceedings of the 46th Annual IEEE Symposium on the Foundations of Computer Science*, 2005.

[47] P. H. Zipkin. *Foundations of Inventory Management*. The McGraw-Hill Companies, Inc, 2000.