Sampling-Based Approximation Schemes for Capacitated Stochastic Inventory Control Models

Citation: Cheung, Wang Chi and David Simchi-Levi. "Sampling-Based Approximation Schemes for Capacitated Stochastic Inventory Control Models." Mathematics of Operations Research 44, 2 (May 2019): 668-692.

As Published: 10.1287/MOOR.2018.0940

Publisher: Institute for Operations Research and the Management Sciences (INFORMS)

Persistent URL: https://hdl.handle.net/1721.1/129792

Version: Original manuscript: author's manuscript prior to formal peer review

Sampling-based Approximation Schemes for Capacitated Stochastic Inventory Control Models

Wang Chi Cheung*, David Simchi-Levi†

Abstract

We study the classical multi-period capacitated stochastic inventory control problems in a data-driven setting. Instead of assuming full knowledge of the demand distributions, we assume that the demand distributions can only be accessed through drawing random samples. Such data-driven models are ubiquitous in practice, where the cumulative distribution functions of the underlying random demand are either unavailable or too complicated to work with.

We apply the Sample Average Approximation (SAA) method to the capacitated inventory control problem and establish an upper bound on the number of samples needed for the SAA method to achieve a near- optimal expected cost, under any level of required accuracy and pre-specified confidence probability. The sample bound is polynomial in the number of time periods as well as the confidence and accuracy parameters. Moreover, the bound is independent of the underlying demand distributions. However, the SAA requires solving the SAA problem, which is $\#P$-hard. Thus, motivated by the SAA analysis, we propose a randomized polynomial time approximation scheme which also uses polynomially many samples. Finally, we establish a lower bound on the number of samples required to solve this data-driven newsvendor problem to near-optimality.

1 Introduction

In this paper, we consider the multi-period Capacitated Stochastic Inventory Control problem in a data-driven setting. This problem encapsulates the dilemma of matching supply with volatile demand for a commodity, in the presence of supply constraints. The multi-period problem can be described as follows. At the start of each period, the decision maker reviews the amount of on-hand inventory or backorders, and decides the amount of additional commodity to order, in anticipation of the random demand in the period. Due to supply constraints, there is an upper limit on the amount she can order. After placing the order, the additional inventory arrives instantaneously, and her on-hand inventory increases accordingly. Then, the random demand for the commodity is realized, and the decision maker satisfies the demand to the fullest extent by her on-hand inventory. In the case of excess inventory, i.e. more on-hand inventory than demand, a holding cost is incurred per unit of unused commodity. Otherwise, in the case of insufficient inventory, a backlog cost is incurred per unit of unsatisfied demand. In addition, the unused commodity or backlogged demand is carried over to the next period, which means that the decision made in a period affects the inventory levels in the future periods. The objective is to minimize the sum of expected holding

*Institute of High Performance Computing, Agency for Science, Technology And Research, Singapore, wangchimit@gmail.com.

†Department of Civil and Environmental Engineering, and Institute for Data, Systems and Society, and Operations Research Center, MIT, dslevi@mit.edu.

and backlog costs incurred in all the periods. The demands across the periods are independent, though not necessarily identically distributed.

In the previous work on inventory models, it is typically assumed that the decision maker knows the cumulative distribution functions (cdfs) of the demand distributions. Under such assumption, the problem is well studied, and can be solved to optimality in polynomial time. In contrast, under the data-driven setting, we assume that the decision maker does not have direct access to the underlying demand distributions. The only information available to the decision maker is a set of random samples from the demand distribution in each period. Such a data-driven setting arises in many real life scenarios, since the decision maker's knowledge on the demand distributions is often gained through historical data or market forecast on the trend of future demands. Moreover, even when the decision maker has access to the true demand distributions, sometimes they could be too complicated to work with.

The *Sample Average Approximation (SAA) method* is an intuitive and popular heuristic for solving stochastic optimization problems in the data-driven setting. The idea is to consider the *SAA problem*, which is formulated by replacing the latent random distributions by their empirical counterparts constructed using the drawn samples. Then, under the SAA method, the decision maker solves the SAA problem to optimality. The rationale behind the heuristic is that, with sufficiently many samples, the SAA problem could serve as a reasonably accurate proxy for the original problem. Thus, the optimal solution for the SAA problem may be nearly optimal for the original problem. In fact, when the number of samples drawn tends to infinity, the set of optimal solutions for the SAA problem converges to the set of optimal solution for the original problem, under certain regularity conditions. A catalogue of such asymptotic results is presented in Shapiro et al. [SDR09]. However, the non-asymptotic performance of the SAA method in multi-stage stochastic optimization problems seems very hard to analyze, as remarked by Shapiro and Nemirovski [SN05].

In this paper, we provide a non-asymptotic analysis on the performance of the SAA method for the data-driven capacitated inventory control problem. We establish an explicit upper bound on the number of samples needed for the SAA method to achieve a near-optimal expected cost with high probability. The sample bound for the SAA method is polynomial in the number of periods T as well as the accuracy and confidence parameters, and the bound is independent of the underlying demand distributions. However, the SAA problem is in general $\#\mathbf{P}-$hard to solve (See Appendix I for the precise definition of $\#\mathbf{P}-$hardness). By harnessing our analysis for the SAA method, we propose a polynomial time approximation scheme to the problem, by introducing a sparsification procedure to the SAA method. One caveat in our polynomial bounds is that these bounds have a pseudo-polynomial dependence on the unit holding and backlog costs. We argue that such dependence is inherent to the data-driven model, by proving a lower bound on the number of samples used any algorithm that solves the data-driven problem to near optimality. In particular, by the results in Levi et al. [LPU15], our sample lower bound is tight for the special case of the *newsvendor problem*, which is the special case when there is only one period, but there is no supply constraint.

1.1 Literature Review

Data-driven multi-stage stochastic optimization problems are actively studied in the realms of Computer Science and Operations Research. Kleywegt et al. [KSHdM02] study the performance

of SAA method on two-stage stochastic discrete optimization problems. They prove that the set of empirically optimal solutions converges to the set of optimal solutions in the original problem as the number of samples tend to infinity. In addition, they provide an upper bound on the number of samples needed for achieving near optimality within an additive error. However, the sample bound provided in [KSHdM02] depends on the variances of the underlying random variables. Shapiro and Nemirovski [SN05] study the computational complexity of solving 2-stage and multi-stage stochastic optimization problems by the SAA methods. On one hand, they show that 2-stage stochastic optimization problems are tractable under regularity assumptions. On the other hand, for multi-stage stochastic optimization problems, they show evidence that the analysis is likely to be hard, and the SAA problems are apparently computationally intractable. The asymptotic analyses of the SAA methods under various stochastic optimization models are presented in the book [SDR09].

Data-driven two-stage stochastic combinatorial optimization problems are well studied in the literature. Gupta et al. [GPRS04] consider such two-stage problems on a variety of $\mathbf{NP}-$hard combinatorial optimization problems, and approximation algorithms with constant ratios are provided. Shmoys and Swamy [SS06] consider the two-stage linear optimization problems under covering constraints, and they propose a fully polynomial randomized approximation scheme (FPRAS). Using the FPRAS, Shmoys and Swamy [SS06] provide approximation algorithms with constant ratios for classical covering problems, such as the minimum vertex cover problems and the facility location problems. Charikar et al. [CCP05] consider a more general version of two-stage stochastic optimization than [SS06], and they show that the SAA method achieves near optimality with high probability. Finally, under the full knowledge of the underlying randomness, various two-stage problems are also studied, for example see the results in [IKMM04, RS06, GRS07, BQ11].

Data-driven multi-stage stochastic optimization problems are significantly more difficult when the number of stages is greater than two. Swamy and Shmoys [SS12] generalizes [SS06] by studying the class of multi-stage stochastic linear optimization problems with covering constraints. Swamy and Shmoys [SS12] shows that the SAA method gives rise to an FPRAS for the problem. Moreover, they provide $O(T)$-approximation algorithms for the T-stage stochastic combinatorial problems considered in [SS06]. However, the number of samples needed grows exponentially in the number of stages T. Gupta et al. [GPRS05] extends the framework in [GPRS04] to the case of multiple stages. Both the number of samples needed and the approximation ratios deteriorates exponentially with the number of stages. Shapiro [Sha06] provide bounds on the sample size required in the SAA method for data-driven multi-stage stochastic optimization problems in a general setting. However, the bounds depend on the identities of the underlying random variables in the problem, and these bounds tend to infinity as the support of the random variables become large. In fact, as stated in [SN05], the analysis of the SAA method in multi-stage stochastic optimization problems seems very hard in geneal.

The works of Levi et al. [LRS07] and Halman [Hal15] are the most relevant to ours. Levi et al. [LRS07] consider the data-driven uncapacitated inventory control problems, which is the special case of our problem when there is no supply constraints. They provide a near optimal ordering policy that has running time polynomial in the number of periods (stages) T as well as the accuracy parameters. This is in contrast with most previous works on multi-stage stochastic optimization problems, where the number of necessary samples and performance guarantee grow rapidly with the number of stages. [Hal15] considers a general class of dynamic programs in a more restricted data-driven setting, where the underlying demand distributions are assumed to satisfy a certain

lower bound property. FPTASs are provided for solving the class of DPs. A detailed comparison of our approach with [LRS07], [Hal15] is provided in the next subsection.

Dynamic programs in a different data-driven model has been considered in Halman et al. [HKM+09], Halman et al. [HKL+14]. In these works, the decision maker has oracle access to the *exact* value of the cdf of a random variable at the queried point. This is a stronger assumption than ours, in which the cdfs cannot be exactly ascertained (cf. Theorem 2.5). Halman et al. [HKM+09] consider single commodity inventory control problems with more general holding and backlog costs, and Halman et al. [HKL+14] consider an even more general class of dynamic programming problems. Halman et al. [HKM+09, HKL+14] show that under their oracle access models, the problems are #**P**-hard, and additionally they propose FPTASs for these problems. In particular, the number of queries to the cdf oracles is polynomial in the number of stages.

The capacitated stochastic inventory control models are well studied in the realm of Operations Research. The research of [AF97, KT98] show that the optimal expected cost can be achieved by a modified base stock policy, when the demands are independent but not necessarily identically distributed. Aviv and Federgruen [AF97] proposes a value iteration algorithm to compute the base stocks, while Kapuscinski and Tayur [KT98] proposes a simulation base method for the computation. Subsequently, Levi et al. [LRST08] propose a 2-approximation algorithm to the problem in the case when the demand distributions across the periods are correlated. In all these works, the underlying demand distributions are fully specified.

1.2 Our Approach

We analyze the performance of the SAA method by comparing the *empirical DP*, i.e. the dynamic program for the SAA problem, and the *original DP*, i.e. the dynamic program for the original problem. A natural approach to prove the convergence of the SAA method is to compare the values of the cost-to-go functions in the empirical DP with those in the original DP. However, for such a zero-order approach, the number of samples required for near-optimality grows with the support sizes of the underlying demand distributions.

We avoid such a dependence on the support size by a first order analysis on the cost-to-go functions. In particular, we establish a sample bound independent of the underlying demand distributions. Our analysis is based on the following approach. First, we demonstrate that, in order to prove the convergence of the SAA method, it suffices to prove that the *right derivatives* of the cost-to-go functions for the empirical DP converge to their original counterparts, when the number of samples is sufficiently large. The aforementioned reduction is performed using a Lemma by Levi et al [LRS07] and our analysis on the sample path of the inventory control problem. Next, we demonstrate that the number of samples required for the convergence in the right derivatives is finite for any given level of required accuracy and pre-specified confidence probability. In particular, the required number of samples is independent of the underlying demand distributions, since the supports of the right derivatives for the cost-to-go functions only depend on the holding and backlog cost coefficients, as well as the number of periods.

In fact, a first order approach is also proposed by [LRS07]. However, our approach is significantly different from [LRS07] in the following aspects. Levi et al. [LRS07] does not compare the original DP with the SAA DP. Instead, they compare the original DP with the *shadow dynamic program*, which is a dynamic program tailored for their inventory model and analysis. In a nutshell, the shadow dynamic program is constructed by suitably perturbing the original DP in a sequential

manner from $t = T$ to $t = 1$. Thus, the shadow dynamic program does not correspond to the SAA problem. The purpose of such an approach is to maintain the convexity of certain cost-to-go functions (namely $\tilde{U}_j, \tilde{V}_j$ in [LRS07]) in their analysis. Such perturbation crucially uses the fact that the function $\tilde{V}_j(x_j)$ (in [LRS07]) takes a constant value when x_j is smaller than a certain threshold $\tilde{R}_j$, which is not true in the presence of supply constraints (cf. §2.2). Thus, we take an alternative approach by directly analyzing the SAA method.

Our proof framework is also different from [LRS07], in the sense that we directly investigate the impact of perturbing the base stock levels to the expected cost. We demonstrate (cf. Lemma 3.4) that the optimal modified base stock policy is robust to perturbation, which paves our way to establishing the required near-optimality by bounding the estimation error in each period. We believe that such robustness result in inventory control models could be of interest in other settings.

In addition, the analysis of the SAA method in their inventory models is raised as an open question in Levi et al. [LRS06]. In this paper, we establish that with polynomially many samples, the SAA method does yield a near optimal ordering policy in the case of capacitated inventory control models (which include their models). However, the SAA method does not immediately lead to a polynomial time algorithm, since the underlying SAA problem is $\#\mathbf{P}-$complete. Thus, we provide a polynomial time approximation scheme *Sample* by introducing a sparsification procedure to the SAA method.

In Halman [Hal15], the author consider a general class of dynamic programs in a data-driven setting, but with the following assumption on the demand distribution. For all demand distributions D_t and $d \in \mathbb{R}^+$, it is assumed that either $\mathbb{P}[D_t = d] = 0$, or $\mathbb{P}[D_t = d] \geq \gamma$, where $\gamma > 0$ is a parameter known to the decision maker. FPTASs are proposed in this particular data-driven setting, and the bounds on running time is proportional to $1/\gamma^2$. Note that the results in [Hal15] is incomparable to ours, since we have a more general data-drvien setting (we do not assume such a γ to be existent and known to the decision maker in our setting) than [Hal15], but the latter considers a more general class of DPs than ours. The running time of the algorithm by [Hal15] and the running time of *Sample* in this paper are also incomparable. One one hand, [Hal15] algorithm's running time has a much better dependence on T than ours. On the other hand, the running time of [Hal15] depends on γ and (mildly) on the support of the underlying probability distribution, but the running time of our algorithm does not depend on these quantities. Finally, we analyze the SAA method by a first order analysis, while the FPTASs proposed in [Hal15] are not SAA methods, and are analyzed through zero-order analyses. (A comparison of our work, [Hal15] and [LRS07] is provided in [Hal15]).

Finally, we note that in inventory control models, the presence of order lead time and linear ordering costs are commonly assumed. Nevertheless, the order lead time can be easily incorporated into our analysis by a suitable time shift in the dynamic program. Linear ordering costs can also be incorporated into our analysis by adding suitable constants in the computations of the left and right derivatives. In the spirit of [LRS07], we assume zero lead time and zero ordering cost for the sake of clarity.

The rest of the paper is organized as follows. In §2.1, we formulate the data driven capacitated inventory control problem. In §2.2, we review the notion of modified base stock policies. In §2.3, we state the main results in the paper. In §3, we outline our first order analysis on the sample average approximation (SAA) method for the data-driven problem. In §4, we provide a polynomial time approximation scheme *Sample* to the problem, by introducing a sparsification procedure to the SAA method. In §5, we provide insights into the proofs for the hardness results. In §6, we compare

the performance of Algorithm *Sample* in simulations with the performance guarantee predicted by our analysis. Finally, in §7, we conclude the paper. The proofs of the Theorems and Lemmas are deferred to the Appendix sections.

2 Problem Model and Our Contributions

2.1 The Data-driven Capacitated Inventory Control Model

We consider the periodic review capacitated inventory control problem in a data-driven setting. The decision maker faces a finite time horizon with T discrete time periods, labeled as $1, \cdots, T$. From period 1 to period T, the decision maker performs the following actions:

1. Observe the starting inventory level x_t.
2. Order up to y_t, where $0 \leq y_t - x_t \leq B_t$. The parameter B_t is the capacity on the inventory that can be ordered in the t^{th} period.
3. Observe the t^{th} period demand D_t.
4. If $y_t > D_t$, it incurs a linear holding cost of $h_t \times (y_t - D_t)$; else if $y_t \leq D_t$, it incurs a linear backlog cost $b_t \times (D_t - y_t)$. In the latter case, the unsatisfied demand is backlogged.
5. Proceed to period $t + 1$, with starting inventory level being $x_{t+1} = y_t - D_t$.

The latent demand distributions $D_1, \cdots, D_T$ are assumed to be independent, though not necessarily identically distributed. The decision maker's objective is to design an ordering policy that minimizes her expected total operational cost

$$\mathbb{E}\left[\sum_{t=1}^{T} h_t(y_t - D_t)^+ + b_t(D_t - y_t)^+\right] \tag{2.1}$$

across the planning horizon, subject to the capacity constraint in each period. The function $(x)^+$ denotes $\max\{x, 0\}$. For each $t = 1, \cdots, T$, we assume that $h_t, b_t > 0$.

The decision maker does not know the explicit demand distributions. Rather, the only information available is a set of independent samples drawn from the true distributions; the decision maker can draw any number N_t of independent samples $d_t^1, \cdots, d_t^{N_t}$ of D_t from its sample generating oracle. This data-driven setting, analogous to the settings in [LRS07, SS06, SS12], is strictly weaker than the model considered by [HKL$^+$14], where the decision maker has an oracle access to the cdfs of the underlying demand distributions.

We assume that the expectation $\mathbb{E}|D_t|$ is finite for all t, which is necessary for the problem to be well-defined. This is the only assumption we make on the underlying demand distributions $D_1, \cdots, D_T$. In particular, we neither assume that the demand distributions are parametrized, nor assume that they have bounded supports.

To consider the problem in a data-driven setting, we use $SAA(T; N_1, \cdots, N_T)$ to denote the *sample average approximation* counterpart of the original capacitated inventory control problem. The empirical problem $SAA(T; N_1, \cdots, N_T)$ is constructed by using N_t samples from D_t for each period t. More precisely, conditional on the N_t samples $d_t^1, \cdots, d_t^{N_t}$ drawn from D_t for each period

t, the capacitated inventory control problem $SAA(T; N_1, \cdots, N_T)$ is the T-period problem where the t^{th} period demand distribution $\hat{D}_t$ is the empirical distribution for D_t:

$$\mathbb{P}[\hat{D}_s = d] = \frac{\sum_{i=1}^{N_s} \mathbf{1}[d = d_s^i]}{N_s}. \tag{2.2}$$

Note that the optimal cost of $SAA(T; N_1, \cdots, N_T)$ is a random variable that depends on the random samples drawn.

2.2 Modified Base Stock Policies

Throughout the paper, we consider a certain class of policies called *modified base stock policies*:

Definition 2.1. Under a modified base stock policy $(R_1, \cdots, R_T)$, at period t the decision maker determines the order-up-to level y_t (in step 2 in §2.1) in the following manner:

$$y_t = \begin{cases} x_t + B_t & \text{if } x_t \in (-\infty, R_t - B_t] \\ R_t & \text{if } x_t \in (R_t - B_t, R_t] \\ x_t & \text{if } x_t \in (R_t, \infty) \end{cases}.$$

In other words, for each period t, the decision maker makes the inventory level y_t as close to R_t as possible, under the supply constraints. Under a modified base stock policy $(R_1, \cdots, R_T)$, the decision made in period t is only dependent on the amount of inventory x_t at hand and the base stock R_t, but it does not depend on other base stocks and observations made in the previous periods.

By the work of [KT98], [Tay93], there exists an optimal modified base stock policy $(R_1^*, \cdots, R_T^*)$ under which the expected cost (2.1) for the original problem is minimized. The derivation of the optimality of modified base stock policy is useful for our analysis on the SAA algorithm. Thus, we review the derivation below. An optimal policy can be found by solving the following Bellman equations from $t = T$ to $t = 1$:

$$V_t(x_t) = \min_{x_t \le y_t \le x_t + B_t} C_t(y_t) + \mathbb{E}\left[V_{t+1}(y_t - D_t)\right], \quad V_{T+1}(x_{T+1}) = 0,$$

the function $C_t(y_t) = \mathbb{E}[h_t(y_t - D_t)^+ + b_t(D_t - y_t)^+]$ is the t^{th} period expected operational cost. To facilitate our subsequent analysis on the SAA algorithm, we introduce the function U_t:

$$U_t(y_t) = C_t(y_t) + \mathbb{E}\left[V_{t+1}(y_t - D_t)\right]. \tag{2.3}$$

Thus, we have

$$V_t(x_t) = \min_{x_t \le y_t \le x_t + B_t} U_t(y_t). \tag{2.4}$$

The function $V_t(x_t)$ represents the expected cost over $t, \cdots, T$ when the starting inventory level in period t is x_t, and the decision maker orders optimally in the periods $t, \cdots, T$. The function $U_t(y_t)$ represents the expected cost over $t, \cdots, T$ when the inventory level after ordering is y_t in period t, and the decision maker orders optimally in the periods $t+1, \cdots, T$.

By a backward induction from $t = T$ to $t = 1$, [AF97, KT98] further show that

1. The functions U_t, V_t are convex for all t,

2. The modified base stock policy $(R_1^*, \cdots, R_T^*)$, where $R_t^* \in \operatorname{argmin}_{y_t \in \mathbb{R}} U_t(y_t)$, is optimal.

The induction is shown as follows. Suppose $V_{t+1}(x_{t+1})$ is a convex function. Then the function $U_t(y_t)$ is also convex, by virtue of (2.3). Now we wish to show that the convexity of U_t implies the convexity of V_t. First, note that $\lim_{|y_t| \to \infty} U_t(y_t) = \infty$, thus $U_t(y_t)$ has a global minimum R_t^* in $\mathbb{R}$, which can be computed when the CDFs of the demand distributions are known. Considering (2.4) for the following ranges of x_t, we have

$$\operatorname{argmin}_{x_t \le y_t \le x_t + B_t} U_t(y_t) \ni \begin{cases} x_t + B_t & \text{if } x_t \in (-\infty, R_t^* - B_t] \\ R_t^* & \text{if } x_t \in (R_t^* - B_t, R_t^*] \\ x_t & \text{if } x_t \in (R_t^*, \infty) \end{cases}$$

by the convexity of U_t. In particular, this shows that it is optimal to follow a modified base stock policy with threshold R_t^* in period t. Finally, applying this in (2.4) yields

$$V_t(x_t) = \begin{cases} U_t(x_t + B_t) & \text{if } x_t \in (-\infty, R_t^* - B_t) \\ U_t(R_t^*) & \text{if } x_t \in [R_t^* - B_t, R_t^*) \\ U_t(x_t) & \text{if } x_t \in [R_t^*, \infty) \end{cases} \tag{2.5}$$

for all $x_t \in \mathbb{R}$. It is easy to verify that $V_t(x_t)$ is also a convex function, which establishes the backward induction. Altogether, the optimality of the modified base stock policy $(R_1^*, \cdots, R_T^*)$ is established.

In the above (zero-order) analysis, we can choose the threshold R_t^* to be any minimizer of U_t. However, in order to facilitate the forthcoming first order analysis for the SAA algorithm in §3 (cf. Theorem 3.8), we will choose R_t^* to be the *smallest* minimizer of U_t for all t.

Similarly, we choose $\hat{R}_t$ to be the *smallest* minimizer of in the empirical cost-to-go function $\hat{U}_t$ for all t in the sample average problem $SAA(T; N_1, \cdots, N_T)$, where $\hat{U}_t$ is defined in an analogous way to U_t:

$$\hat{U}_t(y_t) = \hat{C}_t(y_t) + \mathbb{E}\hat{V}_{t+1}(y_t - \hat{D}_t), \tag{2.6}$$

$$\hat{V}_t(x_t) = \min_{x_t \le y_t \le x_t + B_t} \hat{U}_t(y_t) = \begin{cases} \hat{U}_t(x_t + B_t) & \text{if } x_t \in (-\infty, \hat{R}_t - B_t) \\ \hat{U}_t(\hat{R}_t) & \text{if } x_t \in [\hat{R}_t - B_t, \hat{R}_t) \\ \hat{U}_t(x_t) & \text{if } x_t \in [\hat{R}_t, \infty) \end{cases}, \tag{2.7}$$

where $\hat{D}_t$ is the empirical distribution constructed using N_t samples from D_t (as defined in (2.2)), and $\hat{C}_t(y_t) = \mathbb{E}[h_t(y_t - \hat{D}_t)^+ + b_t(\hat{D}_t - y_t)^+]$ is the empirical operational cost in the t^{th} period.

We say that a set of base stocks $(R_1, \cdots, R_T)$ is $(1+\epsilon)$-*optimal* if the expected cost under the modified base stock policy defined by $(R_1, \cdots, R_T)$ is at most $(1+\epsilon)$ times the optimal expected cost. For example, by definition, $(R_1^*, \cdots, R_T^*)$ is 1-optimal.

2.3 Main Results

Firstly, we show that for the SAA method to output a near optimal modified base stock policy, it is sufficient to use only polynomially many samples:

Theorem 2.2. *Consider the Sample Average Approximation problem $SAA(T; N_1, \cdots, N_T)$, where N_t is defined as follows*

$$N_t = \max\left\{(h_t + b_t)^2, \left(\sum_{s=t+1}^{T} h_s + b_s\right)^2\right\} \frac{144T^4}{\epsilon^2 \min_{s\in\{1,\cdots,T\}}\{\min\{h_s, b_s\}\}^2} \log\frac{4T}{\delta}. \tag{2.8}$$

For $t = T, \cdots, 1$, let $\hat{R}_t$ be the smallest minimizer of the empirical cost-to-go function $\hat{U}_t(y_t)$, i.e. $\hat{R}_t = \min_{y\in\mathbb{R}}\{y : y \in \text{argmin } \hat{U}_t(y)\}$, where $\hat{U}_t$ is defined in (2.6). With probability at least $1-\delta$, the modified base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ is $(1+\epsilon)$-optimal to the original problem.

Note that the base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ in the Theorem is a set of random variables that depend on the samples drawn. The base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ achieves near-optimality in the original problem, in addition to achieving optimality in the empirical problem $SAA(T; N_1, \cdots, N_T)$.

While Theorem 2.2 shows that the SAA method is *informationally efficient*, i.e. only requires a bounded number of samples, the Theorem and its analysis do not imply the *computational efficiency* of the SAA method. In fact, solving the empirical problem exactly is computationally hard, in the sense that even $SAA(T; 2, \cdots, 2)$ can be intractable:

Lemma 2.3. *Consider the stochastic capacitated inventory control problem, where the demand distributions $D_1, \cdots, D_T$ are explicitly given, and each of D_t has a discrete support $\{0, a_t\}$. If there is an algorithm that runs in time polynomial in T and returns an optimal modified base stock policy, then $\mathbf{P} = \#\mathbf{P}$.*

Lemma 2.3 is proven by reducing the problem to the K^{TH} LARGEST SUBSET problem, which is a $\#\mathbf{P}-$hard problem ([HKM+09], also see SP20 on page 225 in [GJ79]). We provide the formal definition of $\#\mathbf{P}-$hardness in Appendix I, and the proof for the Lemma in Appendix J. Thus, the SAA method, which solves the SAA problem exactly, is not computationally efficient. This hardness result is in contrast with [BLQ12, CCP05, SS12], in which the corresponding SAA problems (constructed with polynomially many samples) can be solved in polynomial time.

Thus, we propose a polynomial time randomized approximation scheme that returns a modified base stock $(\tilde{R}_1, \cdots, \tilde{R}_T)$ for the origin problem, by considering a suitable sparsification procedure on the subgradients of the cost-to-go functions. This sparsification procedure is described and analyzed in §4.

Theorem 2.4. *For every $\epsilon > 0, 0 < \delta < 1$, there is a randomized algorithm that produces a set of $(1+\epsilon)-$optimal base stocks $(\tilde{R}_1, \cdots, \tilde{R}_T)$ with probability $1-\delta$. The number of samples needed is polynomial in $\left(\max_{t\in\{1,\cdots,T\}}\left\{\frac{h_t+b_t}{\min\{h_t,b_t\}}\right\}, T, \frac{1}{\epsilon}, \log\left(\frac{1}{\delta}\right)\right)$. In addition, the algorithm has running time polynomial in $\left(\max_{t\in\{1,\cdots,T\}}\left\{\frac{h_t+b_t}{\min\{h_t,b_t\}}\right\}, T, \frac{1}{\epsilon}, \log\left(\frac{1}{\delta}\right), \log(d_{max}c^*)\right)$, where d_{max} denotes the maximum value of the samples drawn, and $c^* = \max_{t=1,\cdots,T}\max\{h_t, b_t\}$.*

Finally, we note that in all the polynomial bounds mentioned above, there is a pseudo-polynomial dependence on the cost parameter $\max_{t\in\{1,\cdots,T\}}\left\{\frac{h_t+b_t}{\min\{h_t,b_t\}}\right\}$, which is the same as the case in [LRS07]. Rather than an artifact of the first order analyses, we show that such pseudo-polynomial dependence is necessary in an information theoretic sense. More precisely, we provide a lower bound on the number of samples needed to solve the *data-driven newsvendor problem* to near optimality.

The data-driven newsvendor problem is the special case of the data-driven capacitated inventory control problem when $T = 1$ and $B_1 = \infty$, i.e. it is the one period problem without any supply constraint. In the Theorem below, we drop the subscript for the period for clarity sake.

Theorem 2.5. *Let $\mathcal{A}$ be an algorithm that returns an $(1+\epsilon)$ optimal base stock to the data-driven newsvendor problem with probability at least $1-\delta$, under any latent demand distribution, where $0 < \epsilon < 1/20, 0 < \delta < 1/4$. Then $\mathcal{A}$ draws at least $\frac{(1-4\delta)(h+b)}{2000\min\{h,b\}\epsilon^2}$ samples.*

The proof of Theorem 2.5 is given in Appendix K. Finally, we remark the lower bound in Theorem 2.5 matches the upper bound provided in [LPU15] for the data-driven newsvendor problem. Thus, combined with [LPU15], it is shown that when the level of confidence probability is high, the number of samples needed for computing a $(1+\epsilon)$ optimal base stock for the data-driven newsvendor problem is precisely $\Theta\left(\frac{h+b}{\min\{h,b\}\epsilon^2}\right)$.

3 A First Order Analysis on the Sample Average Approximation Problem

In this Section, we lay out the roadmap for the proof of Theorem 2.2. The proof is based on comparing the dynamic program for the SAA problem $SAA(T; N_1, \cdots, N_T)$, i.e. the *empirical DP*, and the dynamic program for the original problem (with latent demand distributions), i.e. the *original DP*.

We present the proof in three parts, namely §3.1, §3.2 and §3.3. In §3.1, we demonstrate that the closeness between the right derivatives for the cost-to-go functions for the original and empirical DPs implies Theorem 2.2. Thus, the proof of Theorem 2.2 is reduced to a first order analysis. We establish the reduction using a lemma by [LRS07], as well as a lemma on the robustness of an optimal modified base stock policy (cf. Lemma 3.4). In §3.2, we set up the tools for the first order analysis by providing the expressions for the right derivatives for the cost-to-go functions for the original and empirical DPs. Finally, in §3.3, we argue that our choice of $N_1, \cdots, N_T$ ensures the uniform closeness between the right derivatives required in §3.1. The proof involves a careful analysis on the dynamics of the capacitated inventory control problem. In particular, the choice of $(N_1, \cdots, N_T)$ leads to a $(1+\epsilon)$-approximation to the problem with high probability, hence completing the proof.

We define the following notations. The functions $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}, \hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ denote the right derivatives of $U_t, V_t, \hat{U}_t, \hat{V}_t$, which are defined in (2.3), (2.4), (2.6) and (2.7) respectively.

3.1 Closeness in right derivatives implies near-optimality

In this subsection, we relates the near optimality of modified base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ to the closeness of the right derivatives $U_t^{\mathrm{r}}, \hat{U}_t^{\mathrm{r}}$:

Lemma 3.1. *Suppose that for all y_t, we have $\left|\hat{U}_t^r(y_t) - U_t^r(y_t)\right| \le \eta_t$, where $\eta_t > 0$. Then the modified base stock $(\hat{R}_1, \cdots, \hat{R}_T)$, where $\hat{R}_t$ is the minimal minimizer of $\hat{U}_t$, is $\left(1 + \sum_{t=1}^T \frac{6\eta_t}{\min\{h_t, b_t\}}\right)$ − optimal.*

The Lemma implies that, to prove Theorem 2.2, it suffices to bound $\max_{y\in\mathbb{R}} \left|\hat{U}_t^{\mathrm{r}}(y) - U_t^{\mathrm{r}}(y)\right|$. In other words, we have reduced the proof of Theorem 2.2 to a first order analysis. Equipped with

Lemma 3.1, we avoid the need to estimate the value $\hat{U}_t(y)$ for any given y. In fact, estimating the function value within a fixed additive error is impossible with a finite number of samples in our setting, where we do not have any assumption on the concentration property of the demand distributions.

Lemma 3.1 follows from Claim 3.2 and Lemma 3.4. Claim 3.2 transforms the first order approximation guarantee on U_t to a certain zero-order approximation guarantee on U_t:

Claim 3.2. *Suppose that for all y_t, we have $\left|\hat{U}_t^r(y_t) - U_t^r(y_t)\right| \leq \eta_t$, where $\eta_t > 0$ is a given constant. Then the following inequality holds:*

$$U_t(\hat{R}_t) \leq \left(1 + \frac{3\eta_t}{\min\{h_t, b_t\}}\right) \min_{x\in\mathbb{R}} U_t(x) = \left(1 + \frac{3\eta_t}{\min\{h_t, b_t\}}\right) U_t(R_t^*). \tag{3.1}$$

The proof of Claim 3.2 is provided in Appendix A. The Claim is a straightforward consequence of a Lemma by [LRS07].

Lemma 3.3 (Lemma 3.3 in [LRS07])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function such that for all $x \in \mathbb{R}$ the inequality $f(x) \geq \bar{f}(x) := h(x-d)^+ + b(d-x)^+$ holds, where d is a constant. Suppose that for the real number y there exists a subgradient $s_y \in \partial f(y)$ with the property that $|s_y| \leq (\epsilon/3)\min\{b, h\}$. Then we have*

$$f(y) \leq (1+\epsilon) min_{x\in\mathbb{R}} f(x). \square$$

Equation (3.1) in Claim 3.2 has the following interpretation. Recalling the definition of U_t in (2.3), equation (3.1) is equivalent to

$$Cost_t(x_t; \hat{R}_t, R_{t+1}^*, \cdots, R_T^*) \leq \left(1 + \frac{3\sum_{s=t}^T \alpha_s}{\min\{b_t, h_t\}}\right) Cost_t(x_t; R_t^*, R_{t+1}^*, \cdots, R_T^*).$$

The function $Cost_t(x_t; R_t, \cdots, R_T)$ is defined as the expected cost from period t to period T, when the starting inventory level in period t is x_t, and then the decision maker follows the modified base stock policy defined by $(R_t, \cdots, R_T)$. In particular, note that $V_t(x_t) = Cost_t(x_t; R_t^*, \cdots, R_T^*)$. The inequality above shows that when the optimal base stock R_t^* is replaced by the suboptimal base stock $\hat{R}_t$, the deterioration in the expected cost from period t to T can be bounded. This suggests that the empirical base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ is a candidate for a near optimal policy. In the following Lemma, we quantify the robustness of an optimal modified base stock policy, when the optimal base stock levels are subjected to perturbation.

Lemma 3.4. *Let $(R_1, \cdots, R_T)$ be a set of base stocks. Suppose that for all $t = 1, \cdots, T$, we have*

$$U_t(R_t) \leq \left(1 + \frac{\epsilon_t}{2}\right) U_t(R_t^*),$$

where $\epsilon_1, \cdots, \epsilon_T$ are non-negative real numbers such that $\sum_{t=1}^T \epsilon_t \leq 1$. Then for any starting inventory level x_1 in period 1, we have

$$Cost_1(x_1; R_1, \cdots, R_T) \leq \left(1 + \sum_{t=1}^T \epsilon_t\right) Cost_1(x_1; R_1^*, \cdots, R_T^*) = \left(1 + \sum_{t=1}^T \epsilon_t\right) V_1(x_1).$$

The Lemma is proven by sequentially replacing R_t^* with R_t from $t = T$ to $t = 1$, and by comparing $Cost_t(R_t, \cdots, R_T)$ with the optimal cost $Cost_t(R_t^*, \cdots, R_T^*)$. Its proof is provided in Appendix B.

By combining Claim 3.2 and Lemma 3.4, we have established Lemma 3.1. Altogether, we show that, in order to prove Theorem 2.2, it suffices to bound the distance between $\hat{U}_t^{\mathrm{r}}$ and U_t^{r} for all t. In the next two subsections, we set up the tools and perform a first order analysis on the cost-to-go functions for the original and empirical DPs.

3.2 The Expressions for the Right Derivatives

By §3.1, our proof of Theorem 2.2 requires a first order analysis on the original and empirical DPs. To set up the tools for the analysis, we provide the expressions for $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}, \hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$, the right derivatives of $U_t, V_t, \hat{U}_t, \hat{V}_t$ defined in (2.3), (2.4), (2.6) and (2.7). By the assumption that $\mathbb{E}[|D_t|] < \infty$ for all t, we can apply the dominated convergence theorem to express the right derivatives as follows:

$$U_t^{\mathrm{r}}(y_t) = C_t^{\mathrm{r}}(y_t) + \mathbb{E}V_{t+1}^{\mathrm{r}}(y_t - D_t), \quad \hat{U}_t^{\mathrm{r}}(y_t) = \hat{C}_t^{\mathrm{r}}(y_t) + \mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - \hat{D}_t).$$

The right derivatives $V_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ have the following expressions:

$$V_t^{\mathrm{r}}(x_t) = \begin{cases} U_t^{\mathrm{r}}(x_t + B_t) & \text{if } x_t \in (-\infty, R_t^* - B_t) \\ 0 & \text{if } x_t \in [R_t^* - B_t, R_t^*) \\ U_t^{\mathrm{r}}(x_t) & \text{if } x_t \in [R_t^*, \infty) \end{cases}, \tag{3.2}$$

$$\hat{V}_t^{\mathrm{r}}(x_t) = \begin{cases} \hat{U}_t^{\mathrm{r}}(x_t + B_t) & \text{if } x_t \in (-\infty, \hat{R}_t - B_t) \\ 0 & \text{if } x_t \in [\hat{R}_t - B_t, \hat{R}_t) \\ \hat{U}_t^{\mathrm{r}}(x_t) & \text{if } x_t \in [\hat{R}_t, \infty) \end{cases}. \tag{3.3}$$

Finally, we have the expressions for the right derivatives of the single period costs:

$$C_t^{\mathrm{r}}(y_t) = -b_t + (h_t + b_t)\mathbb{P}[D_t \le y_t], \quad \hat{C}_t^{\mathrm{r}}(y_t) = -b_t + (h_t + b_t)\frac{1}{N_t}\sum_{i=1}^{N_t} \mathbf{1}[d_t^i \le y_t]. \tag{3.4}$$

The thresholds R_t^*, $\hat{R}_t$ are the smallest minimizers of $U_t, \hat{U}_t$ respectively, which satisfy the following:

$$R_t^* = \min\{y : U_t^{\mathrm{r}}(y) \ge 0\}, \quad \hat{R}_t = \min\{y : \hat{U}_t^{\mathrm{r}}(y) \ge 0\}.$$

For every x_t, y_t, the empirical right derivatives $\hat{U}_t^{\mathrm{r}}(x_t), \hat{V}_t^{\mathrm{r}}(y_t)$ are random variables that depends on the empirical distributions $\hat{D}_t, \cdots, \hat{D}_T$; and $\hat{D}_t, \cdots, \hat{D}_T$ are constructed using samples from $D_t, \cdots, D_T$. In addition, observe that $\hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ are right continuous (random) step functions with finitely many break points, while $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}$ could be continuous functions. It is important to note that in general $\mathbb{E}\hat{U}_t^{\mathrm{r}}(y_t) \ne U_t^{\mathrm{r}}(y_t)$ (except when $t = T$), since for $t < T$, $\hat{U}_t, U_t$ are cost-to-go functions for the capacitated inventory control problem with different underlying distributions; the former being the empirical distribution and the latter being the original distribution. Similarly, in general $\mathbb{E}\hat{V}_t^{\mathrm{r}}(y_t) \ne V_t^{\mathrm{r}}(y_t)$, except when $t = T + 1$. Thus, it requires extra work to show that $\hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ uniformly approximate $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}$.

Another important property is that $\hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ are bounded, unlike their value functions $\hat{U}_t, \hat{V}_t$.

We have $\hat{U}_t^{\mathrm{r}}(y_t), \hat{V}_t^{\mathrm{r}}(x_t) \in [-\sum_{s=t}^T b_s, \sum_{s=t}^T h_s]$ for all x_t, y_t with probability 1. These bounds are independent of the underlying demand distributions $D_t, \cdots, D_T$. Thus, we are able to establish sample upper bounds for estimating the right derivatives, and these bounds do not depend on the underlying distributions. This is in contrast to the case for estimating the function values.

3.3 Bounding the Estimation Error $\left|U_t^{\mathrm{r}} - \hat{U}_t^{\mathrm{r}}\right|$

In this subsection, we show that, with our choice of $N_1, \cdots, N_T$, the empirical right derivatives $\hat{U}_t^{\mathrm{r}}$ uniformly approximate the original right derivatives $U_t^{\mathrm{r}}(y_t)$ for all t with probability at least $1-\delta$. Here, we say that a function $\hat{f}: \mathbb{R} \to \mathbb{R}$ *uniformly approximates* another function $f: \mathbb{R} \to \mathbb{R}$ if there exists a constant η such that $\left|\hat{f}(x) - f(x)\right| \leq \eta$ for all $x \in \mathbb{R}$. More precisely, we show the following:

Lemma 3.5. *Consider the empirical problem $SAA(1; N_1, \cdots, N_T)$, where*

$$N_t = \max\left\{(h_t + b_t)^2, \left(\sum_{s=t+1}^T h_s + b_s\right)^2\right\} \frac{4}{\alpha_t^2} \log \frac{4T}{\delta}. \tag{3.5}$$

We have the following bound on the estimation errors of the empirical right derivatives $\hat{U}_1^r, \cdots \hat{U}_T^r$:

$$\mathbb{P}\left[\textit{For all } t \textit{ and } y, \ \left|U_t^r(y) - \hat{U}_t^r(y)\right| \leq \sum_{s=t}^T \alpha_s\right] \geq 1 - \delta.$$

That is, the empirical right derivatives $\hat{U}_T^r, \cdots, \hat{U}_1^r$ uniformly approximate the original right derivatives $U_T^r, \cdots, U_1^r$ with high probability.

The Lemma is proven by a backward induction argument, which involves applying appropriate concentration bounds, as well as a careful analysis on the dynamics of the capacitated inventory control problem. To establish the backward induction, we first provide the following Theorem, which states that if $\hat{V}_{t+1}^{\mathrm{r}}$ uniformly approximates V_{t+1}^{r}, then $\hat{U}_t^{\mathrm{r}}$ uniformly approximates U_t^{r} with high probability, by a suitable choice N_t of number of samples drawn from D_t.

Theorem 3.6. *Suppose we are given an empirical right derivative $\hat{V}_{t+1}^r : \mathbb{R} \to \mathbb{R}$ which uniformly approximates V_{t+1}^r. That is, for all $x_{t+1} \in \mathbb{R}$, it holds that $\left|\hat{V}_{t+1}^r(x_{t+1}) - V_{t+1}^r(x_{t+1})\right| \leq \gamma_t$, where γ_t is a constant. Let $d_t^1, \cdots, d_t^{N_t}$ be independent samples of D_t, where*

$$N_t = \max\left\{(h_t + b_t)^2, \left(\sum_{s=t+1}^T h_s + b_s\right)^2\right\} \frac{4}{\alpha_t^2} \log \frac{4}{\delta_t}.$$

Then the empirical right derivative $\hat{U}_t^r(y_t) = \hat{C}_t^r(y_t) + \mathbb{E}\hat{V}_{t+1}^r(y_t - \hat{D}_t)$ uniformly approximates the original right derivative U_t^r with high probability. In particular, the following inequality holds:

$$\mathbb{P}\left[\textit{For all } y_t, \ \left|\hat{U}_t^r(y_t) - U_t^r(y_t)\right| \leq \gamma_t + \alpha_t\right] \geq 1 - \delta_t.$$

The proof of Theorem is provided in Appendix C. The proof involves the decomposition of $U_t, \hat{U}_t$ according to (2.3), and crucially uses a Theorem of [Mas90], which provides us with a concentration bound on $U_t^{\mathrm{r}}(y_t)$ that holds *uniformly* across all $y_t \in \mathbb{R}$.

Theorem 3.7 ([Mas90])**.** *Let $X_1, \cdots, X_N$ be independent samples of the random variable X, where $N = \frac{1}{\epsilon^2} \log \frac{2}{\delta}$. Then we have*

$$\mathbb{P}\left[\textit{For all } x, \ \left|\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[x \le X_i] - \mathbb{P}[x \le X]\right| \le \epsilon\right] \ge 1 - \delta.$$

□

Next, we demonstrate that if the empirical right derivative $\hat{U}_t^{\mathrm{r}}$ uniformly approximates the original right derivative U_t^{r}, then $\hat{V}_t^{\mathrm{r}}$ also uniformly approximates V_t^{r} with the same additive error.

Theorem 3.8. *Suppose for all y_t, the inequality $\left|\hat{U}_t^r(y_t) - U_t^r(y_t)\right| \le \eta_t$ holds. Then we have*

$$\left|\hat{V}_t^r(x_t) - V_t^r(x_t)\right| \le \eta_t$$

for all $x_t \in \mathbb{R}$.

Theorem 3.8 is proven by considering different cases on R_t^* and $\hat{R}_t$, and a careful analysis on the dynamics of the capacitated inventory control problem in each case. In the analysis, we crucially use the fact that $R_t^*, \hat{R}_t$ are the smallest minimizers of $U_t, \hat{U}_t$ respectively. The proof is provided in Appendix D.

Altogether, Lemma 3.5 is proved by combining Theorem 3.6 and 3.8, and a suitable backward induction argument. The proof is rather routine, and we defer the details to Appendix E. Finally, Theorem 2.2 is proven by putting together Lemma 3.1 and Lemma 3.5, and applying the value

$$\alpha_t = \frac{\epsilon \min_{t \in \{1,\cdots,T\}} \{\min\{h_t, b_t\}\}}{6T^2}. \tag{3.6}$$

Applying (3.6) to (3.5) in Lemma 3.5 recovers our choice of $N_1, \cdots, N_T$ (2.8) stated in Theorem 2.2. The choice of $\{\alpha_t\}_{t=1}^T$ ensures that the numbers of samples $N_1, N_2, \cdots, N_T$ stated in (2.8) are sufficient for achieving $(1+\epsilon)-$optimality. The full proof of Theorem 2.2 based on the Lemmas are found in Appendix F.

4 A Polynomial Time Approximation Scheme via Sparsification

The analysis in §3 shows that the data-driven capacitated inventory control problem is *inforamtionally tractable*, in the sense that a a set of near optimal base stocks $(\hat{R}_1, \cdots, \hat{R}_T)$ for the original problem can be constructed with a bounded number of samples. However, it does not show that the problem is *computationally tractable*, i.e. can be solved in polynomial time. Indeed, by Lemma 2.3, there exists instances for which the SAA problem $\mathrm{SAA}(T; N_1, \cdots, N_T)$ is computationally intractable. Thus, in this Section, we propose a polynomial time algorithm approximation scheme, *Sample*, which also returns a set of near optimal modified base stocks $(\tilde{R}_1, \cdots, \tilde{R}_T)$ for the original problem.

Our polynomial time approximation scheme $Sample(\eta, N_1, \cdots, N_T)$ is based a sparsification procedure to the SAA Method. The parameter $\eta > 0$ is the accuracy parameter for the sparsification procedure, which is crucial for ensuring *Sample* to terminate in polynomial time.

The algorithm $Sample(\eta, N_1, \cdots, N_T)$ can be interpreted as follows. From line 1 to 2, the empirical demand distributions are constructed. From line 4 to line 9, the algorithm constructs the functions $\tilde{U}_t^{\mathrm{r}}$ and $\tilde{V}_t^{\mathrm{r}}$, which turn out to uniformly approximate the original right derivatives U_t^{r} and V_t^{r}. This impliest the near-optimality of $(\tilde{R}_1, \cdots, \tilde{R}_T)$ to the original problem. The sparsification step in line 8 makes the algorithm different from the SAA method. In general we have $\tilde{V}_t^{\mathrm{r}} \neq \hat{V}_t^{\mathrm{r}}$, and $\tilde{U}_t^{\mathrm{r}} \neq \hat{U}_t^{\mathrm{r}}$.

Algorithm 1 Algorithm $Sample(\eta, N_1, \cdots, N_T)$

1: For each $t \in \{1, \cdots, T\}$, draw N_t independent samples $d_t^1, \cdots, d_t^{N_t}$ from D_t.
2: Construct the empirical distribution $\hat{D}_t$:

$$\mathbb{P}[\hat{D}_t = d] = \frac{\sum_{i=1}^{N_t} \mathbf{1}[d = d_t^i]}{N_t}.$$

3: Define $\tilde{V}_{T+1}^{\mathrm{r}}(x) = 0$ for all x.
4: **for** $t = T, \cdots, 1$ **do**
5: Construct the right derivative function $\tilde{U}_t^{\mathrm{r}}(y_t) = \hat{C}_t^{\mathrm{r}}(y_t) + \mathbb{E}\tilde{V}_{t+1}^{\mathrm{r}}(y_t - \hat{D}_t)$.
6: By a binary search on the break points of $\tilde{U}_t^{\mathrm{r}}$, compute the smallest $\tilde{R}_t \in \mathbb{R}$ such that $\tilde{U}_t^{\mathrm{r}}(\tilde{R}_t) \geq 0$.
7: Construct the following right derivative function $\hat{\tilde{V}}_t^{\mathrm{r}} : \mathbb{R} \to [-\sum_{s=t}^{T} b_s, \sum_{s=t}^{T} h_s]$

$$\hat{\tilde{V}}_t^{r}(x_t) = \begin{cases} \tilde{U}_t^{\mathrm{r}}(x_t + B_t) & \text{if } x_t \in (-\infty, \tilde{R}_t - B_t) \\ 0 & \text{if } x_t \in [\tilde{R}_t - B_t, \tilde{R}_t) \\ \tilde{U}_t^{\mathrm{r}}(x_t) & \text{if } x_t \in [\tilde{R}_t, \infty) \end{cases}.$$

8: (Sparsification) Now, for each x_t, define

$$\tilde{V}_t^{\mathrm{r}}(x_t) = \eta \lfloor \frac{1}{\eta} \hat{\tilde{V}}_t^{\mathrm{r}}(x_t) \rfloor.$$

9: **end for**
10: Return the base stocks $(\tilde{R}_1, \cdots, \tilde{R}_T)$.

In line 5, the algorithm construct the functions $\tilde{U}_t^{\mathrm{r}}$ based on the empirical right derivative $\hat{C}_t^{\mathrm{r}}$ as well as the function $\tilde{V}_{t+1}^{\mathrm{r}}$, which serves as a tractable uniform approximation to the original right derivative V_{t+1}^{r}. In line 6, we compute the smallest minimizer $\tilde{R}_t$ of $\tilde{U}_t^{\mathrm{r}}$. As demonstrated proof of Lemma 4.1 (Appendix G), $\tilde{R}_t$ is one of the break points of $\tilde{U}_t^{\mathrm{r}}$, and $\tilde{R}_t$ can be found in polynomial time.

In line 7, the algorithm computes the function $\hat{\tilde{V}}_t^{\mathrm{r}}$. While the function $\hat{\tilde{V}}_t^{\mathrm{r}}$ is a candidate for an approximation to the original right derivative V_t^{r}, we cannot use $\hat{\tilde{V}}_t^{\mathrm{r}}$ to construct $\tilde{U}_{t-1}^{\mathrm{r}}$ in the next iteration, since in doing so the number of breakpoints in the subsequent step functions will grow exponentially in $T - t$. This is further discussed in the proof of Lemma 4.1. Thus, in line 8,

we introduce a sparsification procedure to prevent the number of breakpoints in the step functions from increasing too rapidly. The sparsification step for period t can be interpreted as follows. First, we overlay the grid $\{\eta z : z \in \mathbb{Z}\}$ onto the range $[-\sum_{s=t}^{T} b_s, \sum_{s=t}^{T} h_s]$ of $\hat{\tilde{V}}_t^{\mathrm{r}}$. Next, for each x_t, we define the value of $\tilde{V}_t^{\mathrm{r}}(x_t)$ to be the closest grid point to $\hat{\tilde{V}}_t^{\mathrm{r}}(x_t)$ from below. That is,

$$\tilde{V}_t^r(x_t) \in \{\eta z : z \in \mathbb{Z}\} \cap \left[-\sum_{s=t}^{T} b_s, \sum_{s=t}^{T} h_s\right], \quad 0 \le \hat{\tilde{V}}_t^r(x_t) - \tilde{V}_t^r(x_t) < \eta \text{ for all } x_t.$$

This rounding down procedure keeps the number of breakpoints in control while maintaining a uniform approximation to $\hat{\tilde{V}}_t^{\mathrm{r}}$ (hence also a uniform approximation to V_t^{r}), as proven in Lemma 4.1 and Theorem 4.2.

We remark that if the sparsification step is removed from $Sample(\eta, N_1, \cdots, N_T)$, the algorithm is equivalent to the SAA method, which solves the empirical problem to optimality. Nevertheless, by having the sparsification step, we argue that the algorithm $Sample(\eta, N_1, \cdots, N_T)$ is efficient, while the performance guarantee is only slightly worse than the SAA method (with the same number of samples).

We now proceed to the analysis of $Sample(\eta, N_1, \cdots, N_T)$. We first demonstrate in Lemma 4.1 that it has a polynomial running time by proving that the functions $\tilde{U}_t^{\mathrm{r}}, \tilde{V}_t^{\mathrm{r}}$ (hence also the output $(\tilde{R}_1, \cdots, \tilde{R}_T)$) can be constructed efficiently.

Lemma 4.1. *The algorithm $Sample(\eta, N_1, \cdots, N_T)$ has running time polynomial in the parameters $\left(N_1, \cdots, N_T, T, \frac{\sum_{s=1}^{T} h_s + b_s}{\eta}, \log d_{max}, \log c^*\right)$, where d_{max} is the maximum value of drawn sample, and $c^* = \max_{t=1,\cdots,T} \max\{b_t, h_t\}$.*

The Lemma is proven in Appendix G. The analysis in the proof shows that, by the sparsification procedure, we have sparsify $\hat{\tilde{V}}_t^{\mathrm{r}}$, which has $O(N_t \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta})$ break points, to a simpler function $\tilde{V}_t^{\mathrm{r}}$, which has only $O(\frac{\sum_{s=t}^{T} h_s + b_s}{\eta})$ break points. This prevents the number of breakpoints in the step functions $\tilde{U}_t^{\mathrm{r}}, \tilde{V}_t^{\mathrm{r}}$ from growing exponentially in $T-t$ as we proceed from $t = T$ to $t = 1$, which was the case when we solve the sample average problem exactly.

Next, we establish in Theorem 4.2 the performance guarantee of the algorithm, by proving that $\tilde{U}_t^{\mathrm{r}}, \tilde{V}_t^{\mathrm{r}}$ uniformly approximate the original right derivatives $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}$ via a backward induction on t. This justifies the use of $\tilde{U}_t^{\mathrm{r}}, \tilde{V}_t^{\mathrm{r}}$ in place of $\hat{U}_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$, where the former guarantees the uniform approximations on the origin right derivative while maintaining an efficient run time, unlike the latter.

Theorem 4.2. *For $t = 1, \cdots, T$, let*

$$N_t = \max\left\{(h_t + b_t)^2, \left(\sum_{s=t+1}^{T} h_s + b_s\right)^2\right\} \frac{4}{\alpha_t^2} \log \frac{4T}{\delta},$$

where α_t is as defined in (3.6). Let

$$\eta = \alpha_1 = \frac{\epsilon \min_{t \in \{1,\cdots,T\}} \{\min\{h_t, b_t\}\}}{6T^2}. \tag{4.1}$$

With probability at least $1-\delta$*, the algorithm* $Sample(\eta, N_1, \cdots, N_T)$ *returns a set of* $(1+2\epsilon)$*-optimal base stock* $(\tilde{R}_1, \cdots, \tilde{R}_T)$*.*

The proof of Theorem 4.2 can be found in Appendix H. The proof shares some similarity with the proof of Lemma 3.5. However, in the proof for the Theorem, we need to take into account the sparsification step, and the proof essentially argue that the sparsification does not affect the uniform approximation on the right derivatives of the cost-to-go functions for the original DP. Altogether, Lemma 4.1 and Theorem 4.2 respectively establish the computational efficiency and performance guarantee of *Sample*, hence establishing Theorem 2.4.

5 Insights for the Hardness Results

In this Section, we provide insight into the proofs of Lemma 2.3 and Theorem 2.5. Lemma 2.3 and Theorem 2.5 respectively assert the computational and information theoretic hardness of the data driven capacitated inventory control problem. They complement our analysis on the SAA method in §3, as well as the design and analysis of our algorithm *Sample* in §4.

Lemma 2.3 asserts that the SAA problem is $\#\mathbf{P}-$hard. We first review the notion of $\#\mathbf{P}-$hardness in Appendix I, and then provide the proof for Lemma 2.3 in Appendix J. Lemma 2.3 is proved by reducing the data-driven capacitated inventory control problem to the K^{th} largest subset problem, which is $\#\mathbf{P}$-hard. The K^{th} largest subset problem is stated as follows:

K^{th} largest subset problem: We are given a list of T positive integers $\{a_t\}_{t=1}^T$, an integer capacity R, and an integer $1 \le K \le 2^T$. The objective is to decide whether the assertion

$$\left| \left\{ S \subset \{1, \cdots, T\} : \sum_{t \in S} a_t \le R \right\} \right| \ge K \tag{5.1}$$

is true.

The K^{th} largest subset problem is demonstrated to be $\#\mathbf{P}-$hard in [GJ79] (see SP20 on page 225 in the reference). The problem is also used in [HKM$^+$09] to show that the inventory control problem under oracle access to the cumulative distribution functions of the random demands is $\#\mathbf{P}-$hard.

Next, we discuss the proof for Theorem 2.5, which is provided in Appendix K. The Theorem provides a lower bound on the number of samples necessary for solving the data-driven newsvendor problem to near optimality, for any level of pre-specified confidence probability. The basic idea of the proof is to reduce the data-driven optimization problem to a statistical classification problem.

More precisely, suppose we have an algorithm $\mathcal{A}$ that returns a $(1+\epsilon/20)$-optimal base stock for the data-driven newsvendor problem using m samples, under *any* demand distribution, with probability 0.999. Then, $\mathcal{A}$ still returns a $(1+\epsilon/20)$-optimal base stock when it is provided with m samples from demand distributions D_1 or D_2. Here, D_1 and D_2 are designed such that they have disjoint sets of $(1+\epsilon/20)-$optimal base stocks, but D_1, D_2 are statistically far apart.

Now, consider the following statistical classification problem. We are given m i.i.d. samples, and we are told that these samples are either from D_1 or D_2 (we know the cdfs of D_1, D_2). The problem is: can we correctly identify if these samples are from D_1 or D_2 with probability 0.999?

On one hand, we can identify the distribution as follows. We first run $\mathcal{A}$ on the m samples, which returns a base stock R. If R belongs to the set of $(1+\epsilon/20)-$optimal base stocks for D_1, we declare that the data is from D_1; otherwise, we declare that the data is from D_2. By the disjointness of the sets of optimal base stocks, the decision rule is well defined. Importantly, by the near optimality of $\mathcal{A}$, the decision rule is correct with probabiltiy 0.999. On the other hand, we use the statistical closeness of D_1, D_2 to argue that, in order to differentiate between D_1, D_2 with probability 0.999, the number of samples m must be bounded from below. This establishes the required lower bound.

6 Simulation Results

In this Section, we compare our theoretical results on the performance of Algorithm *Sample* with simulation results. We consider three families of 5 period capacitated inventory control problems. In all these problems, the starting inventory in period 1 is zero, i.e. $x_1 = 0$; the unit holding cost is set to be $h_t = h = 1$ while we vary the unit backlog cost $b_t = b = 1, 5,$ or 9. Finally, the capacities are set to be $B_1 = B_2 = B_3 = 1.4 \times 10^4$, and $B_4 = B_5 = 1.6 \times 10^4$ respectively.

Each problem family has a fixed set of discrete underlying random demand distributions. The set of demand distributions associated with each family is listed below:

1. The first family (U): D_1, D_2, D_3 are distributed as U$[0, 3 \times 10^4]$, D_4, D_5 are distributed as U$[2.5 \times 10^4, 5 \times 10^4]$.

2. The second family (P): D_1, D_2, D_3 are distributed as Poisson(1.5×10^4), D_4, D_5 are distributed as Poisson(3.75×10^4).

3. The third family (B): D_1, D_2 are distributed as U$[0, 3 \times 10^4]$, D_3 is distributed as Poisson(1.5×10^4), D_4, D_5 are distributed as Poisson(3.75×10^4).

The number of samples drawn from the demand in each period is 5×10^4, namely $N_t = 5 \times 10^4$ for $t = 1, \cdots, 5$.

For each family and each choice of unit backlog cost b, we compute the *theoretical relative ratio*, which is the performance ratio predicted by our analysis, as well as the *simulated relative ratio*, which is the performance ratio associated with the simulation.

The theoretical relative ratio is an upper bound on the ratio

$$\frac{\mathbb{E}\text{Cost}_1(0, \hat{R}_1, \cdots, \hat{R}_5)}{\text{Cost}_1(0, R_1^*, \cdots, R_5^*)} \tag{6.1}$$

predicted by our analysis. Recall that the notation $\text{Cost}_1(0, R_1, \cdots, R_5)$, which is defined in Section 5.3, denotes the expected cost under policy $(R_1, \cdots, R_5)$ when the starting inventory level x_1 in period 1 is zero. The numerator is the expected cost under the empirical policy $(\hat{R}_1, \cdots, \hat{R}_5)$ returned by the SAA method using $N_t = 5 \times 10^4$ samples for all t. The denominator is the expected cost under the optimal policy $(R_1^*, \cdots, R_5^*)$.

While Theorem 2.2 implies an upper bound on the relative error $(1 + \epsilon)$ for any given choice of $\{N_t\}_{t=1}^T$, our analysis in fact implies a stronger bound. (The Theorem states a weaker bound for

b	Simulated Rel Ratio			Theoretical Relative Ratio
	U	P	B	-
1	1.0060	1.0009	1.0081	2.1019
5	1.0147	1.0007	1.0312	6.3634
9	1.0165	1.0008	1.0268	14.7103

Table 1: Simulated and Theoretical Relative Ratios.

the sake of clarity.) By a more careful analysis, we have the following stronger bound

$$\frac{\mathbb{E}\text{Cost}_1(0,\hat{R}_1,\cdots,\hat{R}_5)}{\text{Cost}_1(0,R_1^*,\cdots,R_5^*)} \leq \prod_{t=1}^{5}\left(1+\frac{3(b+1)(5-t+1)\max\{1,5-t\}}{\sqrt{50000}}\sqrt{\log 2000}\right). \tag{6.2}$$

We use the bound at the right hand side of (6.2) as the theoretical relative ratio, for each choice of unit backlog cost b. Note that the bound is independent of the choice of the underlying distribution.

The simulated relative ratio is defined as:

$$\frac{\widehat{\text{Cost}}_1(0,\hat{R}_1^{Sam},\cdots,\hat{R}_5^{Sam})}{\text{Cost}_1(0,R_1^*,\cdots,R_5^*)}.$$

The empirical policy $(\hat{R}_1^{Sam},\cdots,\hat{R}_5^{Sam})$ is computed using Algorithm *Sample* defined in §4, with accuracy parameter $\eta = 0.001$. For each period t, $N_t = 5\times 10^4$ samples are drawn from D_t. The numerator $\widehat{\text{Cost}}_1(0,\hat{R}_1,\cdots,\hat{R}_5)$ is the cost under the empirical policy, averaged over 10^4 random realizations of $(D_1,\cdots,D_5)$. The denominator is the expected cost under the optimal policy $(R_1^*,\cdots,R_5^*)$, which is the same as the denominator in (6.1).

Table 1 shows the simulated and theoretical relative ratios of the families U, P and B under different choices of unit backlog cost b. The table suggests that Algorithm *Sample* has good performance in simulation. In general, the relative ratios in simulations are lower than their counterparts predicted by our analysis, implying that the theoretical bounds are quite conservative. This is primarily due to the looseness in our analysis, especially in Claim 3.2 and Lemma 3.4.

Levi et al. [LPU15] also report that the performance of the SAA method on the data-driven newsvendor problem is significantly better in simulation than predicted by the analysis in [LRS07]. [LPU15] provides the following explanation for the difference between theoretical and empirical performance. While [LRS07] draw sufficiently many samples to ensure an accurate estimation on the subgradient of the newsvendor cost function (which implies the near optimality of the empirical base stock), [LPU15] argue that such an accurate estimation may not be always necessary. Instead, [LPU15] correct this conservatism by directly approximating the set of $(1+\epsilon)-$optimal base stocks through the second-order Taylor series expansion on the newsvendor cost function. This results in a sample bound that depends on a distribution parameter. Through extensive simulation experiments, it is demonstrated in [LPU15] that the performance predicted by the second-order approach by [LPU15] is often much closer to the performance in simulation than the performance predicted by [LRS07]. It is interesting to see if the approach by [LPU15] can be extended to our setting.

Another possible reason for the discrepancy is that our analysis of the SAA method needs to be valid for any demand distribution with finite mean. In particular, we do not assume that the

demand distributions belong to any parametric family or even have bounded support. This leads to a more conservative bound, which could have been tighter using additional properties about the underlying demand distributions, on the theoretical performance ratio.

7 Conclusion

In this paper, we considered the capacitated inventory control problem in a data-driven setting. It is shown that with polynomially many samples, the Sample Average Approximation method outputs a near optimal policy with high probability. Nevertheless, solving the underlying SAA problem is computationally intractable, which motivates us to design a polynomial time approximation scheme by modifying the SAA method.

Altogether, our work and the paper [LRS07] demonstrate the tractability of certain class of multi-stage data driven stochastic optimization problems. This is in contrast to [SN05] which argues that in full generality these problems are intractable. We hope that our approach can be generalized to a broader class of dynamic programming problem. However, the major obstacles are the transformation from first order approximation to zero order approximation (cf. Claim 3.2), which seems to only hold in inventory related problems, as well as the validity of Lemma 3.4, Theorem 3.8 in other dynamic programming models. Another interesting direction is to investigate if the machinery in [LPU15] can be used to improve the sample bound in Theorem 2.2, since the sample bounds in Theorem 2.2, 2.4 are found to be too conservative in our simulation.

References

[AF97] Yossi Aviv and Awi Federgruen. Stochastic inventory models with limited production capacity and periodically varying parameters. *Probability in the Engineering and Informational Sciences*, 11:107–135, 1 1997.

[BLQ12] Mehmet A. Begen, Retsef Levi, and Maurice Queyranne. Technical notea sampling-based approach to appointment scheduling. *Operations Research*, 60(3):675–681, 2012.

[BQ11] Mehmet A. Begen and Maurice Queyranne. Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, 36(2):240–257, 2011.

[CCP05] Moses Charikar, Chandra Chekuri, and Martin Pal. Sampling bounds for stochastic optimization. In *PROC. 9TH RANDOM*, pages 257–269. Springer, 2005.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Wiley, 1979.

[GPRS04] Anupam Gupta, Martin Pál, R. Ravi, and Amitabh Sinha. Boosted sampling: Approximation algorithms for stochastic optimization. In *Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing*, STOC '04, pages 417–426, 2004.

[GPRS05] Anupam Gupta, Martin Pl, R. Ravi, and Amitabh Sinha. What about wednesday? approximation algorithms for multistage stochastic optimization. In *in APPROX*, pages 86–98, 2005.

[GRS07] Anupam Gupta, R. Ravi, and Amitabh Sinha. Lp rounding approximation algorithms for stochastic network design. *Mathematics of Operations Research*, 32(2):345–364, 2007.

[HKL+14] N. Halman, D. Klabjan, C. Li, J. Orlin, and D. Simchi-Levi. Fully polynomial time approximation schemes for stochastic dynamic programs. *SIAM Journal on Discrete Mathematics*, 28(4):1725–1796, 2014.

[Hal15] Nir Halman. Provably near-optimal approximation schemes for sample-based dynamic programs with emphasis on stochastic inventory control models. *Manuscript*, 2015.

[HKM+09] Nir Halman, Diego Klabjan, Mohamed Mostagir, Jim Orlin, and David Simchi-Levi. A fully polynomial-time approximation scheme for single-item stochastic inventory control with discrete demand. *Mathematics of Operations Research*, 34(3):674–685, 2009.

[IKMM04] Nicole Immorlica, David Karger, Maria Minkoff, and Vahab S. Mirrokni. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '04, pages 691–700, 2004.

[KT98] Roman Kapuscinski and Sridhar Tayur. A capacitated production-inventory model with periodic demand. *Operations Research*, 46:899–911, 1998.

[KSHdM02] Anton J. Kleywegt, Alexander Shapiro, and Tito Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. on Optimization*, 12(2):479–502, February 2002.

[LPU15] Retsef Levi, Georgia Perakis, and Joline Uichanco. The data-driven newsvendor problem: New bounds and insights. *Operations Research*, 63(6):1294–1306, 2015.

[LRS06] Retsef Levi, Robin O. Roundy, and David B. Shmoys. Provably near-optimal sampling-based algorithms for stochastic inventory control models. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 739–748, 2006.

[LRS07] Retsef Levi, Robin O. Roundy, and David B. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32(4):821–839, 2007.

[LRST08] Retsef Levi, Robin O. Roundy, David B. Shmoys, and Van Anh Truong. Approximation algorithms for capacitated stochastic inventory control models. *Operations Research*, 56(5):1184–1199, 2008.

[Mas90] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283, 07 1990.

[RS06] R. Ravi and Amitabh Sinha. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. *Mathematical Programming*, 108(1):97–114, 2006.

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyski. *Lectures on Stochastic Programming.* Society for Industrial and Applied Mathematics, 2009.

[Sha06] Alexander Shapiro. On complexity of multistage stochastic programs. *Oper. Res. Lett.*, 34(1):1–8, January 2006.

[SN05] Alexander Shapiro and Arkadi Nemirovski. On complexity of stochastic programming problems. In Vaithilingam Jeyakumar and Alexander Rubinov, editors, *Continuous Optimization*, volume 99 of *Applied Optimization*, pages 111–146. Springer US, 2005.

[SS06] David B. Shmoys and Chaitanya Swamy. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *Journal of ACM*, 53(6):978–1012, November 2006.

[SS12] C. Swamy and D. Shmoys. Sampling-based approximation algorithms for multistage stochastic optimization. *SIAM Journal on Computing*, 41(4):975–1004, 2012.

[Tay93] Sridhar R. Tayur. Computing the optimal policy for capacitated inventory models. *Communications in Statistics. Stochastic Models*, 9(4):585–598, 1993.

A Proof of Claim 3.2

We apply Lemma 3.3 to prove the claim. First, recall that U_t is a convex function. Now, we claim that there exists $s \in \partial U_t(\hat{R}_t)$ such that $|s| \leq \eta$, where $\partial U_t(\hat{R}_t)$ is the set of subgradients of U_t at $\hat{R}_t$. By the assumption in the theorem, we know that

$$\left|\hat{U}_t^{\mathrm{r}}(\hat{R}_t) - U_t^{\mathrm{r}}(\hat{R}_t)\right| \leq \eta.$$

Now, by the convexity of U_t, we know that for all $y \in \mathbb{R}$, we have $\lim_{y \uparrow x} U_t^{\mathrm{r}}(y) = U_t^{\mathrm{l}}(x)$. By applying the assumption of the Theorem on an increasing sequence that converges to $\hat{R}_t$, we also have

$$\left|\hat{U}_t^{\mathrm{l}}(\hat{R}_t) - U_t^{\mathrm{l}}(\hat{R}_t)\right| \leq \eta.$$

Now, by the definition of $\hat{R}$, we know that $0 \in [\hat{U}_t^{\mathrm{l}}(\hat{R}_t), \hat{U}_t^{\mathrm{r}}(\hat{R}_t)]$. But this implies that there exists a number s such that $|s| \leq \eta$ and $s \in [U_t^{\mathrm{l}}(\hat{R}_t), U_t^{\mathrm{r}}(\hat{R}_t)] = \partial U_t(\hat{R}_t)$, which proves the the existence of such a subgradient.

Finally, note that $U_t(y_t) \geq h_t(y_t - \mathbb{E}[D_t])^+ + b_t(\mathbb{E}[D_t] - y_t)^+$, thus applying by Lemma 3.3 we have

$$U_t(\hat{R}_t) \leq \left(1 + \frac{3\eta}{\min\{b_t, h_t\}}\right) U_t(R_t^*),$$

which proves the claim. □

B Proof of Lemma 3.4

First, by the definition of a modified base stock policy, the following equation holds:

$$Cost_t(x_t; R_t, \cdots, R_T) = \begin{cases} \mathbb{E}\left[C_t(x_t + B_t - D_t) + Cost_{t+1}(x_t + B_t - D_t; R_{t+1}, \cdots, R_T)\right] & \text{if } x_t \in (-\infty, R_t - B_t] \\ \mathbb{E}\left[C_t(\hat{R}_t - D_t) + Cost_{t+1}(R_t - D_t; R_{t+1}, \cdots, R_T)\right] & \text{if } x_t \in (R_t - B_t, R_t] \\ \mathbb{E}\left[C_t(x_t - D_t) + Cost_{t+1}(x_t - D_t; R_{t+1}, \cdots, R_T)\right] & \text{if } x_t \in (R_t, \infty) \end{cases}. \quad \text{(B.1)}$$

Next, we will prove by a backward induction from $t = T$ to $t = 1$ that, for all starting inventory level x_t in period t the following inequality holds:

$$Cost_t(x_t; R_t, \cdots, R_T) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) Cost_t(x_t; R_t^*, \cdots, R_T^*) = \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t). \quad \text{(B.2)}$$

Now, suppose (B.2) holds for period $t+1$, and we will prove that (B.2) is also true for period t. First, we know that for all $y_t \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}C_t(y_t - D_t) + \mathbb{E}Cost_{t+1}(y_t - D_t; R_{t+1}, \cdots, R_T) \\
\leq &\mathbb{E}C_t(y_t - D_t) + \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) \mathbb{E}Cost_{t+1}(y_t - D_t; R^*_{t+1}, \cdots, R^*_T) \\
\leq &\left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) \left(\mathbb{E}C_t(y_t - D_t) + \mathbb{E}Cost_{t+1}(y_t - D_t; R^*_{t+1}, \cdots, R^*_T)\right) \\
= &\left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(y_t).
\end{aligned}$$

Thus, by (B.1), the following inequality holds:

$$Cost_t(x_t; R_t, \cdots, R_T) \leq \begin{cases} \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t + B_t) & \text{if } x_t \in (-\infty, R_t - B_t] \\ \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(R_t) & \text{if } x_t \in (R_t - B_t, R_t] \\ \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t) & \text{if } x_t \in (R_t, \infty) \end{cases}. \tag{B.3}$$

To prove the induction claim (B.2), we first note that by our assumption on $\epsilon_1, \cdots, \epsilon_T$, the following holds for all $t \in \{1, \cdots, T-1\}$:

$$\left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(R_t) \leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) \left(1 + \frac{\epsilon_t}{2}\right) U_t(R^*_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) U_t(R^*_t)$$

Now, consider the following two cases:

1. **Case 1: We have** $R_t < R^*_t$. We further consider the following 4 subcases:

   (a) We have $x_t \leq R_t - B_t$. Then we also have $x_t \leq R^*_t - B_t$. Therefore,

   $$Cost_t(x_t; R_t, \cdots, R_T) \leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t + B_t) = \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) V_t(x_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t).$$

   (b) We have $R_t - B_t < x_t \leq R_t$. Then we have

   $$Cost_t(x_t; R_t, \cdots, R_T) \leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(R_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) U_t(R^*_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t).$$

(c) We have $R_t < x_t \leq R_t^*$. Then we have

$$\begin{aligned} Cost_t(x_t; R_t, \cdots, R_T) &\leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t) \\ &\overset{(\dagger)}{\leq} \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(R_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) U_t(R_t^*) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t), \end{aligned}$$

where (†) holds since we know that $U_t(R_t) \leq U_t(x_t) \leq U_t(R_t^*)$ by the convexity of U_t.

(d) We have $R_t^* < x_t$. Then we also have $R_t < x_t$. Thus,

$$Cost_t(x_t; R_t, \cdots, R_T) \leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t) = \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) V_t(x_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t).$$

2. **Case 2: We have $R_t > R_t^*$.** We further consider the following 4 subcases:

(a) We have $x_t \leq R_t^* - B_t$. Then we also have $x_t \leq R_t - B_t$, and the induction claim (B.2) holds true by the same reasoning as in Subcase (a) in Case 1.

(b) We have $R_t^* - B_t < x_t \leq R_t - B_t$. Then we have,

$$\begin{aligned} Cost_t(x_t; R_t, \cdots, R_T) &\leq \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(x_t + B_t) \\ &\overset{(\ddagger)}{\leq} \left(1 + \sum_{s=t+1}^{T} \epsilon_s\right) U_t(R_t) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) U_t(R_t^*) \leq \left(1 + \sum_{s=t}^{T} \epsilon_s\right) V_t(x_t), \end{aligned}$$

where in (‡) we know that $U_t(x_t + B_t) \leq U_t(R_t)$, since the subcase assumption clearly implies that $R_t^* < x_t + B_t \leq R_t$, and U_t is convex.

(c) We have $R_t - B_t < x_t \leq R_t$. This is identical to Subcase (b) in Case 1, therefore the induction claim (B.2) is still true.

(d) We have $R_t < x_t$. Then we also have $R_t^* < x_t$, and the induction claim (B.2) holds true by the same reasoning as in Subcase (d) in Case 1.

Altogether, we have established the induction claim (B.2), which proves the Lemma. □

C Proof of Theorem 3.6

By applying triangle inequality twice, we know that

$$\begin{aligned}
&\left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \\
\leq &\left|\hat{C}_t^{\mathrm{r}}(y_t) - C_t^{\mathrm{r}}(y_t)\right| + \left|\mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - \hat{D}_t) - \mathbb{E}V_{t+1}^{\mathrm{r}}(y_t - D_t)\right| \\
\leq &\underbrace{\left|\hat{C}_t^{\mathrm{r}}(y_t) - C_t^{\mathrm{r}}(y_t)\right|}_{(a)} + \underbrace{\left|\mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - \hat{D}_t) - \mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - D_t)\right|}_{(b)} + \underbrace{\left|\mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - D_t) - \mathbb{E}V_{t+1}^{\mathrm{r}}(y_t - D_t)\right|}_{(c)}.
\end{aligned}$$

First, we analyze the first term (a). By the definitions of $\hat{C}_t^{\mathrm{r}}(y_t), C_t^{\mathrm{r}}(y_t)$ in (3.4), we assert the following equality:

$$\begin{aligned}
&\left|\hat{C}_t^{\mathrm{r}}(y_t) - C_t^{\mathrm{r}}(y_t)\right| \\
=&\left|\left\{-b_t + (h_t + b_t)\frac{1}{N_t}\sum_{i=1}^{N_t} \mathbf{1}(d_t^i \leq y_t)\right\} - \{-b_t + (h_t + b_t)\Pr[D_t \leq y_t]\}\right| \\
=&(h_t + b_t)\left|\frac{1}{N_t}\sum_{i=1}^{N_t} \mathbf{1}(d_t^i \leq y_t) - \Pr[D_t \leq y_t]\right|.
\end{aligned}$$

Thus, by invoking Massart's Theorem and our choice of N_t, we have

$$\mathbb{P}\left[\left|\hat{C}_t^r(y_t) - C_t^r(y_t)\right| \leq \frac{\alpha_t}{2}\right] \geq 1 - 2\exp\left[-\frac{\alpha_t^2}{4(h_t + b_t)^2}N_t\right] \geq 1 - \frac{\delta_t}{4}.$$

Next, we analyze the second term (b). As remarked previously, the empirical right derivative $\hat{V}_{t+1}^{\mathrm{r}}$ is a step function with range $[-\sum_{s=t+1}^T b_s, \sum_{s=t+1}^T h_s]$, and it has finitely many break points. Let's denote

$$\hat{V}_{t+1}^{\mathrm{r}}(y) = \sum_{j=0}^{M-1} \ell_j \mathbf{1}(\beta_j \leq y < \beta_{j+1}),$$

where $\beta_1 < \cdots < \beta_{M-1}$ are the breakpoints of $\hat{V}_{t+1}^{\mathrm{r}}(y)$, and for notational convenience we define $\beta_0 = -\infty, \beta_M = \infty$. Note that $\ell_0 = -\sum_{s=t+1}^T b_s$, and $\ell_{M-1} = \sum_{s=t+1}^T h_s$. Thus, the second term

can be bounded as follows:

$$
\begin{aligned}
\text{(b)} &= \left| \sum_{j=0}^{M-1} \ell_j \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{1}(\beta_j \le y - d_t^i < \beta_{j+1}) - \mathbb{P}[\beta_j \le y - D_t < \beta_{j+1}] \right) \right| \\
&= \left| \sum_{j=0}^{M-1} \ell_j \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \big(\mathbf{1}(d_i^t \le y - \beta_j) - \mathbf{1}(d_t^i \le y - \beta_{j+1})\big) - (\mathbb{P}[D_t \le y - \beta_j] - \mathbb{P}[D_t \le y - \beta_{j+1}]) \right) \right| \\
&= \left| \sum_{j=0}^{M-2} (\ell_{j+1} - \ell_j) \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{1}(d_t^i \le y - \beta_{j+1}) - \mathbb{P}[D_t \le y - \beta_{j+1}] \right) \right| \\
&\le \left(\sum_{j=0}^{M-2} \ell_{j+1} - \ell_j \right) \sup_{y \in \mathbb{R}} \left| \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{1}(d_t^i \le y) - \mathbb{P}[D_t \le y] \right| \\
&= \left(\sum_{s=t+1}^{T} h_s + b_s \right) \sup_{y \in \mathbb{R}} \left| \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{1}(d_t^i \le y) - \mathbb{P}[D_t \le y] \right|.
\end{aligned}
$$

By applying Massart's Theorem, the second term satisfies the following probability bound:

$$
\begin{aligned}
\mathbb{P}\left[\text{(b)} \le \frac{\alpha_t}{2}\right] &\ge \mathbb{P}\left[\sup_{y \in \mathbb{R}} \left| \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbf{1}(d_t^i \le y) - \mathbb{P}[D_t \le y] \right| \le \frac{\alpha_t}{2\left(\sum_{s=t+1}^{T} h_s + b_s\right)} \right] \\
&\ge 1 - \exp\left[-\frac{\alpha_t^2}{4\left(\sum_{s=t+1}^{T} h_s + b_s\right)^2} N_t \right] \ge 1 - \frac{\delta_t}{4},
\end{aligned}
$$

where the last inequality holds by our choice of N_t.

Lastly, for the analysis of the third term (c), by the Theorem's assumption we know that $\left|\mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - D_t) - \mathbb{E}V_{t+1}^{\mathrm{r}}(y_t - D_t)\right| \le \gamma_t$ with probability 1.

Altogether, we have the following guarantee for the right derivatives, which prove the Theorem:

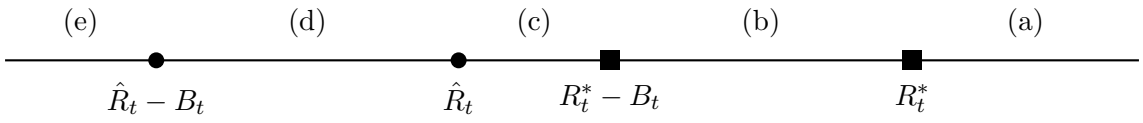
$$
\mathbb{P}\left[\text{For all } y_t,\ \left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \le \gamma_t + \alpha_t\right] \ge \mathbb{P}\left[\text{For all } y_t,\ \text{(a)} \le \frac{\alpha_t}{2},\ \text{(b)} \le \frac{\alpha_t}{2},\ \text{(c)} \le \gamma_t\right] \ge 1 - \delta_t/2.
$$

□

D Proof of Theorem 3.8

First, recall the definitions that R_t^*, $\hat{R}_t$ are the *smallest* minimizers of $U_t(y_t), \hat{U}_t(y_t)$ respectively. This implies that $U_t^{\mathrm{r}}(y_t) < 0$ for all $y_t < R_t^*$, and $U_t^{\mathrm{r}}(y_t) \ge 0$ for all $y_t \ge R_t^*$. Similar inequalities hold for the empirical counterpart. In the following, we will repeatedly recall the right derivatives $V_t^{\mathrm{r}}, \hat{V}_t^{\mathrm{r}}$ in terms of $U_t^{\mathrm{r}}, \hat{U}_t^{\mathrm{r}}$ in (3.2), (3.3). Consider the following 4 cases on $R_t^*, \hat{R}_t$:

1. **Case 1:** We have $\hat{R}_t \le R_t^* - B_t$. We further consider the subcases (a) to (e) as depicted here.



(a) We have $R_t^* \le x_t$. Then we have $V_t^{\mathrm{r}}(x_t) = U_t^{\mathrm{r}}(x_t)$ and $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t)$. Therefore

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \left|\hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t)\right|.$$

(b) We have $R_t^* - B_t \le x_t < R_t^*$. Then we have $V_t^{\mathrm{r}}(x_t) = 0 > U_t^{\mathrm{r}}(x_t)$, and $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t) \ge 0$. Thus

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) \le \hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t) = \left|\hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t)\right|.$$

(c) We have $\hat{R}_t \le x_t < R_t^* - B_t$. On one hand, we know that $V_t^{\mathrm{r}}(x_t) = U_t^{\mathrm{r}}(x_t + B_t)$. Since $x_t \le x_t + B_t < R_t^*$, by the convexity of U_t, we have $U_t^{\mathrm{r}}(x_t) \le U_t^{\mathrm{r}}(x_t + B_t) < 0$. On the other hand, we know that $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t) \ge 0$. Therefore,

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t + B_t) \le \hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t) = \left|\hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t)\right|.$$

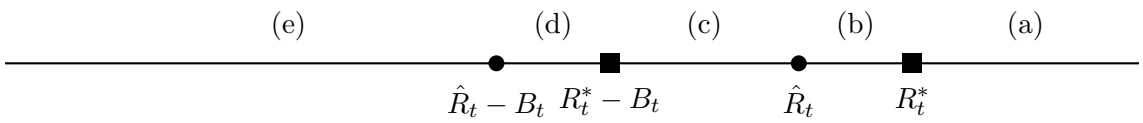
(d) We have $\hat{R}_t - B_t \le x_t < \hat{R}_t$. Then $V_t^{\mathrm{r}}(x_t) = U_t^{\mathrm{r}}(x_t + B_t) \le U_t^{\mathrm{r}}(R_t) \le 0$, and $\hat{V}_t^{\mathrm{r}}(x_t) = 0 \le \hat{U}_t^{\mathrm{r}}(x_t + B_t)$. Therefore

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) \le \hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t) \le \left|\hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t)\right| \le \eta_t.$$

(e) We have $x_t \le \hat{R}_t - B_t$. Then $V_t^{\mathrm{r}}(x_t) = U_t^{\mathrm{r}}(x_t + B_t)$ and $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t + B_t)$. Therefore

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \left|\hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t)\right| \le \eta_t.$$

2. **Case 2:** We have $R_t^* - B_t < \hat{R}_t \le R_t^*$. We further consider the subcases (a) to (e) as depicted



in here. Note that the proofs for subcases (a) and (e) are identical to those in **case 1**, thus we focus on proving the bound for subcases (b), (c) and (d):

(a) Same as (a) in **case 1**.

(b) We have $\hat{R}_t \le x_t < R_t^*$. Then $V_t^{\mathrm{r}}(x_t) = 0 \ge U_t^{\mathrm{r}}(x_t)$, and $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t) \ge 0$. Therefore

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) \le \hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t) = \left|\hat{U}_t^{\mathrm{r}}(x_t) - U_t^{\mathrm{r}}(x_t)\right|.$$

(c) We have $R_t^* - B_t \le x_t < \hat{R}_t$. Then $V_t^{\mathrm{r}}(x_t) = 0 = \hat{V}_t^{\mathrm{r}}(x_t)$.

(d) We have $\hat{R}_t - B_t \le x_t < R_t^* - B_t$. Then $V_t^{\mathrm{r}}(x_t) = U_t^{\mathrm{r}}(x_t + B_t) < 0$, and $\hat{V}_t^{\mathrm{r}}(x_t) = 0 \le \hat{U}_t^{\mathrm{r}}(x_t + B_t)$. Therefore

$$\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) \le \hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t) \le \left|\hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t)\right| \le \eta_t.$$

(e) Same as (e) in **case 1**.

3. **Case 3:** We have $\hat{R}_t - B_t \le R_t \le \hat{R}_t$. Then we can prove this case by interchanging R_t and $\hat{R}_t$ in **Case 2**.

4. **Case 4:** We have $R_t < \hat{R}_t - B_t$. Then we can prove this case by interchanging R_t and $\hat{R}_t$ in **Case 1**.

Finally, we note that the same approximation guarantees can be shown for the left derivatives. Altogether we have considered all the cases on $R_t, \hat{R}_t$, and the Theorem is proven. □

E Proof of Lemma 3.5

For a period t, we say that $\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}$ is $\{\alpha_s\}_{s=t}^{T}$-good if for all $s \in \{t, \cdots, T\}$, the followings hold simultaneously:

$$\left|U_s^{\mathrm{r}}(y_s) - \hat{U}_s^{\mathrm{r}}(y_s)\right| \le \sum_{r=s}^{T} \alpha_r \quad \forall y_s \in \mathbb{R}, \qquad \left|V_s^{\mathrm{r}}(x_s) - \hat{V}_s^{\mathrm{r}}(x_s)\right| \le \sum_{r=s}^{T} \alpha_r \quad \forall x_s \in \mathbb{R}.$$

Now, we have

$$\begin{aligned}
&\mathbb{P}\left[\text{For all } t \text{ and } y,\ \left|U_t^{\mathrm{r}}(y) - \hat{U}_t^{\mathrm{r}}(y)\right| \le \sum_{s=t}^{T} \alpha_s\right] \\
\ge &\mathbb{P}\left[\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=1}^{T} \text{ is } \{\alpha_s\}_{s=1}^{T}\text{-good}\right] \\
= &\prod_{t=1}^{T} \mathbb{P}\left[\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t}^{T} \text{ is } \{\alpha_s\}_{s=t}^{T}\text{-good} \middle| \left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T} \text{ is } \{\alpha_s\}_{s=t+1}^{T}\text{-good}\right].
\end{aligned}$$

By Theorem 3.6, conditioned on the event that $\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}$ is $\{\alpha_s\}_{s=t+1}^{T}$-good, we have

$$\mathbb{P}\left[\text{For all } y_t,\ \left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \le \alpha_t + \sum_{s=t+1}^{T} \alpha_s \mid \left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T} \text{ is } \{\alpha_s\}_{s=t+1}^{T} - \text{good}\right] \ge 1 - \frac{\delta}{T}.$$

Next, by Theorem 3.8, we know that if $\left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \le \sum_{s=t}^{T} \alpha_s$ holds for all $y_t \in \mathbb{R}$, then $\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| \le \sum_{s=t}^{T} \alpha_s$ holds for all $x_t \in \mathbb{R}$ with probability 1. Altogether, this shows that

for all t

$$\mathbb{P}\left[\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t}^{T} \text{ is } \{\alpha_s\}_{s=t}^{T}\text{-good} \middle| \left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T} \text{ is } \{\alpha_s\}_{s=t+1}^{T}\text{-good}\right] \geq 1 - \frac{\delta}{T}$$

and thus $\mathbb{P}\left[\text{For all } t \text{ and } y,\ \left|U_t^{\mathrm{r}}(y) - \hat{U}_t^{\mathrm{r}}(y)\right| \leq \sum_{s=t}^{T} \alpha_s\right] \geq 1 - \delta$. □

F Proof of Theorem 2.2

For each $t = 1, \cdots, T$, recall our choice

$$\alpha_t = \frac{\epsilon \min_{t\in\{1,\cdots,T\}} \{\min\{h_t, b_t\}\}}{6T^2}$$

in (3.6). The following holds:

$$\begin{aligned}
&\mathbb{P}\left[(\hat{R}_1, \cdots, \hat{R}_T) \text{ is a } (1+\epsilon)\text{-optimal modified base stock policy}\right] \\
\geq &\mathbb{P}\left[\text{For all } t \text{ and } y,\ \left|U_t^{\mathrm{r}}(y) - \hat{U}_t^{\mathrm{r}}(y)\right| \leq \frac{\epsilon \min_{t\in\{1,\cdot,T\}}\{\min\{h_t, b_t\}\}}{6T}\right] \qquad \text{(F.1)} \\
\geq &\mathbb{P}\left[\text{For all } t \text{ and } y,\ \left|U_t^{\mathrm{r}}(y) - \hat{U}_t^{\mathrm{r}}(y)\right| \leq \sum_{s=t}^{T} \alpha_s\right] \\
\geq &1 - \delta. \qquad \text{(F.2)}
\end{aligned}$$

Step (F.1) is by Lemma 3.1, with $\eta_t = \frac{\epsilon \min_{t\in\{1,\cdot,T\}}\{\min\{h_t,b_t\}\}}{6T}$. Step (F.2) is by our choice of N_t, cf. (2.8, 3.5). Altogether, Theorem 2.2 is established. □

G Proof of Lemma 4.1

We first note that $\tilde{U}_t^{\mathrm{r}}, \tilde{V}_t^{\mathrm{r}}$ are non-decreasing step functions. Next, we will prove the following additional properties:

- $\tilde{U}_t^{\mathrm{r}}$ has at most $O\left(N_t \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta}\right)$ break points,
- Given an explicit expression for $\tilde{V}_{t+1}^{\mathrm{r}}$, $\tilde{U}_t^{\mathrm{r}}$ can be constructed in time

$$O\left(N_t^2 \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta}\left(\log \frac{N_t \sum_{s=t+1}^{T} h_s + b_s}{\eta}\right) \log d_{\max} c^*\right),$$

- Given an explicit expression for $\tilde{U}_t^{\mathrm{r}}$, $\tilde{V}_t^{\mathrm{r}}$ can be constructed in time $O\left(N_t \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta} \log d_{\max} c^*\right)$,
- $\tilde{V}_t^{\mathrm{r}}$ has at most $O\left(\frac{\sum_{s=t}^{T} h_s + b_s}{\eta}\right)$ break points,

by a backward induction on t. Recall that $d_{\max}$ is the maximum value of the samples drawn, and $c^* = \max_{t=1,\cdots,T} \max\{h_t, b_t\}$.

First, this is clearly true for $t = T+1$, since $\tilde{V}^{\mathrm{r}}_{T+1} = 0$. Now, suppose that it is true for $\tilde{V}^{\mathrm{r}}_{t+1}$, we will show that the induction claim is also true for $\tilde{U}^{\mathrm{r}}_t$.

To prove the induction claims for $\tilde{U}^{\mathrm{r}}_t$, we will describe in details how $\tilde{U}^{\mathrm{r}}_t$ is constructed, given the samples $d^1_t, \cdots, d^{N_t}_t$ from D_t as well as an explicit expression for the step function $\tilde{V}^{\mathrm{r}}_{t+1}$. By the induction hypothesis for $t+1$, we are given the explicit expression

$$\tilde{V}^{\mathrm{r}}_{t+1}(x) = \sum_{j=0}^{M-1} \ell_j \mathbf{1}(\beta_j \le x < \beta_{j+1}),$$

where $\beta_1 < \cdots < \beta_{M-1}$ are the breakpoints of $\tilde{V}^{\mathrm{r}}_{t+1}(x)$. For notational convenience we define $\beta_0 = -\infty, \beta_M = \infty$. In addition, by the hypothesis we know that $M = O\left(\frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta}\right)$, and $\ell_0 < \ell_1 < \cdots, < \ell_{M-1}$.

Recall from line 5 of $Sample(\eta, N_1, \cdots, N_T)$ that $\tilde{U}^{\mathrm{r}}_t$ is defined as follows:

$$\tilde{U}^{\mathrm{r}}_t(y_t) = -b_t + (h_t + b_t)\left(\frac{1}{N_t}\sum_{i=1}^{N_t} \mathbf{1}(y_t \le d^t_i)\right) + \frac{1}{N_t}\sum_{i=1}^{N_t} \tilde{V}^{\mathrm{r}}_{t+1}(y_t - d^t_i).$$

This implies that $\tilde{U}^{\mathrm{r}}_t(y_t)$ has breakpoints $\left\{d^t_i\right\}_{i=1}^{N_t} \cup \left\{d^t_i + \beta_j\right\}_{i=1,j=1}^{N_t, M-1}$. Hence, this demonstrates the upper bound $O\left(N_t \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta}\right)$ on the number of breakpoints of $\tilde{U}^{\mathrm{r}}_t$.

Now, we also argue that an explicit expression of the step function $\tilde{U}^{\mathrm{r}}_t(y_t)$ can be constructed with the stated running time. First, sort the breakpoints of $\tilde{U}^{\mathrm{r}}_t$ in the increasing order $p_1, \cdots, p_N$, where we denote N as the number of breakpoints of $\tilde{U}^{\mathrm{r}}_t$. Next, we take a sequence of $N+1$ points $q_1, \cdots, q_{N+1}$ that sandwiches the breakpoints of $\tilde{U}^{\mathrm{r}}_t$, i.e. $q_1 < p_1 < q_2 < \cdots < p_N < q_{N+1}$. We know that $\tilde{U}^{\mathrm{r}}_t$ can be explicitly expressed as follows:

$$\tilde{U}^{\mathrm{r}}_t(y) = \sum_{j=0}^{N+1} \tilde{U}^{\mathrm{r}}_t(q_j) \mathbf{1}(p_j \le y < p_{j+1}),$$

where $p_0 = -\infty$ and $p_{N+1} = \infty$. Thus, to attain an explicit expression of $\tilde{U}^{\mathrm{r}}_t$, it suffices to evaluate the function $\tilde{U}_t(y_t)$ at the $N+1$ points $q_1, \cdots, q_{N+1}$. These $N+1$ evaluations require the computations of $\{C^{\mathrm{r}}_t(q_j)\}_{j=1}^{N+1}$, as well as the computations of $\tilde{V}^{\mathrm{r}}_{t+1}(q_j - d^1_t), \cdots, \tilde{V}^{\mathrm{r}}_{t+1}(q_j - d^{N_t}_t)$. The computations for C^{r}_t can be done by fist sorting $d^1_t, \cdots, d^{N_t}_t$, which takes time $O(N_t \log N_t \log d_{\max})$ (there are $O(N_t \log N_t)$ comparisons, and each takes $O(\log d_{\max})$), follows by the $(N+1)$ evaluations of C^{r}_t, which takes time $O(N \log N_t (\log d_{\max} + \log c^*))$. Thus, for C^{r}_t it takes time $O((N + N_t) \log N_t \log d_{\max} c^*)$ to compute. Next, for $\tilde{V}^{\mathrm{r}}_{t+1}$, it involves computing the function at $(N+1)N_t$ points, and each evaluation of $\tilde{V}^{\mathrm{r}}_{t+1}$ (when its explicit form is given) takes time $O\left(\log \frac{\sum_{s=t+1}^{T} h_s + b_s}{\eta} \log d_{\max} c^*\right)$, by binary search on the sorted break points of $\tilde{V}^{\mathrm{r}}_{t+1}$. Thus, the total time needed for $\tilde{V}^{\mathrm{r}}_{t+1}$ is

$O\left(NN_t \log \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta} \log d_{\max} c^*\right)$.

It is clear that the running time of the constructing $\tilde{U}_t^{\mathrm{r}}$ is dominated by the running time for the $N+1$ evaluations on $\tilde{U}_t(q_i)$s. Thus, given the explicit expression for $\tilde{V}_{t+1}^{\mathrm{r}}$, the run time for providing an explicit expression for $\tilde{U}_t^{\mathrm{r}}$ is at most

$$\begin{aligned} & O\left(\left((N+N_t)\log N_t + N_t N \log \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta}\right)\log d_{\max} c^*\right) \\ =& O\left(\left(N_t^2 \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta}\left(\log \frac{N_t \sum_{s=t+1}^{T} h_s+b_s}{\eta}\right)\right)\log d_{\max} c^*\right). \end{aligned}$$

Next, given the induction hypothesis for $\tilde{U}_t^{\mathrm{r}}$, we prove the induction claims relevant to $\tilde{V}_t^{\mathrm{r}}$. First, observe that $\tilde{R}_t$ is the smallest break point p_j of $\tilde{U}_t^{\mathrm{r}}$ such that $\tilde{U}_t^{\mathrm{r}}(p_j) \geq 0$, which can be found efficiently by binary search (which takes time $O\left(\log\left(N_t \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta}\right)\log d_{\max}\right)$), since $\tilde{U}_t^{\mathrm{r}}$ is non-decreasing. Now, given $\tilde{R}_t$ and the explicit expression of $\tilde{U}_t^{\mathrm{r}}$, by step 7 we can construct $\hat{\tilde{V}}_t^{\mathrm{r}}$ in $O\left(N_t \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta} \log(d_{\max} c^*)\right)$ time. Note that the number of break points of $\hat{\tilde{V}}_t^{\mathrm{r}}$ is equal to the number of break points of $\tilde{U}_t^{\mathrm{r}}$. Finally, to provide an explicit expression to $\tilde{V}_t^{\mathrm{r}}$ based on $\hat{\tilde{V}}\mathrm{r}_t$, it also requires only $O\left(N_t \frac{\sum_{s=t+1}^{T} h_s+b_s}{\eta} \log(d_{\max} c^*)\right)$ time, since to perform the sparsification procedure, it suffices to perform it on the break points of $\tilde{U}_t^{\mathrm{r}}$.

Finally, since $\hat{V}_t^{\mathrm{r}}$ is non decreasing, and has range $\{\eta z : z \in \mathbb{Z}\} \cap \left[-\sum_{s=t}^{T} b_s, \sum_{s=t}^{T} h_s\right]$, it has at most $O\left(\frac{\sum_{s=t}^{T} h_s+b_s}{\eta}\right)$ many break points. This completes the induction argument. □

H Proof of Theorem 4.2

First, we recall the following definition from the proof in Lemma 3.5 in Appendix E. For any $1 \leq t \leq T$, we say that $\left\{\tilde{U}_s^{\mathrm{r}}, \tilde{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}$ is $\{\alpha_s+\eta\}_{s=t}^{T}$-good if

$$\left|U_s^{\mathrm{r}}(y_s) - \hat{U}_s^{\mathrm{r}}(y_s)\right| \leq \sum_{r=s}^{T}(\alpha_r+\eta) \quad \forall y_s \in \mathbb{R}, \qquad \left|V_s^{\mathrm{r}}(x_s) - \hat{V}_s^{\mathrm{r}}(x_s)\right| \leq \sum_{r=s}^{T}(\alpha_r+\eta) \quad \forall x_s \in \mathbb{R}.$$

We first prove that

$$\Pr\left[\text{For all } 1 \leq t \leq T \text{ and } y \in \mathbb{R},\ \left|U_t^{\mathrm{r}}(y) - \tilde{U}_t^{\mathrm{r}}(y)\right| \leq \sum_{s=t}^{T}(\alpha_s+\eta)\right] \geq 1-\delta, \tag{H.1}$$

which explains our choice of N_t. Indeed, we have

$$\begin{aligned}
&\Pr\left[\text{For all } 1\le t\le T \text{ and } y\in\mathbb{R},\ \left|U_t^{\mathrm{r}}(y)-\tilde{U}_t^{\mathrm{r}}(y)\right|\le\sum_{s=t}^{T}(\alpha_s+\eta)\right]\\
\ge&\Pr\left[\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=1}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=1}^{T}-\text{good}\right]\\
=&\prod_{t=1}^{T}\Pr\left[\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=t}^{T}-\text{good}\mid\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=t+1}^{T}-\text{good}\right]
\end{aligned}$$

We argue that, for any $1\le t\le T$, we have

$$\Pr\left[\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=t}^{T}-\text{good}\mid\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=t+1}^{T}-\text{good}\right]\ge 1-\delta/T.$$

The argument is demonstrated as follows. First, by Theorem 3.6, we have

$$\Pr\left[\text{For all } y_t\ \left|\tilde{U}_t^{\mathrm{r}}(y_t)-U_t^{\mathrm{r}}(y_t)\right|\le\alpha_t+\sum_{s=t+1}^{T}(\alpha_s+\eta)\mid\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}\ \text{is}\ \{\alpha_s+\eta\}_{s=t+1}^{T}-\text{good}\right]\ge 1-\frac{\delta}{T},$$

by the choice of N_t.

By Theorem 3.8, conditional on $\left|\tilde{U}_t^{\mathrm{r}}(y_t)-U_t^{\mathrm{r}}(y_t)\right|\le\alpha_t+\sum_{s=t+1}^{T}(\alpha_s+\eta)$ for all y_t, we have $|\hat{\tilde{V}}_t^{\mathrm{r}}(x_t)-V_t^{\mathrm{r}}(x_t)|\le\alpha_t+\sum_{s=t+1}^{T}(\alpha_s+\eta)$ for all $x_T\in\mathbb{R}$ always. This implies that for all $x_t\in\mathbb{R}$, we have

$$\left|\tilde{V}_t^{\mathrm{r}}(x_t)-V_t^{\mathrm{r}}(x_t)\right|\le\left|\tilde{V}_t^{\mathrm{r}}(x_t)-\hat{\tilde{V}}_t^{\mathrm{r}}(x_t)\right|+\left|\hat{\tilde{V}}_t^{\mathrm{r}}(x_t)-V_t^{\mathrm{r}}(x_t)\right|\le\eta+\alpha_t+\sum_{s=t+1}^{T}(\alpha_s+\eta).$$

Altogether, we show that

$$\begin{aligned}
&\Pr\left[\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}\ \text{is}\ \{(\alpha_s+\eta)\}_{s=t}^{T}-\text{good}\mid\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}\ \text{is}\ \{(\alpha_s+\eta)\}_{s=t+1}^{T}-\text{good}\right]\\
\ge&\Pr\left[|\tilde{U}_t^{\mathrm{r}}(y_t)-U_t^{\mathrm{r}}(y_t)|\le\alpha_t+\sum_{s=t+1}^{T}(\alpha_s+\eta)\mid\left\{\tilde{U}_s^{\mathrm{r}},\tilde{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}\ \text{is}\ \{(\alpha_s+\eta)\}_{s=t+1}^{T}-\text{good}\right]\\
\ge&1-\delta/T,
\end{aligned}$$

which establishes (H.1). Finally, we can conclude the $(1+2\epsilon)$ optimality of $(\tilde{R}_1,\cdots,\tilde{R}_T)$ by Lemma 3.1, with our definitions of $\{\alpha_t\}_{T=1}^{T}$ and η. □

I A Discussion on #P−Hardness

To provide the definition for #**P**−hardness, we first provide the defintion of the class #**P** (for example, see Chapter 7 in [GJ79]):

Definition I.1. A function $f : \{0,1\}^* \to \mathbb{N}$ belongs to the complexity class $\#\mathbf{P}$, if there exists a polynomial $p : \mathbb{N} \to \mathbb{N}$ and a polynomial-time computable function $M : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ such that for every $x \in \{0,1\}^*$,

$$f(x) = \left| y \in \{0,1\}^{p(|x|)} : M(x,y) = 1 \right|.$$

That is, f belongs to the class if it counts the number of satisfying assignments for certain polynomial time Turing machine M. Here is a list of common examples in the literature:

1. Let the binary string x be an encoding of a bipartite graph $G = ((L,R),E)$, where L, R ($|L| = |R|$) are the bipartition and $E \subset L \times R$ is the edge set. Let y encodes a subset of E, and let $M(x,y) = 1$ iff y encodes a perfect matching in G. Then f is the function that counts the number of perfect matching in G. Note that M, which checks if an edge subset is a perfect matching, is polynomial time computable.

2. Let the binary string x be an encoding of an undirected graph $G = (V,E)$, and let y encodes a subset of E. Let $M(x,y) = 1$ iff y encodes a hamiltonian cycle in G. Then f is the function that counts the number of hamiltonian cycles in G. Again, M is evidently polynomial time computable.

3. Let the binary string x be an encoding of the tuple $(\{a_t\}_{t=1}^T, R)$, where $\{a_t\}_{t=1}^T, R$ are non-negative real numbers. Let y encodes a subset S of $\{1, \cdots, T\}$. Let $M(x,y) = 1$ if and only if $\sum_{t \in S} a_t \leq R$, where S is the subset encoded by y. Then f is the function that counts the number of subsets S such that the corresponding knapsack constraint is satisfied. Evidently, M is polynomial time computable.

Next, a function $g : \{0,1\}^* \to \mathbb{N}$ is $\#\mathbf{P}-$*hard* to compute, if for every function f in $\#\mathbf{P}$, there exists a algorithm $\tilde{f}$ that computes $f(n)$ with time polynomial in n and makes poly(n) many oracle calls to g. For example, it is stated in [HKM$^+$09] that Problem 3 is $\#\mathbf{P}-$hard. A function f is $\#\mathbf{P}-$*complete* if it is both $\#\mathbf{P}-$hard and in $\#\mathbf{P}$. It is known that Problems 1, 2 are $\#\mathbf{P}-$complete. [GJ79].

We remark that a $\#\mathbf{P}$ problem, which requires us to count the number of $y \in \{0,1\}^{p(|x|)}$ satisfying $M(x,y) = 1$, is at least as hard as its corresponding **NP** problem, which requires us to ascertain an existence of $y \in \{0,1\}^{p(|x|)}$ satisfying $M(x,y) = 1$. While the $\#\mathbf{P}$ problem requires us to output $f(x)$, the corresponding **NP** problem only requires us to decide if $f(x) > 0$. In example 1, the $\#\mathbf{P}$ problem, which counts the number of perfect matching, is **NP**-hard, while the corresponding **NP** problem, which asks if there exists a perfect matching, is polynomial time solvable.

Finally, suppose that g is $\#\mathbf{P}-$hard to compute. Then, the corresponding decision problem, which is to ascertain whether $g(x) \geq K$ for *any given* $K \in \{1, 2, \cdots, 2^{p(|x|)}\}$, is also $\#\mathbf{P}-$hard. (Though, as remarked above, it could be easier for the special case when $K = 1$. However, we are working with general K here.) Indeed, given an algorithm $\mathcal{A}$ for the decision problem for any K, we can compute $g(x)$ by a binary search on $\{1, 2, \cdots, 2^{p(|x|)}\}$, which requires applying the algorithm $\mathcal{A}$ at most $O(p(|x|))$ times. Therefore, the K^{th}-largest subset problem stated in §5, which is the corresponding decision problem to the knapsack counting problem in example 3, is $\#\mathbf{P}-$hard.

J Proof of Lemma 2.3

Consider the following T period instance:

- Holding and backlog costs: $h_1, \cdots, h_{T-1} = 0$, $b_1, \cdots b_{T-1} = 0$, and h_T, b_T are positive real numbers.
- Demands distributions: For each $t \in \{1, \cdots, T\}$, we have

$$\mathbb{P}[D_t = a_t] = \mathbb{P}[D_t = 0] = \frac{1}{2}.$$

- Capacities: $B_1 = \infty$, $B_2, \cdots, B_T = 0$.

In this instance, the decision maker cannot place orders in periods $2, \cdots, T$. His objective is to order an appropriate amount of goods in period 1 in order to match the cumulative demand $\sum_{t=1}^{T} D_t$ at the end of the time horizon. Thus, this instance can be viewed as the single period uncapacitated inventory control problem (also known as the *newsvendor problem*), where the decision maker faces a random demand with distribution $\sum_{t=1}^{T} D_t$; the holding cost and backlog cost are h_T and b_T respectively. We demonstrate the computational hardness by arguing that it is hard to compute the cumulative distribution function of the random variable $\sum_{t=1}^{T} D_t$.

Let $X_1, \cdots, X_T$ be i.i.d. random variables, each distributed as $Bern(0.5)$. Then we can rewrite $D_t = a_t X_t$ for each t. The optimal modified base stock policy $(R_1^*, \cdots, R_T^*)$ has the following form:

$$\begin{aligned}
R_1^* &= \min_{R \in \mathbb{R}} \left\{ R : \mathbb{P}\left[\sum_{t=1}^{T} D_t \le R \right] \ge \frac{b_T}{h_T + b_T} \right\} \\
&= \min_{R \in \mathbb{R}} \left\{ R : \mathbb{P}\left[\sum_{t=1}^{T} a_t X_t \le R \right] \ge \frac{b_T}{h_T + b_T} \right\} \\
&= \min_{R \in \mathbb{R}} \left\{ R : \mathbb{P}_{S \sim \mathrm{U}(T)}\left[\sum_{t \in S} a_t \le R \right] \ge \frac{b_T}{h_T + b_T} \right\} \qquad \text{(J.1)} \\
&= \min_{R \in \mathbb{R}} \left\{ R : \left| \left\{ S \subset \{1, \cdots, T\} : \sum_{t \in S} a_t \le R \right\} \right| \ge \frac{b_T}{h_T + b_T} 2^T \right\}, \qquad \text{(J.2)}
\end{aligned}$$

and $R_2^*, \cdots R_T^*$ are arbitrary real numbers. The step (J.1) is justified as follows. Since $X_1, \cdots, X_T$ are i.i.d. random variables with distribution Bern(0.5), the set $\{t : X_t = 1\}$ is a uniformly random set in $\{1, \cdots, T\}$ (denoted $\{t : X_t = 1\} \sim \mathrm{U}(T)$); That is, for any $S \in \{1, \cdots, T\}$, we have $\Pr[\{t : X_t = 1\} = S] = 1/2^T$. The line (J.2) follows by multiplying each side in the condition by 2^T.

We demonstrate that the SAA problem is $\#\mathbf{P}-$hard by reducing an instance of the K^{th} largest subset problem to a specific instance of SAA problem. Suppose we are given an instance $(\{a_t\}_{t=1}^{T}, R, K\}$ of the K^{th} largest subset problem. We solve this given problem by the following: we first solve the the SAA problem, with $X_1, \cdots, X_T$, $B_1, \cdots, B_T$, $h_1, \cdots, h_{T-1}, b_1, \cdots, b_{T-1}$ as specified in the beginning of the proof; we also specify $b_t = K$, $h_t + b_t = 2^T$. Then, given the output R_1^* for the SAA problem, we declare that assertion (5.1) is true if $R \ge R_1^*$, and false otherwise. Our

declaration is correct, since R_1^* is the smallest optimal base stock for the SAA problem instance; that is R_1^* is the smallest R such that the assertion (5.1) is true. □

K Proof of Theorem 2.5

Recall h, b respectively denote the holding cost and backlog cost in the newsvendor problem. To establish the sample lower bound, we first consider the case when $h \leq b$. We demonstrate the lower bound by considering the following two demand distributions D_1, D_2:

$$\mathbb{P}[D_1 = 0] = \frac{b-h}{b+h}, \qquad \mathbb{P}[D_1 = 1] = \frac{h+h\epsilon}{b+h}, \qquad \mathbb{P}[D_1 = A] = \frac{h-h\epsilon}{b+h},$$

$$\mathbb{P}[D_2 = 0] = \frac{b-h}{b+h}, \qquad \mathbb{P}[D_2 = 1] = \frac{h-h\epsilon}{b+h}, \qquad \mathbb{P}[D_2 = A] = \frac{h+h\epsilon}{b+h},$$

where $0 < \epsilon < 1/20$ is a small accuracy parameter, and $A = 2000 \max\left\{1, \frac{h}{b\epsilon}\right\}$. These two distributions are handcrafted so that the following two properties hold:

- **Claim 1:** D_1, D_2 have disjoint sets of $(1 + \frac{\epsilon}{20})-$optimal base stocks,
- **Claim 2:** D_1, D_2 have small statistical distance: $\mathrm{KL}(D_1||D_2) \leq \frac{8h\epsilon^2}{b+h}$.

Recall that the KL divergence $\mathrm{KL}(X||Y)$ between the integral random variables X, Y is defined as

$$\mathrm{KL}(X||Y) = \sum_{i \in \mathbb{Z}} \mathbb{P}[X = i] \log \frac{\mathbb{P}[X = i]}{\mathbb{P}[Y = i]}.$$

The proofs for these two claims are by technical calculations, thus postponed to the end.

Suppose that we are given an algorithm $\mathcal{A}$ be an algorithm which draws m i.i.d. samples and returns a $(1 + \frac{\epsilon}{20})-$optimal base stock with probability at least $1 - \delta$ under any latent distribution D. Then, the approximation guarantee of $\mathcal{A}$ still holds if the m i.i.d. samples are from D_1 or D_2. We consider the following reduction from solving the data-driven newsvendor problems on D_1, D_2 to solving the statistical classification problem on D_1, D_2.

Consider the following classification problem. We are given m i.i.d. samples $d^m = (d_1, \cdots, d_m)$. We know that these samples are drawn from D_1 or D_2, and we also know the CDFs of D_1 and D_2. However, the identity of the distribution is unknown. The goal of the classification problem is to correctly identify the latent distribution with high probability. We claim that the following algorithm $\mathcal{B}$, which calls $\mathcal{A}$ as a subroutine, achieves our goal.

Algorithm 2 Algorithm $\mathcal{B}$

1: INPUT: m i.i.d. samples d^m from D, where the latent distribution D can be D_1 or D_2.
2: Run $\mathcal{A}$ on the m samples d^m.
3: **if** $\mathcal{A}$ returns a $\left(1 + \frac{\epsilon}{20}\right)-$optimal base stock for D_1 **then**
4: Return 1. ▷ Identify $D = D_1$.
5: **else**
6: Return 0. ▷ Identify $D = D_2$.
7: **end if**

Note that the condition in Line 3 can be checked, since we have the access to the CDFs of D_1, D_2.

We first argue that, by Claim 1, the classifier $\mathcal{B}$ correctly identifies the true underlying distribution D with probability $1-\delta$. Now, by an abuse of notation, we use $\mathcal{B}(d^m)$ to denote the output of $\mathcal{B}$ on the sequence of m samples d^m. Note that $\mathcal{B}(d^m) = 1$ if $\mathcal{B}$ identifies $D = D_1$, and $\mathcal{B}(d^m) = 0$ if $\mathcal{B}$ identifies $D = D_2$. Therefore, when $D = D_1$, the probability that $\mathcal{B}$ correctly identifies the underlying distribution is equal to

$$\mathbb{E}_{d^m \sim D_1^m} \mathcal{B}(d^m) = \mathbb{P}_{d^m \sim D_1^m} \left[\mathcal{A} \text{ outputs a } \left(1 + \frac{\epsilon}{20}\right) \text{ base stock for } D_1 \right] \geq 1 - \delta,$$

where the first equality is by Claim 1, and the second inequality is by our assumption on the approximation guarantee by $\mathcal{A}$. Likewise, when $D = D_2$, the probability that $\mathcal{B}$ correctly identifies the underlying distribution is equal to

$$1 - \mathbb{E}_{d^m \sim D_2^m} \mathcal{B}(d^m) \geq \mathbb{P}_{d^m \sim D_2^m} \left[\mathcal{A} \text{ outputs a } \left(1 + \frac{\epsilon}{20}\right) \text{ base stock for } D_2 \right] \geq 1 - \delta.$$

where the first inequality is by Claim 1, and the second inequality is by our assumption on the approximation guarantee by $\mathcal{A}$.

In particular, the two inequalities result in the following:

$$1 - 2\delta \leq \mathbb{E}_{d^m \sim D_1^m} \mathcal{B}(d^m) - \mathbb{E}_{d^m \sim D_2^m} \mathcal{B}(d^m). \tag{K.1}$$

Next, we argue that the number of samples m must be sufficiently large in order for $\mathcal{B}$ to succeed in identifying D with high probability. We demonstrate the argument by bounding the right hand side of (K.1) from above by information theoretic techniques and Claim 2:

$$\begin{aligned}
\mathbb{E}_{d^m \sim D_1^m} \mathcal{B}(d^m) - \mathbb{E}_{d^m \sim D_2^m} \mathcal{B}(d^m) &= \sum_{d^m \in \{0,1,A\}^m} \mathcal{B}(d^m) \left(\mathbb{P}[D_1^m = d^m] - \mathbb{P}[D_2^m = d^m] \right) \\
&\leq \sum_{d^m \in \{0,1,A\}^m} \left| \mathbb{P}[D_1^m = d^m] - \mathbb{P}[D_2^m = d^m] \right| && \text{(K.2)} \\
&\leq \sqrt{(2 \log 2) \mathrm{KL}\left(D_1^m || D_2^m\right)} && \text{(K.3)} \\
&= \sqrt{(2m \log 2) \mathrm{KL}\left(D_1 || D_2\right)} && \text{(K.4)} \\
&\leq \sqrt{\frac{(16m \log 2) h \epsilon^2}{h + b}} && \text{(K.5)}
\end{aligned}$$

The inequality (K.2) is true because $|\mathcal{B}(d^m)| \leq 1$ for all d^m. Note that the quantity in (K.2) is the total variation distance between D_1^m and D_2^m. The inequality (K.3) is an application of Pinsker Inequality (for example see [CT06]) on D_1^m and D_2^m. The equality (K.4) is by applying the Chain Rule for KL-Divergence (see Theorem 2.5.3 in [CT06]). The inequality (K.5) is by Claim 2.

Altogether, by combining (K.1) and (K.5), we conclude that if there exists an algorithm $\mathcal{A}$ that returns a $(1 + \frac{\epsilon}{20})$-optimal base stock with probability at least $1 - \delta$ using m samples, the quantity

m must satisfy the following inequality:

$$\sqrt{\frac{(16m\log 2)h\epsilon^2}{h+b}} \geq 1-2\delta \Rightarrow m \geq \frac{(1-4\delta)(h+b)}{(16\log 2)h\epsilon^2}.$$

Finally, by replacing ϵ with 20ϵ, we see that in order to output a $(1+\epsilon)$-optimal base stock with probability at least $1-\delta$, an algorithm must draw at least $\frac{(1-4\delta)(h+b)}{(6400\log 2)h\epsilon^2}$ many samples.

Thus, we have proven Theorem 2.5 for the case $h \leq b$. The complementary case $h > b$ can be argued by the following symmetry argument. For any demand distribution D with support $[0, A]$, denote $\bar{D} = A - D$. Then we have, for all $x \in [0, A]$:

$$\begin{aligned} C(x) &= \mathbb{E}[h(x-D)^+ + b(D-x)^+] = \mathbb{E}[h(\bar{D}-(A-x))^+ + b((A-x)-\bar{D})^+] \\ &= \mathbb{E}[\bar{h}((A-x)-\bar{D})^+ + \bar{b}(\bar{D}-(A-x))^+] = \bar{C}(A-x), \end{aligned}$$

where $\bar{h} = b, \bar{b} = h$, and $\bar{C}$ is the newsvendor cost with unit holding and backlog costs $\bar{h}, \bar{b}$, under distribution $\bar{D}$. By applying the reduction argument with $\bar{h}, \bar{b}$ (note that $\bar{h} < \bar{b}$) in place of h, b, we see that $\frac{(1-4\delta)(\bar{h}+\bar{b})}{(6400\log 2)\bar{h}\epsilon^2} = \frac{(1-4\delta)(h+b)}{(6400\log 2)b\epsilon^2}$ samples are necessary for obtaining a $(1+\epsilon)$-optimal solution for $\bar{C}$ with confidence probability $1-\delta$. Hence, the same sample bound is also true for C, which proves Theorem 2.5 for the case when $h > b$.

Finally, we return to the proof of Claim 1 and 2:

Proof of Claim 1: For any distribution D, it is a classical result (for example, see [LRS07], or deduce from §2.2) that the newsvendor cost function $C(x) = \mathbb{E}[h(x-D)^+ + b(D-x)^+]$ is minimized at the $\frac{b}{b+h}$ quantile R:

$$R = \min_x \left\{ x : \mathbb{P}[D \leq x] \geq \frac{b}{b+h} \right\}.$$

Let C_1, C_2 denote the newsvendor cost functions under D_1, D_2 respectively, and let R_1, R_2 denote the $\frac{b}{b+h}$ quantiles of D_1, D_2 respectively. By our definitions of D_1, D_2, we know that $R_1 = 1$, $R_2 = A$. To show that D_1, D_2 have disjoints sets of $(1+\frac{\epsilon}{20})-$optimal base stocks, it suffices to show that

$$C_1\left(\frac{A+1}{2}\right) \geq \left(1+\frac{\epsilon}{10}\right) C_1(1), \qquad C_2\left(\frac{A+1}{2}\right) \geq \left(1+\frac{\epsilon}{10}\right) C_2(A),$$

since C_1, C_2 are convex functions, which implies that their sets of minima are intervals. To prove these inequalities, we first provide the expressions for $C_1(x), C_2(x)$ in the domain $x \in [1, A]$:

$$C_1(x) = \frac{hb-h^2}{b+h}x + \frac{h^2(1+\epsilon)}{b+h}(x-1) + \frac{bh(1-\epsilon)}{b+h}(A-x),$$

$$C_2(x) = \frac{hb-h^2}{b+h}x + \frac{h^2(1-\epsilon)}{b+h}(x-1) + \frac{bh(1+\epsilon)}{b+h}(A-x).$$

This means

$$\begin{aligned}
C_1(1) &= \frac{hbA}{h+b}\left(1-\epsilon-\frac{h}{bA}+\frac{\epsilon}{A}\right) &&\in \left[\frac{hbA}{h+b}\left(1-\epsilon-\frac{\epsilon}{1000}\right), \frac{hbA}{h+b}\left(1-\epsilon+\frac{\epsilon}{1000}\right)\right],\\
C_1(A) &= \frac{hbA}{h+b}\left(1+\frac{h\epsilon}{b}-\frac{h}{bA}-\frac{h\epsilon}{bA}\right) &&\geq \frac{hbA}{h+b}\left(1+\frac{h\epsilon}{b}-\frac{\epsilon}{1000}\right),\\
C_2(1) &= \frac{hbA}{h+b}\left(1+\epsilon-\frac{h}{bA}-\frac{\epsilon}{A}\right) &&\geq \frac{hbA}{h+b}\left(1+\epsilon-\frac{\epsilon}{1000}\right),\\
C_2(A) &= \frac{hbA}{h+b}\left(1-\frac{h\epsilon}{b}-\frac{h}{bA}+\frac{h\epsilon}{bA}\right) &&\in \left[\frac{hbA}{h+b}\left(1-\frac{h\epsilon}{b}-\frac{\epsilon}{1000}\right), \frac{hbA}{h+b}\left(1-\frac{h\epsilon}{b}+\frac{\epsilon}{1000}\right)\right].
\end{aligned}$$

The bounds are justified by our choice of A to be sufficiently large (recall $A = 2000\max\left\{1, \frac{h}{b\epsilon}\right\}$), which implies that $\frac{h}{bA}, \frac{\epsilon}{A}, \frac{h\epsilon}{bA} \leq \frac{\epsilon}{2000}$. Then we have the following bounds on $C_1(\frac{A+1}{2}), C_2(\frac{A+1}{2})$:

$$C_1\left(\frac{A+1}{2}\right) = \frac{1}{2}(C_1(1)+C_1(A)) \geq \left(1+\frac{\epsilon}{10}\right)\frac{hbA}{h+b}\left(1-\epsilon-\frac{\epsilon}{1000}\right) \geq \left(1+\frac{\epsilon}{10}\right)C_1(1),$$

$$C_2\left(\frac{A+1}{2}\right) = \frac{1}{2}(C_2(1)+C_2(A)) \geq \left(1+\frac{\epsilon}{10}\right)\frac{hbA}{h+b}\left(1-\frac{h\epsilon}{b}-\frac{\epsilon}{1000}\right) \geq \left(1+\frac{\epsilon}{10}\right)C_2(A).$$

This proves Claim 1.

Proof of Claim 2: The KL divergence $\mathrm{KL}(D_1||D_2)$ can be expressed as follows:

$$\begin{aligned}
\mathrm{KL}(D_1||D_2) &= \frac{b-h}{b+h}\log\frac{b-h}{b-h} + \frac{h+h\epsilon}{b+h}\log\frac{h+h\epsilon}{h-h\epsilon} + \frac{h-h\epsilon}{b+h}\log\frac{h-h\epsilon}{h+h\epsilon}\\
&= \frac{2h\epsilon}{b+h}\log\left(1+\frac{2\epsilon}{1-\epsilon}\right) \leq \frac{4h\epsilon^2}{(b+h)(1-\epsilon)} \leq \frac{8h\epsilon^2}{b+h}.
\end{aligned}$$

where the last inequality is by the assumption that $\epsilon < 1/2$. This proves Claim 2, which concludes the proof of Theorem 2.5. □