Data-driven algorithms for Operational Problems

by

Wang Chi Cheung

B.A., University of Cambridge (2010)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

Author..... Signature redacted .....
Sloan School of Management
August 26, 2016

Certified by.... Signature redacted .....
David Simchi-Levi
Professor of Engineering Systems
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by ..... Signature redacted .....
Dimitris Bertsimas
Boeing Professor of Operations Research
Co-director, Operations Research Center

Data-driven algorithms for Operational Problems

by

Wang Chi Cheung

Submitted to the Sloan School of Management
on August 26, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

In this thesis, we propose algorithms for solving revenue maximization and inventory control problems in data-driven settings. First, we study the choice-based network revenue management problem. We propose the Approximate Column Generation heuristic (ACG) and Potential Based algorithm (PB) for solving the Choice-based Deterministic Linear Program, an LP relaxation to the problem, to near-optimality. Both algorithms only assume the ability to approximate the underlying single period problem. ACG inherits the empirical efficiency from the Column Generation heuristic, while PB enjoys provable efficiency guarantee. Building on these tractability results, we design an earning-while-learning policy for the online problem under a Multinomial Logit choice model with unknown parameters. The policy is efficient, and achieves a regret sublinear in the length of the sales horizon.

Next, we consider the online dynamic pricing problem, where the underlying demand function is not known to the monopolist. The monopolist is only allowed to make a limited number of price changes during the sales horizon, due to administrative constraints. For any integer m, we provide an information theoretic lower bound on the regret incurred by any pricing policy with at most m price changes. The bound is the best possible, as it matches the regret upper bound incurred by our proposed policy, up to a constant factor.

Finally, we study the data-driven capacitated stochastic inventory control problem, where the demand distributions can only be accessed through sampling from offline data. We apply the Sample Average Approximation (SAA) method, and establish a polynomial size upper bound on the number of samples needed to achieve a near-optimal expected cost. Nevertheless, the underlying SAA problem is shown to be #P hard. Motivated by the SAA analysis, we propose a randomized polynomial time approximation scheme which also uses polynomially many samples. To complement our results, we establish an information theoretic lower bound on the number of samples needed to achieve near optimality.

Thesis Supervisor: David Simchi-Levi
Title: Professor of Engineering Systems
Professor of Civil and Environmental Engineering

Acknowledgments

This thesis marks the finishing point in my five year journey at MIT. I am fortunate and privileged to be accompanied and supported by many incredible people, who made this journey exciting and rewarding.

First, I would like to thank my adviser, David Simchi-Levi, who has been a great and supportive mentor. Apart from his research guidance, David provided valuable help and suggestions for scientific writing and presentation, which are also very important for my development as a researcher. David gave me a lot of freedom in trying out what I wanted to do; while the research outcome might not be what we expected (often in a bad way), David always gave suggestions and encouragement to turn things around, and make the most out of what had been done. I guess one thing that I did not manage to learn from David is his multi-tasking skill, and I am always amazed by how David handled so many things at the same time.

Next, I would like to thank my thesis committee members, James Orlin and John Tsitsiklis, for their time and effort. I had learned a lot from John since my first year probability class; and John's ability to cast applied problems into beautiful mathematical frameworks had always been a source of inspiration for my research. I was fortunate enough to be a teaching assistant for James in his optimization class. Through the teaching, I learned about how to manage a huge class of students and make sure everyone learns. Most importantly, James showed me how to make people appreciate the power and beauty of optimization. Outside the class, we also talked about research from time to time, and his witty comments often offered me alternative viewpoints for my research.

I have my full gratitude to the support from Operations Research Center. I would like to thank the ORC co-directors, Dimitris Bertsimas and Patrick Jaillet, for their great job creating a comfortable and conducive environment. I am indebted to many staff, especially Laura Rose, Janet Kerrigan and Andrew Carvalho, for handling all the administrative issues. Moreover, I would like to thank a few other MIT faculties – David Gamarnik,

Michel Goemans and Andreas Schulz, who I had the privilege to work with or talk to in my early stage in the PhD program.

Chapter 3 in the thesis is a result of collaboration with my fellow ORC student, He Wang, and undoubtedly all chapters in the thesis had been benefited from the discussions (which were sometimes senseless) with many other ORC students, during David's weekly seminar, and sometimes with students from EECS.

I am so lucky to have made many amazing friends at MIT: He, Will, Peng, Yaron, Kris, Louis, Joel, Daniel, Nataly, Zach, Yehua, Zach, Clark, Sam, Alice, Nicole, Godine, Lawrence, and many others. I learned to do better research and be a better person from all of them, and I would like to thank them all!

It would not be possible for me to come to MIT without the generous support from Singapore, for which I am always thankful for. Finally, I owe my deepest gratitude to my family and Bobby the Scottish fold, who always provided me with emotional comfort and warm encouragement.

– W.C.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Optimization under uncertainty has been a recurring theme in Operations Research and Management Science. When an inventory manager determines the order quantity for a product, she has to anticipate its random demand in the future in order to minimize the amount of inventory excess or shortage. When an online retailer offers a product for sales in her e-store, the posted price has to strike a fine balance between the maximization of sales volume, which is *a priori* random, and the maximization of profit margin per product. The omnipresence of uncertainty while optimizing motivates the concept of *Stochastic Optimization* in 1950s, pioneered by [Charnes and Cooper, 1959, Dantzig, 1955], and surveyed in [Shapiro et al., 2009].

In a nutshell, stochastic optimization problems are minimization or maximization problems, where a subset of input parameters are random variables. We refer to the probability distribution for these random variables as the *model uncertainty*. For example, consider the classical newsvendor problem with random demand $D \sim \mathcal{D}$. The decision maker, who begins with no inventory, aims to determine the order up to level y that minimizes the

expected operational cost

$$\mathbb{E}_{D\sim\mathcal{D}}[h(y-D)^+ + b(D-y)^-],$$

where h and b are the unit holding and backlog cost. In this stochastic optimization problem, the model uncertainty is $\mathcal{D}$, the probability distribution for D.

Conventionally, stochastic optimization problems are solved under the assumption of perfect knowledge on the model uncertainty. For example, the classical works on the inventory control models (surveyed in [Scarf, 2002, Simchi-Levi et al., 2013]) assume the knowledge of the demand distributions across the periods. The seminal works on single product revenue management [Gallego and van Ryzin, 1994], network revenue management under independent demand model [Gallego and Van Ryzin, 1997], as well as choice-based network revenue management [Gallego et al., 2004] assume the knowledge of the customers' purchase probability distributions.

While these assumptions of perfect knowledge grant us valuable insights into the problem models, the resulting solution may not be immediately applicable. Indeed, the decision maker hardly has the complete knowledge of the model uncertainty. For example, an inventory manager only has access to the historical sales volume of a product, rather than the product's demand distribution; an online retailer only observes whether a customer purchases the product at the posted price, but not the probability distribution of the customer's willingness-to-pay.

In the recent decade, there has been an increasing interest in *data-driven optimization*. Data-driven optimization concerns the solution of stochastic optimization problems, where we only assume access to the samples of the underlying random variables, *but not their probability distributions*. With the advent of "Big Data" era, the decision makers are capable of collecting massive amount data in a timely manner, which makes the concept of data-driven optimization practically appealing and relevant.

In the thesis, we study data-driven optimization problems for revenue management and

supply chain management. We consider both *online* and *offline* settings. In the online setting, the data is revealed to the decision maker sample by sample. For a concrete example, consider the *online dynamic pricing* problem [Besbes and Zeevi, 2009, Babaioff et al., 2012]: the decision maker (a.k.a. seller) aims to maximize her profit during the sales horizon, where the purchase decisions of the customers are revealed sequentially. The online data-driven setting is very closely related to the stochastic multi armed bandit setting [Bubeck and Cesa-Bianchi, 2012]. A stochastic multi armed bandit problem is essentially the following multi-stage stochastic optimization problem. At every stage, the decision maker attempts to acquire new information (explore) about the model uncertainty, while maximizing the reward gained based on current knowledge (exploit).

In the offline setting, the pool of data is available to the decision maker in the outset. For instance, when an inventory manager determines the ordering quantity of a certain product in the offline data-driven setting, such decision is based on the entire pool of historical sales data of the product. Data driven optimization in the offline setting has been actively studied. [Charikar et al., 2005, Shmoys and Swamy, 2006] consider the two-stage covering problem in the data driven setting, and [Swamy and Shmoys, 2012] consider the multi-stage covering problem. In these works, the number of samples required for near-optimality does not depend on the underlying distribution. [Kleywegt et al., 2002] consider a more general framework for two stage stochastic optimization problem in the offline data-driven setting, but the number of samples required for near optimality (essentially) depends in the variance of the underlying distribution. A more thorough survey on multi-stage stochastic optimization in the offline data-driven setting could be found in [Shapiro et al., 2009].

Compared to the online setting, the optimization problem solved in each stage in the offline setting is substantially harder (computationally and information theoretically). However, most works on multi-stage stochastic optimization in the offline data-driven setting only consider the case when the number of stages is constant, as the number of samples required for near optimality grows exponentially with the number of stages. This is in con-

trast with the online setting (such as multi-armed bandit), where the number of stages is often assumed to be large, but the number of samples required for near optimality is polynomial in the number of stages when the underlying randomness in each stage is assumed to be independently and identically distributed.

1.2 Outline

In the thesis, we consider various operational problems in online and offline data-driven settings, as outlined below.

1.2.1 Solving Choice-based Network Revenue Management Problems

In the first part of the thesis, we consider the choice-based network revenue management (NRM) problem, which is first proposed by [Gallego et al., 2004]. It is a resource constrained revenue maximization problem, where the decision maker (a.k.a. monopolist) has the flexibility to select an assortment of products to offer to a customer. The customer then chooses a product (or none) from the offered assortment, where her selection depends on her choice behavior. This model is of fundamental importance to online retailers, who have the ability to personalize an offered assortment to an arriving customer based on the customer's attributes.

However, the choice-based NRM problem is intractable in the following two aspects. The first is computational intractability. While it is natural to cast the problem as a dynamic program, such a program has exponential size state and action spaces, and an optimal strategy is likely to be complicated. The second is informational intractability. Indeed, the purchase probability of a customer changes when the offered assortment varies. Even with the high level of sales data availability, it is impossible to estimate the purchase probability assortment by assortment, since the cardinality of all assortments increases exponentially with the number of products.

In this chapter, we first tackle the computational intractability by providing efficient algorithms for solving the Choice-based Deterministic Linear Program (CDLP-P), a linear program relaxation to the choice-based NRM problem. A salient feature of our algorithms is that they only assume the ability to approximately solve the underlying Single Period Problem (SPP) without resource constraints. This is in contrast to existing literature, such as [Gallego et al., 2004, Liu and van Ryzin, 2008, Bront et al., 2009], which requires the ability to solve the underlying SPP exactly.

We then consider the online choice-based NRM problem, where the parameters for the underlying MultiNomial Logit (MNL) choice model are unknown to the monopolist. The problem is in an online data-driven setting, where the realizations of the customers' choices are revealed sequentially to the monopolist. We design an earning-while-learning policy for the online problem. We overcome the informational intractability of learning the choice probability over exponentially many assortments by focusing on learning the unknown parameters. The policy runs in polynomial time by our tractability results, and the revenue earned through the policy converges to the optimum when the sales horizon increases.

1.2.2 Dynamic Pricing with Limited Price Experimentation

In the next chapter, we consider the online dynamic pricing problem, which is also a fundamental problem in the realm of e-commerce. In this problem, a revenue maximizing seller has an unlimited supply of a product for sales. However, the demand distribution (i.e. the distribution of the willingness-to-pay of a typical customer) is not known to the seller; she only knows that the demand function belongs to a finite set of demand function family Ψ. In addition to the model uncertainty, the seller also faces the challenge of limited price experimentation; throughout the sales horizon, the seller is only allowed to change the offered price at most m times. Similar to the online assortment problem in the previous part, the dynamic pricing problem is an online problem, where the purchase decisions of

the customers are revealed to the seller one by one.

We consider the following question: Given a demand function family Ψ and an upper bound on the number of price changes m, what is the smallest *regret* incurred by any non-anticipatory policy in a T period sales horizon? Here, the notion of regret refers to the difference between the expected optimal revenue and the total revenue achieved by the seller.

Via an information theoretic argument, we establish a regret lower bound for any family Ψ, integers m and T, under certain mild assumptions on Ψ. The regret lower bound matches the regret upper bound of an non-anticipatory policy proposed by [Cheung et al., 2015], which shows that the regret lower bound is tight up to a constant factor. Apart from showing the optimality of the proposed policy, our analysis also sheds light onto the structure of any policy that achieves the best possible regret order bound.

1.2.3 Data-driven Capacitated Inventory Control Models

In the final part of the thesis, we study the classical multi-period capacitated stochastic inventory control problems in an offline data-driven setting. The objective is to match the on-hand inventory level with the random demand in each period, subject to supply constraints and while minimizing expected cost over finite horizon. Instead of assuming the full knowledge on the demand distributions, we only assume the access to the random demand data across the periods. Such data-driven models are ubiquitous in practice, where the cumulative distribution functions of the underlying random variables are either unavailable or too complicated to work with.

We apply the Sample Average Approximation (SAA) method to the data-driven capacitated inventory control problem. We 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. We crucially use an inequality of Massart, which is a stronger form of Chernoff inequality, to ensure a uniform approximation in the estimation of the right derivatives of the cost-to-go functions in the associated dynamic program.

While the SAA method uses polynomially many samples, we show that the underlying SAA problem is in fact #P-hard. Thus the SAA method, which solves the SAA problem to optimality, is not a polynomial time algorithm in general, unless #P=P. Nevertheless, motivated by the SAA analysis, we propose a randomized polynomial time approximation scheme which also uses polynomially many samples. The approximation scheme involves a sparsification procedure on the right derivatives of the cost-to-go functions, which maintains the tractability of the underlying dynamic program. Finally, we establish an information theoretic lower bound on the number of samples required to solve the data-driven newsvendor problem to near-optimality.

1.3 Overview

The remaining parts of this thesis are organized as follows.

In Chapter 2, we first define the Choice-based Deterministic Linear Program (CDLP-P). Then, we propose the algorithms, the Approximate Column Generation heuristic (ACG) and the Potential Based Algorithm, for solving CDLP-P to near optimality, assuming the ability to approximate the underlying Single Period Problem (SPP). We establish the performance guarantees of both algorithms. More importantly, we show that PB only requires carrying out polynomially many arithmetic operations and solving polynomially many SPPs. Building on the tractability result, we then propose an efficient online policy for the online choice based NRM problem with an unknown MNL choice model. The policy incurs a regret of $\tilde{O}(T^{2/3})$, where T is the length of the sales horizon.

In Chapter 3, we consider an online dynamic pricing problem for a single product. The demand function is not known *a priori*, and the seller is only allowed to change the offered

price m times, for a given integer m. We establish a lower bound on the regret incurred by any pricing policy that carries out at most m price changes. In particular, we develop a change-of-measure Lemma, which could be of independent interest. Our analysis provides important structural insights into the optimal pricing strategies.

In Chapter 4, we first define the capacitated inventory control problem in an offline data-driven setting. Then we establish a polynomial upper bound on the number of samples needed for the Sample Average Approximation (SAA) method to achieve near-optimality with any prescribed success probability. The bound is proven through a first order analysis. Then, we show the computational hardness in solving the underlying SAA problem, and complement the hardness result by proposing a randomized polynomial time approximation scheme to the data driven problem. Finally, we demonstrate an information theoretic sample lower bound to complement our established sample upper bound by SAA.

We provide some concluding remarks in Chapter 5, and we also indicate some interesting future directions. The technical proofs for Chapters 2, 3, 4 are included in the Appendices A, B, C respectively.

Chapter 2

Solving Choice-based Network Revenue Management Problems

2.1 Introduction

With e-commerce sales reaching USD 1.672 trillion globally [INT, 2015], major retailers are rapidly moving to the online space. As a result, online decision making is a central problem for many retailers. In the e-commerce industry, an online retailer often has the flexibility to select a subset of products, namely an *assortment*, and present it to an arriving customer. By offering an appropriate assortment, the retailer channels the incoming demand to the most profitable products. Moreover, when certain products are unavailable, similar alternatives can be included in the assortment to prevent lost sales. Therefore, assortment optimization is an important decision process for online retailers, as recognized by academics [Kök et al., 2015, Talluri and Van Ryzin, 2006] and practitioners [TCS, 2016, SAP, 2015].

Unfortunately, assortment optimization is often hindered by the following two obstacles. The first obstacle is the presence of *resource constraints*; this requires a revenue-maximizing retailer to carefully select an assortment based on her current resource levels, since resources have to be reserved for future customers. However, it leads to a highly intractable problem,

since the number of possible inventory levels and assortments is astronomical.

The second obstacle is the *uncertainty in the underlying choice model.* Apart from her utility, a customer's purchase depends on her offered assortment. Indeed, the customer potentially substitutes to an available product if her most preferred one is unavailable. Such substitution behavior is captured by her choice model, which is often unknown due to the introduction of new products or limited historical data. This motivates an online learning approach to the assortment optimization problem. We attempt to address the two issues by making the following two contributions:

Efficiency in Solving CDLP-P. We propose two algorithms, the Potential Based algorithm (PB) and the Approximate Column Generation heuristic (ACG), that solve the Choice-based Deterministic Linear Program (CDLP-P) to near optimality efficiently. ACG generalizes the classical Column Generation (CG) heuristic; both are practically efficient heuristics, but they are not known to be provably efficient. In contrast, PB solves CDLP-P to near optimality with provable efficiency. Both ACG and PB assume an oracle that solves the Single Period Problem (SPP) exactly or approximately for the underlying choice model. The algorithms apply to a wide range of choice models; in particular, it applies when the underlying choice model satisfies the *substitutability assumption* (See Assumption 2.3.8).

Table 2.1 displays the approximation guarantees and running times of PB for solving CDLP-Ps for different choice models. N is the number of products, and K is the number of resource constraints. To understand the table, consider for example the cell ($S \subset \mathcal{N}$, Mixed MNL). PB provides a $(1-\epsilon)$-approximation to CDLP-P for the mixed MNL choice model where κ is the number of mixtures; there is an additive error of δ in the approximation. The offered assortment S can be any subset of the product set $\mathcal{N}$. The running time is $\tilde{O}\left(\frac{N^{2\kappa+2}K}{\delta\epsilon^{4\kappa+2}}\right)$, and the approximation is obtained by calling an $(1-\epsilon)$-*approximate oracle* provided by [Désir and Goyal, 2014]. Table 2.1 is further discussed in §2.3, and the abbreviations for the references are defined in Appendix A.1. We reiterate that PB is not limited to the displayed choice models. Rather, it provides a black box transformation from approxi-

mation algorithms for SPPs to approximation algorithms for the corresponding CDLP-Ps, when Assumption 2.3.8 is satisfied.

$\mathbb{S}\backslash$ Model		MNL	Nested Logit	Mixed MNL	Markov Chain
$S \subset \mathcal{N}$	App Ratio	$(1-\epsilon)$ (†)	$(1-\epsilon)$	$(1-\epsilon)$	$(1-\epsilon)$ (‡)
	Run Time	$\tilde{O}\left(\frac{NK}{\delta\epsilon^2}\right)$	$\tilde{O}\left(\frac{\text{LP}(N+1,N+1)+NK}{\delta\epsilon^2}\right)$[1]	$\tilde{O}\left(\frac{N^{2\kappa+2}K}{\delta\epsilon^{4\kappa+2}}\right)$	$\tilde{O}\left(\frac{\text{LP}(N,2N)+NK}{\epsilon^2\delta}\right)$
	Based on	|TVR04|	|DGT13|	|DG14|	|FT14|
$\lvert S\rvert \leq C$	App Ratio	$(1-\epsilon)$	$(1-\epsilon)$	$(1-\epsilon)$	$(1/2-\epsilon)$
	Run Time	$\tilde{O}\left(\frac{N^2K}{\delta\epsilon^2}\right)$	$\tilde{O}\left(\frac{\text{LP}(\kappa N,N^4/\kappa^2)+NK}{\delta\epsilon^2}\right)$	$\tilde{O}\left(\frac{N^{2\kappa+2}K}{\delta\epsilon^{4\kappa+2}}\right)$	$\tilde{O}\left(\frac{CN^3+NK}{\delta\epsilon^3}\right)$
	Based on	|RSS10|	|FT15|	|DG14|	|DGSY15|

Table 2.1: Approximating CDLP-P on specific choice models by PB. Provably efficient algorithms are previously known in (†) by [Gallego et al., 2015b] and (‡) by [Feldman and Topaloglu, 2014]

The tractability of solving CDLP-Ps has immediate applications in the realm of choice-based NRM, since efficient implementations are now possible for many proposed algorithms for various online and stochastic choice-based NRM problems, as surveyed in the **Importance of CDLP-P** in §2.1.1. In particular, we apply our algorithmic framework in our second contribution:

Near-optimal Policy for Online Choice-based NRM. Next, we consider the online choice based network revenue management problem with an unknown MultiNomial Logit (MNL) choice model. We propose an efficient and non-anticipatory policy achieving a regret of $\tilde{O}(T^{2/3})$, where T is the length of the sales horizon. As far as we know, this is the first sublinear regret result for online choice-based NRM problems with resource constraints. Our regret bound is polynomial in the number of products, which is much smaller than the number of assortments. Moreover, our policy only requires solving the corresponding CDLP-P once in the whole sales horizon. Thus, our policy achieves both near optimality and computational efficiency.

We avoid the curse of dimensionality in the learning process by estimating the parameters in the underlying MNL choice model, instead of learning the choice probabil-

[1] $\text{LP}(n,m)$ is the time complexity of solving an LP with n variables and m constraints.

ities assortment by assortment. The estimation crucially uses the strong convexity of the negative log-likelihood for MNL choice models, which is used by [Chen et al., 2015, Kallus and Udell, 2015] in the offline setting. Apart from efficient learning, we also demonstrate that confidence bounds can be incorporated in CDLP-P without losing computational tractability.

2.1.1 Literature Review

The literature on assortment optimization is vast, so we survey the most relevant research here.

Solving Choice-based Deterministic Linear Programs CDLP-Ps. CDLP-P is first introduced by [Gallego et al., 2004] as a tractable approach to the choice-based NRM problem. Although CDLP-P has a size exponential in the number of products, the works by [Gallego et al., 2004, Liu and van Ryzin, 2008] demonstrate that Column Generation (CG) is a particularly efficient heuristic for solving CDLP-P. [Gallego et al., 2004] also consider flexible products in their formulation. [Liu and van Ryzin, 2008] investigate the structural property of an optimal solution to CDLP-P.[Bront et al., 2009] consider solving CDLP-P for the Mixed MNL model by CG, which solves the CG subproblem using a mixed integer LP. Finally, [Gallego et al., 2015a] complement [Bront et al., 2009] by demonstrating that a PTAS for the CG subproblem can guarantee a $(1+\epsilon)$-approximation to CDLP-P when CG terminates. However, CG is not known to be a polynomial time algorithm. [Feldman and Topaloglu, 2014, Gallego et al., 2015b] report that CG could be inefficient when the number of products is large. Thus, [Feldman and Topaloglu, 2014], [Gallego et al., 2015b] propose polynomial size reformulations for CDLP-P for the MNL and Markov Chain choice models respectively. [Topaloglu, 2013] proposes a polynomial size reformulation for a variant of CDLP-P associated with the MNL choice model. Nevertheless, the reformulation techniques in [Feldman and Topaloglu, 2014, Gallego et al., 2015b, Topaloglu, 2013] are specialized to the underlying choice models; it is not known if it is

possible to have a compact formulation for any choice model.

Our Potential Based algorithm in §2.4 is inspired by the Multiplicative Update Method [Arora et al., 2012], which solves online optimization problems in adversarial settings. It can also be used to solve packing LP with exponentially many variables [Plotkin et al., 1995, Young, 1995]. Our proposed algorithm is remarkably different from [Arora et al., 2012, Plotkin et al., 1995, Young, 1995], since CDLP-P is not a packing LP, and we incorporate approximation algorithms into our algorithmic framework for PB.

Finally, a stream of research focuses on different relaxations to the choice-based NRM problem to achieve tractability, such as approximate dynamic programming by the work [Zhang and Adelman, 2009], approximate LP [Kunnumkal and Talluri, 2016], Lagrangian relaxations [Topaloglu, 2009]. We remark that apart from the works by [Topaloglu, 2013], [Feldman and Topaloglu, 2014], [Gallego et al., 2015b], other approaches focus on empirically efficient heuristics, but the running times are not known to be polynomially bounded.

$\mathbb{S}\backslash$ Model		MNL	Nested Logit[2]	Mixed MNL	Markov Chain
$S \subset \mathcal{N}$	App Ratio	Exact	Exact	$(1-\epsilon)$	Exact
	Reference	[TVR04]	[DGT13]	[DG14], [MS13], [RSS09]	[FT14]
$\lvert S\rvert \leq C$	App Ratio	Exact	Exact	$(1-\epsilon)$	$(1/2-\epsilon)$
	Reference	[RSS10]	[FT15]	[DG14], [MS13], [RSS09]	[DGSY15]

Table 2.2: Performance Guarantees for solving SPPs

Solving Single Period Problems. Unlike the case of choice-based NRM problems, which are multi-period models, SPPs are well studied in the literature. Table 2.2 summarizes the existing algorithms for solving SPPs for a selection of common choice models, and the abbreviations for the references are defined in Appendix A.1. When the App Ratio for a cell is Exact, the SPP for the corresponding choice model is polynomial time solvable. For example, the MNL model with assortment family $\{S : |S| \leq C\}$ is polynomial time solvable, by [Rusmevichientong et al., 2010]. The SPP for Mixed MNL model is NP-hard

[2]Assuming dissimilarity parameter $\gamma \in [0, 1]$ and a customer never leaves a nest

even with the unrestricted family $S \subset \mathcal{N}$, but PTASes [Rusmevichientong et al., 2009, Mittal and Schulz, 2013] and FPTAS [Désir and Goyal, 2014] have been proposed. The SPP for Markov Chain model with the family $\{S : |S| \leq C\}$ is APX-hard [Désir et al., 2015], but a $(1/2 - \epsilon)-$approximation algorithm is proposed by [Désir et al., 2015]. Finally, we remark that algroithms for solving SPPs for many other choice models are proposed, for example in [Aouad et al., 2015, Guang et al., 2015].

Importance of CDLP-P. The solution of CDLP-P can be interpreted as a probability distribution over assortments, which leads to the *greedy strategy* that offers a random assortment according to the distribution. [Gallego et al., 2004], [Liu and van Ryzin, 2008] show that the revenue gained by the greedy strategy converges to the optimum in the asymptotic setting. [Jasin and Kumar, 2012] consider solving (CDLP-P) repeatedly, which yields a constant regret in the asymptotic setting. [Gallego et al., 2015a] propose an approximation algorithm for the choice-based NRM problem with heterogeneity in customer types and arrival rates; the algorithm involves solving the personalized version of CDLP-P. [Ciocan and Farias, 2012] consider a resource allocation problem that encompasses the choice-based NRM problem, and their algorithm involves repeated solving of CDLP-P. [Golrezaei et al., 2014] study the online choice-based NRM problem with hetereogeneous customer arrivals, which requires solving CDLP-P to achieve a $(1 + \epsilon)$ competitive ratio in the stochastic setting.

Online choice-based NRM. [Besbes and Zeevi, 2012] consider an online NRM problem with resource constraints, in which the online retailer optimizes the prices for multiple products. They assume an independent demand model, which is a simplification of the choice-based model. The online choice-based NRM without resource constraints is first considered in [Rusmevichientong et al., 2010], which propose an online policy that achieves a $O(\log^2 T)$ regret, where T is the length of the sales horizon, and the big-O notation hides a problem dependent constant. [Saure and Zeevi, 2013] improve the regret to $O(\log T)$, and their online policy applies to more general choice models. As far as we

know, online choice-based NRM in the presence of resource constraints is only considered in [Golrezaei et al., 2014]. In this paper, the authors consider an adversarial setting; their policy achieves a multiplicative guarantee, which translates to a linear regret.

2.2 Problem Definition

We consider the choice-based network revenue management problem with flexible products, which is proposed by [Gallego et al., 2004]. In this problem, the monopolist maximizes her revenue by dynamically adjusting the assortments offered to sequentially arriving customers. We decompose the model description into four parts for clarity.

Products and Resources Consumption Model. The monopolist has a collection of *specific products* $\mathcal{N}$ and *flexible products* $\mathcal{F}$ for sale, where each flexible product $\mathcal{F} \in \mathcal{F}$ is a subset of $\mathcal{N}$. She also has K types of resources, $\mathcal{K} = \{1, \cdots, K\}$, for manufacturing the specific products. When one specific product $i \in \mathcal{N}$ is sold, the monopolist earns a revenue of $r(i)$. She needs to supply i immediately, which consumes $a(i,k)$ units of resource k for each $k \in \mathcal{K}$. When one flexible product $\mathcal{F} \in \mathcal{F}$ is sold, the monopolist earns a revenue of $r(\mathcal{F})$. However, different from the case of specific products, the monopolist has the flexibility to deliver $\mathcal{F}$ by supplying one specific product $I(\mathcal{F}) \in \mathcal{F}$ of her choice when sales end. Supplying $I(\mathcal{F})$ consumes $a(I(\mathcal{F}), k)$ units of each resource $k \in \mathcal{K}$.

Sale Dynamics. Starting with $C(k) = Tc(k)$ units of resource k for each $k \in \mathcal{K}$, the monopolist offers assortments to T sequentially arriving customers in a T-period sales horizon. From period $t = 1$ to T, the following sequence of events happens. First, the period-t customer arrives. Second, the monopolist offers an assortment S_t of products. Third, the period-t customer purchases a subset of products $\Sigma_t \subset S_t$. Fourth, the monopolist earns a revenue of $\sum_{j \in \Sigma_t} r(j)$. Fifth, she delivers the purchased specific products $\mathcal{N} \cap \Sigma_t$, which consumes the corresponding resources for producing the specific products: $C(k) \leftarrow C(k) - \sum_{i \in \mathcal{N} \cap \Sigma_t} a(i,k)$, for all $k \in \mathcal{K}$.

Finally, at the end of period T: For each t and each purchased flexible product $\mathcal{F} \in$

$\mathcal{F} \cap \Sigma_t$, the monopolist selects one specific product $I_t(\mathcal{F}) \in \mathcal{F}$ and supplies it to the period-t customer, using the remaining resources. That is, for all $k \in \mathcal{K}$, it requires $\sum_{t=1}^{T} \sum_{\mathcal{F} \in \mathcal{F} \cap \Sigma_t} a(I_t(\mathcal{F}), k) \leq C(k)$.

We emphasize that the monopolist cannot renege on the sales of specific and flexible products. That is, she has to ensures that $C(k) \geq 0$ from period 1 to T; moreover, at the end of period T, there exists selections $\{I_t(\mathcal{F})\}_{\mathcal{F} \in \Sigma_t \cap \mathcal{F}}^{t \in \{1,\ldots T\}}$ such that $\sum_{t=1}^{T} \sum_{\mathcal{F} \in \mathcal{F} \cap \Sigma_t} a(I_t(\mathcal{F}), k) \leq C(k)$. This, thus, requires the monopolist to carefully plan the usage of the resources while offering assortments.

Choice Model. We use $\mathbb{S} \subset 2^{\mathcal{N} \cup \mathcal{F}}$ to denote the family of assortments, namely subsets of products, that the monopolist is allowed to offer. We assume that the empty set $\emptyset$ belongs to $\mathbb{S}$, meaning that the monopolist can reject a customer by offering no product. A common example of $\mathbb{S}$ is the cardinality constrained family $\mathbb{S} = \{S \subset \mathcal{N} \cup \mathcal{F} : |S| \leq C\}$, where C represents the shelf space constraint on the assortment. For other interesting choices of $\mathbb{S}$, please consult [Davis et al., 2013], for example.

The purchase decision of a customer is modeled by the choice probability $\varphi : \mathcal{N} \cup \mathcal{F} \times \mathbb{S} \to [0, 1]$. For an assortment $S \in \mathbb{S}$ and a generic product $j \in \mathcal{N} \cup \mathcal{F}$, $\varphi(j, S)$ represents the probability of a customer purchasing j when S is offered. For $S \in \mathbb{S}$, $j \notin \mathbb{S}$, we have $\varphi(j, S) = 0$. A customer can purchase more than one product from an assortment S (of size larger than 1); likewise, a customer can choose to purchase nothing. Altogether, we call $(\varphi, \mathbb{S})$ a choice model.

Objective. The monopolist's objective to maximize her expected total revenue,

$$\mathbb{E}\left[\sum_{t=1}^{T} \sum_{j \in \mathcal{N} \cup \mathcal{F}} r(j)\varphi(j, S_t)\right],$$

subject to the resource constraints. The expectation is taken over the randomness of the customers' choices as well as the randomness in the adopted policy.

2.2.1 Choice-based Deterministic Linear Program

The choice-based network revenue management problem with flexible products can be cast as a dynamic programming (DP) problem [Gallego et al., 2004], with the state space consisting of all possible resource levels, and the action space being $\mathbb{S}$. Therefore, the action and state spaces could be exponential in the number of products and resources respectively. This makes solving the underlying DP intractable. [Gallego et al., 2004] proposes the Choice-based Deterministic Linear Program (CDLP-P) to mitigate the curse of dimensionality. CDLP-P is an LP relaxation to the problem; the linear program is defined as follows:

$$\max \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y(S)$$

$$\text{s.t.} \sum_{S\in\mathbb{S}} \sum_{i\in\mathcal{N}} a(i,k)\varphi(i,S)y(S) + \sum_{i\in\mathcal{N}} \sum_{\mathcal{F}\ni i} a(i,k)z(\mathcal{F},i) \le c(k) \quad \forall k \in \mathcal{K} \tag{2.1a}$$

$$\sum_{S\in\mathbb{S}} \varphi(\mathcal{F},S)y(S) = \sum_{i\in\mathcal{F}} z(\mathcal{F},i) \quad \forall \mathcal{F} \in \mathcal{F} \tag{2.1b}$$

$$\sum_{S\in\mathbb{S}} y(S) \le 1, \qquad y(S), z(\mathcal{F},i) \ge 0 \quad \forall S \in \mathbb{S}, \mathcal{F} \in \mathcal{F}, i \in \mathcal{F} \tag{2.1c}$$

In this relaxation, the objective corresponds to the expected revenue per period. The decision variable $y(S)$ represents the probability of offering the assortment S to a customer. For each flexible product $\mathcal{F} \in \mathcal{F}$ and $i \in \mathcal{F}$, the variable $z(\mathcal{F}, i)$ represents the probability of selecting i to deliver $\mathcal{F}$. Constraint (2.1a) stipulates that the expected per period consumption of resource k cannot exceed the per period capacity $c(k)$. Constraint (2.1b) stipulates that every purchased flexible product $\mathcal{F}$ must be delivered by choosing some $i \in \mathcal{F}$ and supplying it to the customer. Constraint (2.1c) stipulates that at most one assortment can be offered in a period. Note that the optimal value of CDLP-P does not change even if we strengthen (2.1c) to $\sum_{S\in\mathbb{S}} y(S) = 1$, since $\emptyset \in \mathbb{S}$.

CDLP-P has an optimal solution of sparse support, as observed in [Gallego et al., 2004]. More precisely, for an extreme point optimal solution (y^*, z^*) to CDLP-P, at most $N+F+1$

variables in $\{y^*(S), z^*(\mathcal{F}, i)\}_{S,\mathcal{F},i}$ are non zero. This makes the Column Generation heuristic (CG) practically efficient for solving CDLP-P, as further discussed in §2.4.

2.3 Solving CDLP-Ps by Solving Polynomially Many SPPs

In this Section, we demonstrate that CDLP-P can be solved or approximated efficiently, assuming the ability to solve or approximate the underlying Single Period Problem, denoted as SPP$-(r, \varphi, \mathbb{S})$.

Definition 2.3.1 (SPP$-(r, \varphi, \mathbb{S})$)**.** The Single Period Problem, denoted SPP$-(r, \varphi, \mathbb{S})$, is the following revenue maximization problem:

$$\max_{S \in \mathbb{S}} \sum_{j \in \mathcal{N} \cup \mathcal{F}} r(j)\varphi(j, S).$$

SPP is a special case of CDLP-P when resource constraints are absent. While the SPPs are polynomial time solvable for the MNL choice model with a wide variety assortment families [Davis et al., 2013], SPPs are NP-hard for general choice models, as surveyed in §2.1.1.

We first discuss the efficiency in solving CDLP-P, assuming an *exact oracle* $\mathcal{A}$ for the underlying choice model $(\varphi, \mathbb{S})$. That is, $\mathcal{A}$ outputs an optimal assortment for SPP-$(r, \varphi, \mathbb{S})$, for any choice of $r \in \mathbb{R}^{\mathcal{N} \cup \mathcal{F}}$. Under this assumption, [Gallego et al., 2004], [Liu and van Ryzin, 2008] show that the Column Generation heuristic (CG) is empirically efficient. However, CG is not known to terminate in time polynomial in N, F, K. Nevertheless, under the same assumption, the Ellipsoid Algorithm [Khachiyan, 1980] solves CDLP-P efficiently:

Theorem 2.3.2 ([Khachiyan, 1980])**.** *Consider the Choice-based Deterministic Linear Program (CDLP-P), and assume an exact oracle $\mathcal{A}$ for the underlying choice model $(\varphi, \mathbb{S})$. The Ellipsoid Algorithm can be adapted to solve CDLP-P by invoking $\mathcal{A}$ for $O(poly(N, F, K))$*

many times, and performing $O(poly(N, F, K))$ many elementary operations.

An elementary operation refers to either an addition of, a multiplication of, or a comparison between two numbers; an oracle call does not need any elementary operation. The proof of Theorem 2.3.2 is discussed in Appendix A.2. While Theorem 2.3.2 guarantees provable efficiency for solving CDLP-P, it assumes an exact oracle, which is equivalent to assuming the ability to solve an NP-hard problem for general choice models. Moreover, Theorem 2.3.2 involves the Ellipsoid Method, which is known to be practically inefficient. These motivate the design of alternative algorithms for solving CDLP-P efficiently with a weaker oracle assumption.

We thus propose the Approximate Column Generation heuristic (ACG) and the Potential Based algorithm (PB), which compute near optimal solutions to CDLP-P. Both ACG and PB assume an $\alpha-$*approximate oracle* for the underlying choice model.

Definition 2.3.3 ($\alpha-$Approximate Oracle)**.** Let $\alpha \in [0, 1]$. We say $\mathcal{A}$ is an $\alpha-$*approximate oracle* for the choice model $(\varphi, \mathbb{S})$, if for any revenue coefficients $r \in \mathbb{R}^{\mathcal{N}\cup\mathcal{F}}$, the oracle $\mathcal{A}$ returns an assortment $S_{\mathcal{A}} \in \mathbb{S}$ for SPP$-(r, \varphi, \mathbb{S})$ such that

$$\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j, S_{\mathcal{A}}) \geq \alpha \max_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j, S). \tag{2.2}$$

We remark that the assumption of an $\alpha-$approximate oracle is weaker than that of an exact oracle; in particular, an exact oracle is equivalent to a 1-approximate oracle. In Definition 2.3.3, we allow $r(j) < 0$ for any product j. However, the optimum of SPP$-(r, \varphi, \mathbb{S})$ is always non-negative, since we assume $\emptyset \in \mathbb{S}$. If it is the case that $\max_{S\in\mathbb{S}\setminus\emptyset} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j, S) \leq 0$, then an $\alpha-$approximate oracle $\mathcal{A}$ returns an assortment with zero expected revenue, for example by returning $S_{\mathcal{A}} = \emptyset$. Otherwise, it returns an $\alpha-$approximate assortment, which has a positive revenue.

We formally state the performance guarantees of the algorithms, assuming the access to an $\alpha-$approximate oracle $\mathcal{A}$. For this purpose, we define the following notations: (y^*, z^*)

denotes an optimal solution to CDLP-P, B denotes the maximum number of products purchased by a customer, and we denote $\text{Val}(y) = \sum_{j \in \mathcal{N} \cup \mathcal{F}} \sum_{S \in \mathbb{S}} r(j)\varphi(j,S)y(S)$.

PB returns a near optimal solution to CDLP-P with provable efficiency:

Theorem 2.3.4. *Consider the Choice-based Deterministic Linear Program (CDLP-P), and assume an $\alpha-$approximate oracle $\mathcal{A}$ for the underlying choice model $(\varphi, \mathbb{S})$. For any $\epsilon \in [0, 1/2], \delta \in [0, 1]$, the Potential Based Algorithm (PB) returns a feasible solution $(\hat{y}, \hat{z})$ to CDLP-P satisfying*

$$Val(\hat{y}) \geq (1 - 2\epsilon)\alpha\, Val(y^*) - \delta. \tag{2.3}$$

PB calls the oracle $\mathcal{A}$ at most $\mathsf{T} = O\left(\frac{B\log(K+1)}{\epsilon^2\delta}\right)$, and performs at most $O(NFK\mathsf{T})$ elementary operations.

We note that that T, the number of oracle calls, does not increase with the number of products $N + F$. Rather, T only increases linearly with B, which is equal to 1 for most choice models. This contrasts with the exponential increase in the number of variables in CDLP-P as $N + F$ increases.

Next, ACG also returns a near optimal solution to CDLP-P.

Lemma 2.3.5. *Consider the Choice-based Deterministic Linear Program (CDLP-P), and assume an $\alpha-$approximate oracle $\mathcal{A}$ for the underlying choice model $(\varphi, \mathbb{S})$. The Approximate Column Generation heuristic (ACG) returns an $\alpha-$approximate solution $(\bar{y}, \bar{z})$ to CDLP-P, that is,*

$$Val(\bar{y}) \geq \alpha\, Val(y^*), \tag{2.4}$$

at termination.

We remark that approximation guarantee of ACG is slightly better than that of PB. However, PB has the key advantage of being provably efficient. In contrast, ACG is only guaranteed to return a near optimal solution *at termination*, and it is unknown if ACG terminates in polynomial time.

Remark on Approximate Oracle. The notion of an approximate oracle, which applies for all choice of r, is more general than the convention notion of an approximation algorithm, which only applies when $r \geq 0$.

Definition 2.3.6 ($\alpha-$Approximation Algorithm)**.** Let $\alpha \in [0,1]$. We say the algorithm $\mathcal{A}'$ is an *$\alpha-$approximation algorithm* for $(\varphi, \mathbb{S})$, if for all $r \in \mathbb{R}_{\geq 0}^{\mathcal{N}\cup\mathcal{F}}$, the algorithm $\mathcal{A}'$ returns an assortment $S_{\mathcal{A}'} \in \mathbb{S}$ for SPP$-(r, \varphi, \mathbb{S})$ that satisfies $\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j, S_{\mathcal{A}'}) \geq \alpha \max_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)$.

Nevertheless, we can construct an α-approximate oracle with an $\alpha-$approximation algorithm, when the choice model satsifies the substitutability assumption, see Assumption 2.3.8.

Claim 2.3.7. *Let $r \in \mathbb{R}^{\mathcal{N}\cup\mathcal{F}}$, and $(\varphi, \mathbb{S})$ be a choice model that satisfies Assumption 2.3.8. We can compute an assortment $S' \in \mathbb{S}$ such that*

$$\sum_{j\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j, S') \geq \alpha \max_{S\in\mathbb{S}} \sum_{j\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)$$

with one invocation of an α-approximation algorithm $\mathcal{A}'$ and polynomial time computation.

Assumption 2.3.8 is known as the subsititutability assumption, which is satisfied by many choice models, such as MNL, mixed MNL and Markov Chain models.

Assumption 2.3.8 (Substitutability)**.** We say that the choice model $(\varphi, \mathbb{S})$ satisfies the substitutability assumption, if $\mathbb{S}$ is downward closed, i.e. for all $S \in \mathbb{S}$, $S' \subset S$, we have $S' \in \mathbb{S}$, and for all distinct $j, j' \in S \in \mathbb{S}$, we have $\varphi(j, S) \leq \varphi(j, S \setminus j')$.

Proof of Claim 2.3.7. Suppose $(\varphi, \mathbb{S})$ satisfies Assumption 2.3.8. Given $r \in \mathbb{R}^{\mathcal{N}\cup\mathcal{F}}$, define $\hat{r}(j) = r(j)\mathbf{1}(j \in S_{\geq 0})$, where $S_{\geq 0} = \{j \in \mathcal{N} \cup \mathcal{F} : r(j) \geq 0\}$. Now, apply $\mathcal{A}'$ on SPP$-(\hat{r}, \varphi, \mathbb{S})$, which returns an assortment $S_{\mathcal{A}'} \in \mathbb{S}$. Lastly, output the assortment $S' = S_{\mathcal{A}'} \cap S_{\geq 0}$. The assortment S' satisfies the claimed properties. First, $S' \in \mathbb{S}$, since $\mathbb{S}$ is

downward closed. Next, we have

$$\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S') = \sum_{j\in\mathcal{N}\cup\mathcal{F}} \hat{r}(j)\varphi(j,S') \underset{(\dagger)}{\geq} \sum_{j\in\mathcal{N}\cup\mathcal{F}} \hat{r}(j)\varphi(j,S_{\mathcal{A}'}) \geq \alpha \max_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} \hat{r}(j)\varphi(j,S),$$

where inequality (†) is by Assumption 2.3.8. Finally, note

$$\max_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} \hat{r}(j)\varphi(j,S) \geq \max_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S).$$

□

Implications for Specific Choice Models. Existing approximation algorithms for SPPs for various choice models can be used to construct the approximation algorithms for the corresponding CDLP-Ps, as displayed in Table 2.1.The table focuses on the cases when $\mathcal{F} = \emptyset$ for clarity; the cases with both specific and flexible products can be deduced from Theorem 2.3.4. In Table 2.2, ($S \subset \mathcal{N}$, Mixed MNL), ($|S| \leq C$, Mixed MNL), ($|S| \leq C$, Markov Chain) are the cases when the underlying SPPs are NP-hard. The complexity results for the Nested Logit model hold under the following assumptions. First, we assume the dissimilarity parameter $\nu \in [0,1]$ (see [Davis et al., 2013] for the definition). Second, if a customer chooses a nest, she must make a purchase. κ is the number of nests. The assumption $\nu \in [0,1]$ is necessary for applying our potential based algorithm. When $\nu > 1$, there is complementarity among products, which violates Assumption 2.3.8. In this case ($\nu > 1$), an efficient approximation algorithm for the corresponding CDLP-P is still an open problem.

2.4 Design and Analysis of ACG and PB

In this Section, we discuss the design and analysis of the two proposed algorithms, Approximate Column Generation heuristic (ACG) and the Potential Based algorithm (PB). Both ACG and PB involve solving a series of SPPs, $\{\text{SPP}-(r_\tau,\varphi,\mathbb{S})\}_{\tau=1,2,\ldots}$, with suitably

chosen r_τ , which return a sequence of assortments $\{S_\tau\}_{\tau=1,2,\ldots}$. However, the two algorithms differ in the definitions of r_τs, and the two algorithms constructs their respective near optimal solutions by combining the respective sequences of assortments in different manners. We first discuss ACG, which is reminiscent to the classical CG.

2.4.1 Approximate Column Generation Heuristic

Similar to CG, the Approximate Column Generation heuristic involves solving the dual of CDLP-P restricted on a small collections of assortments $\mathbb{S}_\tau$. Based on the solution to the dual, the heuristic either terminates or augments the collection of assortments. To state ACG formally, we define the following pair of linear programs, CDLP-P[$\mathbb{S}_\tau$] and CDLP-D[$\mathbb{S}_\tau$]. For $\mathbb{S}_\tau \subset \mathbb{S}$, the linear program CDLP-P[$\mathbb{S}_\tau$] is the CDLP-P restricted to the variables for assortments $\mathbb{S}_\tau$, as defined below:

$$\max \sum_{S\in\mathbb{S}_\tau}\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y(S)$$

$$\text{s.t.} \sum_{S\in\mathbb{S}_\tau}\sum_{i\in\mathcal{N}} a(i,k)\varphi(i,S)y(S) + \sum_{i\in\mathcal{N}}\sum_{\mathcal{F}\ni i} a(i,k)z(\mathcal{F},i) \le c(k) \quad \forall k\in\mathcal{K} \tag{2.5a}$$

$$\sum_{S\in\mathbb{S}_\tau} \varphi(\mathcal{F},S)y(S) = \sum_{i\in\mathcal{F}} z(\mathcal{F},i) \quad \forall \mathcal{F}\in\mathcal{F} \tag{2.5b}$$

$$\sum_{S\in\mathbb{S}_\tau} y(S) \le 1, \quad y(S), z(\mathcal{F},i) \ge 0 \quad \forall S\in\mathbb{S}_\tau, \mathcal{F}\in\mathcal{F}, i\in\mathcal{F} \tag{2.5c}$$

We remark that CDLP-P[$\mathbb{S}$] is equal to CDLP-P. Next, CDLP-D[$\mathbb{S}_\tau$] is the dual of CDLP-P[$\mathbb{S}_\tau$]:

$$\min \eta + \sum_{k=1}^{K} c(k)\rho(k)$$

$$\text{s.t. } \eta \geq \sum_{i=1}^{N} \left(r(i) - \sum_{k=1}^{K} a(i,k)\rho(k) \right) \varphi(i,S) + \sum_{\mathcal{F} \in S} \left(r(\mathcal{F}) - \sigma(\mathcal{F}) \right) \varphi(\mathcal{F},S) \quad \forall S \in \mathbb{S}_\tau \tag{2.6a}$$

$$\sum_{k \in \mathcal{K}} a(i,k)\rho(k) \geq \sigma(\mathcal{F}) \qquad \forall \mathcal{F} \in \mathcal{F}, i \in \mathcal{F} \tag{2.6b}$$

$$\eta, \rho(k) \geq 0, \sigma(\mathcal{F}) \text{ free} \qquad \forall k \in \mathcal{K}, \mathcal{F} \in \mathcal{F} \tag{2.6c}$$

ACG assumes an $\alpha-$approximate oracle, and is formally stated in the following:

Algorithm 1 Approximate Column Generation ACG($\mathbb{S}_1$)

1: **for** $\tau = 1, 2, \cdots$ **do**

2: Solve CDLP-D[$\mathbb{S}_\tau$] for an optimal solution $(\bar{\eta}_\tau, \bar{\rho}_\tau, \bar{\sigma}_\tau)$.

3: Define revenue coefficients

$$\bar{r}_\tau(i) = r(i) - \sum_{k=1}^{K} a(i,k)\bar{\rho}_\tau(k) \text{ for all } i \in \mathcal{N}, \quad \bar{r}_\tau(\mathcal{F}) = r(\mathcal{F}) - \bar{\sigma}_\tau(\mathcal{F}) \text{ for all } \mathcal{F} \in \mathcal{F}.$$

4: Apply the approximate oracle $\mathcal{A}$ on SPP($\bar{r}_\tau, \varphi, \mathbb{S}$), which returns an assortment $S_\tau \in \mathbb{S}$.

5: **if** $\bar{\eta}_\tau \geq \sum_{j \in S_\tau} \bar{r}_\tau(j)\varphi(j, S_\tau)$ **then**

6: **Break** the **for** loop.

7: **end if**

8: $\mathbb{S}_{\tau+1} \rightarrow \mathbb{S}_\tau \cup \{S_\tau\}$.

9: **end for**

10: **return** $(\bar{y}, \bar{z})$, an optimal solution for CDLP-P[$\mathbb{S}_\tau$].

We note that ACG is almost identical to the classical CG; if we apply an exact oracle in Line 4 to solve the SPP$-(\bar{r}_\tau, \varphi, \mathbb{S})$ instead of applying $\mathcal{A}$, the algorithm would precisely be

the CG heuristic, for example in the case of MNL choice model. However, we only assume the ability to approximate an SPP for the underlying choice model, which deviates from the CG framework.

Recall from §2.2 that CDLP-P has an optimal solution of sparse support. Similar to the classical CG, ACG strives to construct a collection $\mathbb{S}_\tau$ of assortments containing the support of a sparse near-optimal solution. ACG augments the collection $\mathbb{S}_\tau$ in the τ^{th} iteration by first solving the corresponding dual problem CDLP-D[$\mathbb{S}_\tau$] in Line 2, followed by defining the reduced revenue $\bar{r}_\tau$ in Line 3 based on the dual solution. Then, it applies the approximate oracle $\mathcal{A}$ on SPP$-(\bar{r}_\tau, \varphi, \mathbb{S})$ in Line 4, and check a certain inequality in Line 5. Line 5 is equivalent to checking if the current collection $\mathbb{S}_\tau$ contains the support of a near-optimal solution to CDLP-P. If it is the case, it breaks and outputs an optimal solution to CDLP-P[$\mathbb{S}_\tau$], which would be a near optimal solution to CDLP-P. Otherwise, it augments the collection $\mathbb{S}_\tau$ to form $\mathbb{S}_{\tau+1}$. Lemma 2.3.5, which demonstrates the performance guarantee of ACG, is proved in Appendix A.3

2.4.2 Potential Based Algorithm

Next, we discuss the design and analysis of the Potential Based Algorithm (PB). In a nutshell, PB constructs a sequence of assortments $S_1, \cdots, S_\mathcal{T}$ such that the averaged solution $\hat{y}(S) = \sum_{\tau=1}^{\mathcal{T}} 1(S_\tau = S)/\mathcal{T}$, along with an appropriately defined $\hat{z}$, constitutes a near optimal solution to CDLP-P. The assortments $S_1, \cdots, S_\mathcal{T}$ are generated one by one. In iteration τ, PB generates an assortment S_τ that achieves a substantial revenue, while avoids using resources that are heavily consumed by $S_1, \cdots, S_{\tau-1}$. Such S_τ is generated by solving an SPP$-(\tilde{r}_\tau, \varphi, \mathbb{S})$, where the definition of the *net revenue* $\tilde{r}_\tau(j)$ takes into account the revenue $r(j)$ by selling product j and the amount of resource consumed by the assortments in the previous iterations.

Design of PB. The following definition is crucial in the design of PB:

Definition 2.4.1. For a given $W \in \mathbb{R}_+$, we say that a solution $(\tilde{y}, \tilde{z})$ is *W-successful* if

$(\tilde{y}, \tilde{z})$ is feasible to (CDLP-P), and $\text{Val}(\tilde{y}) = \sum_{j \in \mathcal{N} \cup \mathcal{F}} \sum_{S \in \mathbb{S}} r(j) \varphi(j, S) \tilde{y}(S) \geq W$.

The Potential Based algorithm consists of Algorithm 2 and 3. Algorithm 3 is a binary search algorithm which calls Algorithm 2, the Potential Based subroutine, and returns a near optimal solution $(\hat{y}, \hat{z})$ satisfying Theorem 2.3.4. The Potential Based subroutine takes as an input a target objective value $Z \leq B \max_{j \in \mathcal{N} \cup \mathcal{K}} r(j)$ and an accuracy parameter ϵ. $B \max_{j \in \mathcal{N} \cup \mathcal{F}} r(j)$ is an upper bound to the optimal value to CDLP-P. It outputs a *candidate* solution $(\tilde{y}, \tilde{z})$. Algorithm 3, namely BINSEARCH(ϵ, δ), computes the largest target value Z such that the output $(\tilde{y}, \tilde{z})$ of FEASIBILITY(ϵ, Z) is $(1-2\epsilon)Z-$successful, within an additive error of δ.

The Potential Based subroutine FEASIBILITY(ϵ, Z) solves a series of suitably defined SPPs in $\mathcal{T}$ iterations. At the end of iteration τ, it outputs an indicator variable v_τ and a variable ζ_τ that assigns a flexible product $\mathcal{F}$ to a specific product $I_\tau(\mathcal{F}) \in \mathcal{F}$ (Line 6). Line 7 updates the potential functions Ξ, Ψ_k, which account for the revenue attained and resources consumed so far. The final output $(\tilde{y}, \tilde{z})$ is a scaled average of $\{(v_\tau, \zeta_\tau)\}_{\tau=1}^{\mathcal{T}}$ (Line 9). The number of iterations $\mathcal{T}$ and its associated scaling parameter γ are:

$$\mathcal{T} = \left\lceil \frac{3\gamma \log(1+K)}{\epsilon^2} \right\rceil, \text{ where } \gamma = B \max \left\{ \max_{i \in \mathcal{N}, k \in \mathcal{K}} \left\{ \frac{a(i,k)}{c(k)} \right\}, \frac{\max_{j \in \mathcal{N} \cup \mathcal{F}} r(j)}{Z} \right\}.$$

At the start of iteration τ, for each specific product $i \in \mathcal{N}$, the *net revenue* $\tilde{r}_\tau(i)$ in Line 2 takes the following form: $\tilde{r}_\tau(i) = A_{\tau-1} r(i) - \sum_{k=1}^{K} B_{\tau-1}(k) a(i,k)$. Essentially, the net revenue can be understood as the difference between the scaled original revenue $r(i)$ and the scaled opportunity cost in producing i, which consumes resources in $\mathcal{K}$.

More precisely, A_τ and $B_\tau(k)$ are defined by the following four parameters maintained by the Potential Based subroutine. The definitions of these parameters are crucial in the analysis:

$$\Xi(\tau) = (1-\epsilon)^{-(1-\epsilon)\frac{\tau}{\gamma}} \left(1 - \frac{\epsilon}{\gamma}\right)^{\mathcal{T}-\tau} \prod_{s=1}^{\tau} (1-\epsilon)^{\frac{V(s)}{\gamma}}, \tag{2.7}$$

$$\Psi_k(\tau) = (1+\epsilon)^{-(1+\epsilon)\frac{\mathcal{T}}{\gamma}} \left(1+\frac{\epsilon}{\gamma}\right)^{\mathcal{T}-\tau} \prod_{s=1}^{\tau} (1+\epsilon)^{\frac{U_k(s)}{\gamma}}, \tag{2.8}$$

$$V(s) = \sum_{S\in\mathbb{S}} \left(\sum_{i\in\mathcal{N}} \frac{r(i)}{Z}\varphi(i,S) + \sum_{\mathcal{F}\in\mathcal{F}} \frac{r(\mathcal{F})}{Z}\varphi(\mathcal{F},S) \right) v_s(S), \tag{2.9}$$

$$U_k(s) = \sum_{S\in\mathbb{S}} \left(\sum_{i\in\mathcal{N}} \frac{a(i,k)}{c(k)}\varphi(i,S) + \sum_{\mathcal{F}\in\mathcal{F}} \frac{a(I_s(\mathcal{F}),k)}{c(k)}\varphi(\mathcal{F},S) \right) v_s(S). \tag{2.10}$$

For each resource $k \in \mathcal{K}$, the potential function $B_{\tau-1}(k)$ is non-negative, and increases exponentially with the expected amount of resource k consumed by assortments $\{S_s\}_{s=1}^{\tau-1}$. Thus, the term $B_{\tau-1}(k)a(i,k)$ could be interpreted as the scaled opportunity cost of using $a(i,k)$ units of resource k to produce i in iteration τ. Indeed, if resource k is consumed for iteration τ, this will constrain the supply of other products that require resource k in future iterations. The potential function $A_{\tau-1}$ serves as a scaling factor that allows us to compare the revenue $r(i)$ of a product i to the opportunity cost of producing i. The precise definitions of $A_{\tau-1}, B_{\tau-1}(k)$ are motivated by the Multiplicative Weight Update (MWU) Method [Arora et al., 2012].

The net revenue $\tilde{r}_{\mathcal{F}}$ of a flexible product $\mathcal{F}$ can be interpreted as follows. First, in Line 3, we assign a flexible product $\mathcal{F}$ to a specific product $I_\tau(\mathcal{F}) \in \mathcal{F}$ that has the lowest scaled opportunity cost. That is, in the delivery of the flexible product (if purchased), the monopolist prefers to select a specific $I_\tau(\mathcal{F})$ which uses the least scarce resources. Then, in Line 4, we define the net revenue $\tilde{r}(\mathcal{F})$ as the difference between the scaled revenue of $\mathcal{F}$ and the scaled opportunity cost of $I_\tau(\mathcal{F})$.

By applying an $\alpha-$approximate oracle to $\mathrm{SPP}(\tilde{r}_\tau, \varphi, \mathbb{S})$ in Line 5, it returns the assortment S_τ that achieves a near optimal total net revenue. The definition of net revenue ensures that the assortment S_τ achieves a near optimal balance between maximizing the revenue in iteration τ and reserving resources for future iteration.

Analysis of PB. Now, we demonstrate that if the target $Z \leq \alpha\mathrm{Val}(y^*)$, then Algorithm 2 returns a solution $(\tilde{y}, \tilde{z})$ that is $(1-2\epsilon)Z-$successful. We use the potential function

Ω defined in Lemma 2.4.2 to quantify the progress of Algorithm 2.

Algorithm 2 Potential Based subroutine FEASIBILITY(ϵ, Z)

1: **for** $\tau = 1, \cdots, \mathcal{T}$ **do**

2: For each specific product $i \in \mathcal{N}$, compute its reduced revenue: ▷ Note $\gamma \geq 1$

$$\tilde{r}_\tau(i) = \frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} \frac{r(i)}{Z} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)}, \tag{2.11}$$

where the variables $\Xi(\tau-1)$, $\Psi_k(\tau-1)$ are defined in (2.7), (2.8).

3: For each flexible product $\mathcal{F}$, select a specific product

$$I_\tau(\mathcal{F}) = \operatorname*{argmin}_{i \in \mathcal{F}} \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)} \tag{2.12}$$

that maximizes the reduced revenue within the subset $\mathcal{F}$.

4: Define the reduced revenue of $\mathcal{F}$ as follows:

$$\tilde{r}_\tau(\mathcal{F}) = \frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} \frac{r(\mathcal{F})}{Z} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(I_\tau(\mathcal{F}),k)}{c(k)}. \tag{2.13}$$

5: Apply the approximate oracle $\mathcal{A}$ on SPP$(\tilde{r}_\tau, \varphi, \mathbb{S})$, which returns an assortment $S_\tau \in \mathbb{S}$.

6: Define the variables $\{v_\tau(S)\}_{S\in\mathbb{S}} \cup \{\zeta_\tau(\mathcal{F},i)\}_{\mathcal{F}\in\mathcal{F}, i\in\mathcal{F}}$ as follows

$$v_\tau(S) = \begin{cases} 1 & \text{if } S = S_\tau \\ 0 & \text{otherwise} \end{cases}, \quad \zeta_\tau(\mathcal{F},i) = \begin{cases} \varphi(\mathcal{F}, S_\tau) & \text{if } \mathcal{F} \in S_\tau \text{ and } i = I_\tau(\mathcal{F}) \\ 0 & \text{otherwise} \end{cases}. \tag{2.14}$$

7: Update the variables $\Xi(\tau)$, $V(\tau)$, $\Psi_k(\tau)$, $U_k(\tau)$ in (2.7), (2.9), (2.8), (2.10) respectively.

8: **end for**

9: **return** the solution $\{\tilde{y}(S)\}_{S\in\mathbb{S}} \cup \{\tilde{z}(\mathcal{F},i)\}_{\mathcal{F}\in\mathcal{F}, i\in\mathcal{F}}$, where

$$\tilde{y}(S) = \frac{1}{(1+\epsilon)\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} v_\tau(S), \quad \tilde{z}(\mathcal{F},i) = \frac{1}{(1+\epsilon)\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} \zeta_\tau(\mathcal{F},i). \tag{2.15}$$

Lemma 2.4.2. *Consider the potential function* $\Omega(\tau) = \Xi(\tau) + \sum_{k=1}^{K} \Psi_k(\tau)$ *for Algorithm 2. Suppose that the inequality* $Z \leq \alpha\, Val(y^*)$ *holds. Then we have* $\Omega(\tau) \leq \Omega(\tau-1)$ *for all*

τ.

The analysis in the proof of Lemma 2.4.2, given in Appendix A.4, explains the definition of the net revenue $\tilde{r}_\tau$ in Algorithm 2 (Line 5), as well as the definitions of the relevant parameters.

Algorithm 3 Binary Search Algorithm BINSEARCH(ϵ, δ) for approximating CDLP-P

1: Initialize $(\hat{y}, \hat{z}) \leftarrow (0, 0)$.
2: Initialize $[\text{LB}(1), \text{UB}(1)] \leftarrow [0, B \max_{j \in \mathcal{N} \cup \mathcal{F}} r(j)]$.
3: For all ℓ, maintain $\text{MP}(\ell) = 0.5(\text{LB}(\ell) + \text{UB}(\ell))$.
4: **for** $\ell = 1, 2, \cdots, M = \left\lceil \log_2 \left(\frac{\text{UB}(0)}{\delta} \right) \right\rceil$ **do**
5: Perform FEASIBILITY$(\epsilon, \text{MP}(\ell))$, which returns $(\tilde{y}, \tilde{z})$.
6: **if** $(\tilde{y}, \tilde{z})$ is $(1 - 2\epsilon)\text{MP}(\ell)$successful, **then**
7: Set $(\hat{y}, \hat{z}) \leftarrow (\tilde{y}, \tilde{z})$.
8: Set $[\text{LB}(\ell + 1), \text{UB}(\ell + 1)] \leftarrow [\text{MP}(\ell), \text{UB}(\ell)]$.
9: **else**
10: Set $[\text{LB}(\ell + 1), \text{UB}(\ell + 1)] \leftarrow [\text{LB}(\ell), \text{MP}(\ell)]$.
11: **end if**
12: **end for**
13: **return** $(\hat{y}, \hat{z})$.

Next, we show that the monotonicity in the potential function Ω implies that Algorithm 2 returns a feasible solution to (CDLP-P) with strong guarantee:

Lemma 2.4.3. *For a given $\epsilon \in [0, 1/2]$ and $Z \geq 0$, suppose the potential function Ω satisfies $\Omega(\mathcal{T}) \leq \Omega(0)$, where $\mathcal{T} \geq \frac{3\gamma \log(K+1)}{\epsilon^2}$. Then the output solution $(\tilde{y}, \tilde{z})$ is $(1 - 2\epsilon)Z-$successful.*

The proof, given in Appendix A.5, follows from unraveling the definition of $\Omega(\mathcal{T}), \Omega(0)$; the assumption that $\Omega(\mathcal{T}) \leq \Omega(0)$ turns out to implies that the returned solution is $(1 - 2\epsilon)Z-$successful. To motivate the proof, the inequality $\Omega(\mathcal{T}) \leq \Omega(0)$ implies the following bound for the potential function of resource k: $\Psi_k(\mathcal{T}) \leq \Omega(0)$. Note that $\Psi_k(\mathcal{T})$ is an increasing function of $\sum_{s=1}^{\mathcal{T}} U_k(s)$, which is the ratio of the total amount of resource k consumed to the capacity $c(k)$. The parameters are designed in such a way that unraveling the inequality $\Psi_k(\mathcal{T}) \leq \Omega(0)$ yields the feasibility for resource k constraint. A similar argument for $\Xi(\mathcal{T}) \leq \Omega(0)$ demonstrates that $\text{Val}(\hat{y})$ approximately achieves the target Z.

Using Lemmas 2.4.2, 2.4.3, we prove Theorem 2.3.4 in Appendix A.6 by tallying the total number of oracle calls invoked and elementary operations performed.

Finally, we provide empirical enhancement to PB in Appendix A.15, and compare the performance of ACG, CG and PB with simulation in Appendix A.16. While comparatively CG is often inefficient on small number of products, ACG is very efficient for moderate size problem instances, and PB sometimes perform better than ACG when the instances are large.

2.5 Online Models with Unknown MNL

Apart from computational tractability, model uncertainty is also a challenge in choice-based network revenue management. In this Section, we consider the online choice-based NRM problem with the MultiNomial Logit (MNL) choice model $(\varphi(\cdot,\cdot|\theta^*),\mathbb{S})$, where the underlying mean utility parameter θ^* is unknown. The problem can be cast as a Bandit-with-Knapsack problem [Badanidiyuru et al., 2013] with the number of arms exponential in N.

The Online Model. We assume the "Products and Resource Consumption Model" in § 2.2, with the following additional assumptions. We restrict $\mathcal{F}=\emptyset$, thus we refer to $\mathcal{N}$ as the set of all products. For normalization sake, we assume $r(i)\in[0,1]$ for all $i\in\mathcal{N}$, $a(i,k)\in\{0,1\}$ for all $i\in\mathcal{N},k\in\mathcal{K}$, $Tc(k)\in\mathbb{Z}^+$ for all $k\in\mathcal{K}$. Additionally, we assume the "Sales Dynamics" in § 2.2.

Since θ^* is unknown, the choice probability φ is now uncertain, different from § 2.2. The monopolist knows that $\varphi(\cdot,\cdot|\theta^*):\mathcal{N}\times\mathbb{S}\to[0,1]$ follows the MNL choice probability [Luce, 1959, McFadden, 1974, Placket, 1975] for some mean utility parameter $\theta^*=\{\theta^*(i)\}_{i\in\mathcal{N}}$, which is unknown to the monopolist. She only knows that $\theta^*(i)\in[-R,R]$ for all $i\in\mathcal{N}$.

In the MNL choice model $(\varphi(\cdot,\cdot|\theta),\mathbb{S})$ with parameter θ, a customer purchases at most 1 product; thus $B=1$. We denote $C=\max\{|S|:S\in\mathbb{S}\}$. The probability of purchasing

$i \in S$ when the customer is offered $S \in \mathbb{S}$ is

$$\varphi(i, S|\theta) = \frac{\exp[\theta(i)]}{1 + \sum_{\ell \in S} \exp[\theta(\ell)]},$$

and the probability of purchasing nothing is $\varphi(\emptyset, S|\theta) = \frac{1}{1+\sum_{\ell \in S} \exp[\theta(\ell)]} = 1 - \sum_{i \in S} \varphi(i, S|\theta)$. $\theta(\ell)$ can be interpreted as the appeal of product ℓ. Recall that $\emptyset \in \mathbb{S}$, as stated in §2.2. We make the following mild assumption on $\mathbb{S}$ to allow a fast learning rate.

Assumption 2.5.1. For all $i \in \mathcal{N}$, the single product assortment $\{i\}$ belongs to $\mathbb{S}$.

Regret. The monopolist's **Objective** is still to maximize the total revenue under the resource constraints. Following the online RM literature (for e.g., in [Besbes and Zeevi, 2012, Rusmevichientong et al., 2010, Saure and Zeevi, 2013]), we formulate the objective as the minimization of *regret*. The monopolist's objective is to design a *non-anticipatory* policy that minimizes

$$\text{REG} = T\text{OPT}(\theta^*) - \sum_{t=1}^{T} r(\Sigma_t), \tag{2.16}$$

subject to the resource constraints: for all $k \in \mathcal{K}$, we require $\sum_{t=1}^{T} a(\Sigma_t, k) \leq Tc(k)$. Note that the objective REG is a random variable depending on S_t, Σ_t. The quantity $\text{OPT}(\theta^*)$ is the optimal value of CDLP-P(θ^*). For any $\theta \in \mathbb{R}^{\mathcal{N}}$, CDLP-P($\theta$) denotes the CDLP-P for the MNL choice model with parameter θ, as displayed. The benchmark $T\text{OPT}(\theta^*)$ is motivated by the following Theorem due to [Badanidiyuru et al., 2013, Gallego et al., 2004]:

$$\begin{aligned}
\max \quad & \sum_{S \in \mathbb{S}} \sum_{i \in \mathcal{N}} r(i) \varphi(i, S \mid \theta)\, y(S) \\
\text{s.t.} \quad & \sum_{S \in \mathbb{S}} \sum_{i \in \mathcal{N}} a(i,k) \varphi(i, S \mid \theta) y(S) \leq c(k) \quad \forall k \in \mathcal{K} \\
& \sum_{S \in \mathbb{S}} y(S) = 1, \quad y(S) \geq 0 \qquad \forall S \in \mathbb{S}
\end{aligned}$$

CDLP-P(θ); Optimal value = $\text{OPT}(\theta)$.

Theorem 2.5.2 ([Badanidiyuru et al., 2013, Gallego et al., 2004]). *For any given non-*

anticipatory policy that satisfies the resource constraints with probability 1, the following inequality holds:

$$T\textsc{Opt}(\theta^*) \geq \mathbb{E}\left[\sum_{t=1}^{T} r\left(\Sigma_t\right)\right].$$

S_t denotes the assortment offered by the monopolist under her policy at period t, and $\Sigma_t \in S_t \cup \{\emptyset\}$ denotes the purchased product. $\Sigma_t = \emptyset$ implies that nothing is purchased at period t. A policy is *non-anticipatory* if the offered assortment S_t depends only on the sales history as well as the monopolist's randomness U_t in period t, i.e. $S_t = \pi(U_t, \{S_s, \Sigma_s, U_s\}_{s=1}^{t-1})$.

Near-optimal Online Policy. We propose the non-anticipatory policy $\textsc{Online}(\tau)$, where τ is the length of the learning phase. $\textsc{Online}(\tau)$ enjoys the following performance guarantee:

Theorem 2.5.3. *Suppose that Assumptions 2.5.1 holds, and τ satisfies Assumption 2.5.4. The policy $\textsc{Online}(\tau)$ (Algorithm 4) satisfies all resource constraints and achieves the regret at most*

$$\textsc{Bound}(\tau) = \tau + O\left(TCe^R\sqrt{\frac{N}{\tau}\log\frac{N}{\delta}}\right) + O\left(\sqrt{T\log\frac{K+1}{\delta}}\right) \tag{2.18}$$

with probability $1-\delta$. Furthermore, assuming an oracle for $(\varphi(\cdot,\cdot|\theta), \mathbb{S})$ for any choice of $\theta \in [-R, R]^N$, $\textsc{Online}(\tau)$ runs in polynomial time.

While our online problem is an instance of the Bandits-with-Knapsack problem proposed by [Badanidiyuru et al., 2013], an application of the policy by [Badanidiyuru et al., 2013] yields a regret bound of $O(\sqrt{|\mathbb{S}|T})$, which holds only when $T = \Omega(|\mathbb{S}|)$. Note that both bounds are exponential in N when $|\mathbb{S}| = \Omega(2^N)$. However, by exploiting the dependence among the choice probabilities under different assortments, we reduce the dependence on N from being exponential to polynomial in these bounds. Next, Assumption 2.5.4 on τ is stated in the following:

Assumption 2.5.4. The length τ of the learning phase τ satisfies the following the following two inequalities for all $k \in \mathcal{K}$: (i) We have $\tau\sqrt{\log\frac{4NK}{\delta}} \leq Tc(k)$. (ii) We have

$C\epsilon(\tau) \leq \frac{1}{2}c(k)$, where

$$\epsilon(\tau) = 4e^R\sqrt{\frac{N}{\tau}\log\frac{4N}{\delta}}. \tag{2.19}$$

The bound in Assumption 2.5.4 (i) ensures that none of the resource is depleted during the learning phase of length τ, while the bound in (ii) ensures that the learning phase is long enough for a non-trivial performance guarantee.

The choice of $\tau = (CTe^R)^{2/3}N^{1/3}$, which optimizes the regret order bound in (2.18), yield the bound $\tilde{O}((CTe^R)^{2/3}N^{1/3})$. Another natural choice of τ is $T^{2/3}$, which yields a regret bound of $\tilde{O}(Ce^R\sqrt{NT})$. These choices are valid when T is sufficiently large, as discussed in Appendix A.8.

ONLINE(τ) is presented in Algorithm 4. Period 1 to period τ are the learning phase, and period $\tau+1$ to period T are the earning phase. During the learning phase, the monopolist offers single item assortments to learn $\theta^\star$. In Line 7, she computes the MLE $\hat{\theta}(i)$:

$$\hat{\theta}(i) = \underset{\theta(i)\in[-R,R]}{\operatorname{argmin}}\ \mathcal{L}^i(\theta(i))\ \text{, where } \mathcal{L}^i(\theta(i)) = \sum_{s=((i-1)\tau/N)+1}^{i\tau/N} -\log(\varphi(\Sigma_s,\{i\}|\theta(i))). \tag{2.20}$$

$\mathcal{L}^i(\theta(i))$ has the following simpler form:

$$\mathcal{L}^i(\theta(i)) = N(i)\log\left(1+\exp\left[-\theta(i)\right]\right) + \left(\frac{\tau}{N} - N(i)\right)\log\left(1+\exp\left[\theta(i)\right]\right), \tag{2.21}$$

where $N(i) = \sum_{s=((i-1)\tau/N)+1}^{i\tau/N} 1(\Sigma_s = i)$ is the frequency of successful sales for product i.

After that, in Line 9, we solve the CDLP-P($\hat{\theta}$) for $\hat{y}$, which can be interpreted as a probability distribution over the assortment family $\mathbb{S}$. Finally, we enter the earning phase, and offer an assortment S with probability $\hat{y}(S)$. At the end of a period, the policy signals ABORT to stop the sales process when some resource is depleted. This ensures that the resource constraints are satisfied with probability 1. By (i) in Assumption 2.5.4, all resources are never depleted in the learning phase, hence ABORT *is never signaled during the learning phase.*

Algorithm 4 Online Policy ONLINE(τ)

Initialize $C(k) = Tc(k)\ \forall k \in \mathcal{K}$.
for $i = 1, \cdots, N$ **do** ▷ Learning
  for $t = (i-1)\tau/N + 1$ to $i\tau/N$ **do**
    Offer assortment $S_t = \{i\}$, and observe the purchased product Σ_t.
    For all $k \in \mathcal{K}$, $C(k) \leftarrow C(k) - a(\Sigma_t, k)$.
  end for
  Compute the MLE $\hat{\theta}(i)$ in (2.20).
end for
Solve CDLP-P($\hat{\theta}$) for the solution $\hat{y}$.
for $t = \tau + 1, \cdots, T$ **do** ▷ Earning
  Offer an assortment S_t with probability $\hat{y}(S_t)$.
  Update resource levels. For all $k \in \mathcal{K}$, $C(k) \leftarrow C(k) - a(\Sigma_t, k)$.
  if $\exists k \in \mathcal{K}$ s.t. $C(k) = 0$ **then**
    Signal ABORT, break the **for**-loop and offer $S = \emptyset$ to the remaining customers.
  end if
end for

Analysis. A challenge in proving the regret bound is that the period t_{stop} when the policy signals ABORT is a random variable depending on θ^* and τ in a contrived manner. This makes a direct analysis on the total revenue $\sum_{t=1}^{t_{\text{stop}}} r(\Sigma_t)$ difficult. We overcome this technical challenge by first showing that $t_{\text{stop}} \leq T - \rho$ with high probability, where ρ is defined as follows:

$$\rho = \frac{TC\epsilon(\tau)}{\min_{k\in\mathcal{K}} c(k)} + \frac{\sqrt{T \log \frac{4(K+1)}{\delta}}}{\min_{k\in\mathcal{K}} c(k)}, \tag{2.22}$$

where $\epsilon(\tau)$ is defined in (2.19).Thus, REG is upper bounded by the sum of the following three regret terms: (1) The regret due to learning in the first τ periods, (2) the regret due to estimation error on θ^* during the earning periods $\tau + 1, \cdots, T - \rho$, where no ABORT is signaled, and (3) the regret due to ABORTING (possibly) the sales process in the last ρ periods. The regret due to (1, 3) is at most $\tau + \rho$. Next, the error to the estimation on the choice probability is $\epsilon(\tau)$, which translates to a regret of at most $\tilde{O}(T\epsilon(\tau))$ for (2). Altogether, this leads to the $\tau + \rho + \tilde{O}(T\epsilon(\tau)) = \tilde{O}(\tau + T\tau^{-1/2} + \sqrt{T})$ regret bound.

To formalize the argument, we consider the following sales process $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$ gen-

erated by Procedure 5, which resembles Algorithm 4. In Procedure 5, the notation $\tilde{\Sigma}_t \sim \tilde{S}_t$ denotes sampling a product $\tilde{\Sigma}_t$ from $\tilde{S}_t \cup \{\emptyset\}$ with the underlying choice probability $\varphi(\tilde{\Sigma}_t, \tilde{S}_t|\theta^*)$. We emphasize that $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$ is only used for the analysis; the monopolist does not need to generate such a process. (Such generation is not needed in Algorithm 4).

The sales process $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$ is closely related to the process $\{S_t, \Sigma_t\}_{t=\tau+1}^{T-\rho}$ generated by Algorithm 4. If $t_{\text{stop}} > T-\rho$, the processes $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$ and $\{S_t, \Sigma_t\}_{t=\tau+1}^{T-\rho}$ are identically distributed. Otherwise, we have $t_{\text{stop}} \leq T-\rho$, which means ABORT is signaled before the end of period $T-\rho$. Then, $S_t = \emptyset = \Sigma_t$ for $t = t_{\text{stop}}+1, \cdots, T-\rho$, which in general is distributed differently from $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=t_{\text{stop}}+1}^{T-\rho}$. For further comparisons, recall that Algorithm 4 satisfies the resource constraints with probability 1, i.e. $\sum_{t=1}^{T} a(\Sigma_t, k) \leq Tc(k)$ with certainty; but $\sum_{t=1}^{\tau} a(\Sigma_t, k) + \sum_{t=\tau+1}^{T} a(\tilde{\Sigma}_t, k) > Tc(k)$ with positive (despite being exponentially small) probability.

Procedure 5 Generation of $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$

1: **for** $t = \tau+1, \cdots, T-\rho$ **do**
2: Sample the assortment $\tilde{S}_t$ with probability $\hat{y}(\tilde{S}_t)$, where $\hat{y}$ is an opt solution to CDLP-P($\hat{\theta}$).
3: Sample $\tilde{\Sigma}_t \sim \tilde{S}_t$
4: **end for**

We prove the regret bound in Theorem 2.5.3 as follows:

$$\mathbb{P}\left[\text{REG} \leq \text{BOUND}(\tau)\right] = \mathbb{P}\left[T\text{OPT}(\theta^*) - \sum_{t=1}^{t_{\text{stop}}} r(\Sigma_t) \leq \text{BOUND}(\tau)\right]$$

$$\geq \mathbb{P}\left[T\text{OPT}(\theta^*) - \sum_{t=1}^{T-\rho} r(\Sigma_t) \leq \text{BOUND}(\tau), t_{\text{stop}} > T-\rho\right]$$

$$\overset{(\dagger)}{=} \mathbb{P}\left[\left\{T\text{OPT}(\theta^*) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t) \leq \text{BOUND}(\tau)\right\} \cap \left\{\sum_{t=1}^{\tau} a(\Sigma_t, k) + \sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) < Tc(k)\ \forall k\right\}\right]$$

$$\overset{(\ddagger)}{\geq} \mathbb{P}\left[\underbrace{\left\{T\text{OPT}(\theta^*) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t) \leq \text{BOUND}(\tau)\right\}}_{\mathcal{E}_{\text{REG}}} \cap \bigcap_{k=1}^{K} \underbrace{\left\{\tau + \sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) \leq Tc(k)\right\}}_{\mathcal{E}_k} \mid \mathcal{E}_{\hat{\theta}}\right] \mathbb{P}[\mathcal{E}_{\hat{\theta}}].$$

The equality (†) is justified as follows. The event $\{t_{\text{stop}} > T - \rho\}$ is equivalent to the event that the resource level $C(k)$ is positive at the end of period $T - \rho$. This is in turn equivalent to the event that $\{S_t\}_{t=\tau+1}^{T-\rho}$ are i.i.d. according to the probability $y^\star$ (instead of forcing $S_t = \emptyset$), since no ABORT is signaled. Thus, by our preceding discussion, we can replace $\{S_t, \Sigma_t\}_{t=\tau+1}^{T-\rho}$ with $\{\tilde{S}_t, \tilde{\Sigma}_t\}_{t=\tau+1}^{T-\rho}$. In Line (‡), the event $\mathcal{E}_{\hat{\theta}}$ concerns the accuracy of the MLE $\hat{\theta}$:

$$\mathcal{E}_{\hat{\theta}} = \left\{ \hat{\theta} \text{ satisfies } \left|\hat{\theta}(i) - \theta^\star(i)\right| \leq \epsilon(\tau) = 4e^R \sqrt{\frac{N}{\tau} \log \frac{4N}{\delta}} \text{ for all } i. \right\} \tag{2.23}$$

In the following, we will argue that the event $\mathcal{E}_{\hat{\theta}}$ happens with high probability, and conditional on $\mathcal{E}_{\hat{\theta}}$, the events $\mathcal{E}_{\text{REG}}$, $\{\mathcal{E}_k\}_{k=1}^K$ happen with high probability. More formally, we first demonstrate that $\theta^\star$ is estimated with the stipulated accuracy with high probability:

Lemma 2.5.5. *For any $\tau \geq N$, $\mathbb{P}[\mathcal{E}_{\hat{\theta}}] \geq 1 - \delta/2$.*

The proof of Lemma 2.5.5, provided in Appendix A.9, crucially uses the strong convexity of the negative log likelihood $\mathcal{L}^i(\theta)$. Next, we translate the accuracy in estimating $\theta^\star$ to the choice probability:

Lemma 2.5.6. *The inequality $\sum_{i \in S} b(i) \left(\varphi(i, S|\theta) - \varphi(i, S|\theta')\right) \leq \sum_{i \in S} |\theta(i) - \theta'(i)|$ holds for all $\theta, \theta' \in \mathbb{R}^{\mathcal{N}}$, for all $b \in [0,1]^N$, and for all $S \subset \mathcal{N}$.*

The proof of Lemma 2.5.6 is given in Appendix A.10. After that, conditional on the accuracy in estimating $\theta^\star$, we show that the event $\mathcal{E}_k$ happens with high probability, by using Lemma 2.5.6 and applying Chernoff inequality:

Lemma 2.5.7. *For every $k \in \mathcal{K}$, we have $\mathbb{P}\left[\mathcal{E}_k \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{2(K+1)}$.*

The proof of Lemma 2.5.7 is given in Appendix A.11, which explains the choice of ρ in (2.22). Finally, we show that conditional on $\mathcal{E}_{\hat{\theta}}$, the regret bound holds with high probability. For the argument we relates the optimal value to CDLP-P($\theta^\star$), namely to $\text{OPT}(\hat{\theta})$, to $\text{OPT}(\hat{\theta})$. Note that $\text{OPT}(\theta) \in [0,1]$ for all θ.

Lemma 2.5.8. *Condition on the event $\mathcal{E}$, we have* $\textsc{Opt}(\hat{\theta}) \geq \left[1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right] \textsc{Opt}(\theta^*) - C\epsilon(\tau)$.

Assumption 2.5.4 (ii) ensures that $\frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}} < 1$. The proof of Lemma 2.5.8 is given in Appendix A.12. Using Lemma 2.5.8 and Chernoff Inequality, we prove the following Lemma in Appendix A.13:

Lemma 2.5.9. *We have* $\mathbb{P}\left[\mathcal{E}_{\textsc{Reg}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{2(K+1)}$

Altogether, the regret bound in Theorem 2.5.3 is proved. Finally, we remark that $\textsc{Online}(\tau)$ is can be implemented in polynomial time as long as the underlying SPP is polynomial time solvable, by Theorem 2.3.2. Nevertheless, it in fact suffices to approximate CDLP-P(θ) within a factor of $(1 - 1/T^{1/3})$, for example by the Potential Based Algorithm.

2.6 Numerical Results on $\textsc{Online}(\tau)$

We numerically test the performance of $\textsc{Online}(\tau)$ proposed in § 2.5 on random instances.

Generating an online problem instance. Suppose we are given the *instance class tuple* $\Gamma = (\mathbb{S}, N, K, R)$, where $\mathbb{S}$ is the assortment family, N is the number of products, K is the number of resources, R is the upper bound on $\{|\theta^*(i)|\}_{i=1}^N$. We propose Procedure 8, displayed in Appendix A.14, to generate a *random output tuple* $\Lambda = (r, (\varphi(\cdot,\cdot|\theta^*), \mathbb{S}), A, c)$ for the online problem with $c(k) \geq 0.1$, where $r \in [0,1]^N$ is the revenue coefficients, $(\varphi(\cdot,\cdot|\theta^*)$ is the underlying MNL choice model, $A \in \{0,1\}^{N\times K}$ is the resource consumption matrix and $c \in [0,1]^K$ is the per period capacity vector. Given the output tuple Λ from Procedure 8, we can define an instance $(\Lambda, T) = (r, (\varphi(\cdot,\cdot|\theta^*), \mathbb{S}), A, c, T)$ for the choice based NRM with MNL choice model for any length of sales horizon T, where the amount of resource k available at the beginning is $Tc(k)$.

Simulation Results - Revenue and Regret. We evaluate the performance of the online policy with varying assortment families and parameters. For setting up, we let $\mathbb{S}_1(C) = \{S \subset \mathcal{N} : |S| \leq C\}$ denote the cardinality constrained family, and let $\mathbb{S}_2(p, \Pi) =$

$\{S \subset \mathcal{N} : |S \cap \mathcal{N}_j| \leq \Pi \text{ for all } 1 \leq j \leq p\}$ denote the partition matroid family. Here, $\mathcal{N}_1, \cdots, \mathcal{N}_p$ is a partition of the set of product $\mathcal{N}$ into p disjoint subsets of cardinality N/p. (Thus, we implicitly assume that N is divisible by p). By [Davis et al., 2013], $\text{SPP}(r, \varphi(\cdot, \cdot|\theta), \mathbb{S}_1(C))$ and $\text{SPP}(r, \varphi(\cdot, \cdot|\theta), \mathbb{S}_2(p, \Pi))$ are both polynomial time solvable, hence their respective CDLPs could be solved efficiently, by §2.4.

The online policy Algorithm 4 is evaluated on random instances generated based on the following 6 instance class tuples $\{\Gamma_\ell\}_{\ell=1}^6$ and 8 choices $\{\mathcal{T}(q)\}_{q=1}^8$ of sales horizon length:

$$\Gamma_1 = (\mathbb{S}_1(6), 10, 5, 3), \quad \Gamma_2 = (\mathbb{S}_1(9), 15, 6, 5), \quad \Gamma_3 = (\mathbb{S}_1(15), 25, 8, 7),$$
$$\Gamma_4 = (\mathbb{S}_2(2, 3), 10, 5, 3), \quad \Gamma_5 = (\mathbb{S}(3, 3), 15, 6, 5), \quad \Gamma_6 = (\mathbb{S}_2(5, 3), 25, 8, 7),$$
$$\mathcal{T} = [250, 500, 750, 100, 1500, 2000, 5000, 10000].$$

For each instance class tuple Γ_ℓ, we generate 5 random output tuples $\{\Lambda_{\ell,w}\}_{w=1}^5$ independently. We then run Online$(\mathcal{T}(q)^{2/3})$ for 200 times on the instance $(\Lambda_{\ell,w}, \mathcal{T}(q))$, for each $\ell \in \{1, \cdots, 6\}, w \in \{1, \cdots, 5\}, q \in \{1, \cdots, 8\}$.

For each ℓ, w, q, $\textsc{Revenue}(\ell, w, q)$ denotes the average revenue of the 200 trials of Online$(\mathcal{T}(q)^{2/3})$ over the instance $(\Lambda_{\ell,w}, \mathcal{T}(q))$; $\textsc{Optimum}(\ell, w, q)$ denotes the optimal expected revenue for the instance $(\Lambda_{\ell,w}, \mathcal{T}(q))$, i.e. $\textsc{Optimum}(\ell, w, q) = \mathcal{T}(q)\textsc{Opt}(\theta^*_{\ell,w})$. The choice of $\tau = \mathcal{T}(q)^{2/3}$ may not always satisfy Assumption 2.5.4 for the instance $(\Lambda_{\ell,w}, \mathcal{T}(q))$; however, our simulation results indicate that the online policy performs well even when the assumption is violated.

Figure 2-1 depicts the trend of the average revenue to optimum ratio,

$$\frac{1}{5}\sum_{w=1}^{5} \frac{\textsc{Revenue}(\ell, w, q)}{\textsc{Optimum}(\ell, w, q)},$$

for the cardinality constrained family instance class tuples $(\Gamma_1, \Gamma_2, \Gamma_3)$ and the partition matroid family instance class tuples $(\Gamma_4, \Gamma_5, \Gamma_6)$. It is observed that the average revenue to optimum ratio achieved by $\textsc{Online}(\mathcal{T}(q)^{2/3})$ converges to 1 as $\mathcal{T}(q)$ increases. Apart from asymptotic optimality, $\textsc{Online}(\mathcal{T}(q)^{2/3})$ also performs well when $\mathcal{T}(q)$ is small. For

example, in the random instances generated based on Γ_3, the monopolist has to select an assortment from $\mathbb{S}_1(15)$, which contains about 3.17×10^7 assortments. Despite the sheer size of $\mathbb{S}_1(15)$, Online$(250^{2/3})$ is able to achieve a revenue to optimum ratio of 0.68 in the instances where $T = 250$. This demonstrates the benefit and effectiveness of learning θ^*. If the monopolist are to estimate the choice probability assortment by assortment, it will require many more periods to learn an asymptotically optimal policy.

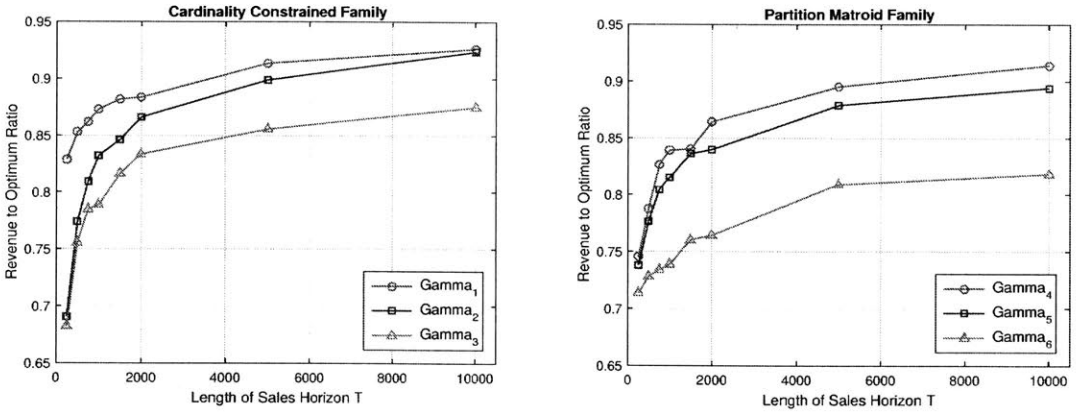


Figure 2-1: Revnue to optimum ratios for $\mathbb{S}_1$ (left) and $\mathbb{S}_2$ (right).

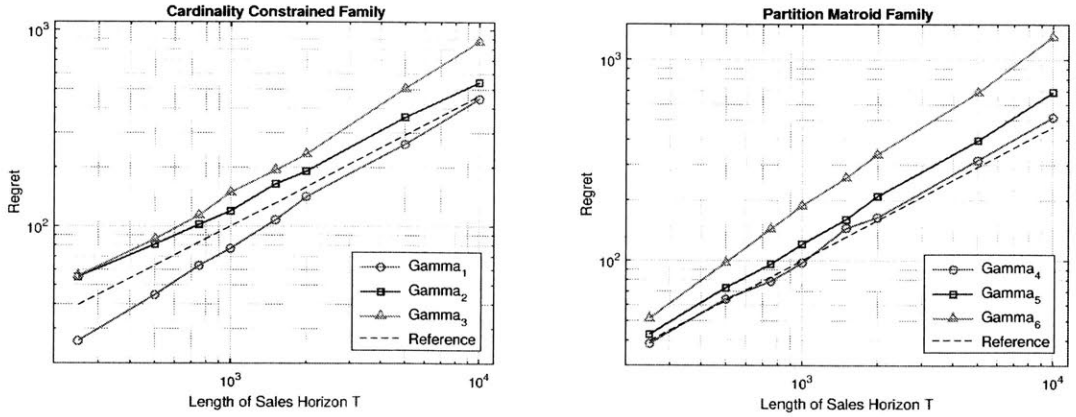


Figure 2-2: Regret for $\mathbb{S}_1$ (left) and $\mathbb{S}_2$ (right), displayed in log-log scale.

Class \ T	500	1000	2000	5000	10000
$(\mathbb{S}_1(4), 6, 5, 3)$	0.41	0.40	0.43	0.55	0.51
$(\mathbb{S}_1(6), 10, 5, 3)$	0.10	0.12	0.17	0.27	0.42
$(\mathbb{S}_1(9), 15, 5, 3)$	0.06	0.10	0.08	0.12	0.14
$(\mathbb{S}_1(10), 20, 9, 5)$	0.09	0.05	0.01	0.03	0.04

Table 3: Fraction of instances where $\hat{\mathbb{S}} = \mathbb{S}^*$ for $\mathbb{S}_1$.

Class \ T	500	1000	2000	5000	10000
$(\mathbb{S}_2(2, 2), 6, 5, 3)$	0.47	0.45	0.52	0.56	0.65
$(\mathbb{S}_2(2, 2), 10, 5, 3)$	0.29	0.30	0.34	0.4	0.43
$(\mathbb{S}_2(3, 3), 15, 5, 3)$	0.09	0.12	0.22	0.28	0.26
$(\mathbb{S}_2(4, 3), 20, 9, 5)$	0.02	0.02	0.09	0.12	0.10

Table 4: Fraction of instances where $\hat{\mathbb{S}} = \mathbb{S}^*$ for $\mathbb{S}_2$.

After witnessing the convergence to optimum, we investigate the growth rate of regret of the online policy as the length of sales horizon increases. Figure 2-2 depicts the trend of the average regret,

$$\frac{1}{5}\sum_{w=1}^{5}\left[\textsc{Optimum}(\ell, w, q) - \textsc{Revenue}(\ell, w, q)\right],$$

for the cardinality constrained family instance class tuple $(\Gamma_1, \Gamma_2, \Gamma_3)$ and the partition matroid family instance class tuples $(\Gamma_4, \Gamma_5, \Gamma_6)$. The graphs in Figure 2-2 are plotted in log-log scale; the regret curves are compared against the reference curve $\text{ref}(T) = T^{2/3}$, the dashed black curve. Under the log-log scaling, the reference curve is a straight line with gradient $2/3$.

By Theorem 2.5.3, the regret of $\textsc{Online}(\mathcal{T}(q)^{2/3})$ is at most $O(\mathcal{T}(q)^{2/3})$. Thus, the Theorem implies that the curves for the regret should have gradients at most $2/3$ when plotted in a log-log scale. Figure 2-2 demonstrates that the curves for the regret is essentially parallel to the reference curve; this implies that in the simulated examples, the growth rate of the regret is essentially $\Theta(T^{2/3})$. The vertical distances between the regret curves and the reference curves is the logarithm of the multiplicative constant in $\Theta(T^{2/3})$.

Simulation Results - Learning the Right Assortment Combination. Apart from evaluating the revenue earned, we measure the accuracy of the online policy by comparing the empirical support $\hat{\mathbb{S}} := \{S \in \mathbb{S} : \hat{y}(S) > 0\}$ with the optimal support $\mathbb{S}^* := \{S \in \mathbb{S} : y^*(S) > 0\}$. The solutions $\hat{y}, y^*$ are the optimal solutions to $\text{CDLP}(\hat{\theta})$, $\text{CDLP}(\theta^*)$ respectively, where $\text{CDLP}(\hat{\theta})$ is constructed in Line 9 in Algorithm 4. $\hat{y}, y^*$ are chosen to be extreme point solutions, so that they have sparse supports, i.e. $|\hat{\mathbb{S}}|, |\mathbb{S}^*| \leq K + 1$.

For each instance class tuple (labeled 'Class') and each length of sales horizon (labeled 'T') displayed in Tables 3, 4, we generate 100 choice NRM problem instances at random, using Procedure 8. Then we apply ONLINE($T^{2/3}$) once on each of the random instances. After that, we check if $\hat{\mathbb{S}} = \mathbb{S}^*$, that is, the online policy correctly identify the optimal support $\mathbb{S}^*$. An incorrect identification, i.e. $\hat{\mathbb{S}} \neq \mathbb{S}^*$, can be due to a poor estimation on θ^* or to the multiplicity of optimal solutions to CDLP(θ^*) and CDLP($\hat{\theta}$) (or both). Finally, for each displayed 'Class' and each 'T', we compute the fraction of random instances where $\hat{\mathbb{S}} = \mathbb{S}^*$. These fractions are displayed in Table 3, 4.

The results in Table 3, 4 show that the online policy often succeeds in identifying $\mathbb{S}^*$ when T is long enough. The results also indicate that online policy is more effective in correctly identifying $\mathbb{S}^*$ in the case of partition matroid family $\mathbb{S}_2$ than in the case of cardinality constrained family $\mathbb{S}_1$.

2.7 Conclusions

Efficient algorithms are proposed for solving the Choice-based Deterministic Linear Program (CDLP), a central tool for the choice-based network revenue management problem. Assuming the ability to approximately solve a Single Period Problem (SPP), we propose the Approximate Column Generation heuristic (ACG), which generalizes the classical Column Generation heuristic and solves CDLP to near optimality. We also the Potential Based algorithm (PB) that solves (CDLP-P) to near optimality with provable time efficiency. The key implication of our results is that, to design a computational efficient algorithm for solving a CDLP, it suffices to solves the underlying SPP efficiently. This directly translates the efficient algorithms for solving SPPs for various choice models in the literature to efficient algorithms for solving their corresponding CDLPs, many of which are not previously known. Our results also imply that many assortment planning strategies in the literature, which rely on solving the corresponding CDLP, could be implemented efficiently for a wide variety of choice models.

Building on the tractability results, we also design a non-anticipatory policy ONLINE(τ) for the online choice-based network revenue management problem with an unknown Multinomial Logit choice model. The regret incurred by the policy is sublinear in the length of time horizon, and the policy can be implemented in polynomial time for a variety of assortment families. Experiment results shows that the policy is effective even when the length of the sales horizon is small compared to the number of products.

Chapter 3

Dynamic Pricing with Limited Price Experimentation

3.1 Introduction.

The online retail environment provides a platform for a seller to conduct her sale through a posted price mechanism, which determines a product offered price to an incoming customer based on the previous sale outcomes. Such an ability to price a product dynamically is particularly relevant when the functional relationship between the price and mean product demand rate is not completely known. Such uncertainty arises when the seller introduces a new product to the market, or when she interacts with a new population of customers in an online environment for example. Under such partial information, the seller is motivated to conduct price experimentation in order to learn consumer sensitivity and maximize revenue, and these two goals can be achieved by harnessing the power of dynamic pricing. This online revenue maximization problem under partial information on consumer preference is in contrast to the conventional revenue maximization problem, which assumes that the price-demand relationship is completely known to the seller [Talluri and Van Ryzin, 2004, Ozer and Philips, 2012].

Certain online retails, such as Groupon, specify an upper limit on the number of permitted price changes due to administrative constraints, customer satisfaction, as well as the need to prevent strategic customer behavior [Aviv et al., 2009], which may lead to a revenue decrease. For a more detailed discussion, see [Aviv et al., 2009, Cheung et al., 2015, Wang, 2016]. This is in contrast to the airline industry, where frequent price change is the norm. Such price change limit gives rise to a tension between *constrained learning* and *earning*. On one hand, to ascertain the optimal price with high confidence under limited number of price changes, the seller needs to have a sufficiently long learning phase in order to collect enough sale data. On the other hand, the seller also desires to minimize her potential loss in revenue during the learning phase, where her earning could have been compromised for the sake of extracting customer information. These two contradicting objectives beg the following questions: How much should the seller learn about the customers' perceived product values? What is the highest revenue the seller can obtain under both her information uncertainty and the price change constraint? How should the seller quantify the value of customers' information?

In this chapter, we shed light on the abovementioned questions by considering a single product dynamic pricing problem under limited experimentation. For every price change limit m, we prove a lower bound on the regret, defined as the loss in revenue due to partial information on demand distribution, achievable by any pricing strategy that changes price m times. Our analysis provides insights into the structure of any optimal pricing strategy, and quantify the loss in revenue when the seller deviates from the optimal strategy.

3.1.1 Literature Review.

Our point of departure is the following prototypical single product dynamic pricing problem, which is widely considered in the computer science, economics and operation research communities. Suppose a group of customers arrive at the market in an online manner during a finite time interval; each customer demands one unit of the product, and the cus-

tomers' perceived product values, i.e. the demand curves, are identically and independently distributed random variables. At any point of time in the sale season, the seller decides the price offered to the next customer, based on her sale history and her previously offered prices. The question is: How should the prices be offered in order to maximize the revenue?

In the traditional revenue management literature, this single product dynamic pricing problem is solved when the full knowledge of the demand curve is assumed. For the case of unlimited supply of the product, the optimal revenue is achieved by identifying the revenue optimal price for one customer and offering this price to every customer, for example see [Myerson, 1981]. For the case of limited supply, [Gallego and van Ryzin, 1994] characterize the optimal pricing strategy when the customer arrival process is a price-sensitive Poisson Process, by considering the Hamilton-Jacobi-Bellman equation of a stochastic control problem. Since an optimal pricing strategy does not always admit a closed form, they propose a fixed price hueristics with revenue converging to the optimal revenue as the length of sale horizon tends to infinity. Subsequently, [Feng and Gallego, 1995] consider the same continuous time arrival model, where at any time the seller is restricted to offer one of the two given prices, and there can be at most one price change. They give a characterization on the optimal time of price change. Furthermore, [Feng and Xiao, 2000] provide a characterization of the optimal policy when the offered prices belong to a finite set, and the prices are required to be monotonic (i.e. the pricing strategy has to be either a mark-up or mark-down strategy) across the time horizon. Note that the pricing strategies proposed in the abovementioned papers crucially depend on the underlying demand distribution.

Ever since the dawn of e-commerce, retailers have more opportunity to reach out to new populations of customers, whose demand profiles may not be completely known to the retailers. This leads a paradigm shift in revenue management, which now requires pricing strategies to extract information on the underlying demand curve while maximizing the total revenue. The performance of such a pricing strategy is often measured by its *regret*, which is defined as the loss in revenue compared to an oracle who has full knowledge on

the underlying demand curve.

For the case of unlimited supply, the single product dynamic pricing problem has been considered under different level of uncertainty on the demand curve. The research of [Kleinberg and Leighton, 2003] and [Harrison et al., 2012] are closest to ours. The work [Kleinberg and Leighton, 2003] considers the single product dynamic pricing problem where n customers arrive at the market in the discrete time model, and their common demand curve is completely unknown to the seller. By considering a suitable Multi-Armed Bandit problem, the work [Kleinberg and Leighton, 2003] provides a pricing strategy achieving a regret of $O(\sqrt{n \log n})$, among other results. The work [Harrison et al., 2012] consider the same dynamic pricing problem, except that the seller knows that the underlying demand curves has two possible identities (demand curves), instead of having no information on the demand curve at all–this corresponds to the "Hypothesis Testing" version of the problem in [Kleinberg and Leighton, 2003]. [Harrison et al., 2012] shows that a modified myopic Bayesian policy achieves a constant regret, but the number of price changes is in general unbounded.

For the case of limited supply, [Besbes and Zeevi, 2009] consider the continuous time arrival model which is the same as [Gallego and van Ryzin, 1994], except that now the underlying demand curve is unknown. They consider two different type of uncertainty: the non-parametric case where the demand curve is completely unknown, and the parametric case where the demand curve is known to belong to a single parameter family. Under a high-volume-of-sale regime, that is, when the initial supply level is in the same order of magnitude as the length of horizon T, [Besbes and Zeevi, 2009] proposes a pricing strategy that achieves a regret of $O(T^{3/4}(\log T)^{1/2})$ in the non-parametric case. In addition, they also propose a pricing strategy achieving a regret of $O(T^{2/3}(\log T)^{1/2})$ in the parametric case. [Wang et al., 2011] also consider the continuous time model, and they propose a pricing strategy that achieves a $O(n^{1/2}(\log n)^{9/2})$ regret in a general non-parametric setting, where n is the order of magnitude of the initial inventory and demand rate (For a more precise

definition on n, see page 6 in [Wang et al., 2011]). In the high-volume-of-sale regime, i.e. $n = \Theta(T)$, this implies a regret upper bound $O(T^{1/2}(\log T)^{9/2})$, which improves upon the results of [Besbes and Zeevi, 2009]. [Babaioff et al., 2012] considers the same model as [Kleinberg and Leighton, 2003], but with an additional constraint of having limited supply. They propose a pricing strategy that achieves a regret upper bound $O((k \log n)^{2/3})$, where k is the number of available products and n is the number of customers, by transforming the problem into a Multi-Armed Bandit problem.

In the pursuit to understand the value of demand curve information in dynamic pricing with unknown demands, numerous papers provide lower bounds on the regret achievable by any pricing strategy. For the case of unlimited supply, [Kleinberg and Leighton, 2003] proves a regret lower bound $o(\sqrt{n})$ in the discrete time model when there is no information about the demand curve. Their lower bound is proved by an involved information theoretic argument on a family of suitably chosen demand curves. The work by [Broder and Rusmevichientong, 2012] also proves a regret lower bound $\Omega(\sqrt{n})$ in a multi-dimensional parametric model in the discrete time model, with different assumptions on the family of possible demand curves from [Kleinberg and Leighton, 2003]. For the case of limited supply, the lower bound $\Omega(T^{1/2})$ is provided by [Besbes and Zeevi, 2009] and [Wang et al., 2011] under high-volume-of-sale regime in the continuous time model, and these two results on the lower bound differ in their assumption on the underlying demand curve. [Babaioff et al., 2012] proves a different lower bound $o(k^{2/3})$ in the discrete time model, where k is the amount of inventory, by a black box reduction from the lower bound in [Kleinberg and Leighton, 2003]. Note that in the abovementioned list of works, in order to match the regret lower bound, the underlying demand curve has to satifies numerous assumptions, such as regularity [Kleinberg and Leighton, 2003], [Besbes and Zeevi, 2009], [Babaioff et al., 2012]; Lipschitz continuity [Besbes and Zeevi, 2009], the research work by [Broder and Rusmevichientong, 2012], as well as [Wang et al., 2011]; and differentiability of the mean demand rate function [Broder and Rusmevichientong, 2012], [Wang et al., 2011],

[Babaioff et al., 2012], as well as [Kleinberg and Leighton, 2003]; as well as some other technical assumptions. Surprisingly, in the current chapter, the regret lower bound is matched by a broader class of demand curves.

Apart from the single product dynamic pricing problem, different versions of dynamic pricing problems are considered in the literature. [Besbes and Zeevi, 2011] considers the single product dynamic pricing problem where there is one demand curve change at a random time during the sale horizon. Pricing strategies for multi-product dynamic pricing problems are proposed in [Besbes and Zeevi, 2012] as well as in [den Boer, 2011]. For a more thorough survey on various versions of dynamic pricing problems under unknown demands, see [den Boer, 2014].

3.2 Problem Formulation

We consider a seller offering a single product with unlimited supply for T periods. The set of allowable prices is denoted by $\mathcal{P}$. For example, $\mathcal{P}$ can either be an interval $[\underline{p}, \overline{p}]$ or a finite set $\{p_1, \ldots, p_k\}$, although no restriction on $\mathcal{P}$ is assumed here. In the t^{th} period $(t = 1, \ldots, T)$, the seller offers a unit price $P_t \in \mathcal{P}$, and observes a random customer demand X_t, i.e. the number of units purchased by customers. Given $P_t = p$, the distribution of X_t is only determined by price p, and is independent of previous prices and demands $\{P_1, X_1, \ldots, P_{t-1}, X_{t-1}\}$. We use $D(p) \sim X_t$ to denote a random variable distributed as X_t given $P_t = p$. The corresponding *mean demand function* $d : \mathcal{P} \to \mathbb{R}_+$ is defined as $d(p) = \mathbb{E}[D(p)]$.

The distribution of $D(p)$ is unknown to the seller. However, the seller knows that the distribution belongs to a finite set of *demand models*, or demand distributions as a function of p. The demand models are indexed by $i = 1, \ldots, K$. We use $\mathbb{P}_i(\cdot)$ and $\mathbb{E}_i(\cdot)$ to denote the probability measure and expectation under demand model i. In particular, the mean demand function $d(p)$ belongs to a finite set of K demand functions, denoted by $\Phi = \{d_1(p), \ldots, d_K(p)\}$, where $d_i(p) = \mathbb{E}_i[D(p)]$. For each demand function $d_i \in$

Φ, $(i = 1, \ldots, K)$, the expected revenue per period is $r_i(p) = pd_i(p)$. We also denote the optimal revenue for demand function d_i by $r_i^* = \max_{p \in \mathcal{P}} r_i(p)$ and an optimal price by $p_i^* \in \arg\max_{p \in \mathcal{P}} r_i(p)$. The seller does not necessarily know the distribution of demand model i apart from the mean $d_i(p)$.

For all $p \in \mathcal{P}$ and $i = 1, \ldots, K$, the probability distribution of $D(p)$ is assumed to be *light-tailed* with parameters (σ, b), where $\sigma, b > 0$. That is, we have $\mathbb{E}_i[e^{\lambda(D(p)-d_i(p))}] \leq \exp(\lambda^2\sigma^2/2)$ for all $|\lambda| < 1/b$. Note that the class of light-tailed distributions includes all sub-Gaussian distributions. Some common light-tailed distributions include normal, Poisson and Gamma distributions, as well as all distributions with bounded support, such as binomial and uniform distributions.

3.2.1 Pricing Policies

We say that π is a non-anticipating pricing policy if the price P_t offered at period t is determined by the realized demand $(X_1, \ldots, X_{t-1})$ and previous prices $(P_1, \ldots, P_{t-1})$, but does not depend on future demand. Mathematically, it is expressed as

$$P_t = \pi(P_1, X_1, \cdots, P_{t-1}, X_{t-1}).$$

For a given non-anticipating policy π and a time period t, we say that $h_t = \{(P_s, X_s)\}_{s=1}^t$ is a history that is *feasible* under π, if for all $s \in \{1, \cdots, t\}$, we have

$$P_s = \pi(P_1, X_1, \ldots, P_{s-1}, X_{s-1}).$$

It is important to note that the definition of a feasible history h_t does not involve any assumption on the underlying demand function. Rather, the feasibility of a history for a given policy π depends on how π sets a new price given previous sales history.

For $i = 1, \ldots, K$, let $\mathbb{P}_i^\pi(\cdot)$ and $\mathbb{E}_i^\pi(\cdot)$ be the probability measure and expectation induced by policy π if the underlying demand model is i. In this case, the seller's expected revenue

in T periods under policy π is given by

$$R_i^\pi(T) = \mathbb{E}_i^\pi \left[\sum_{t=1}^{T} P_t X_t \right] = \mathbb{E}_i^\pi \left[\sum_{t=1}^{T} P_t \mathbb{E}_i^\pi [X_t \mid P_t] \right] = \mathbb{E}_i^\pi \left[\sum_{t=1}^{T} r_i(P_t) \right]. \tag{3.1}$$

As motivated earlier, in many revenue management applications, the seller faces a constraint on the number of price changes. In the model, we assume that the seller can make at most m changes to the price over the course of the sales event, where m is a fixed integer. A feasible policy π should therefore satisfy the following condition:

$$\mathbb{P}_i^\pi \left(\sum_{t=2}^{T} I(P_t \neq P_{t-1}) \leq m \right) = 1, \quad \forall i = 1, \ldots, K,$$

where $I(\cdot)$ is the indicator function. We refer to a policy with at most m price changes as an *m-change policy.*

The performance of a pricing policy is measured against the optimal policy in the full information case. If the true demand is d_i, then a clairvoyant with full knowledge of the demand function would offer price p_i^* and obtain expected revenue r_i^* for every period. The *regret* with respect to demand d_i is defined as the gap between the expected revenue achieved by the clairvoyant and the one achieved by policy π, namely

$$\text{Regret}_i^\pi(T) = T r_i^* - R_i^\pi(T) = \mathbb{E}_i^\pi \left[\sum_{t=1}^{T} (r_i^* - r_i(P_t)) \right]. \tag{3.2}$$

Finally, we define the minimax regret for the demand set, $\Phi = \{d_1, \ldots, d_K\}$, as

$$\text{Regret}_\Phi^\pi(T) = \max_{i=1,\ldots,K} \text{Regret}_i^\pi(T).$$

When there is no ambiguity of which policy we are referring to, we suppress the superscript "π" in the notation for clarity: $\mathbb{E}_1 := \mathbb{E}_1^\pi$, $\mathbb{P}_1 := \mathbb{P}_1^\pi$.

3.2.2 Notations

We use $\log^{(m)} T$ to represent m iterations of the logarithm, $\log(\log(\ldots \log(T)))$, where m is the number of price changes. For convenience, we let $\log(x) = 0$ for all $0 \leq x < 1$, so the function $\log^{(m)} T$ is defined for all $T \geq 1$. Similarly, we define $e^{(0)} := 1$ and $e^{(\ell)} := \exp(e^{(\ell-1)})$ for $\ell \geq 1$. As mentioned earlier, function $\log^* T$ denotes the smallest nonnegative integer m such that $\log^{(m)} T \leq 1$. For any real number x, $\lceil x \rceil$ denotes the minimum integer greater than or equal to x. For any finite set S, $|S|$ is the cardinality of S.

3.3 Lower Bounds on the Regret of any m-change Policy

In this section we prove the main results of the chapter: a lower bound on regret as a function of the number of price changes. To appreciate the lower bound, we first state a non-anticipating pricing policy mPC that changes price no more than m times and achieves a regret of $O(\log^{(m)} T)$. Then, we show that the regret of any non-anticipating policy with at most m price changes is at least $\Omega(\log^{(m)} T)$. Thus, the lower bound is tight in the sense that it matches the regret upper bound of the proposed pricing policy, up to a constant factor.

3.3.1 $O(\log^{(m)} T)$ Regret via mPC

We present the policy mPC (which stands for "m-price change") that achieves a regret of $O(\log^{(m)} T)$ with at most m price changes. An important feature of policy mPC is that it applies a *discriminative* price for every period. A price p is *discriminative* if demands $d_1(p), \cdots, d_K(p)$ are mutually distinct. For the analysis of mPC and its implementation for Groupon, the readers are welcome to consult [Cheung et al., 2015, Wang, 2016]

We make the following assumption on the set of demand functions Φ:

Assumption 3.3.1. For all $d_i \in \Phi = \{d_1, \cdots, d_K\}$, there exists a corresponding revenue-optimal price $p_i^* \in \text{argmax}_{p \in \mathcal{P}} r_i(p)$ such that p_i^* is a discriminative price for Φ, that is,

$d_1(p_i^*), \cdots, d_K(p_i^*)$ are distinct. Moreover, such price p_i^* can be efficiently computed.

Assumption 3.3.1 ensures that the seller is able to learn the underlying demand curve while maximizing its revenue for any given demand function $d_i \in \Phi$.

Algorithm 6 m-change policy mPC

1: INPUT:
   - A set of demand functions $\Phi = \{d_1, \cdots, d_K\}$.
   - A discriminative price P_0^*.
2: (Learning) Set $\tau_0 = 0$.
3: **for** $\ell = 0, \cdots, m-1$ **do**
4: **if** $\log^{(m-\ell)} T = 0$ **then**
5: Set $\tau_{l+1} = 0$ and $P_{\ell+1}^* = P_\ell^*$.
6: **else**
7: From period $\tau_\ell + 1$ to $\tau_{\ell+1} := \tau_\ell + \left\lceil M_\Phi(P_\ell^*) \log^{(m-\ell)} T \right\rceil$, set the offered price as P_ℓ^*.
8: At the end of period $\tau_{\ell+1}$, compute the sample mean $\bar{X}^\ell$ from period $\tau_\ell + 1$ to $\tau_{\ell+1}$:
$$\bar{X}^\ell := \frac{\sum_{j=\tau_\ell+1}^{\tau_{\ell+1}} X_j}{\tau_{\ell+1} - \tau_\ell}, \text{ where } X_j = \text{Number of items sold in period } j.$$
9: Choose an index $i_\ell \in \{1, \cdots, K\}$ which solves
$$\min_{i \in \{1,\cdots,K\}} \left| \bar{X}^\ell - d_i(P_\ell^*) \right|.$$
10: Set the next offered price as $P_{\ell+1}^* = p_{i_\ell}^*$, where $p_{i_\ell}^*$ is the optimal price for demand d_{i_ℓ}.
11: **end if**
12: **end for**
13: (Earning) From period $\tau_m + 1$ to period $\tau_{m+1} = T$, set the selling price as P_m.

Algorithm 6 describes the mPC policy. The policy partitions the finite time horizon $1, \cdots, T$ into $m+1$ phases. For each $0 \le \ell \le m$, a single price P_ℓ^* is offered through Phase ℓ, which starts at period $\tau_\ell + 1$ and ends at $\tau_{\ell+1}$. Phase 0 to Phase $m-1$ are called the *learning phases*, and Phase m is referred to as the *earning phase*. Except for a constant factor $M_\Phi(P_\ell^*)$, which is to be defined later, the lengths of phases are iterated-exponentially (tetrationally) increasing, which ensures an optimal balance between exploration and ex-

ploitation.

At the end of learning phase ℓ, policy mPC computes the sample mean $\bar{X}^{\ell}$ of the sales under price P_{ℓ}^{*} (in line 8 of the algorithm). Since price P_{ℓ}^{*} is discriminative, the seller gains new information about the underlying demand in this learning phase. She then updates her belief on the true demand distribution to be $d_{i_{\ell+1}}$ (in line 9), and sets the offered price $P_{\ell+1}^{*}$ to be $p_{i_{\ell+1}}^{*}$ in the next phase. In going through all the learning phases, the seller progressively refines her estimate on the optimal price, which enables her to establish the choice of optimal price in the earning phase.

The function $M_{\Phi}(P)$ in line (7) of the mPC algorithm is defined as follows.

Definition 3.3.2. Let $p \in \mathcal{P}$ be a discriminative price. We define $M_{\Phi}(p)$ as

$$M_{\Phi}(p) := \frac{16\sigma^2}{\min_{i\neq j}\left(d_i(p) - d_j(p)\right)^2} \vee \frac{8b}{\min_{i\neq j}\left|d_i(p) - d_j(p)\right|}, \tag{3.3}$$

where the minimum is taken over distinct pairs of indices $i, j \in \{1, \cdots, K\}$.

Since we assume p to be discriminative, $M_{\Phi}(p)$ is well defined. The function $M_{\Phi}(p)$ measures the distinguishability of the demand functions $d_1, \cdots, d_k$ under the discriminative price p. We explain the definition of $M_{\Phi}(p)$ further in the analysis of mPC.

Define $M_{\Phi}^{*} = \max_{i\in\{1,\cdots,K\}} M_{\Phi}(p_i^{*})$ and $r^{*} = \max_{i\in\{1,\cdots,K\}} r_i^{*}$. The following result shows that the regret of mPC is bounded by $O(\log^{(m)} T)$.

Theorem 3.3.3 ([Cheung et al., 2015, Wang, 2016]). *Suppose the demand set Φ satisfies Assumption 3.3.1. For all $T \geq 1$, the regret of mPC is bounded by*

$$\text{Regret}_{\Phi}^{mPC}(T) \leq C_{\Phi}(P_0^{*}) \max\{\log^{(m)} T, 1\} + 4(M_{\Phi}^{*} + 1)r^{*},$$

where $C_{\Phi}(P_0^{*}) = \max_{i\in\{1,\cdots,K\}}\{M_{\Phi}(P_0^{*})(r_i^{*} - r_i(P_0^{*}))\}$.

3.3.2 Lower Bound

We show next that for a family of problem instances, any m-change policy incurs a regret of $\Omega(\log^{(m)} T)$. Thus, the regret achieved by the m-change policy mPC is optimal up to a constant factor.

Consider a problem instance (Γ) that satisfies the following conditions:

1. There exists a constant $Q_\Gamma > 0$, such that $\sum_{i=1}^{K}(r_i^* - r_i(p)) \geq Q_\Gamma$ for all $p \in \mathcal{P}$.

2. The demand $D(p) \in \mathbb{N}$ for any price $p \in \mathcal{P}$.

3. Given $p \in \mathcal{P}$, there exists a subset $\mathcal{B}_p \subset \mathbb{N}$, such that for all i, $\mathbb{P}_i(D(p) = d) > 0$ if and only if $d \in \mathcal{B}_p$.

4. There exists a constant $0 < \kappa_\Gamma < 1$, such that $\mathbb{P}_i(D(p) = d)/\mathbb{P}_j(D(p) = d) \geq \kappa_\Gamma$ for all $i, j \in \{1, \ldots, K\}$, $p \in \mathcal{P}, d \in \mathcal{B}_p$.

The first condition states that there is no price $p \in \mathcal{P}$ that simultaneously maximizes the revenue of all demand functions in Φ. This ensures that the problem instance is nontrivial and a learning process is necessary for maximizing the revenue when the demand function is unknown. The second condition is that demand must be integers. The third condition states that all demand functions have the same support for a given price. The fourth condition states that the ratios of probability mass functions of different demand models are bounded. We provide an example of hypothesis demand functions set Φ satisfying these properties in Appendix B.1.

The key step in the proof of the lower bound theorem is to quantify the performance of a pricing policy under different demand functions. This is made precise by following lemma.

Lemma 3.3.4 (Change-of-Measure Lemma). *Let $H_t = (P_1, X_1, \cdots, P_t, X_t)$ be the history observed by the end of period t, and let h_t be a realization of H_t. For any non-anticipating*

pricing policy π, we have

$$\mathbb{P}_i^\pi(H_t = h_t) \geq \kappa_\Gamma^t \mathbb{P}_{i'}^\pi(H_t = h_t),$$

for all $i, i' \in \{1, \ldots, K\}$. The constant κ_Γ is defined in the condition (Γ).

The proof of Lemma 3.3.4 can be found in Appendix B.2.

The regret lower bound of any m-change policy is formally stated in the following.

Theorem 3.3.5 (Lower Bound Theorem). *For any m-change policy π on problem instance Γ, there exists a constant $\theta_m > 0$ such that for any $T > \theta_m$, we have*

$$\mathrm{Regret}_\Phi^\pi(T) \geq \frac{1}{K} C_\Gamma Q_\Gamma \log^{(m)} T,$$

where $C_\Gamma := (-8 \log \kappa_\Gamma)^{-1} \wedge 1$ and Q_Γ is given by the first condition of (Γ).

Proof Idea of Theorem 3.3.5. We consider the time period τ when the first price change occurs, and compare it with $C_\Gamma \log^{(m)} T$. If $\tau > C_\Gamma \log^{(m)} T$, the seller spends at least $C_\Gamma \log^{(m)} T$ periods on learning with price P_1, which is determined without any observation. This implies that the seller must incur a regret of at least $\Omega(\log^{(m)} T)$. Otherwise, if we have $\tau \leq C_\Gamma \log^{(m)} T$, we argue that the seller has not extracted enough information about the underlying demand function, using the Change-of-Measure Lemma. In addition, the seller can perform at most $m - 1$ price changes after $C_\Gamma \log^{(m)} T$ periods. It turns out that these two facts cause the seller to incur a regret of at least $\Omega(\log^{(m)} T)$.

Proof. Proof of Theorem 3.3.5. Without loss of generality, we restrict π to be a deterministic policy, since the regret of a randomized policy is the expectation of the regret of corresponding deterministic policies. In other words, we restrict price P_t to be a deterministic function of the history $h_{t-1} := (P_1, X_1, \ldots, P_{t-1}, X_{t-1})$.

We prove the Theorem by induction on m. We show that for every non-negative integer m, both induction claims IC$-(m)$ and IC'$-(m)$ hold. These induction claims are displayed below:

Induction Claim IC-(m) There exists $\theta_m > 0$ such that for any m-change non-anticipating policy π and any $T > \theta_m$, we have

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi}(T) \geq C_\Gamma Q_\Gamma \log^{(m)} T,$$

where $C_\Gamma := (-8 \log \kappa_\Gamma)^{-1} \wedge 1$ and Q_Γ is given by the first condition of (Γ).

To define the induction claim IC'-(m), which is a strengthened version of IC-(m), we define the notations $\pi(h_{t_0})$ and $\text{Regret}_i^{\pi(h_{t_0})}(t_0+1, T)$. Let π be a non-anticipating policy, and let h_{t_0} be an arbitrary but fixed history over the periods $[1, t_0]$ that is feasible under π. (See Section 3.1 for the definitions.) We use $\pi(h_{t_0})$ to denote the following derived policy from period $t_0 + 1$ to period T:

$$P_t = \pi(h_{t_0})(P_{t_0+1}, X_{t_0+1}, \ldots, P_{t-1}, X_{t-1}) = \pi(h_{t_0}, P_{t_0+1}, X_{t_0+1}, \ldots, P_{t-1}, X_{t-1}). \quad (3.4)$$

For a demand model i, we define the regret of policy π in the interval $[t_0+1, T]$ conditioned on the realized history h_{t_0} as following:

$$\text{Regret}_i^{\pi(h_{t_0})}(t_0+1, T) = \mathbb{E}_i^{\pi(h_{t_0})}\left[\sum_{t=t_0+1}^{T} (r_i^* - r_i(P_t))\right] = \mathbb{E}_i^{\pi}\left[\sum_{t=t_0+1}^{T} (r_i^* - r_i(P_t)) \mid h_{t_0}\right]. \quad (3.5)$$

It is important to note that, in the definition (3.5), the sales history h_{t_0} is a fixed sequence of offered price and demand pairs $\{(P_s, X_s)\}_{s=1}^{t_0}$. Moreover, the definition does not involve the probability that h_{t_0} is generated under demand model i; we only require that the sales history h_{t_0} is feasible under π.[1]

[1] Note that (3.5) is different from the expected regret over the interval $[t_0+1, T]$, which is equal to

$$\mathbb{E}_i^{\pi}\left[\sum_{t=t_0+1}^{T} (r_i^* - r_i(P_t))\right] = \mathbb{E}_i^{\pi}\left[\mathbb{E}_i^{\pi}[\sum_{t=t_0+1}^{T} (r_i^* - r_i(P_t)) \mid h_{t_0}]\right].$$

The outer expectation is taken over all feasible histories under π up to time t_0. Unlike (3.5), this quantity does involve the probability that h_{t_0} is generated under demand model i.

Induction Claim IC'-(m) There exists a constant $\theta_m > 0$ such that the following is true. For any time interval $[t_0+1, T]$ with $T - t_0 > \theta_m$, given a non-anticipating policy π and a feasible sales history h_{t_0} over $[1, t_0]$ under policy π, let $\pi(h_{t_0})$ be the derived policy over $[t_0+1, T]$ defined in (3.4). If the policy $\pi(h_{t_0})$ changes price at most m times in the interval $[t_0+1, T]$, it holds that

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi(h_{t_0})}(t_0+1, T) \geq C_\Gamma Q_\Gamma \log^{(m)}(T - t_0),$$

where $C_\Gamma := (-8 \log \kappa_\Gamma)^{-1} \wedge 1$ and Q_Γ is given by the first condition of (Γ).

While the induction claim IC$-(m)$ is already sufficient for the proof, we will need to prove the stronger statement IC'$-(m)$ to complete the induction argument. We first show that IC$-(m)$ implies IC'$-(m)$, and next prove that IC'$-(m)$ implies IC$-(m+1)$.

Note that the constants C_Γ, Q_Γ and θ_m in the two induction claims are the same. Contants C_Γ, Q_Γ are defined in the Theorem, and θ_m is defined in the analysis below. They are dependent on Φ, but are independent of the length of time horizon, the number of price changes, and the choice of pricing policy.

If the two claims are established for all $m \geq 0$, the theorem is then proved by invoking IC$-(m)$:

$$\text{Regret}_\Phi^\pi(T) = \max_{i=1,\ldots,K} \text{Regret}_i^\pi(T) \geq \frac{1}{K} \sum_{i=1}^{K} \text{Regret}^\pi(T) \geq \frac{1}{K} C_\Gamma Q_\Gamma \log^{(m)} T.$$

First we prove the basic induction hypothesis:

Proving IC$-(0)$**:** In this case, the seller must use a fixed price throughout the sales horizon, i.e., $P_t = P_1$ for all $t = 1, \ldots, T$, where the price P_1 is determined by π. By the first condition of (Γ), the regret under any 0-change policy π is at least

$$\sum_{i=1}^{K} \text{Regret}_i^\pi(T) = \sum_{i=1}^{K} \sum_{t=1}^{T} \Delta_i(P_1) = \sum_{t=1}^{T} \left(\sum_{i=1}^{K} \Delta_i(P_1) \right) \geq Q_\Gamma T \geq C_\Gamma Q_\Gamma T,$$

where $\Delta_i(P_1) = r_i^* - r_i(P_1)$. This proves the case for $m = 0$ by setting $\theta_0 = 0$.

Next, we prove the induction argument below, where m is a positive integer.

Proving that IC$-(m-1)$ implies IC'$-(m-1)$: Suppose we are given a non-anticipating policy π over the periods $[1,T]$, and an arbitrary but fixed history h_{t_0} feasible under π over the periods $[1,t_0]$. Consider the following policy π' over $[1,T-t_0]$:

$$\pi'(P'_1, X'_1, \ldots, P'_{s-1}, X'_{s-1}) = P'_s = \pi(h_{t_0})(P'_1, X'_1, \ldots, P'_{s-1}, X'_{s-1}). \tag{3.6}$$

Recall that $\pi(h_{t_0})$ is the derived policy over $[t_0+1,T]$, defined in (3.4). Essentially, π' is a policy over $[1,T-t_0]$ that mimics the policy $\pi(h_{t_0})$ over $[t_0+1,T]$. We reiterate that h_{t_0} is an arbitrary but fixed sales history that is feasible under policy π, so the right hand side of (3.6) is well defined. This means that π' is non-anticipating policy (as defined in Section 3.1) over the periods $[1,T-t_0]$.

By the definition of π' in (3.6), we observe that, for all demand models $i = 1,\ldots,K$, we have

$$\text{Regret}_i^{\pi(h_{t_0})}(t_0+1,T) = \text{Regret}_i^{\pi'}(1,T-t_0).$$

Therefore, we have

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi(h_{t_0})}(t_0+1,T) = \sum_{i=1}^{K} \text{Regret}_i^{\pi'}(1,T-t_0).$$

Furthermore, we recall the assumption that $\pi(h_{t_0})$ carries out at most $m-1$ price changes over $[t_0+1,T]$ with probability 1. This implies that (by (3.6)) π' also carries out at most at most $m-1$ price changes over $[1,T-t_0]$ with probability 1. Invoking the induction claim IC$-(m-1)$ and recalling the condition $T - t_0 > \theta_m$, we have

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi'}(T-t_0) \geq C_\Gamma Q_\Gamma \log^{(m)}(T-t_0).$$

Altogether, this establishes the induction claim IC'$-(m-1)$.

Proving that IC'$-(m-1)$ implies IC$-(m)$:

Without loss of generality, we assume $\log^{(m)}(T) > 0$, otherwise the induction claim trivially holds. For a given m-change policy π, let τ be the time period when the first price change occurs, i.e., $\tau = \min_{1\le t\le T}\{t : P_t \neq P_{t-1}\}$. Let $T_m = \lceil C_\Gamma \log^{(m)}(T)\rceil$, where $C_\Gamma = (-8\log\kappa_\Gamma)^{-1} \wedge 1$. Note that the constant $\kappa_\Gamma \in (0,1)$, so $C_\Gamma > 0$. We use $\mathcal{L}$ to denote the event $\mathcal{L} = \{\tau > T_m\}$.

We decompose the regret $\sum_i \mathrm{Regret}_i^\pi(T)$ and bound it from below as follows:

$$\begin{aligned}
&\sum_{i=1}^K \mathrm{Regret}_i^\pi(T) = \sum_{i=1}^K \mathbb{E}_i^\pi\left[\sum_{t=1}^T \Delta_i(P_t)\right] \\
&= \sum_{i=1}^K \mathbb{E}_i^\pi\left[\sum_{t=1}^T \Delta_i(P_t)\middle|\mathcal{L}\right]\mathbb{P}_i^\pi(\mathcal{L}) + \sum_{i=1}^K \mathbb{E}_i^\pi\left[\sum_{t=1}^T \Delta_i(P_t)\middle|\mathcal{L}^C\right]\mathbb{P}_i^\pi(\mathcal{L}^C) \\
&\ge \underbrace{\sum_{i=1}^K \mathbb{E}_i^\pi\left[\sum_{t=1}^{T_m} \Delta_i(P_t)\middle|\mathcal{L}\right]\mathbb{P}_i^\pi(\mathcal{L})}_{(\dagger)} + \underbrace{\sum_{i=1}^K \mathbb{E}_i^\pi\left[\sum_{T_m+1}^{T} \Delta_i(P_t)\middle|\mathcal{L}^C\right]\mathbb{P}_i^\pi(\mathcal{L}^C)}_{(\ddagger)}.
\end{aligned}$$

Consider the regret term (†). Conditioned on the event $\mathcal{L}$, the first price change only occurs after the T_m^{th} period. Thus, we have $P_t = P_1$ for all $1 \le t \le T_m$, and the term (†) can bounded by

$$(\dagger) \ge \sum_{i=1}^K C_\Gamma \log^{(m)}(T)\Delta_i(P_1)\mathbb{P}_i^\pi(\mathcal{L}). \tag{3.7}$$

Next, we analyze the regret term (‡). Recall that $H_t = (P_1, X_1, \cdots, P_t, X_t)$ is the history observed by the seller at the end of period t, and let h_t be a specific realization of H_t. We define the set

$$\mathcal{H}_m^\Delta = \{h_{T_m} = (P_1, X_1, \cdots, P_{T_m}, X_{T_m}) : P_s \neq P_{s+1} \text{ for some } 1 \le s \le T_m - 1\}$$

as the set of history for which a price change occurs before the end of period T_m. By the

definition, we have $\mathbb{P}_i(\mathcal{L}^C) = \mathbb{P}_i(H_{T_m} \in \mathcal{H}_m^\Delta)$.

Thus, term (‡) is bounded by

$$
\begin{aligned}
(\ddagger) &= \sum_{i=1}^{K} \mathbb{E}_i^\pi \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i^\pi(\mathcal{L}^C) \\
&= \sum_{i=1}^{K} \mathbb{E}_i^\pi \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i^\pi(H_{T_m} \in \mathcal{H}_m^\Delta) \\
&= \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \mathbb{E}_i^\pi \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| H_{T_m} = h_{T_m} \right] \mathbb{P}_i^\pi \left(H_{T_m} = h_{T_m} \right) && (3.8) \\
&= \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \mathrm{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) \mathbb{P}_i^\pi \left(H_{T_m} = h_{T_m} \right) && (3.9) \\
&\geq \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \mathrm{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota^\pi \left(H_{T_m} = h_{T_m} \right) \right) && (3.10) \\
&= \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota^\pi \left(H_{T_m} = h_{T_m} \right) \right) \sum_{i=1}^{K} \mathrm{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) && (3.11)
\end{aligned}
$$

In step (3.8), we decompose the previous expression into a summation of conditional expectations over realized history $h_{T_m} \in \mathcal{H}_m^\Delta$. Note that the set $\mathcal{H}_m^\Delta$ is countable, since the demand X_t is integer and the price P_t is completely determined by the previous history. In step (3.9), the pricing policy $\pi(h_{T_m})$ denotes the policy adopted by the seller from period $T_m + 1$ to period T, after she observes the history h_{T_m} from period 1 to period T_m. Note that the policy $\pi(h_{T_m})$ is determined after the history h_{T_m} is realized. Thus, the expression in (3.9) is a weighted sum of regret under strategies $\{\pi(h_{T_m}), h_{T_m} \in \mathcal{H}_m^\Delta\}$, where each regret term is weighted by the probability of the corresponding history. Step (3.10) applies the Change-of-Measure Lemma.

We lower bound the expression in (3.11) in the following. First, let θ'_{m-1} the threshold

such that $\left\lceil C_\Gamma \log^{(m)} T \right\rceil \leq 2C_\Gamma \log^{(m)} T$ for all $T > \theta'_{m-1}$. We have

$$\begin{aligned}\kappa_\Gamma^{T_m} &\geq \kappa_\Gamma^{2C_\Gamma \log^{(m)} T} \geq \exp(-\frac{1}{4}\log^{(m)} T) \\ &\geq \exp\left(-\left(1 - \frac{\log(2\log^{(m)} T)}{\log^{(m)} T}\right)\log^{(m)} T\right) = \frac{2\log^{(m)} T}{\log^{(m-1)} T}. \end{aligned} \tag{3.12}$$

The first inequality uses the definition of T_m, the second inequality applies the definition of C_Γ, and the third inequality uses the fact that $\log(2x)/x < 3/4$ for all $x > 0$.

Next, we bound the regret term $\sum_{i=1}^K \text{Regret}_i^{\pi(h_{T_m})}(T - T_m)$ by invoking the induction claim IC'$-(m-1)$. For any sales history $h_{T_m} \in \mathcal{H}_m^\Delta$, the policy $\pi(h_{T_m})$ changes price no more than $m-1$ times during the time interval $[T_m + 1, T]$, because at least one price change is exhausted before period T_m. Applying the claim IC'$-(m-1)$, we have

$$\sum_{i=1}^K \text{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) \geq C_\Gamma Q_\Gamma \log^{(m-1)}(T - T_m)$$

for T such that $T - T_m \geq \theta_{m-1}$. Furthermore, let $\theta''_{m-1} \geq \theta_{m-1}$ be a threshold such that $\log^{(m-1)}\left(T - \left\lceil C_\Gamma \log^{(m)} T \right\rceil\right) \geq \frac{1}{2}\log^{(m-1)} T$ for all $T \geq \theta''_{m-1}$. Then, for $T \geq \theta''_{m-1}$, we have:

$$\sum_{i=1}^K \text{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) \geq \frac{1}{2} C_\Gamma Q_\Gamma \log^{(m-1)} T. \tag{3.13}$$

Combining (3.12) and (3.13), we have the following:

$$\begin{aligned}(\ddagger) &\geq \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota^\pi\left(H_{T_m} = h_{T_m}\right)\right) \sum_{i=1}^K \text{Regret}_i^{\pi(h_{T_m})}(T_m + 1, T) \\ &\geq \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \left(\frac{2\log^{(m)} T}{\log^{(m-1)} T} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota^\pi\left(H_{T_m} = h_{T_m}\right)\right) \frac{1}{2} C_\Gamma Q_\Gamma \log^{(m-1)} T \\ &= C_\Gamma Q_\Gamma \log^{(m)} T \sum_{h_{T_m} \in \mathcal{H}_m^\Delta} \left(\max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota^\pi\left(H_{T_m} = h_{T_m}\right)\right)\end{aligned}$$

$$
\begin{aligned}
&\geq C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\ldots,K\}} \sum_{h_{T_m}\in\mathcal{H}_m^\Delta} \mathbb{P}_\iota^\pi\left(H_{T_m}=h_{T_m}\right)\\
&= C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\ldots,K\}} \mathbb{P}_\iota^\pi(\mathcal{L}^C). \qquad (3.14)
\end{aligned}
$$

Altogether, by both (3.7) and (3.14), we have

$$
\begin{aligned}
&\sum_{i=1}^K \mathrm{Regret}_i^\pi(T) \geq (\dagger)+(\ddagger)\\
\geq &\sum_{i=1}^K C_\Gamma \log^{(m)}(T)\Delta_i(P_1)\mathbb{P}_i^\pi(\mathcal{L}) + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\ldots,K\}} \mathbb{P}_\iota^\pi(\mathcal{L}^C)\\
\geq &C_\Gamma \log^{(m)}(T)\left(1-\max_{\iota\in\{1,\ldots,K\}}\mathbb{P}_\iota^\pi(\mathcal{L}^C)\right)\sum_{i=1}^K \Delta_i(P_1) + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\ldots,K\}} \mathbb{P}_\iota^\pi(\mathcal{L}^C)\\
\geq &C_\Gamma \log^{(m)}(T)\left(1-\max_{\iota\in\{1,\ldots,K\}}\mathbb{P}_\iota^\pi(\mathcal{L}^C)\right) Q_\Gamma + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\ldots,K\}} \mathbb{P}_\iota^\pi(\mathcal{L}^C)\\
= &C_\Gamma Q_\Gamma \log^{(m)} T,
\end{aligned}
$$

for all $T \geq \max\{\theta'_{m-1}, \theta''_{m-1}\}$. By setting $\theta_m := \max\{\theta'_{m-1}, \theta''_{m-1}\}$, the induction step is established. This completes the proof. □ □

Taken together, the proofs of the upper and lower bounds provide important insights into the structure of any optimal m-change policy. With high probability, an optimal m-change policy has $m-1$ learning phases of lengths $\Theta(\log^{(m)} T), \cdots, \Theta(\log T)$. They are followed by the last phase, which is the earning phase on the last $T - \Theta(\log T)$ time periods, see Fig 3-1.

The lengths of the learning phases are set in a way to ensure an optimal balance between learning and earning. If any of the learning phases is shortened significantly, such lack of learning will incur a large regret in the subsequent phases. In general, for each $\ell \in \{1, \cdots, m\}$, if the ℓ^{th} learning phase is of length $o(\log^{(m-\ell+1)} T)$, then a regret of $\Omega(\log^{(m-\ell)} T)$ is incurred in the subsequent phases. This quantifies the value of learning in

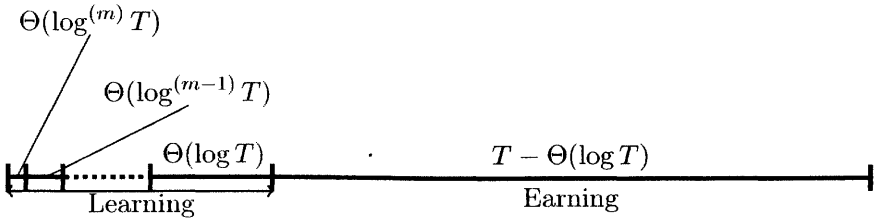


Figure 3-1: The structure of an optimal m-change policy.

any m-change policy.

3.4 Conclusion.

In conclusion, we have derived tight regret lower bound on a single product dynamic pricing problem when the number of price change is limited. The regret lower bound holds for a wide variety of demand function sets, showing that a regret of $\Theta(\log^{(m)}(T))$ is often necessary when we are only allowed to change prices at most m times. Moreover, the lower bound matches the regret upper bound (within a constant factor) achieved by the proposed policy in [Cheung et al., 2015], which shows that the lower bound is the best possible, up to scaling.

Through the analysis, we gain important insights into the structure of any optimal pricing policy. One crucial observation is that, for any asymptotically optimal m-change policy, the length of a learning phase is exponential in the length of the previous learning phase. This explicit quantifies the optimal trade off between exploration and exploitation. We hope that such insight could be useful in designing dynamic pricing policy with limited experimentation in other settings.

Chapter 4

Data-driven Capacitated Inventory Control Models

4.1 Introduction

In this chapter, 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, 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 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 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. [Shapiro et al., 2009]. 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 [Shapiro and Nemirovski, 2005].

In this chapter, we provide a non-asymptotic analysis on the performance of SAA method in the data-driven capacitated inventory control problems. 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. 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-drive 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. [Levi et al., 2014], our sample lower bound is tight for the special case on the *newsvendor problem*, which is the special case when there is only one period, but there is no supply constraint.

4.1.1 Literature Review

Data-driven multi-stage stochastic optimization problems are actively studied in the realms of Computer Science and Operations Research. Kleywegt et al. [Kleywegt et al., 2002] 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 [Kleywegt et al., 2002] depends on the variances of the underlying random variables. [Shapiro and Nemirovski, 2005]

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 [Shapiro et al., 2009].

Data-driven two-stage stochastic combinatorial optimization problems are well studied in the literature. Gupta et al. [Gupta et al., 2004] consider such two-stage problems on a variety of **NP**−hard combinatorial optimization problems, and approximation algorithms with constant ratios are provided. Shmoys and Swamy [Shmoys and Swamy, 2006] 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, 2006] 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. [Charikar et al., 2005] consider a more general version of two-stage stochastic optimization than [Shmoys and Swamy, 2006], and they show that the SAA method achieves near optimality with high probability. Finally, under the full knowledge on the underlying randomness, various two-stage problems are also studied, for example see the results in [Immorlica et al., 2004, Ravi and Sinha, 2006, Gupta et al., 2007, Begen and Queyranne, 2011].

On the other hand, data-driven multi-stage stochastic optimization problems are significantly more difficult when the number of stages is greater than two. Swamy and Shmoys [Swamy and Shmoys, 2012] generalizes [Shmoys and Swamy, 2006] by studying the class of multi-stage stochastic linear optimization problems with covering constraints. The research [Swamy and Shmoys, 2012] 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 [Shmoys and Swamy, 2006]. However, the number of samples needed grows exponentially in the number of stages T. Gupta et al. [Gupta et al., 2005] extends the framework in [Gupta et al., 2004] to the case of multiple stages. Both the number of samples needed and the approximation ratios deteriorates exponentially with the number of stages. Shapiro [Shapiro, 2006] 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 [Shapiro and Nemirovski, 2005], the analysis of the SAA method in multi-stage stochastic optimization problems seems very hard in geneal.

The works of Levi et al. [Levi et al., 2007], [Halman, 2015] are the most relevant to ours. Levi et al. [Levi et al., 2007] 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. [Halman, 2015] 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 [Levi et al., 2007], [Halman, 2015] is provided in the next subsection.

Dynamic programs in a different data-driven model has been considered in Halman et al. [Halman et al., 2009], Halman et al. [Halman et al., 2014]. 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 can-

not be exactly ascertained (cf. Theorem 4.3.4). [Halman et al., 2009] consider single commodity inventory control problems with more general holding and backlog costs, and [Halman et al., 2014] consider an even more general class of dynamic programming problems. [Halman et al., 2009, Halman et al., 2014] 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 [Aviv and Federgruen, 1997, Kapuscinski and Tayur, 1998] 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 [Aviv and Federgruen, 1997] proposes a value iteration algorithm to compute the base stocks, while Kapuscinski and Tayur [Kapuscinski and Tayur, 1998] proposes a simulation base method for the computation. Subsequently, Levi et al. [Levi et al., 2008] 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.

4.1.2 Our Approach

We analyze the performance of the SAA method by considering the *SAA DP*, the dynamic program for the SAA problem, and comparing it with the *Original DP*, the dynamic program for the original problem, i.e., the DP associated with full knowledge of the underline demand distribution. A natural approach is to try and compare the values of the cost-to-go functions in the SAA DP with those in the Original DP. However, since we do not assume that the demand distributions are bounded or sub-Gaussian, it does not seem that the empirical function values are concentrated around their means under any finite pool of samples.

Thus, we consider the *right derivatives* of the cost-to-go functions involved in the SAA

DP, and compare them with their corresponding right derivatives for the original DP. For convenient sake, we call the former *SAA derivatives*, and the latter *Original derivatives*. The rationale behind this approach is that, unlike the function values, the SAA derivatives are bounded random variables, and their bounds can be expressed solely in terms of the unit holding and backlog costs. More importantly, these bounds are independent of the latent demand distributions. This important property allows us to utilize concentration inequalities, and establish the number of samples that ensures that the SAA derivatives approximate their original counterparts within any level of additive error with high confidence probability. The number of samples drawn has a pseudo-polynomial dependence on the cost parameters, but it is independent of the latent demand distributions. A crucial ingredient here is the Massart inequality (cf. Theorem 4.5.2), which is a stronger version of Hoeffding inequality. The former ensures that the approximation guarantee holds for the derivatives at all points in the domain, rather than just for subgradients at certain arguments (which is the case in [Levi et al., 2007]). Using this uniform approximation guarantee, we show that in a certain sense the SAA-optimal policy does not deviate too much from the truly optimal policy. Finally, by considering the sample path of the inventory levels across the periods, we show that the total relative error of the SAA method is bounded when there is a small error made in each period, and thus establishing the near optimality of the SAA method.

In fact, a first order approach is also proposed by [Levi et al., 2007]. However, our approach is significantly different from [Levi et al., 2007] in the following aspects. The research by [Levi et al., 2007] 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 [Levi et al., 2007]) in their analysis. Such perturbation crucially uses the fact that the function $\tilde{V}_j(x_j)$ (in [Levi et al., 2007]) 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. §4). Thus, we take an alternative approach by directly analyzing the SAA method.

In fact, the analysis of the SAA method in their inventory models is raised as an open question in [Levi et al., 2006]. In this chapter, 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.

To complement our results above, we establish a lower bound on the number of samples needed to solve the data-driven newsvendor problem to near optimality in Theorem 4.3.4. The basic idea is to reduce the data-driven optimization problem to a statistical classification problem. More precisely, we construct two demand distributions D_1 and D_2, with the properties that they have disjoint sets of $(1+\epsilon)$-optimal bases stocks, but their statistical distance is small. By the disjointness property, if there exists an algorithm that returns a $(1+\epsilon)$-optimal base stocks under both D_1 or D_2 using m samples, then there also exists an algorithm that distinguishes between D_1 and D_2 using m samples. Nevertheless, since the statistical distance between D_1, D_2 is small, we can provide a lower bound on the number of samples needed for the classification. This provides a lower bound on the number of samples needed for solving the data-driven newsvendor problem.

In Halman [Halman, 2015], 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 [Halman, 2015] is incomparable to ours, since we have a more general data-drvien setting (since we do not assume such a γ to be existent and known to the decision maker) than [Halman, 2015], but the latter considers a more general class of DPs than ours. Moreover, we analyze the SAA method by a first order analysis, while the FPTASs proposed in [Halman, 2015] are not SAA methods, and are analyzed through zero-order analyses.

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 [Levi et al., 2007], we assume zero lead time and zero ordering cost for the sake of clarity.

The rest of the chapter is organized as follows. In §4.2, we provide the data driven model for the capacitated inventory system. In §4.3, we state the main results in the chapter. In §4.4. we provide an review on the optimality of modified base stock policies, in order to facilitate the analysis in the subsequent sections. In §4.5, we prove Theorem 4.3.1 by a first order analysis on the sample average approximation (SAA) problem. In §4.6, we provide a polynomial time approximation scheme *Sample* to the data driven capacitated inventory control, by introducing a sparsfication procedure to the SAA method. In §4.7, we provide insights into the proofs for the hardness results. In §4.8, we compare the performance of Algorithm *Sample* in simulations with the performance guarantee implied by our analysis. Finally, in §4.9, we conclude the chapter.

4.2 The Data-driven Capacitated Inventory Control Model

The Inventory Control Model. We consider a 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

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

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 $h_t, b_t > 0$.

Modified Base Stock Policies. Throughout the chapter, we consider a certain class of policies called *modified base stock policies*:

Definition 4.2.1. 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 above) 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 [Kapuscinski and Tayur, 1998], [Tayur, 1993], there exists an optimal modified base stock policy $(R_1^*, \cdots, R_T^*)$ under which the expected cost is minimized. This is derived by using dynamic programming, which is reviewed in §3.

We say that a set of base stocks $(R_1, \cdots, R_T)$ is $(1+\epsilon)$-*optimal* if the expected cost under the modified based 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.

The Data-driven Model. 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 is analogous to [Levi et al., 2007, Shmoys and Swamy, 2006, Swamy and Shmoys, 2012]. As observed in the literature review, the data driven model considered in this chapter is strictly weaker than the model considered by [Halman et al., 2014] where the decision maker has access to the cumulative probability distributions (CDFs) of the demands by calling an oracle.

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 de-

mand distributions. 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 capacitated inventory control problem, 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:

$$\mathbb{P}[\hat{D}_s = d] = \frac{\sum_{i=1}^{N_s} \mathbf{1}[d = d_s^i]}{N_s}. \tag{4.1}$$

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

4.3 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 4.3.1. *Consider the Sample Average Approximation problem $SAA(T; N_1, \cdots, N_T)$, where N_t is defined as follows*

$$N_t = \max\left\{(h_t + b_t)^2, \left(\sum_{s=t+1}^{T} h_s + b_s\right)^2\right\} \frac{144T^4}{\epsilon^2 \min_{t\in\{1,\cdots,T\}} \{\min\{h_t, b_t\}\}^2} \log\frac{2T}{\delta}. \tag{4.2}$$

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 (4.6). With probability at least $1 - \delta$, the modified base stocks policy $(\hat{R}_1, \cdots, \hat{R}_T)$ is both optimal to the problem $SAA(T; N_1, \cdots, N_T)$, and $(1 + \epsilon)$-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 Theorem is proven in §4.5 by a careful first order analysis on the dynamic programs for the empirical problem $SAA(T; N_1, \cdots, N_T)$ and the original problem. Theorem 4.3.1 does not imply a polynomial time approximation scheme for the data-driven capacitated inventory control problem. Rather, it only concerns the number of samples needed to guarantee the existence of the near optimal policy $(\hat{R}_1, \cdots, \hat{R}_T)$ with probability $1 - \delta$. In fact, by a modification of the hardness result in [Halman et al., 2009], solving the empirical problem exactly is computationally hard, in the sense that even $SAA(T; 2, \cdots, 2)$ can be intractable:

Lemma 4.3.2. *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 4.3.2 is proven by reducing the problem to the Knapsack Counting problem. The Lemma implies that as long as the the support of each of the empirical distributions is larger than one, the SAA problem is intractable in general. This hardness result is in contrast with [Charikar et al., 2005, Swamy and Shmoys, 2012, Begen et al., 2012], 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 §4.6.

Theorem 4.3.3. *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 [Levi et al., 2007]. 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 4.3.4. *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.*

Theorem 4.3.4 shows that the pseudo-polynomial dependence on the cost parameters in Theorem 4.3.3 is an issue of informational complexity rather than computational complexity. Finally, we remark the lower bound in Theorem 4.3.4 matches the upper bound

provided in [Levi et al., 2014] for the data-driven newsvendor problem. Thus, combined with [Levi et al., 2014], 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 $\Theta\left(\frac{h+b}{\min\{h,b\}\epsilon^2}\right)$.

4.4 A Review on the Optimality of Modified Base Stock Policy

To describe the data-driven algorithm, we review the optimality of modified base stock policies, as demonstrated in [Aviv and Federgruen, 1997, Kapuscinski and Tayur, 1998]. When the CDFs of $D_1, \cdots, D_T$ are explicitly given, the 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, we introduce the function U_t:

$$U_t(y_t) = C_t(y_t) + \mathbb{E}\left[V_{t+1}(y_t - D_t)\right]. \tag{4.3}$$

Thus, we have

$$V_t(x_t) = \min_{x_t \le y_t \le x_t + B_t} U_t(y_t). \tag{4.4}$$

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 $t+1, \cdots, T$.

[Aviv and Federgruen, 1997, Kapuscinski and Tayur, 1998] show that, by a backward induction from $t = T$ to $t = 1$,

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

2. The modified base stock policy $(R_1^*, \cdots, R_T^*)$, where $R_t^* \in \text{argmin}_{y_t \in \mathbb{R}} U_t(y_t)$, 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 (4.3). 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 (4.4) for the following ranges of x_t, we have

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

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 (4.4) yields

$$V_t(x_t) = \begin{cases} U_t(x_t + B_t) & \text{if } x_t \in (-\infty, R_t^* - B_t) \\ U_t(R_t^*) & \text{if } x_t \in [R_t^* - B_t, R_t^*) \\ U_t(x_t) & \text{if } x_t \in [R_t^*, \infty) \end{cases} \tag{4.5}$$

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.

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, 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:

$$\hat{U}_t(y_t) = \hat{C}_t(y_t) + \mathbb{E}\hat{V}_{t+1}(y_t - \hat{D}_t), \tag{4.6}$$

$$\hat{V}_t(x_t) = \min_{x_t \le y_t \le x_t + B_t} \hat{U}_t(y_t) = \begin{cases} \hat{U}_t(x_t + B_t) & \text{if } x_t \in (-\infty, \hat{R}_t - B_t) \\ \hat{U}_t(\hat{R}_t) & \text{if } x_t \in [\hat{R}_t - B_t, \hat{R}_t) \\ \hat{U}_t(x_t) & \text{if } x_t \in [\hat{R}_t, \infty) \end{cases}, \tag{4.7}$$

where $\hat{D}_t$ is the empirical distribution constructed using N_t samples from D_t (as defined in (4.1)), 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.

4.5 A First Order Analysis on the Sample Average Approximation Problem

In this Section, we prove Theorem 4.3.1 by performing a first order analysis on the dynamic program for the empirical problem $SAA(T; N_1, \cdots, N_T)$, and comparing it with the dynamic program for the (latent) original problem. We present the first order analysis in three subsections. First, we provide the expressions for the right derivatives of the relevant cost-to-go functions. Second, we show that by our choice of the number of samples $N_1, \cdots, N_T$, the right derivatives of the cost-to-go functions for the empirical problem are good approximations to their original counterparts in a very strong sense, with high probability. This is proven by a careful backward induction on the period t. Third, we show that the approximation guarantees for the right derivatives imply that, for each period t, the expected cost under the optimal policy $(R_1^*, \cdots, R_T^*)$, but with R_t^* replaced by $\hat{R}_t$, is still close to the optimal cost. By bounding the error of using a suboptimal base stock in a certain period to the later periods, we show that the optimal base stock $\hat{R}_1, \cdots, \hat{R}_T$ is a near-optimal policy to the original problem.

4.5.1 The Expressions for the Right Derivatives

To embark on our first order analysis, we first 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 (4.3), (4.4), (4.6) and (4.7). 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:

$$V_t^{\mathrm{r}}(x_t) = \begin{cases} U_t^{\mathrm{r}}(x_t + B_t) & \text{if } x_t \in (-\infty, R_t^* - B_t) \\ 0 & \text{if } x_t \in [R_t^* - B_t, R_t^*) \\ U_t^{\mathrm{r}}(x_t) & \text{if } x_t \in [R_t^*, \infty) \end{cases}, \tag{4.8}$$

$$\hat{V}_t^{\mathrm{r}}(x_t) = \begin{cases} \hat{U}_t^{\mathrm{r}}(x_t + B_t) & \text{if } x_t \in (-\infty, \hat{R}_t - B_t) \\ 0 & \text{if } x_t \in [\hat{R}_t - B_t, \hat{R}_t) \\ \hat{U}_t^{\mathrm{r}}(x_t) & \text{if } x_t \in [\hat{R}_t, \infty) \end{cases}. \tag{4.9}$$

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

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

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

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

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 $\hat{D}_t, \cdots, \hat{D}_T$, which 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 $\mathbb{E}\hat{U}_t^{\mathrm{r}}(y_t) \neq U_t^{\mathrm{r}}(y_t)$ (except when $t = T$), since for $t < T$, $\hat{U}_t, U_t$ are cost-to-go functions for the inventory problem with different underlying distributions; the former being the empirical distribution and the latter being the original distribution. Similarly, in general $\mathbb{E}\hat{V}_t^{\mathrm{r}}(y_t) \neq V_t^{\mathrm{r}}(y_t)$, 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. Note that the bounds are independent of the underlying demand distributions $D_t, \cdots, D_T$, which allows us to establish sample upper bounds that do not depend on the underlying distributions.

4.5.2 Approximating the Right Derivatives $U_T^{\mathrm{r}}, \cdots, U_1^{\mathrm{r}}$ by $\hat{U}_T^{\mathrm{r}}, \cdots, \hat{U}_1^{\mathrm{r}}$

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 $x \in \mathbb{R}$. The uniform approximation is proven by a backward induction on t.

To establish the backward induction, we 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 4.5.1. *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:

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

A crucial ingredient to the proof is a Theorem of [Massart, 1990], which provides us with a concentration bound on $U_t^r(y_t)$ that holds *uniformly* across all $y_t \in \mathbb{R}$.

Theorem 4.5.2 ([Massart, 1990]). *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*

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

□

The proof of Theorem 4.5.1 is provided in Appendix C.1. Next, we demonstrate that if the empirical right derivative $\hat{U}_t^r$ uniformly approximates the original right derivative U_t^r, then $\hat{V}_t^r$ also uniformly approximates V_t^r with the same additive error.

Theorem 4.5.3. *Suppose for all y_t, the inequality $\left|\hat{U}_t^r(y_t) - U_t^r(y_t)\right| \leq \eta_t$ holds. Then we have*

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

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

Theorem 4.5.3 is proven by considering different cases on R_t^* and $\hat{R}_t$. 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 C.2.

Combining Theorem 4.5.1 and 4.5.3, we conclude that the right derivatives $\hat{U}_t^r$ of the empirical cost functions uniformly approximate the right derivatives U_t^r for all t with high probability in the following Corollary:

Corollary 4.5.4. *Consider the empirical problem $SAA(1; N_1, \cdots, N_T)$, where*

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

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

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

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 proof of Corollary is provided in Appendix C.3.

4.5.3 From First Order to Zero Order Approximation

Next, we show that if the empirical right derivatives $\hat{U}_T^r, \cdots, \hat{U}_1^r$ uniformly approximate the original right derivatives , then the optimal base stock $\hat{R}_t$ for the empirical problem is 'close' to the optimal base stock R_t^* for the original problem, in the sense that $U_t(\hat{R}_t) \leq (1 + \epsilon_t)U_t(R_t^*)$ for some constant ϵ_t. This suggests that the empirical modified base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ is a good candidate for a near-optimal policy in the original problem. We demonstrate this in Lemma 4.5.7 by showing that, when the sub-optimal base stock $\hat{R}_t$ in period t instead of the optimal base stock R_t^* in each period t, the total accumulated error can be bounded from above.

Now, we transform the first order approximation guarantee to our desired zero-order approximation guarantee by the following claim:

Claim 4.5.5. *Suppose that for all y_t, we have $\left|\hat{U}_t^T(y_t) - U_t^T(y_t)\right| \leq \eta$, where $\eta > 0$ is a given constant. Then the following inequality holds:*

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

The Claim is proven in Appendix C.4. The Claim is a straightforward consequence of a Lemma by [Levi et al., 2007]:

Lemma 4.5.6 (Lemma 3.3 in [Levi et al., 2007]). *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*

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

From the previous Section, we know that $\left|U_t^T(y) - \hat{U}_t^T(y)\right| \leq \sum_{s=t}^T \alpha_s$ with high probability. By applying Claim 4.5.5 for U_t with $\eta = \sum_{s=t}^T \alpha_s$, we have the inequality

$$U_t(\hat{R}_t) \leq \left(1 + \frac{3\sum_{s=t}^T \alpha_s}{\min\{b_t, h_t\}}\right) U_t(R_t^*). \tag{4.10}$$

By the definition of U_t, (4.10) is equivalent to the following inequality, which compares the expected costs under the modified base stock policy $(\hat{R}_t, R_{t+1}^*, R_{t+2}^*, \cdots, R_T^*)$ with the optimal policy $(R_t^*, R_{t+1}^*, \cdots, R_T^*)$ in the subproblem from period t to period T:

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

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 $(R_t, \cdots, R_T)$. It is important to 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 (in the subproblem) 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 demonstrate that for any modified base stock policy $(R_1, \cdots, R_T)$ such that $U_t(R_t) \leq (1+\epsilon_t)U_t(R_t^*)$ for all t, the policy $(R_1, \cdots, R_T)$ is near optimal:

Lemma 4.5.7. *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 C.5.

Finally, Theorem 4.3.1 is proven by piecing together Corollary 4.5.4, inequality (4.10) and Lemma 4.5.7. The detailed argument in provided in Appendix C.6.

4.6 A Polynomial Time Approximation Scheme via Sparsification

In this Section, we prove Theorem 4.3.3 by providing a polynomial time approximation scheme to the data-driven capacitated inventory control problem. This is achieved by introducing a sparsification procedure to the SAA Method. Note that the SAA method does not immediately lead to a polynomial time algorithm. Indeed, by Lemma 4.3.2, if

there exists an efficient algorithm that computes an optimal modified base stock policy $(\hat{R}_1, \cdots, \hat{R}_T)$ for $SAA(T; N_1, \cdots, N_T)$, then $\mathbf{P} = \#\mathbf{P}$. Nevertheless, we provide the algorithm $Sample(\eta, N_1, \cdots, N_T)$, which achieves near optimality while has an efficient run time, unlike the SAA method. The parameter $\eta > 0$ is the accuracy parameter for the sparsification procedure, which is crucial for ensuring $Sample(\eta, N_1, \cdots, N_T)$ in running 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 uniformly approximate the original right derivatives U_t^{r} and V_t^{r}. This ensures the near optimality of $(\tilde{R}_1, \cdots, \tilde{R}_T)$ in the original problem. Note that the sparsfication 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 $\tilde{U}_t^{\mathrm{r}} \neq \hat{U}_t^{\mathrm{r}}$.

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 $\tilde{U}_t^{\mathrm{r}}$. In the proof of Lemma 4.6.1, we show that $\tilde{R}_t$ is one of the break points of $\tilde{U}_t^{\mathrm{r}}$, and that $\tilde{R}_t$ can be found in polynomial time.

In line 7, the algorithm computes the function $\hat{\tilde{V}}_t^{\mathrm{r}}$. While this function is a candidate for an approximation to the original right derivative V_t^{r}, we cannot use it 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 4.6.1. Thus, in line 8, we introduce a sparsfication procedure to prevent the number of breakpoints in the step functions from increasing too rapidly. The sparsfication 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

Algorithm 7 Algorithm *Sample*$(\eta, N_1, \cdots, N_T)$

For each $t \in \{1, \cdots, T\}$, draw N_t independent samples $d_t^1, \cdots, d_t^{N_t}$ from D_t.

Construct the empirical distribution $\hat{D}_t$:

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

Define $\tilde{V}_{T+1}^{\mathrm{r}}(x) = 0$ for all x.

for $t = T, \cdots, 1$ **do**

Construct the right derivative function $\tilde{U}_t^{\mathrm{r}}(y_t) = \hat{C}_t^{\mathrm{r}}(y_t) + \mathbb{E}\tilde{V}_{t+1}^{\mathrm{r}}(y_t - \hat{D}_t)$.

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 $\tilde{U}_t^{\mathrm{r}}(\tilde{R}_t) \geq 0$.

Construct the following right derivative function $\hat{\tilde{V}}_t^{\mathrm{r}} : \mathbb{R} \to [-\sum_{s=t}^{T} b_s, \sum_{s=t}^{T} h_s]$

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

(Sparsification) Now, for each x_t, define

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

end for

Return the base stocks $(\tilde{R}_1, \cdots, \tilde{R}_T)$.

closest grid point to $\hat{\tilde{V}}_t^{\mathrm{r}}(x_t)$ from below. That is,

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

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 4.6.1 and Theorem 4.6.2.

We remark that if the Sparsfication 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 sparsfication 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 4.6.1 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 4.6.1. *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 C.7. The analysis in the proof shows that, by the sparsfication 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 redestablish in Theorem 4.6.2 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 $U_t^{\mathrm{r}}, V_t^{\mathrm{r}}$ by 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 4.6.2. *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* $(\tilde{R}_1, \cdots, \tilde{R}_T)$, *where*

$$\eta = \frac{\epsilon \min_{t\in\{1,\cdots,T\}} \{\min\{h_t, b_t\}\}}{6T^2} \tag{4.11}$$

and the choices of $\{N_t\}_{t=1}^T$ *are the same as in Theorem 4.3.1.*

The proof of Theorem 4.6.2 is given in Appendix C.8. Altogether, we have established Theorem 4.3.3.

4.7 Insight into the Hardness Results

In this Section we provide insights on the proofs (see Appendix C.9, C.10) for the two hardness results, Lemma 4.3.2 and Theorem 4.3.4. Lemma 4.3.2 is proven by reducing the problem to the **#KNAPSACK** problem, which is **#P**-complete. The **#KNAPSACK** problem is stated as follows:

> **#KNAPSACK**: We are given a list of T positive integers $\{a_t\}_{t=1}^T$ and an integer capacity R. The objective is to count the number of subsets $S \subset \{1, \cdots, T\}$ that satisfies the knapsack constraint $\sum_{t\in S} a_t \leq R$.

Next, Theorem 4.3.4 provides a lower bound on the number of samples necessary for solving the data-driven newsvendor problem to near optimality, with 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, we construct two demand distributions D_1 and D_2, with the properties that they have disjoint sets of $(1+\epsilon)$-optimal bases stocks, but their statistical distance is small. By the disjointness property, if there exists an algorithm that returns a $(1+\epsilon)$-optimal base stocks under both

D_1 or D_2 using m samples, then there also exists an algorithm that distinguishes between D_1 and D_2 using m samples. Nevertheless, since the statistical distance between D_1, D_2 is small, we can provide a lower bound on the number of samples needed for the classfication. This provides a lower bound on the number of samples needed for solving the data-driven newsvendor problem. The proof of Theorem 4.3.4 can be found in Appendix C.10.

4.8 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 sets of demand 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

$$\frac{\mathbb{E}\text{Cost}_1(0, \hat{R}_1, \cdots, \hat{R}_5)}{\text{Cost}_1(0, R_1^*, \cdots, R_5^*)} \tag{4.12}$$

implied 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 4.3.1 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 the sake of clarity.) By combining Corollary 4.5.4, Claim 4.5.5 and Lemma 4.5.7, we have the following stronger bound

$$\frac{\mathbb{E}\text{Cost}_1(0, \hat{R}_1, \cdots, \hat{R}_5)}{\text{Cost}_1(0, R_1^*, \cdots, R_5^*)} \leq \prod_{t=1}^{5} \left(1 + \frac{3(b+1)(5-t+1)\max\{1, 5-t\}}{\sqrt{50000}} \sqrt{\log 2000}\right). \tag{4.13}$$

We use the bound at the right hand side of (4.13) 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 §4.6, 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 (4.12).

b	Simulated Rel Ratio			Theoretical Rel 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 4.1: Simulated and Theoretical Relative Ratios.

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 4.5.5 and Lemma 4.5.7. 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.

4.9 Conclusion

In this chapter, 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. Finally, our algorithms are complemented by an information theoretic lower bound on the number of samples needed to solve the data-driven newsvender problem to near optimality.

Altogether, our work and the paper [Levi et al., 2007] demonstrate the tractability of

certain class of multi-stage data driven stochastic optimization problems. This is in contrast to [Shapiro and Nemirovski, 2005] 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 4.5.5), which seems to only hold in inventory related problems, as well as the validity of Lemma 4.5.7 in other dynamic programming models.

Chapter 5

Concluding Remarks and Future Directions

In the previous chapters, we proposed several algorithms for solving revenue maximization problems and inventory control problems in data-driven settings. In this final chapter, we briefly summarize the results, and propose future directions for continuing the research.

5.1 Summary and Future Work for Chapter 2

In Chapter 2, we consider solving the Choice-based Deterministic Linear Program (CDLP-P) to near-optimality, assuming the ability to approximately solve the underlying Single Period Problem in the absence of resource constraints. We propose two algorithms, namely Approximate Column Generation heuristic (ACG) and Potential-based Algorithm (PB). ACG generalizes the classical Column Generation heuristic; the former provides an $\alpha-$approximation to CDLP-P at termination, assuming an $\alpha-$approximate oracle to the underlying SPP. Under the same assumption, PB essentially provides an $(\alpha+\epsilon)-$approximation to CDLP-P. Different from ACG, PB involves only polynomially many arithmetic operations and polynomially many invocations to the $\alpha-$approximate oracle.

Building on the tractability results, we propose an efficient online policy ONLINE(τ) to the online choice-based NRM problem with an unknown MNL choice model. The policy ONLINE(τ) achieves a regret of $\tilde{O}(T^{2/3})$ when the length of the learning phase τ is chosen to be $T^{2/3}$. Numerical results show that the proposed policy is effective even when the length T of the sales horizon is small.

Future work. Although PB is a provably efficient algorithm, it is still desirable to design an algorithm that solves CDLP-P with both theoretical and empirical efficiency. An attentive reader would have observed that there are much similarity between ACG and PB; both involves solving a series of SPP with appropriately defined revenue coefficients, and then augmenting the collection of candidate assortments. Thus, it is conceivable that the two algorithms could be merged to form a 'hybrid' algorithm. We believe that this would be a very interesting future direction, that is of both theoretical and practical interest to the community.

Apart from efficiency in solving CDLP-P, online assortment planning is in general an unexplored area in revenue management. An immediate open question to our work is a computationally efficient non-anticipatory policy to online choice-based NRM problem with unknown MNL model that achieves a $\tilde{O}(\sqrt{T})$ regret. Improving the dependence on N, B, R in the regret bound is an equally interesting question. Finally, it is very desirable to generalize our approach to other choice models.

5.2 Summary and Future Work for Chapter 3

In Chapter 3, we consider a dynamic pricing problem where the underlying demand function is unknown, but belongs to a finite set of demand functions. The seller aims to maximize the total revenue earned in a sales horizon, subject to the constraint that price can be changed at most m times. We investigate the best possible regret, which is the difference between the expected optimum revenue and the revenue achieves by the seller, who employs a non-anticipatory policy. Our key finding is that any m-change policy (a non-anticipatory policy

that implements at most m price changes) must incur a regret of $\Omega(\log^{(m)} T)$, where T is the length of the sales horizon. This regret lower bound demonstrates that the pricing policy proposed by [Cheung et al., 2015] is the best possible, and our regret analysis provides important insights into the structure of a policy that achieves the optimal regret order bound.

Future work. One interesting direction is to relax the assumption on the finiteness of the demand function family Ψ. For a general Ψ, the regret bound increases when the number of hypothesis demand functions increases. Thus, it is interesting to investigates the cases for Ψ when the optimal regret does not increase with the size of Ψ in the limited price experimentation setting.

An interesting future direction is to generalize our results to resource constrained setting. Currently, a $O(\sqrt{T})$ regret bound has been proved, where the multiplicative factor in the big Oh notation depends on the number of price changes m. It is not known if such bound is the best possible, and if the optimal regret bound for the resource constrained case is different from the unconstrained case.

5.3 Summary and Future Work for Chapter 4

In Chapter 4, we consider the capacitated inventory control problem in an offline data-driven setting. First, we show that the Sample Average Approximation (SAA) method is *sample efficient*, as the SAA method only requires polynomially many samples to achieve near-optimality. However, the SAA method is *computationally inefficient*, in the sense that the underlying SAA problem is #P-hard. Thus, we propose a randomized polynomial time approximation scheme to the data-driven problem. Finally, an information theoretic sample lower bound to the problem is established.

Future Work. An immediate interesting direction is to generalize the approach to other inventory control model. Indeed, a recent work by [Qin et al., 2016] has already consider the multi-period joint pricing and inventory control problem in an offline data-

driven setting.

Another interesting direction is to explore the possibility of data-driven algorithms for dynamic program. More precisely, is it possible to construct an algorithm for a certain class of dynamic programs in the offline data-driven setting, where the amount of sample used is independent of the underlying random distributions? This question is partially answered in [Halman et al., 2014]. They consider a setting where the decision makers only has the oracle access to the probability density functions (pdfs) of the underlying random variables, and they provide a near optimal algorithm which invokes the oracle for a polynomial number of times. Moreover, the number of oracle calls is almost independent of the underlying random variables; the only dependence is on the logarithm of the supports. To generalize from the case of oracle access to pdfs to the case of oracle access to samples in the setting of [Halman et al., 2014] seems a very interesting question.

5.4 A Final Remark

Altogether, we have studied various multi-stage data-driven problems, which have applications in revenue management and supply chain management. However, the quest for data-driven algorithms is still at its very beginning, and there are many other operational problems to be solved in data driven settings. It is also interesting to move beyond the stochastic optimization setting; indeed, stochastic optimization is not the only approach to optimization under uncertainty. In the last decade, robust optimization [Tal and Nemirovski, 1998, Bertsimas et al., 2011] has become another major paradigm for solving operational problems with model uncertainty. While robust optimization has become a rather mature field, robust optimization in an online setting seems unexplored. We believe that it will be an interesting avenue for future exploration.

Appendix A

Technical Results in Chapter 2

A.1 Shorthands for References in Tables 2.1 and 2.2

For Table 2.1, the shorthands are tabulated as follows:

Shorthand	Full Reference
[TVR04]	[Talluri and Van Ryzin, 2006]
[DGT13]	[Davis et al., 2013]
[DG14]	[Désir and Goyal, 2014]
[FT14]	[Feldman and Topaloglu, 2014]
[RSS10]	[Rusmevichientong et al., 2010]
[FT15]	[Feldman and Topaloglu, 2015]
[DG14]	[Désir and Goyal, 2014]
[DGSY15]	[Désir et al., 2015]

For Table 2.2, the shorthands are tabulated as follows:

Shorthand	Full Reference
[TVR04]	[Talluri and Van Ryzin, 2006]
[DGT13]	[Davis et al., 2013]
[DG14]	[Désir and Goyal, 2014]
[MS13]	[Mittal and Schulz, 2013]
[RSS09]	[Rusmevichientong et al., 2009]
[FT14]	[Feldman and Topaloglu, 2014]
[RSS10]	[Rusmevichientong et al., 2010]
[FT15]	[Feldman and Topaloglu, 2015]
[DGSY15]	[Désir et al., 2015]

A.2 A Discussion on the Proof of Theorem 2.3.2

In the following, we provide a discussion on the proof of Theorem 2.3.2, which uses standard optimization techniques. Recall the linear program CDLP-D = CDLP-D[$\mathbb{S}$], which is the dual of CDLP-P defined in §2.4:

$$\min \eta + \sum_{k=1}^{K} c(k)\rho(k)$$

$$\text{s.t. } \eta \geq \sum_{i=1}^{N} \left(r(i) - \sum_{k=1}^{K} a(i,k)\rho(k) \right) \varphi(i,S) + \sum_{\mathtt{f}\in S} \left(r(\mathtt{f}) - \sigma(\mathtt{f}) \right) \varphi(\mathtt{f},S) \quad \forall S \in \mathbb{S} \quad \text{(A.1a)}$$

$$\sum_{k\in\mathcal{K}} a(i,k)\rho(k) \geq \sigma(\mathtt{f}) \qquad \forall \mathtt{f} \in \mathcal{F}, i \in \mathtt{f} \qquad \text{(A.1b)}$$

$$\eta, \rho(k) \geq 0, \sigma(\mathtt{f}) \text{ free} \qquad \forall k \in \mathcal{K}, \mathtt{f} \in \mathcal{F} \qquad \text{(A.1c)}$$

Given the access to the exact oracles $\mathcal{A}$, we can solve the separation problem on the feasible region defined by (A.1a) with one oracle call to $\mathcal{A}$ and polynomial time computation. This

is demonstrated as follows. For a given (η, ρ, σ), we first solve the SPP$-(\tilde{r}, \varphi, \mathbb{S})$, where $\tilde{r}(i) = r(i) - \sum_{k=1}^{K} a(i,k)\rho(k)$, and $\tilde{r}(\mathtt{f}) = r(\mathtt{f}) - \sigma(k)$. After that, if the optimal value of SPP is less than or equal to η, we establish that (η, ρ, σ) satisfies all constraints in (A.1a). Otherwise, we identify an optimal solution S to the SPP, and establish that the inequality corresponding to S in (A.1) defines a separating hyperplane for (η, ρ, σ). Next, since there are only $O(|\mathcal{F}||\mathcal{N}|)$ many constraints in (A.1b, A.1c), their separation problems can also be solved in polynomial time.

This implies that CDLP-D can be solved by the Ellipsoid Algortithm [Khachiyan, 1980], by calling the exact oracle polynomially many times, in addition to performing a polynomial number of elementary operations. This fact is due to the computational equivalence between separation and optimization problems (see [Khachiyan, 1980], [Grötschel et al., 1993]). Given an optimal solution $(\eta^*, \rho^*, \sigma^*)$ to CDLP-D, it is known that an optimal solution to CDLP-P can be constructed by solving polynomially many separation problems on the feasible region of CDLP-D. This fact is stated in Corollary 14.1g(v) in [Schrijver, 1986], page 188.

A.3 Proof of of Lemma 2.3.5

First, note that ACG terminates in $|\mathbb{S}|$ iterations, i.e. finite time. Indeed, in any iteration τ, if we do not "break" the **for** loop in Line 6, then an assortment S_τ from $\mathbb{S} \setminus \mathbb{S}_\tau$ is included in $\mathbb{S}_\tau$ to form $\mathbb{S}_{\tau+1}$. We argue that $S_\tau \in \mathbb{S} \setminus \mathbb{S}_\tau$ by contradiction. Suppose $S_\tau \in \mathbb{S}_\tau$, say $S_\tau = S_{\tau'}$ for some $\tau' < \tau$. Then by the constraint for $S_{\tau'}$ in (2.6a) in CDLP-D[$\mathbb{S}_\tau$] (which is added in iteration τ' before iteration τ), it would be the case that $\bar{\eta}_\tau \geq \sum_{j \in S_\tau} \bar{r}_\tau(j)\varphi(j, S_\tau)$, meaning that a "break" should have occurred in iteration τ, contrary to our assumption.

Next, we show that ACG returns an $\alpha-$approximation when terminates. When ACG signals a "Break" in iteration τ, we have

$$\bar{\eta}_\tau \geq \sum_{j \in S_\tau} \bar{r}_\tau(j)\varphi(j, S_\tau) \geq \alpha \max_{S \in \mathbb{S}} \sum_{j \in S} \bar{r}_\tau(j)\varphi(j, S).$$

Along with the feasibility of $(\bar{\eta}_\tau, \bar{\rho}_\tau, \bar{\sigma}_\tau)$ to the dual constraints (2.6b, 2.6c), we deduce that $(\bar{\eta}_\tau/\alpha, \bar{\rho}_\tau, \bar{\sigma}_\tau)$ is feasible to CDLP-D[$\mathbb{S}$] =CDLP-D. Therefore,

$$\frac{\bar{\eta}_\tau}{\alpha} + \sum_{k=1}^{K} c(k)\bar{\rho}_\tau(k) \geq \eta^* + \sum_{k=1}^{K} c(k)\rho^*(k) = \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y^*(S)$$

by LP duality, OPT(CDLP-P) = OPT(CDLP-D). We thus prove the theorem by LP duality, OPT(CDLP-P[$\mathbb{S}_\tau$]) = OPT(CDLP-D[$\mathbb{S}_\tau$]):

$$\frac{1}{\alpha} \sum_{S\in\mathbb{S}_\tau} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)\bar{y}_\tau(S) \geq \frac{\bar{\eta}_\tau}{\alpha} + \sum_{k=1}^{K} c(k)\bar{\rho}_\tau(k),$$

where $(\bar{y}_\tau, \bar{z}_\tau)$ is an optimal solution to CDLP-P[$\mathbb{S}_\tau$].

A.4 Proof of Lemma 2.4.2

We first assert the following key inequality for the proof:

$$\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}V(\tau) - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}U_k(\tau) \geq \frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}. \tag{A.2}$$

Given (A.2), we prove the required inequality $\Omega(\tau) \leq \Omega(\tau-1)$ as follows:

$$\begin{aligned}
\Omega(\tau) &= \Xi(\tau) + \sum_{k=1}^{K} \Psi_k(\tau) \\
&= \Xi(\tau-1)\frac{(1-\epsilon)^{\frac{V(\tau)}{\gamma}}}{1-\frac{\epsilon}{\gamma}} + \sum_{k=1}^{K} \Psi_k(\tau-1)\frac{(1+\epsilon)^{\frac{U_k(\tau)}{\gamma}}}{1+\frac{\epsilon}{\gamma}} \\
&\leq \frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}\left(1-\frac{V(\tau)}{\gamma}\epsilon\right) + \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}\left(1+\frac{U_k(\tau)}{\gamma}\epsilon\right) \qquad \text{(A.3)} \\
&= \left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} + \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}\right) - \frac{\epsilon}{\gamma}\left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}V(\tau) - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}U_k(\tau)\right)
\end{aligned}$$

$$\leq \left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} + \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \right) - \frac{\epsilon}{\gamma} \left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \right) \tag{A.4}$$

$$= \Xi(\tau-1) + \sum_{k=1}^{K} \Psi_k(\tau-1) = \Omega(\tau-1).$$

The inequality (A.3) is justified in the following. First, notice that $\frac{V(\tau)}{\gamma} \in [0,1]$, and also that $\frac{U_k(\tau)}{\gamma} \in [0,1]$ for all k, by our choice of γ in Algorithm 2. By applying Holder's Inequality, for any $0 \leq a, \epsilon \leq 1$, we have $(1-\epsilon)^a \leq 1 - a\epsilon$ and $(1+\epsilon)^a \leq 1 + a\epsilon$. This explain why we can bring down the exponents. The bound (A.4) is by the claim (A.2).

The key inequality (A.2) is proved in the following steps:

$$\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} V(\tau) - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} U_k(\tau) = \sum_{S \in \mathbb{S}} \sum_{j \in \mathcal{N} \cup \mathcal{F}} \tilde{r}_\tau(j) \varphi(j,S) \upsilon_\tau(S) \tag{A.5}$$

$$\geq \alpha \sum_{S \in \mathbb{S}} \sum_{j \in \mathcal{N} \cup \mathcal{F}} \tilde{r}_\tau(j) \varphi(j,S) y^*(S), \tag{A.6}$$

$$\geq \frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}. \tag{A.7}$$

Equality (A.5) is essentially by our definition of the net revenue $\tilde{r}$ in (2.11), (2.13). Inequality (A.6) is by the application of an $\alpha-$approximate oracle. Inequality (A.7) is by the definition of $I_\tau(\mathtt{f})$ in (2.12) as well as the feasibility of an optimal solution (y^*, z^*) and our assumption on Z.

In the remaining, we prove the asserted equation and inequalities. First, the equality (A.5) holds by our definitions of V, U_k and the net revenue coefficients in (2.11), (2.13):

$$\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} V(\tau) - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} U_k(\tau)$$

$$= \sum_{S \in \mathbb{S}} \left[\sum_{i \in \mathcal{N}} \left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}} \frac{r(i)}{Z} - \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)} \right) \varphi(i,S) + \right.$$

$$\sum_{\mathtt{f}\in\mathcal{F}}\left(\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}\frac{r(\mathtt{f})}{Z}-\sum_{k=1}^{K}\frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}\frac{a(I_\tau(\mathtt{f}),k)}{c(k)}\right)\varphi(\mathtt{f},S)\Bigg]v_\tau(S)$$
$$=\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S)v_\tau(S)=\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S_\tau).$$

Next, we prove (A.6). We recall the observation that $\tilde{r}_\tau(i)$, $\tilde{r}_\tau(\mathtt{f})$ need not be non-negative. By the definition of an $\alpha-$approximation orcale, we have the following:

$$\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S_\tau)\geq\alpha\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S) \tag{A.8}$$

for all $S\in\mathbb{S}$. By taking a weighted sum of (A.8) with coefficients $\{y^*(S)\}_{S\in\mathbb{S}}$, we have

$$\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S_\tau)\geq\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S_\tau)\sum_{S\in\mathbb{S}}y^*(S)\geq\alpha\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S)y^*(S).$$

Hence demonstrating (A.6).

Finally, we prove (A.7). Putting back the definition of $\tilde{r}_\tau(i)$, $\tilde{r}_\tau(\mathtt{f})$, we have

$$\alpha\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}}\tilde{r}_\tau(j)\varphi(j,S)y^*(S)$$
$$=\underbrace{\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}\frac{\alpha}{Z}\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}}r_\tau(j)\varphi(j,S)y^*(S)}_{(\dagger)}$$
$$\underbrace{-\alpha\sum_{k=1}^{K}\sum_{S\in\mathbb{S}}\frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}\left[\sum_{i\in\mathcal{N}}\frac{a(i,k)}{c(k)}\varphi(i,S)+\sum_{\mathtt{f}\in\mathcal{F}}\frac{a(I_\tau(\mathtt{f}),k)}{c(k)}\varphi(\mathtt{f},S)\right]y^*(S)}_{(\ddagger)}.$$

Observe that by our assumption on the target value Z in the Lemma, we have

$$(\dagger)\geq\frac{\Xi(\tau-1)}{1-\frac{\epsilon}{\gamma}}. \tag{A.9}$$

We then upper bound (‡) as follows:

$$(\ddagger) \leq \sum_{k=1}^{K} \sum_{S \in \mathbb{S}} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \left[\sum_{i \in \mathcal{N}} \frac{a(i,k)}{c(k)} \varphi(i,S) + \sum_{\mathtt{f} \in \mathcal{F}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)} \varphi(\mathtt{f},S) \right] y^*(S)$$

$$= \sum_{k=1}^{K} \sum_{S \in \mathbb{S}} \sum_{i \in \mathcal{N}} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)} \varphi(i,S) y^*(S) + \sum_{\mathtt{f} \in \mathcal{F}} \sum_{i \in \mathtt{f}} \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)} z^*(\mathtt{f},i) \tag{A.10}$$

$$= \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \left[\sum_{S \in \mathbb{S}} \sum_{i \in \mathcal{N}} \frac{a(i,k)}{c(k)} \varphi(i,S) y^*(S) + \sum_{\mathtt{f} \in \mathcal{F}} \sum_{i \in \mathtt{f}} \frac{a(i,k)}{c(k)} z^*(\mathtt{f},i) \right]$$

$$\leq \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}}. \tag{A.11}$$

The bound (A.11) is by the feasibility of y^* for the constraints (2.1a), and the bound (A.10) is by the following. For all $\mathtt{f} \in \mathcal{F}$, we have:

$$\sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)} \sum_{S \in \mathbb{S}} \varphi(\mathtt{f},S) y^*(S)$$

$$= \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)} \sum_{i \in \mathtt{f}} z^*(\mathtt{f},i) = \sum_{i \in \mathtt{f}} \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)} z^*(\mathtt{f},i) \tag{A.12}$$

$$\leq \sum_{i \in \mathtt{f}} \sum_{k=1}^{K} \frac{\Psi_k(\tau-1)}{1+\frac{\epsilon}{\gamma}} \frac{a(i,k)}{c(k)} z^*(\mathtt{f},i). \tag{A.13}$$

The equality (A.12) is by the constraint (2.1b), and the inequality (A.13) is by the definition of $I_\tau(\mathtt{f})$ in (2.12) in Algorithm 2. By applying the bounds (A.9) and (A.11) for (†) and (‡), we have proved (A.7), hence the Lemma is proved.

A.5 Proof of Lemma 2.4.3

We first claim that the following upper bound holds:

$$\Omega(0) \leq (K+1)\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right] \tag{A.14}$$

We defer the proof of (A.14) until the end of the proof.

Proving $\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)\tilde{y}(S) \geq (1-2\epsilon)Z$. Given (A.14), we have

$$\exp\left[-\log(1-\epsilon)\left(\frac{\mathcal{T}(1-\epsilon)}{\gamma} - \sum_{\tau=1}^{\mathcal{T}}\frac{V(\tau)}{\gamma}\right)\right] = \Xi(\mathcal{T}) \leq \Omega(\mathcal{T}) \leq \Omega(0) \leq (K+1)\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right].$$

By taking logarithm on each side and the inequality, this simplifies to

$$(1-\epsilon) - \frac{1}{\mathcal{T}}\sum_{\tau=1}^{\mathcal{T}} V(\tau) \leq \frac{\gamma\log(K+1)}{2\epsilon\mathcal{T}} - \frac{\epsilon}{6} \leq 0.$$

The first inequality is by the fact that $-\log(1-\epsilon) \leq 2\epsilon$ for $\epsilon \in [0, 1/2)$, and the second inequality is by the assumption on $\mathcal{T}$. Substituting the definition of $V(s)$ from (2.9) yields

$$\begin{aligned}&\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)\tilde{y}(S) \qquad\qquad (A.15)\\ =&\frac{1}{1+\epsilon}\sum_{S\in\mathbb{S}}\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)\left(\frac{1}{\mathcal{T}}\sum_{\tau=1}^{\mathcal{T}}\dot{v}_\tau(S)\right) \geq \frac{1-\epsilon}{1+\epsilon}Z \geq (1-2\epsilon)Z.\end{aligned}$$

The step (A.15) is by the definition of $\tilde{y}$ in Line 9 in Algorithm 2. This proves the performance guarantee.

Proving that $(\tilde{y},\tilde{z})$ is feasible to CDLP-P. First, we check the constraints (2.1c). It is clear that $(\tilde{y},\tilde{z})$ are non-negative; by the definition (9), we have

$$\sum_{S\in\mathbb{S}}\tilde{y}(S) = \frac{1}{(1+\epsilon)\mathcal{T}}\sum_{S\in\mathbb{S}}\sum_{\tau=1}^{\mathcal{T}} v_\tau(S) \leq \frac{1}{(1+\epsilon)\mathcal{T}}\sum_{\tau=1}^{\mathcal{T}} 1 = \frac{1}{1+\epsilon} \leq 1.$$

Then, we check the constraints (2.1b). We claim that for any given $\mathtt{f} \in \mathcal{F}$ and $\tau \in \{1, \cdots, \mathcal{T}\}$, it holds that

$$\sum_{S \in \mathbb{S}} \varphi(\mathtt{f}, S) v_\tau(S) = \sum_{i \in \mathtt{f}} \zeta_\tau(\mathtt{f}, i). \tag{A.16}$$

Indeed, if $\mathtt{f} \notin S_\tau$, we have

$$\sum_{S \in \mathbb{S}} \varphi(\mathtt{f}, S) v_\tau(S) = \varphi(\mathtt{f}, S_\tau) = 0, \text{ and } \quad \zeta_\tau(\mathtt{f}, i) = 0 \text{ for all } i \in \mathtt{f}.$$

Otherwise, we have $\mathtt{f} \in S_\tau$, and

$$\sum_{S \in \mathbb{S}} \varphi(\mathtt{f}, S) v_\tau(S) = \varphi(\mathtt{f}, S_\tau) = \zeta_\tau(\mathtt{f}, I_\tau(\mathtt{f})) = \sum_{i \in \mathtt{f}} \zeta_\tau(\mathtt{f}, i).$$

This proves (A.16). Finally, the constraint (2.1b) for $\mathtt{f}$ is shown as follows:

$$\sum_{S \in \mathbb{S}} \varphi(\mathtt{f}, S) \tilde{y}(S) = \frac{1}{(1+\epsilon)\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} \sum_{S \in \mathbb{S}} \varphi(\mathtt{f}, S) v_\tau(S) = \frac{1}{(1+\epsilon)\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} \sum_{i \in \mathtt{f}} \zeta_\tau(\mathtt{f}, i) = \sum_{i \in \mathtt{f}} \tilde{z}(\mathtt{f}, i).$$

Finally, we verify that $(\tilde{y}, \tilde{z})$ satisfies the constraints in (2.1a). Now, observe that, for each $k \in \{1, \cdots, K\}$, we have

$$\exp\left[-\log(1+\epsilon)\left(\frac{\mathcal{T}(1+\epsilon)}{\gamma} - \sum_{\tau=1}^{\mathcal{T}} \frac{U_k(\tau)}{\gamma}\right)\right] = \Psi_k(\mathcal{T}) \leq \Omega(\mathcal{T}) \leq \Omega(0) \leq (K+1)\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right].$$

Similarly to before, taking logarithm on each side yields

$$(1+\epsilon) - \frac{1}{\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} U_k(\tau) \geq \frac{\epsilon}{3} - \frac{\gamma \log(K+1)}{\epsilon \mathcal{T}} \geq 0. \tag{A.17}$$

The first inequality is by the fact that $\log(1+\epsilon) \leq \epsilon$ for $\epsilon \in [0, 1]$, and the second inequality

is by the assumption on $\mathcal{T}$. Recall the definition of $U_k(\tau)$:

$$U_k(\tau) = \sum_{S\in\mathbb{S}}\sum_{i\in\mathcal{N}} \frac{a(i,k)}{c(k)}\varphi(i,S)v_\tau(S) + \sum_{\mathtt{f}\in\mathcal{F}}\sum_{S\in\mathbb{S}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)}\varphi(\mathtt{f},S)v_\tau(S). \tag{A.18}$$

We unravel the second term as follows:

$$\sum_{\mathtt{f}\in\mathcal{F}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)}\sum_{S\in\mathbb{S}}\varphi(\mathtt{f},S)v_\tau(S) = \sum_{\mathtt{f}\in\mathcal{F}} \frac{a(I_\tau(\mathtt{f}),k)}{c(k)}\zeta_\tau(\mathtt{f},I_\tau(\mathtt{f})) = \sum_{\mathtt{f}\in\mathcal{F}}\sum_{i\in\mathtt{f}} \frac{a(i,k)}{c(k)}\zeta_\tau(\mathtt{f},i). \tag{A.19}$$

Combining (A.17, A.18, A.19), we have

$$\sum_{S\in\mathbb{S}}\sum_{i\in\mathcal{N}} a(i,k)\varphi(i,S)\tilde{y}(S) + \sum_{i\in\mathcal{N}}\sum_{\mathtt{f}\ni i} a(i,k)\tilde{z}(\mathtt{f},i) = \frac{1}{(1+\epsilon)\mathcal{T}}\sum_{\tau=1}^{\mathcal{T}} c(k)U_k(\tau) \le c(k).$$

We return to the proof of the inequality (A.14). Observe that, at $\tau = 0$,

$$\Xi(0) = \exp\left[-(1-\epsilon)\log(1-\epsilon)\frac{\mathcal{T}}{\gamma}\right]\left(1-\frac{\epsilon}{\gamma}\right)^{\mathcal{T}} \le \exp\left[-\frac{\mathcal{T}}{2\gamma}\epsilon^2\right],$$

$$\sum_{k=1}^{K}\Psi_k(0) = K\exp\left[-(1+\epsilon)\log(1+\epsilon)\frac{\mathcal{T}}{\gamma}\right]\left(1+\frac{\epsilon}{\gamma}\right)^{\mathcal{T}} \le K\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right].$$

This implies that

$$\Omega(0) = \exp\left[-\frac{\mathcal{T}}{2\gamma}\epsilon^2\right] + K\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right] \le (K+1)\exp\left[-\frac{\mathcal{T}}{3\gamma}\epsilon^2\right],$$

hence proving (A.14). This proves Lemma 2.4.3.

A.6 Proof of Theorem 2.3.4

We demonstrate in the following that Algorithm 3 maintains the following three properties. For all $\ell \in \{1,\cdots,M+1\}$, where M is defined in Line 4 in Algorithm 3,

1. It holds that $\mathrm{UB}(\ell) > \alpha \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y^*(S)$.

2. The solution $(\check{y},\check{z})$ output by FEASIBILITY($\mathrm{LB}(\ell),\epsilon$) is $(1-2\epsilon)\mathrm{LB}(\ell)$−successful.

3. At the end of the ℓ^{th} iteration, the solution $(\hat{y},\hat{z})$ maintained by BINSEARCH(ϵ,δ) is the output of FEASIBILITY($\mathrm{LB}(\ell+1),\epsilon$).

We use the notation $(\check{y},\check{z})$ in Property 2 to avoid confusion from the notation $(\tilde{y},\tilde{z})$, which is an output of FEASIBILITY($\mathrm{MP}(\ell),\epsilon$) in iteration ℓ. These claims are proved by induction on ℓ, which are deferred to the end of the proof. The performance guarantee of Algorithm 3 is implied by these claims. Now, the final output $(\hat{y},\hat{z})$ is the output of FEASIBILITY($\epsilon\mathrm{LB}(M+1)$) by Property 3. Then, by property 2, we know that $(\hat{y},\hat{z})$ is feasible to (CDLP-P) with performance guarantee

$$\sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)\hat{y}(S) \geq (1-2\epsilon)\mathrm{LB}(M+1).$$

Finally, recall that by the choice of M, we have $\mathrm{UB}(M+1) - \mathrm{LB}(M+1) = \delta$. Thus, by Property 1, we have $\mathrm{LB}(M+1) \geq \alpha \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y^*(S) - \delta$, thus proving (2.3).

Finally, we go back to the proofs of the 3 properties:

Proving Item 1. This is proved by induction on ℓ. For the case $\ell = 1$, $\mathrm{UB}(1) = \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)$, which clearly serves as an upper bound. Now, suppose that the inequality

$$\mathrm{UB}(\ell-1) > \alpha \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)\varphi(j,S)y^*(S)$$

is true, where $\ell > 1$. If the statement in Line 6 is true, then the output $(\tilde{y},\tilde{z})$ is $(1-2\epsilon)\mathrm{MP}(\ell-1)$−successful. This means that $\mathrm{UB}(\ell) = \mathrm{UB}(\ell-1)$ by Line 8, so the claim clearly holds. Otherwise, Lemma 2.4.2, 2.4.3 imply that $\mathrm{MP}(\ell-1) > \alpha \sum_{S\in\mathbb{S}} \sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)$, and Line 10 stipulates that $\mathrm{UB}(\ell) = \mathrm{MP}(\ell-1)$, so our claim for ℓ holds. This completes the induction.

Proving Item 2. Again, this is demonstrated by induction on ℓ. For the case $\ell = 1$, $\mathrm{LB}(1) = 0$, and the claim is true since any solution feasible to (CDLP-P) is 0-successful. Now, suppose that the output $(\check{y}, \check{z})$ of FEASIBILITY($\mathrm{LB}(\ell-1), \epsilon$) is $(1-2\epsilon)\mathrm{LB}(\ell-1)$−successful, where $\ell > 1$. If the case in Line 6 is true, then we have $\mathrm{LB}(\ell) = \mathrm{MP}(\ell-1)$ by Line 8, hence the claim for ℓ is true, by the condition in Line 6. Otherwise, we have $\mathrm{LB}(\ell) = \mathrm{LB}(\ell-1)$, which means that the claim for ℓ is true. This completes the induction.

Proving Item 3. This is again proved by induction on ℓ. Suppose the claim is true for ℓ. Now, if the first case is true in iteration $\ell+1$, then we would have defined $\mathrm{LB}(\ell+2) = \mathrm{MP}(\ell+1)$, and defined $(\hat{y}, \hat{z})$ as the output of FEASIBILITY($\epsilon, \mathrm{MP}(\ell+1)$. Otherwise, we would have defined $\mathrm{LB}(\ell+2) = \mathrm{LB}(\ell+1)$, and the solution $(\hat{y}, \hat{z})$ is unchanged, so the claim follows from our induction hypothesis.

Total number of oracle call is upper bounded by

$$\sum_{\ell=1}^{M} \mathcal{T}(\ell) \leq \frac{6\log(K+1)}{\epsilon^2}\left[\frac{\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)}{\delta} + B \max_{i\in\mathcal{N},k\in\mathcal{K}}\left\{\frac{a(i,k)}{c(k)}\right\}\log\left(\frac{\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)}{\delta}\right)\right], \tag{A.20}$$

where we recall B is the maximum number of product purchased by a customer. Total number of elementary operations is upper bounded by

$$10NKF\sum_{\ell=1}^{M}\mathcal{T}(\ell) = O\left(NKF\frac{\sum_{j\in\mathcal{N}\cup\mathcal{F}} r(j)}{\epsilon^2\delta}\right). \tag{A.21}$$

A.7 Extension to Personalized Assortment Models

The Choice-based Deterministic Linear Program with choice models $(\varphi_\iota, \mathbb{S}_\iota)$ is defined as follows. (β_ι is the probability that a type ι customer appears)

$$\max \sum_{\iota\in\mathcal{I}}\sum_{S_\iota\in\mathbb{S}_\iota}\sum_{j\in\mathcal{N}\cup\mathcal{F}} \beta_\iota r(j)\varphi_\iota(j, S_\iota)y_\iota(S_\iota)$$

$$\text{s.t.} \sum_{\iota\in\mathcal{I}}\sum_{S\in\mathbb{S}_\iota}\sum_{i\in\mathcal{N}} a(i,k)\varphi_\iota(i,S_\iota)y_\iota(S_\iota) + \sum_{\iota\in\mathcal{I}}\sum_{i\in\mathcal{N}}\sum_{\mathtt{f}\ni i} a(i,k)z_\iota(\mathtt{f},i) \leq c(k) \quad \forall k\in\mathcal{K}$$

$$\sum_{S_\iota \in \mathbb{S}_\iota} \varphi_\iota(\mathtt{f}, S_\iota) y_\iota(S_\iota) = \sum_{i \in \mathtt{f}} z_\iota(\mathtt{f}, i) \qquad \forall \iota \in \mathcal{I}, \mathtt{f} \in \mathcal{F}$$

$$\sum_{S_\iota \in \mathbb{S}_\iota} y_\iota(S_\iota) \leq 1, \qquad y_\iota(S_\iota), z_\iota(\mathtt{f}, i) \geq 0 \qquad \forall \iota, S_\iota, \mathtt{f}, i$$

We call this formulation PERSONALIZED-CDLP-P for the choice models $\{(\varphi_\iota, \mathbb{S}_\iota)\}_{\iota \in \mathcal{I}}$ with probability $\{\beta_\iota\}_{\iota \in \mathcal{I}}$. Our approximation framework in §2.4 can be easily generalized to the personalized setting. More precisely, given an α_ι-approximate oracle $\mathcal{A}_\iota$ for $(\varphi_\iota, \mathbb{S}_\iota)$, we can provide (with additive error δ) a $(\sum_{\iota \in \mathcal{I}} \beta_\iota \alpha_\iota + \epsilon)$-approximation to the PERSONALIZED-CDLP-P by PB, and an $(\sum_{\iota \in \mathcal{I}} \beta_\iota \alpha_\iota)$-approximation to the PERSONALIZED-CDLP-P by ACG.

Indeed, consider the following instance of one customer type. The set of products is $\bigcup_{\iota \in \mathcal{I}} (\mathcal{N}_\iota \cup \mathcal{F}_\iota)$, where $\mathcal{N}_i$, $\mathcal{F}_i$ are copies of $\mathcal{N}, \mathcal{F}$. The assortment family is $\bigcup_{\iota \in \mathcal{I}} \mathbb{S}_\iota$, where $\mathbb{S}_\iota$ is a subset of $2^{\mathcal{N}_\iota \cup \mathcal{F}_\iota}$. A customer is endowed with the following choice model $\varphi : \bigcup_{\iota \in \mathcal{I}} (\mathcal{N}_\iota \cup \mathcal{F}_\iota) \times \bigcup_{\iota \in \mathcal{I}} \mathbb{S}_\iota \to [0, 1]$, defined by $\varphi(j_\iota, \bigcup_{\iota \in \mathcal{I}} S_\iota) = \beta_\iota \varphi_\iota(j_\iota, \Sigma_\iota)$, where j_ι denotes the copy of product j in $\mathcal{N}_\iota \cup \mathcal{F}_\iota$. By defining $\varphi_\iota(j_{\iota'}, S_\iota) =$ for all $\iota \neq \iota'$, we see that $\varphi(j, \bigcup_{\iota \in \mathcal{I}} S_\iota) = \sum_{\iota \in \mathcal{I}} \varphi_\iota(j, S_\iota)$. It is clear that the CDLP-P for the choice model $(\varphi, \mathbb{S})$ is the same as the PERSONALIZED-CDLP-P for the choice models $(\beta_\iota, \varphi_\iota, \mathbb{S}_\iota)_{\iota \in \mathcal{I}}$.

Moreover, it is easy to see that an $\sum_{\iota \in \mathcal{I}} \beta_\iota \alpha_\iota$-approximate oracle for $(\varphi, \mathbb{S})$ can be constructed using the oracles $\{\mathcal{A}_\iota\}_{\iota \in \mathcal{I}}$. Suppose we are given the SPP-$(r, \varphi, \mathbb{S})$, where $r \in \mathbb{R}^{\bigcup_{\iota \in \mathcal{I}} \mathcal{N}_\iota \cup \mathcal{F}_\iota}$, let r_ι denotes the restriction of r to the components for customer type ι. Now, apply $\mathcal{A}_\iota$ on SPP-$(r_\iota, \varphi_\iota, \mathbb{S}_\iota)$, which return a set $S_{\mathcal{A}_\iota}$, a subset of $\mathcal{N}_\iota \cup \mathcal{F}_\iota$. Then it is clear that $\bigcup_{\iota \in \mathcal{I}} S_{\mathcal{A}_\iota}$ is an $\sum_{\iota \in \mathcal{I}} \beta_\iota \alpha_\iota$-approximate solution to SPP-$(r, \varphi, \mathbb{S})$. Finally, we remark that an extreme point feasible solution to PERSONALIZED-CDLP-P has a support of size at most $|\mathcal{N}| + |\mathcal{F}| + |\mathcal{I}|$. In the case when $\mathcal{I}$ is very large compared to $|\mathcal{N}| + |\mathcal{F}|$, the asymptotically optimal greedy policy ([Gallego et al., 2004], also see the discussion in **The importance of CDLP**) essentially offers one type of assortment to each customer type.

A.8 A Remark on Assumption 2.5.4

We remark that the choices of $\tau = (Ce^R)^{2/3}N^{1/3}T^{2/3}$ and $\tau = T^{2/3}$ satisfy Assumption 2.5.4 when T is sufficiently large. Indeed, for the case of $\tau = (Ce^R)^{2/3}N^{1/3}T^{2/3}$, Assumption 2.5.4 (i, ii) are equivalent to

$$T \geq \frac{C^2e^{2R}}{c(k)^3}\log^{3/2}\frac{4NK}{\delta}, \quad T \geq 512\frac{C^2e^{2R}N}{c(k)^3}\log^{3/2}\frac{4N}{\delta}$$

for all $k \in \mathcal{K}$. For the case of $\tau = T^{2/3}$, Assumption 2.5.4 (i, ii) are equivalent to

$$T \geq \frac{1}{c(k)^3}\log^{3/2}\frac{4NK}{\delta}, \quad T \geq 512\frac{C^3e^{3R}N^{3/2}}{c(k)^3}\log^{3/2}\frac{4N}{\delta}$$

for all $k \in \mathcal{K}$. Again for the case of $\tau = T^{2/3}$, our numerical results in §2.6 shows that ONLINE(τ) is effective even when the assumption is violated.

A.9 Proof of Lemma 2.5.5

Recall the expression:

$$\mathcal{L}^i(\theta(i)) = N(i)\log\left(1+\exp[-\theta(i)]\right) + \left(\frac{\tau}{N} - N(i)\right)\log\left(1+\exp[\theta(i)]\right),$$

where $N(i) = \sum_{s=((i-1)\tau/N)+1}^{i\tau/N} 1(\Sigma_s = i)$, that is, the frequency of sampling $\Sigma_s = i$ in the learning phase. By Taylor Series Expansion, we have

$$\mathcal{L}^i(\hat{\theta}(i)) = \mathcal{L}^i(\theta^*(i)) + \frac{\mathrm{d}\mathcal{L}^i(\theta^*(i))}{\mathrm{d}\theta(i)}(\hat{\theta}(i)-\theta^*(i)) + \frac{1}{2}\frac{\mathrm{d}^2\mathcal{L}^i(\breve{\theta}(i))}{\mathrm{d}\theta(i)^2}(\hat{\theta}(i)-\theta^*(i))^2.$$

where $\breve{\theta}(i) = \gamma\theta^*(i) + (1-\gamma)\hat{\theta}(i)$ for some $\gamma \in (0,1)$. Now, note that $\mathcal{L}^i(\hat{\theta}(i)) \leq \mathcal{L}^i(\theta^*(i))$, which gives

$$0 \geq \frac{\mathrm{d}\mathcal{L}^i(\theta^*(i))}{\mathrm{d}\theta(i)}(\hat{\theta}(i)-\theta^*(i)) + \frac{1}{2}\frac{\mathrm{d}^2\mathcal{L}^i(\breve{\theta}(i))}{\mathrm{d}\theta(i)^2}(\hat{\theta}(i)-\theta^*(i))^2. \tag{A.23}$$

We first have:

$$\left|\frac{\mathrm{d}\mathcal{L}^i(\theta^*(i))}{\mathrm{d}\theta(i)}\right| = \left|\sum_{t=((i-1)\tau/N)+1}^{i\tau/N} \varphi(\Sigma_t, \{i\}|\theta^*) - \mathbf{1}(\Sigma_t = i)\right| \leq \sqrt{\frac{\tau}{N}\log\frac{4N}{\delta}} \tag{A.24}$$

with probability at least $1 - \delta/2N$, by Chernoff Inequality.

Next, we bound the second derivative as follows:

$$\frac{\mathrm{d}^2\mathcal{L}^i(\breve{\theta}(i))}{\mathrm{d}\theta(i)^2} = \frac{\tau e^{\breve{\theta}(i)}}{N(1+e^{\breve{\theta}(i)})^2} \geq \frac{\tau e^R}{N(1+e^R)^2}. \tag{A.25}$$

Combining (A.23, A.24, A.25), we have

$$\frac{\tau e^R}{2N(1+e^R)^2}\left|\tilde{\theta}(i) - \theta^*(i)\right|^2 - \sqrt{\frac{\tau}{N}\log\frac{4N}{\delta}}\left|\tilde{\theta}(i) - \theta^*(i)\right| \leq 0.$$

with probability at least $1 - \delta/2N$. This implies that for all $i \in \mathcal{N}$, we have

$$\left|\tilde{\theta}(i) - \theta^*(i)\right| \leq \frac{2(1+e^R)^2}{e^R}\sqrt{\frac{\tau}{N}\log\frac{4N}{\delta}} \leq 4e^R\sqrt{\frac{\tau}{N}\log\frac{4N}{\delta}} \tag{A.26}$$

with probability at least $1 - \delta/2$, which proves the Lemma.

A.10 Proof of Lemma 2.5.6

Consider the function $f : [0,1] \to \mathbb{R}$ defined by $f(\gamma) = \sum_{i\in S} b(i)\left(\varphi(i, S|\theta' + \gamma(\theta - \theta'))\right)$. By mean value theorem,

$$\begin{aligned}
&\sum_{i\in S} b(i)\left(\varphi(i,S|\theta) - \varphi(i,S|\theta')\right)\\
=&f(1) - f(0)\\
=&f'(\gamma) \quad \text{for some } \gamma \in (0,1)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i \in S} \frac{b(i) e^{\theta'(i) + \gamma(\theta(i) - \theta'(i))}}{(1 + \sum_{\ell \in S} e^{\theta'(\ell) + \gamma(\theta(\ell) - \theta'(\ell))})} (\theta(i) - \theta'(i)) \\
&\quad - \sum_{\ell \in S} e^{\theta'(\ell) + \gamma(\theta(\ell) - \theta'(\ell))} \frac{\sum_{i \in S} b(i) e^{\theta'(i) + \gamma(\theta(i) - \theta'(i))}}{(1 + \sum_{\ell \in S} e^{\theta'(\ell) + \gamma(\theta(\ell) - \theta'(\ell))})^2} (\theta(\ell) - \theta'(\ell)) \qquad \text{(A.27)} \\
&\leq \sum_{i \in S} \left| \theta(i) - \theta'(i) \right|.
\end{aligned}
$$

The inequality (A.27)is due to the observation that the coefficients of $(\theta(i) - \theta'(i))$ in the two sums belong to $[0, 1]$.

A.11 Proof of Lemma 2.5.7

Recall that $\mathbb{P}[\mathcal{E}_k \mid \mathcal{E}_{\hat{\theta}}] = \mathbb{P}[\sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) \leq Tc(k) - \tau \mid \mathcal{E}_{\hat{\theta}}]$. We decompose the sum $\sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k)$ into 4 terms, $(\diamondsuit_k), (\clubsuit_k), (\heartsuit_k), (\spadesuit_k)$, as follows:

$$
\begin{aligned}
\sum_{\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) = &\underbrace{\sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) - \sum_{t=\tau+1}^{T-\rho} \sum_{i \in \tilde{S}_t} a(i,k) \varphi(i, \tilde{S}_t | \theta^*)}_{(\diamondsuit_k)} \\
&+ \underbrace{\sum_{t=\tau+1}^{T-\rho} \sum_{i \in \tilde{S}_t} a(i,k) \varphi(i, \tilde{S}_t | \theta^*) - \sum_{t=\rho+1}^{T-\rho} \sum_{i \in \tilde{S}_t} a(i,k) \varphi(i, \tilde{S}_t | \hat{\theta})}_{(\clubsuit_k)} \\
&+ \underbrace{\sum_{t=\rho+1}^{T-\rho} \sum_{i \in \tilde{S}_t} a(i,k) \varphi(i, \tilde{S}_t | \hat{\theta}) - \sum_{t=\tau+1}^{T-\rho} \sum_{S \in \mathbb{S}} \sum_{i \in S} a(i,k) \varphi(i, S | \hat{\theta}) \hat{y}(S)}_{(\heartsuit_k)} \\
&+ \underbrace{\sum_{t=\tau+1}^{T-\rho} \sum_{S \in \mathbb{S}} \sum_{i \in S} a(i,k) \varphi(i, S | \hat{\theta}) \hat{y}(S)}_{(\spadesuit_k)}
\end{aligned}
$$

We bound each term from above, conditional on $\mathcal{E}_{\hat{\theta}}$, as follows:

To bound $(\diamondsuit_k)$: For any choice of $\hat{\theta}$, the sequence $\{a(\tilde{\Sigma}_t, k) - a(i,k)\varphi(i, \tilde{S}_t | \theta^*)\}_{t=\tau+1}^{T-\rho}$

of random variables are i.i.d.. Indeed, each assortment $\tilde{S}_t$ is independently sampled based on the distribution specified by $\hat{y}$ (an optimal solution to CDLP-P($\hat{\theta}$)), and each $\tilde{\Sigma}_t$ is sampled from $\tilde{S}_t \cup \{0\}$ according to the choice probability $\varphi(\cdot, \tilde{S}_t|\theta^*)$, independently of other $(\tilde{S}_{t'}, \tilde{\Sigma}_{t'})_{t' \neq t}$. Moreover, for any choice of $\hat{\theta}$, each of the random variables $a(\tilde{\Sigma}_t, k) - a(i,k)\varphi(i, \tilde{S}_t|\theta^*)$ has mean zero, and lies in the range $[-1, 1]$. By Chernoff inequality, we have

$$\mathbb{P}\left[\sum_{t=\tau+1}^{T-\rho} a(\tilde{\Sigma}_t, k) - \sum_{t=\tau+1}^{T-\rho} \sum_{i \in S_t} a(i,k)\varphi(i, \tilde{S}_t|\theta^*) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}}\right] \geq 1 - \frac{\delta}{4(K+1)}$$

for any choice of $\hat{\theta}$. In particular, this is true when we condition that $\hat{\theta}$ satisfies the confidence bound stated in $\mathcal{E}_{\hat{\theta}}$. By conditioning on $\mathcal{E}_{\hat{\theta}}$, this proves that

$$\mathbb{P}\left[(\diamondsuit_k) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{4(K+1)}.$$

To bound ($\clubsuit_k$): Recall the assumption that $a(i,k) \in \{0,1\}$, and $|\tilde{S}_t| \leq C$ for all t. Conditional on $\mathcal{E}_{\hat{\theta}}$, we have $\left|\hat{\theta}(i) - \theta^*(i)\right| \leq \epsilon(\tau)$. By applying Lemma 2.5.6, this implies that for all i, S, we have

$$\sum_{i \in S} a(i,k)\varphi(i, S|\theta^*) - \sum_{i \in S} a(i,k)\varphi(i, S|\hat{\theta}) \leq C\epsilon(\tau).$$

Thus, we have $(\clubsuit_k) \leq TC\epsilon(\tau)$.

To bound ($\heartsuit_k$): By the definition of the sampling procedure, $\mathbb{P}[\tilde{S}_t = S] = \hat{y}(S)$. Similar to the case in $(\diamondsuit_k)$, for any fixed $\hat{\theta}$ the random variables

$$\left\{\sum_{i \in \tilde{S}_t} a(i,k)\varphi(i, \tilde{S}_t|\hat{\theta}) - \sum_{S \in \mathbb{S}} \sum_{i \in S} a(i,k)\varphi(i, S|\hat{\theta})\hat{y}(S)\right\}_{t=\rho+1}^{T-\rho}$$

are i.i.d., mean 0, and lie in the interval $[-1, 1]$. By Chernoff inequality, we have

$$\mathbb{P}\left[(\heartsuit_k) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{4(K+1)}.$$

To bound ($\spadesuit_k$): Recall that $(\spadesuit_k) \leq (T - \rho - \tau)c(k)$, since $\hat{y}$ is a feasible solution to CDLP-P($\hat{\theta}$).

Combining the bounds, the following inequality (conditional on $\mathcal{E}_{\hat{\theta}}$) holds with probability $1 - \delta/2(K+1)$:

$$(\heartsuit_k) + (\clubsuit_k) + (\diamondsuit_k) + (\spadesuit_k) \leq \sqrt{2T \log \frac{K+1}{\delta}} + TC\epsilon(\tau) + \sqrt{2T \log \frac{K+1}{\delta}} + (T - \rho - \tau)c(k).$$

By the definition of ρ in (2.22), we have

$$\sqrt{2T \log \frac{K+1}{\delta}} + TC\epsilon(\tau) + \sqrt{2T \log \frac{K+1}{\delta}} + (T - \rho - \tau)c(k) \leq Tc(k) - \tau.$$

Altogether, we have $\mathbb{P}\left[(\heartsuit_k) + (\clubsuit_k) + (\diamondsuit_k) + (\spadesuit_k) \leq Tc(k) - \tau \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{2(K+1)}$, hence proving the Lemma.

A.12 Proof of Lemma 2.5.8

Consider the following linear program S-LP:

$$\max \sum_{S \in \mathbb{S}} \left[\left(\sum_{i \in S} r(i)\varphi(i, S \mid \hat{\theta})\right) + C\epsilon(\tau)\right] y(S) \tag{A.28a}$$

$$\text{s.t.} \sum_{S \in \mathbb{S}} \left[\left(\sum_{i \in S} a(i,k)\varphi(i, S \mid \hat{\theta})\right) - C\epsilon(\tau)\right] y(S) \leq c(k) - C\epsilon(\tau) \quad \forall k \in \mathcal{K} \tag{A.28b}$$

$$\sum_{S \in \mathbb{S}} y(S) = 1, \quad y(S) \geq 0 \quad \forall S \in \mathbb{S}, \tag{A.28c}$$

and let OPT(S-LP) denote its optimal value. We claim the following, conditional on the event $\mathcal{E}_{\hat{\theta}}$:

$$\begin{aligned} \text{OPT}(\hat{\theta}) + C\epsilon(\tau) &= \text{OPT}(\text{S-LP}) && \text{(A.29)} \\ &\geq \left(1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right) \text{OPT}(\theta^*). && \text{(A.30)} \end{aligned}$$

Proving (A.29): By the constraint $\sum_{S\in\mathbb{S}} y(S) = 1$, rearranging the constraint for resource k yields $\sum_{S\in\mathbb{S}} \sum_{i\in S} a(i,k)\varphi(i,S|\hat{\theta})y(S) \leq c(k)$, which is the resource k constraint for CDLP-P($\hat{\theta}$). Similarly, the objective of S-LP is equal to the objective of CDLP-P($\hat{\theta}$) plus $C\epsilon(\tau)$. This proves (A.29).

Proving (A.30): We first claim that the solution

$$\breve{y}(S) = \begin{cases} \left(1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right) y^*(S) & \text{if } S \in \mathbb{S} \setminus \emptyset \\ \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}} + \left(1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right) y^*(\emptyset) & \text{if } S = \emptyset \end{cases}$$

is feasible to S-LP, where we recall that y^* is an optimal solution to CDLP-P(θ^*). Given the feasibility of $\breve{y}$ to S-LP, we have

$$\begin{aligned} \text{OPT}(\text{S-LP}) &\geq \sum_{S\in\mathbb{S}} \left[\left(\sum_{i\in S} r(i)\varphi(i,S \mid \hat{\theta})\right) + C\epsilon(\tau)\right] \breve{y}(S) \\ &\geq \sum_{S\in\mathbb{S}} \sum_{i\in S} r(i)\varphi(i,S \mid \theta^*)\breve{y}(S) && \text{(A.31)} \\ &= \left(1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right) \sum_{S\in\mathbb{S}} \sum_{i\in S} r(i)\varphi(i,S \mid \theta^*)y^*(S) \\ &= \left(1 - \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\right) \text{OPT}(\theta^*), \end{aligned}$$

where step (A.31) is justified as follows. In the statement of the Lemma, we condition on

the event $\mathcal{E}_{\hat{\theta}}$. By Lemma 2.5.6, for all $i \in S, S \in \mathbb{S}$ we have

$$\left| \sum_{i \in S} a(i,k)\varphi(i,S|\hat{\theta}) - \sum_{i \in S} a(i,k)\varphi(i,S|\theta^*) \right| \leq C\epsilon(\tau). \tag{A.32}$$

This justifies the step (A.31).

Finally, we return to checking the feasibility $\breve{y}$. First, the constraints in (A.28c) hold; in particular, the equality $\sum_{S \in \mathbb{S}} \breve{y}(S) = 1$ holds by our definition of $\breve{y}(\emptyset)$. Note that the factor $\left(1 - \frac{C\epsilon(\tau)}{\min_{k \in \mathcal{K}}\{c(k)\}}\right)$ is non-negative, by Assumption 2.5.4 (ii).

To check the constraints in (A.28b), we have

$$\begin{aligned}
&\sum_{S \in \mathbb{S}} \left[\left(\sum_{i \in S} a(i,k)\varphi(i,S \mid \hat{\theta}) \right) - C\epsilon(\tau) \right] \breve{y}(S) \\
=&\left(1 - \frac{C\epsilon(\tau)}{\min_{k \in \mathcal{K}}\{c(k)\}}\right) \sum_{S \in \mathbb{S}} \left[\left(\sum_{i \in S} a(i,k)\varphi(i,S \mid \hat{\theta}) \right) - C\epsilon(\tau) \right] y^*(S) \\
\leq&\left(1 - \frac{C\epsilon(\tau)}{\min_{k \in \mathcal{K}}\{c(k)\}}\right) \sum_{S \in \mathbb{S}} \sum_{i \in S} a(i,k)\varphi(i,S \mid \theta^*) y^*(S) \qquad \text{(A.33)} \\
\leq&\left(1 - \frac{C\epsilon(\tau)}{\min_{k \in \mathcal{K}}\{c(k)\}}\right) c(k) \qquad \text{(A.34)} \\
\leq&c(k) - C\epsilon(\tau),
\end{aligned}$$

where (A.33) is by (A.32), and (A.34) is by the feasibility of y^* to CDLP-P(θ^*). Altogether, $\breve{y}$ is feasible to S-LP, and this finishes the proof of the Lemma.

A.13 Proof of Lemma 2.5.9

To demonstrate the regret bound, we have the following conditional on $\mathcal{E}_{\hat{\theta}}$:

$$T\text{OPT}(\theta^*) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t)$$

$$\leq \rho + \tau + (T-\rho-\tau)\left(\textsc{Opt}(\hat{\theta}) + \frac{C\epsilon(\tau)}{\min_{k\in\mathcal{K}}\{c(k)\}}\textsc{Opt}(\theta^*) + C\epsilon(\tau)\right) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t) \quad \text{(A.35)}$$

$$= \rho + \tau + (T-\rho-\tau)C\epsilon(\tau)\left(\frac{\textsc{Opt}(\theta^*)}{\min_{k\in\mathcal{K}}\{c(k)\}} + 1\right) + \underbrace{(T-\rho-\tau)\textsc{Opt}(\hat{\theta}) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t)}_{(\textsc{Regret})}. \quad \text{(A.36)}$$

The inequality (A.35) is by Lemma 2.5.8. We decompose the term REGRET and bound it from above in the following:

$$\begin{aligned}
\textsc{Regret} = & \underbrace{(T-\rho-\tau)\textsc{Opt}(\hat{\theta}) - \sum_{t=\tau+1}^{T-\rho}\sum_{i\in\tilde{S}_t} r(i)\varphi(i,\tilde{S}_t|\hat{\theta})}_{(\heartsuit_0)} \\
& + \underbrace{\sum_{t=\tau+1}^{T-\rho}\sum_{i\in\tilde{S}_t} r(i)\varphi(i,\tilde{S}_t|\hat{\theta}) - \sum_{t=\tau+1}^{T-\rho}\sum_{i\in\tilde{S}_t} r(i)\varphi(i,\tilde{S}_t|\theta^*)}_{(\clubsuit_0)} \\
& + \underbrace{\sum_{t=\tau+1}^{T-\rho}\sum_{i\in\tilde{S}_t} r(i)\varphi(i,\tilde{S}_t|\theta^*) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t)}_{(\diamondsuit_0)}.
\end{aligned}$$

To bound $(\heartsuit_0)$**:** Recall that $\textsc{Opt}(\hat{\theta}) = \sum_{S\in\mathbb{S}}\sum_{i\in S} r(i)\varphi(i,S|\hat{\theta})\hat{y}(S)$. By the definition of the sampling procedure, $\mathbb{P}[\tilde{S}_t = S] = \hat{y}(S)$. Similar to the case in $(\heartsuit_k)$, for any fixed $\hat{\theta}$ the random variables

$$\left\{\sum_{S\in\mathbb{S}}\sum_{i\in S} r(i)\varphi(i,S|\hat{\theta})\hat{y}(S) - \sum_{i\in\tilde{S}_t} r(i)\varphi(i,\tilde{S}_t|\hat{\theta})\right\}_{t=\tau+1}^{T-\rho}$$

are i.i.d., mean 0, and lie in the interval $[-1, 1]$. By Chernoff inequality, we have

$$\mathbb{P}\left[(\heartsuit_0) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{4(K+1)}.$$

To bound ($\clubsuit_0$)**:** Recall the assumption that $r(i) \in [0, 1]$, and $|\tilde{S}_t| \leq C$ for all t. Conditional on $\mathcal{E}_{\hat{\theta}}$, we have $\left|\hat{\theta}(i) - \theta^*(i)\right| \leq \epsilon(\tau)$. By applying Lemma 2.5.6, this implies that for all i, S, we have

$$\sum_{i \in S} r(i)\varphi(i, S|\hat{\theta}) - \sum_{i \in S} r(i)\varphi(i, S|\theta^*) \leq C\epsilon(\tau).$$

Thus, we have $(\clubsuit_0) \leq TC\epsilon(\tau)$.

To bound ($\diamondsuit_0$)**:** For any choice of $\hat{\theta}$, the sequence $\{r(i)\varphi(i, \tilde{S}_t|\theta^*) - r(\tilde{\Sigma}_t)\}_{t=\tau+1}^{T-\rho}$ of random variables are i.i.d.. Indeed, each assortment $\tilde{S}_t$ is independently sampled based on the distribution specified by $\hat{y}$ (an optimal solution to CDLP-P($\hat{\theta}$)), and each $\tilde{\Sigma}_t$ is sampled from $\tilde{S}_t \cup \{0\}$ according to the choice probability $\varphi(\cdot, \tilde{S}_t|\theta^*)$, independently of other $(\tilde{S}_{t'}, \tilde{\Sigma}_{t'})_{t' \neq t}$. Moreover, for any choice of $\hat{\theta}$, each of the random variables $r(i)\varphi(i, \tilde{S}_t|\theta^*) - r(\tilde{\Sigma}_t)$ has mean zero, and lies in the range $[-1, 1]$. By Chernoff inequality, we have

$$\mathbb{P}\left[\sum_{t=\tau+1}^{T-\rho} \sum_{i \in S_t} r(i)\varphi(i, \tilde{S}_t|\theta^*) - \sum_{t=\tau+1}^{T-\rho} r(\tilde{\Sigma}_t) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}}\right] \geq 1 - \frac{\delta}{4(K+1)}$$

for any choice of $\hat{\theta}$. In particular, this is true when we condition that $\hat{\theta}$ satisfies the confidence bound stated in $\mathcal{E}_{\hat{\theta}}$. By conditioning on $\mathcal{E}_{\hat{\theta}}$, this proves that

$$\mathbb{P}\left[(\diamondsuit_0) \leq \sqrt{2T \log \frac{4(K+1)}{\delta}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{4(K+1)}.$$

Altogether, we have shown the following:

$$\mathbb{P}\left[\textsc{Regret} \leq TC\epsilon(\tau) + 2\sqrt{2T\log\frac{4(K+1)}{\delta}} \mid \mathcal{E}_{\hat{\theta}}\right] \geq 1 - \frac{\delta}{2(K+1)}.$$

Finally, combining with (A.36), conditional on $\mathcal{E}_{\hat{\theta}}$ the bound

$$\begin{aligned}
&T\textsc{Opt}(\theta^*) - \sum_{t=r+1}^{T-\rho} r(\tilde{\Sigma}_t) \\
\leq&\rho + \tau + (T-\rho-\tau)C\epsilon(\tau)\left(\frac{\textsc{Opt}(\theta^*)}{\min_{k\in\mathcal{K}}\{c(k)\}} + 1\right) + TC\epsilon(\tau) + 2\sqrt{2T\log\frac{4(K+1)}{\delta}} \\
\leq&\rho + \tau + TC\epsilon(\tau)\left(\frac{1}{\min_{k\in\mathcal{K}}\{c(k)\}} + 2\right) + 2\sqrt{2T\log\frac{4(K+1)}{\delta}} \\
=&\tau + TC\epsilon(\tau)\left(\frac{2}{\min_{k\in\mathcal{K}}\{c(k)\}} + 2\right) + \left(2 + \frac{1}{\min_{k\in\mathcal{K}} c(k)}\right)\sqrt{2T\log\frac{4(K+1)}{\delta}}. \qquad \text{(A.37)}
\end{aligned}$$

holds with probability $1 - \delta/2(K+1)$. The step (A.37) is by the definition of ρ in (2.22).

Finally, by optimizing the choice of τ and assuming the time horizon is sufficiently long, the bound (A.37) is equal to

$$N^{1/3}(4TCe^R)^{2/3}\left(\frac{2}{\min_{k\in\mathcal{K}}\{c(k)\}} + 3\right) + \left(2 + \frac{1}{\min_{k\in\mathcal{K}} c(k)}\right)\sqrt{2T\log\frac{4(K+1)}{\delta}},$$

which demonstrate the $\tilde{O}\left(C^{2/3}N^{1/3}e^{2R/3}T^{2/3}\right)$ order bound. In particular, this choice of τ optimizes the order bound in (A.37).

Another natural choice of τ, which is also the choice in our numerical simulation, is $\tau = T^{2/3}$. With this τ, the bound (A.37) is equal to

$$T^{2/3}\left(1 + 4\left(2 + \frac{1}{\min_{k\in\mathcal{K}} c(k)}\right)Ce^R\sqrt{N\log\frac{4N}{\delta}}\right) + \left(2 + \frac{1}{\min_{k\in\mathcal{K}} c(k)}\right)\sqrt{2T\log\frac{4(K+1)}{\delta}},$$

which establishes a slightly worse regret order bound $\tilde{O}(C\sqrt{N}e^RT^{2/3})$.

A.14 Details on Procedure 8

Procedure 8 is displayed below:

Procedure 8 Generating $\Lambda = (r, (\varphi(\cdot,\cdot|\theta^*), \mathbb{S}), A, c)$, given $\Gamma = (\mathbb{S}, N, K, R)$

1: Sample the revenue coefficients r uniformly at random from $[0.3, 1]^N$.
2: Sample the utility parameter θ^* uniformly at random from $[-R, R]^N$.
3: Generate the resource consumption matrix $A = \{a(i,k)\}_{i\in\mathcal{N},k\in\mathcal{K}} \in \{0,1\}^{N\times K}$ as follows:
   (i) Generate the matrix $A_1 \in \{0,1\}^{N\times K}$, where each of the N rows in A_1 is independently drawn from the K standard basis vectors in $\mathbb{R}^K$.
   (ii) Generate A_2, where $A_2(i,k) \sim \text{Bern}(0.35)$ for each i, k.
   (iii) Define $A(i,k) = \min\{A_1(i,k) + A_2(i,k), 1\}$ for each i, k.
4: Generate the per period capacity $\{c(k)\}_{k\in\mathcal{K}} \in [0,1]^K$ in the following two steps:
   (i) Solve SPP-$(r, \varphi(\cdot,\cdot|\theta^*), \mathbb{S})$ for an optimal assortment S^*.
   (ii) Let $c(k) = 0.75\sum_{i\in\mathcal{N}} a(i,k)\varphi(i, S^*)$.
5: If $c(k) \geq 0.1$ for all $k \in \mathcal{K}$, output $(r, (\varphi(\cdot,\cdot|\theta^*), \mathbb{S}), A, c)$. Otherwise, go back to step 1.

The choices of A and c in Lines 3, 4 closely follows [Feldman and Topaloglu, 2014]. Line 3 ensures that each product consumes at least one type of resources, and on average consumes roughly $0.35K$ types of resources. The former property helps to avoid trivial instances where the monopolist can earn a revenue without consuming any resource. Line 4 involves scaling down the capacities, which ensures that the resource constraints are active. Line 5 ensures that the optimal revenue $\text{OPT}(\theta^*)$ per period is not too small.

A.15 Enhancing the FEASIBILITY(Z, ϵ)

When PB is a polynomial time algorithm, it requires invoking an α-approximate oracle for $\Theta(1/\epsilon^2)$ times in order to (essentially) achieve a $(\alpha + \epsilon)$-approximation. In this section, we provide an enhanced version, namely FEASIBILITY'$(\epsilon, Z, \mathcal{T})$, of the Potential Based subroutine FEASIBILITY(ϵ, Z). This gives rise to an enhanced PB, which consists of the binary

search subroutine BINSEARCH calling FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ instead of FEASIBILITY(ϵ, Z). We modify in three parts to construct FEASIBILITY'$(\epsilon, Z, \mathcal{T})$, which requires substantially less oracle calls. The three modifications are:

1. We introduce two **if** loops that allow FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ to quit earlier than iteration $\mathcal{T}$:

   (a) In the first **if** loop (Line 9 to Line 14 in Algorithm 9), we solve a restricted CDLP for a solution $(\tilde{y}_\tau, \tilde{z}_\tau)$. If the return value is at least Z, then we break the while loop and return the solution $(\tilde{y}_\tau, \tilde{z}_\tau)$. Otherwise, we proceed to iteration $\tau + 1$.

   (b) For the second **if** loop (Line 18 to Line 20 in Algorithm 9), we evaluate the potential function $\Omega(\tau)$. If $\Omega(\tau) > \Omega(\tau + 1)$, then we break the while loop and return $(\tilde{y}, \tilde{z}) = (0, 0)$ as the solution. Otherwise, we proceed to iteration τ.

2. Instead of defining the maximum number of iterations $\mathcal{T}$ to be $\Theta(1/\epsilon^2)$ in Line 1 in Algorithm 2, we are allowed to set $\mathcal{T}$ to be any number. (We will choose $\mathcal{T} = 200$ in our simulation).

The full pseudocode for the enhanced subroutine, namely FEASIBILITY'$(\epsilon, Z, \mathcal{T})$, is displayed in Algorithm 9. Modification 1 (a) allows FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ to possibly terminate earlier when the target value Z is small. Indeed, if Z is small, then FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ might have identify the support of a solution that has value at least Z. Thus, by solving the CDLP restricted to the identified assortments $S_1, \cdots, S_\tau$, FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ could have terminated at iteration τ (Line 12), instead of iteration $\mathcal{T}$. Note that the solving of restricted CDLP-P is reminiscent to CG and ACG, but in FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ the 'column' (or more precisely the new assortment) entering the restricted CDLP-P is defined differently from ACG or CG.

Algorithm 9 Enhanced Potential Based subroutine FEASIBILITY'$(\epsilon, Z, \mathcal{T})$

1: Initialize $\mathbb{S}_1 = \{\{i\}\}_{i\in\mathcal{N}} \cup \{\{\mathtt{f}\}\}_{\mathtt{f}\in\mathcal{F}} \cup \{\emptyset\}$, which is the collection of single item assortments and the empty assortment.

2: Start with $\tau = 1$.

3: **while** $\tau \leq \mathcal{T}$ **do**

4: For each $i \in \mathcal{N}$, compute its reduced revenue, as defined in Line 2 in Algorithm 2.

5: For each $\mathtt{f} \in \mathcal{F}$, select $I_\tau(\mathtt{f}) \in \mathtt{f}$ in the same way as Line 3 in Algorithm 2.

6: For each $\mathtt{f} \in \mathcal{F}$, compute its reduced revenue, as defined in Line 4 in Algorithm 2.

7: Apply the approximate oracle $\mathcal{A}$ on SPP$(\tilde{r}_\tau, \varphi, \mathbb{S})$, which returns an assortment $S_\tau \in \mathbb{S}$.

8: Define $\{v_\tau(S)\}_{S\in\mathbb{S}} \cup \{\zeta_\tau(\mathtt{f}, i)\}_{\mathtt{f}\in\mathcal{F}, i\in\mathtt{f}}$ in the same way as Line 6 in Algorithm 2.

9: **if** $S_\tau \in \mathbb{S} \setminus \mathbb{S}_\tau$ **then**

10: Solve CDLP$[\mathbb{S}_\tau \cup \{S_\tau\}]$ for a solution $(\tilde{y}_\tau, \tilde{z}_\tau)$.

11: **if** $\mathrm{Val}(\tilde{y}_\tau) \geq Z$ **then**

12: Break the While Loop, and **return** $(\tilde{y}, \tilde{z}) = (\tilde{y}_\tau, \tilde{z}_\tau)$.

13: **end if**

14: **end if**

15: $\mathbb{S}_{\tau+1} = \mathbb{S}_\tau \cup \{S_\tau\}$.

16: Update the variables $\Xi(\tau)$, $V(\tau)$, $\Psi_k(\tau)$, $U_k(\tau)$ in (2.7), (2.9), (2.8), (2.10) respectively.

17: Compute $\Omega(\tau) = \Xi(\tau) + \sum_{k=1}^{K} \Psi_k(\tau)$, as defined in Lemma 2.4.2.

18: **if** $\Omega(\tau) > \Omega(\tau - 1)$ **then**

19: Break the While loop, and **return** $(\tilde{y}, \tilde{z}) = (0, 0)$.

20: **end if**

21: $\tau \leftarrow \tau + 1$.

22: **end while**

23: **return** the solution $(\tilde{y}, \tilde{z})$ as defined in Line 9 in Algorithm 2.

Modification 1 (b) allows FEASIBILITY'$(\epsilon, Z, \mathcal{T})$ to possibly terminate earlier when the target value Z is too large. Recall that By Lemma 2.4.2, when $Z \leq \alpha \text{Val}(y^*)$, we have $\Omega(0) \geq \Omega(1) \geq \cdots, \Omega(T)$. Moreover, the aim of the binary search algorithm (Algorithm 3) is to identify a target value Z such that $Z \leq \alpha \text{Val}(y^*)$ but $\alpha \text{Val}(y^*) - Z \leq \delta$. Thus, if we discover that $\Omega(\tau) > \Omega(\tau - 1)$ for some iteration τ, it is a certification that the current target value Z is larger than $\alpha \text{Val}(y^*)$. This implies that the binary search algorithm should disregard the current target value Z, and perform the PB subroutine on smaller target values instead. Therefore, the modification allows us to identify the case $Z > \alpha \text{Val}(y^*)$ in some iteration τ by considering Ω, instead of running all $\mathcal{T}$ iterations and checking if the final solution is $(1 - 2\epsilon)Z$ successful.

Modification 2 is essentially an empirical improvement; for a fixed choice of (ϵ, δ), the decision maker could run PB on random instances, and then determine an upper bound on $\mathcal{T}$ such that either a break by the first **if** loop or the second **if** loop occurs.

A.16 Simulation Results for Markov Chain based Choice Models

After enhancing the Potential Based subroutine, we compare the running time and performance of ACG, CG and the enhanced PB on randomly generated problem instances. For clarity sake, we assume that $\mathcal{F} = \emptyset$. We evaluate the algorithms on the CDLP-Ps for the Markov Chain based Choice Model (MC choice model) with cardinality constrained assortment family $\mathbb{S}(C) = \{S \subset \mathcal{N} : |S| \leq C\}$.

An MC choice model is specified by a Markov Chain (λ, ρ), where $\lambda \in \mathbb{R}_{\geq 0}^{N+1}$ is the initial probability distribution and $\rho \in \mathbb{R}_{\geq 0}^{(N+1)\times(N+1)}$ is the transition matrix. The choice behavior of a customer under the choice model is as follows. Suppose a customer is offered with an assortment $S \subset \mathcal{N}$. Initially, the customer prefers to purchase product $i_0 \in \mathcal{N} \cup \{0\}$ with probability $\lambda(i_0)$. If $i_0 \in S \cup \{0\}$, she will purchase i_0; otherwise, her preference will

transit from i_0 to $i_1 \in \mathcal{N} \cup \{0\}$ with probability $\rho(i_0, i_1)$. Such transition goes on until the customer's preferred products is in $S \cup \{0\}$; in a certain iteration τ, suppose the customer prefers to purchase product $i_\tau \in \mathcal{N} \cup \{0\}$. If $i_\tau \in S \cup \{0\}$, she will purchase i_τ; otherwise, her preference will transit from i_τ to $i_{\tau+1} \in \mathcal{N} \cup \{0\}$ with probability $\rho(i_\tau, i_{\tau+1})$. In a nutshell, the customer's path of preferences $i_0, i_1, i_2, \cdots$ is distributed according to the Markov Chain (λ, ρ), where the path stops when it hits a state in $S \cup \{0\}$. For more details on the MC model and its relationship with other common choice models, the readers are welcome to consult [Blanchet et al., 2013].

The generation procedure is time efficient, since the SPP in Line 5 is polynomial time solvable, by [Blanchet et al., 2013, Feldman and Topaloglu, 2015]. The way we generate C in Line 6 ensures that the cardinality constraint $|S| \leq C$ will be active when SPPs are solved in the proposed algorithms, and we require a lower bound on C to ensure that there are sufficiently many possible assortments.

A.16.1 Generating a random instance

The generation of a random instance for CDLP-P for the MC Model is similar to Appendix A.14. Suppose we are given the *class instance tuple* (LB, N, K), where LB is the lower bound on the maximum cardinality of the assortment family, N is the number of products, and K is the number of constraints. We propose Procedure 10, displayed in 10, to generate a *random output tuple* $(\lambda, \rho, \mathbb{S}(C), r, A, c)$. (λ, ρ) is the Markov Chain associated with the model. C is the maximum cardinality of an assortment in the assortment family, and $C \geq \text{LB}$. r is the revenue coefficient, A is the resource consumption matrix, and c is the per period capacity of the K resources, where $\min_{k=1,\cdots K} c_k \geq 0.1$.

Procedure 10 Generating $(\lambda, \rho, \mathbb{S}(C), r, A, c)$, given $\Gamma = (\text{LB}, N, K)$

1: Sample λ uniformly from the $(N+2)$-dimensional simplex.

2: Generate the transition matrix ρ as follows:

(i) Generate a random matrix $\rho_0 \in [0,1]^{N\times N}$, where each $\rho_0(i,j)$ is independently and uniformly sampled from $[0,1]$.

(ii) Define $\rho_1 \in [0,1]^{N\times N}$ by $\rho_1(i,j) = \frac{\rho_0(i,j)}{\sum_{\ell=1}^{N}\rho_0(i,\ell)}$.

(iii) Define $\rho \in [0,1]^{(N+1)\times(N+1)}$, by having $\rho(i,j) = 0.85\rho_1(i,j)$ if $1 \le i,j \le N$, $\rho(i,0) = 0.15$ if $1 \le i \le N$, $\rho(0,0) = 1$, and $\rho(0,j) = 0$ if $1 \le j \le N$.

3: Generate the resource consumption matrix $A = \{a(i,k)\}_{i\in\mathcal{N},k\in\mathcal{K}} \in \{0,1\}^{N\times K}$ as follows:

(i) Generate the matrix $A_1 \in \{0,1\}^{N\times K}$, where each of the N rows in A_1 is independently drawn from the K standard basis vectors in $\mathbb{R}^K$.

(ii) Generate A_2, where $A_2(i,k) \sim \text{Bern}(0.35)$ for each i,k.

(iii) Define $A(i,k) = \min\{A_1(i,k) + A_2(i,k), 1\}$ for each i,k.

4: Sample the revenue coefficients r uniformly at random from $[0.3,1]^N$.

5: Generate the per period capacity $\{c(k)\}_{k\in\mathcal{K}} \in [0,1]^K$ in the following two steps:

(i) Solve SPP-$(r, \varphi(\cdot,\cdot|\lambda,\rho), 2^{\mathcal{N}})$ for an optimal assortment S^*.

(ii) Let $c(k) = 0.75\sum_{i\in\mathcal{N}} a(i,k)\varphi(i,S^*)$.

6: Sample the maximum cardinality C uniformly from $\left\{\lfloor\frac{|S^*|}{3}\rfloor, \lfloor\frac{|S^*|}{3}\rfloor + 1, \cdots, \lceil\frac{2|S^*|}{3}\rceil\right\}$.

7: If $c(k) \ge 0.1$ for all $k \in \mathcal{K}$ and $C \ge \text{LB}$, output $(r, (\varphi(\cdot,\cdot|\theta^*), \mathbb{S}), A, c)$. Otherwise, go back to Line 5.

A.16.2 Simulation Setup

We apply CG, ACG($\mathbb{S}_1$) and the enhanced PB on random instances generated based on the instance class tuple $\Gamma_1 = (\text{LB} = 10, N = 50, K = 30)$. The generation of random instances

is in accordance with Procedure 10. Since the conventional CG usually does not terminate within 2 hours on instances with 50 products, we only apply ACG($\mathbb{S}_1$) and the enhanced PB on random instances generated based on the class instance tuples

$$\Gamma_2 = (\text{LB} = 30, N = 150, K = 80), \Gamma_3 = (80, 350, 180), \Gamma_4 = (100, 750, 250)$$

by Procedure 10.

For ACG($\mathbb{S}_1$), we define the initial collection of columns as $\mathbb{S}_1 = \{\{i\}\}_{i \in \mathcal{N}} \cup \{\emptyset\}$. Thus, ACG is warm-started in the same way as the enhanced PB. For the enhanced PB, we fix $\mathcal{T} = 200$, $\epsilon = 0.1$ and $\delta = 0.05$.

A.16.3 Simulation Results for Moderate Size Instances

Table A.1 displays the simulation results of CG, ACG($\mathbb{S}_1$) and the enhanced PB on 8 random instances based on Γ_1:

	Running time (seconds)				Objective Value			
Instance	PB	ACG meets PB	ACG	CG	PB	ACG meets PB	ACG	CG
1	2.261	**0.227**	0.449	9.551	0.625	0.630	0.641	**0.657**
2	11.320	**0.227**	7.319	> 2 hrs	0.550	0.556	0.644	**0.647**
3	14.433	**0.331**	6.536	> 2 hrs	0.543	0.554	**0.594**	0.593
4	18.686	**1.757**	5.427	> 2 hrs	0.572	0.572	0.589	**0.592**
5	9.974	**0.267**	3.922	> 2 hrs	0.513	0.520	0.558	**0.559**
6	9.299	**0.281**	5.374	> 2 hrs	0.536	0.541	0.579	**0.583**
7	13.020	**0.363**	6.903	> 2 hrs	0.505	0.505	0.541	**0.541**
8	7.656	**0.254**	10.003	13.670	0.601	0.602	0.646	**0.660**

Table A.1: Simulations Results for the case of 50 products.

The first four columns display the running time (in seconds) of the algorithms. The first column shows the time taken for the enhanced PB to terminate. The second column shows the time taken for ACG to achieve at least the value returned by the enhanced PB.

Thus, these two columns compares the running time of the enhanced PB and ACG in a fair manner, in the sense that we compare the time taken for each of the algorithms to achieve a common value. The third column is the time taken for ACG to terminate. The fourth column is the time taken for the conventional CG algorithm to terminate. In six out of the eight instances, CG does not terminate within 2 hours, so we stop CG at the 2 hours time limit and return the current solutions.

The next four columns display the objective values achieved by the algorithms. The fifth column shows the objective value achieved by the enhanced PB. The sixth column shows the objective value achieved by ACG during the iteration when ACG first achieves an objective value larger than or equal to the objective value returned by the enhanced PB. Thus, the values in the sixth column are a little higher than their counterparts in the fifth column. The seventh column shows the objective value achieved by ACG when it terminates. Again, these values are higher than the respective values in the fifth and sixth columns. Finally, the eighth column shows the objective value achieved by CG. If CG terminates within 2 hours, it shows the objective value at termination. Otherwise, it shows the objective value achieve by CG when the 2 hour benchmark has been reached.

It is demonstrated that ACG terminates very quickly, and returns high quality solutions. The enhanced PB is slower than ACG, but the former is still much faster than the conventional CG in most cases. For six out of the eight random instances, CG does not terminate within 2 hours (although in Instances 1 and 8, CG terminates quite quickly). Moreover, for Instances $2, \cdots, 7$, the objective values returned by CG after 2 hours of computation does not improves substantially upon the objective values returned by ACG. Somewhat surprisingly, in Instance 4, the objective values returned by ACG at termination (which takes 1.8 seconds) could even be higher than the objective value by CG (when is forced to stop at the 2 hour time limit). Altogether, this shows that ACG not only achieves time efficiency, but also achieves high objective values in moderate size problem instances. Finally, it is interesting to note that ACG appears to always return a solution with a higher

objective value than the enhanced PB in our simulation, when ACG is allowed to run until termination.

A.16.4 Simulation Results on Large Instances

Next, we evaluate the performance of ACG and the enhanced PB on larger instances. Note that CG is not performed, since the CG subproblems, which involves solving integer programs with at least 200 variables, appear to be very computationally intensive. The difficulty in solving the integer programs for large instances of MC choice models with cardinality constrained assortment family is also reported in [Désir et al., 2015].

Table A.2 displays the simulation results of ACG and the enhanced PB applied on 8 random instances generated based on the tuple $\Gamma_2 = (\text{LB} = 30, N = 150, K = 80)$.

	Running time (seconds)		Objective Value		
Instance	PB	ACG meets PB	PB	ACG meets PB	ACG after 20
1	40	**9**	0.527	0.529	0.569
2	19	**8**	0.549	0.552	0.575
3	164	**11**	0.556	0.565	0.605
4	375	**15**	0.525	0.527	0.570
5	**9**	13	0.550	0.557	0.591
6	128	**16**	0.578	0.578	0.596
7	34	**9**	0.536	0.540	0.569
8	91	**12**	0.527	0.528	0.570

Table A.2: Simulations Results for the case of 150 products.

Column 1,2 concerns the running time of the algorithms. Column 1 shows the time taken for the enhanced PB to output a solution, and Column 2 shows the time taken for ACG to obtain a solution with value at least the value of the enhanced PB's solution. Column 3, 4, 5 concerns the objective values obtained by the algorithms. Column 3 shows the objective values achieved by the enhanced PB. Column 4 shows the objective values achieved by ACG during the iteration when ACG first achieves an objective value larger than or equal to the objective value by PB. Finally, Column 5 shows the objective value by

ACG after 20 additional iterations.

While the running time of both algorithms increases significantly, ACG is still much more efficient than the enhanced PB. The running time of the enhanced PB is rather uneven; while it often has a longer running time than ACG, there is an instance when the enhance PB is a little faster than ACG.

Next, Table A.3 displays the simulation results of ACG and the enhanced PB applied on 8 random instances generated based on Γ_3. While ACG is in general faster than the

	Running time (seconds)		Objective Value		
Instance	PB	ACG meets PB	PB	ACG meets PB	ACG after 20
1	606	**296**	0.573	0.576	0.585
2	543	**201**	0.528	0.531	0.582
3	**69**	217	0.528	0.552	0.576
4	595	**266**	0.552	0.557	0.591
5	**72**	187	0.551	0.556	0.577
6	**69**	205	0.576	0.577	0.592
7	3262	**422**	0.528	0.533	0.580
8	**86**	168	0.550	0.552	0.575

Table A.3: Simulations Results for the case of 350 products.

enhanced PB, the gap in running time between the two algorithms starts to narrow.

Finally, Table A.4 displays the simulation results of ACG and the enhanced PB applied on 8 random instances generated based on the tuple Γ_4, which are rather large instances consisting of 750 products. Compared to the case of Γ_3, where $N = 350$, the running time of both ACG and the enhanced PB increases significantly. However, there are instances for which the enhanced PB performs significantly faster than ACG. Nevertheless, there are instances (notably 5, 6) for which the enhanced PB has a very long run time (For instance 5, it takes 16 hours!).

Overall, we witness that ACG is very time efficient, and it has very strong performance in terms of the achieved objective values. However, the enhanced PB could be useful when the problem instance is very large, and sometimes the enhanced PB can deliver a near

	Running time (seconds)		Objective Value		
Instance	PB	ACG meets PB	PB	ACG meets PB	ACG after 20
1	**464**	6282	0.575	0.577	0.560
2	13385	**6010**	0.550	0.555	0.585
3	16004	**6038**	0.551	0.553	0.580
4	**573**	6306	0.574	0.574	0.590
5	57652	**7688**	0.551	0.554	0.578
6	28412	**7195**	0.528	0.543	0.588
7	**455**	4671	0.574	0.579	0.600
8	**506**	5303	0.575	0.577	0.591

Table A.4: Simulations Results for the case of 750 products.

optimal solution very efficiently.

The running time of PB seems to vary a lot on different random instances with the same number of products, as demonstrated in the 4 tables. It could be due to our three modifications on PB, which offer the opportunity for the enhanced PB to terminate early sometimes, but not always. The conditions for the enhanced PB to have a fast empirical running time is not a prior clear though. Comparatively, the running time for ACG to achieves the objective value obtained by PB is rather uniform. It is an interesting question to design a heuristic that achieves best of two worlds in ACG and the enhanced PB.

Appendix B

Technical Results for Chapter 3

B.1 An example of Φ satisfying the properties in Section 3.3.2

Consider a price set $\mathcal{P} = \{1, 2\}$ and two demand functions $d_1(1) = 0.6, d_1(2) = 0.25; d_2(1) = 0.4, d_2(2) = 0.3$. Demand per period has a Bernoulli distribution. It is readily verified that this example satisfies the four properties of (Γ) for Theorem 3.3.5.

B.2 Proof of Lemma 3.3.4

Let $h_t = (p_1, x_1, \ldots, p_t, x_t)$ be a realization of $H_t = (P_1, X_1, \cdots, P_t, X_t)$. We first assume $\mathbb{P}_i^\pi(H_t = h_t) > 0$, so we have

$$\begin{aligned}
\mathbb{P}_i^\pi(H_t = h_t) &= \prod_{s=1}^{t} \mathbb{P}_i^\pi(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \\
&= \prod_{s=1}^{t} \left(\mathbb{P}_{i'}^\pi(D(p_s) = x_s) \cdot \frac{\mathbb{P}_i^\pi(D(p_s) = x_s)}{\mathbb{P}_{i'}^\pi(D(p_s) = x_s)} \right) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad \text{(B.1)} \\
&\geq \prod_{s=1}^{t} (\mathbb{P}_{i'}^\pi(D(p_s) = x_s) \cdot \kappa_\Gamma) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad \text{(B.2)}
\end{aligned}$$

$$
\begin{aligned}
&= \kappa_\Gamma^t \prod_{s=1}^{t} \mathbb{P}_{i'}^{\pi}(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_i^{\pi}(P_{s+1} = p_{s+1} \mid H_s = h_s) \\
&= \kappa_\Gamma^t \prod_{s=1}^{t} \mathbb{P}_{i'}^{\pi}(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_{i'}^{\pi}(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad \text{(B.3)} \\
&= \kappa_\Gamma^t \mathbb{P}_{i'}^{\pi}(H_t = h_t).
\end{aligned}
$$

Step (B.1) uses the third condition of (Γ), which states that all demand functions have the same support under a given price, so $\mathbb{P}_{i'}^{\pi}(D(p_s) = x_s) \neq 0$. Step (B.2) uses the fourth condition of (Γ). Step (B.3) holds because price P_{s+1} is determined by policy π and realized history h_s, and is independent of the underlying demand model. Note that if π is a deterministic policy, we always have $\mathbb{P}_i^{\pi}(P_{s+1} = p_{s+1} \mid H_s = h_s) = 1$ for all i.

Finally, if $\mathbb{P}_i^{\pi}(H_t = h_t) = 0$, we have $\mathbb{P}_{i'}^{\pi}(H_t = h_t) = 0$, too. This is again due to the third condition of (Γ), which states that all demand functions have the same support under a given price. □

Appendix C

Technical Results for Chapter 4

C.1 Proof of Theorem 4.5.1

Proof. Proof of Theorem 4.5.1 We will first prove for the right derivatives. By triangle inequality, we know that

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

First, we analyse the first term (a):

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

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

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

Next, we analyze the second term (b). As remarked previously, the empirical right derivative $\hat{V}_{t+1}^r$ 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

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

where $\beta_1 < \cdots < \beta_{M-1}$ are the breakpoints of $\hat{V}_{t+1}^r(y)$, 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:

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

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

bound:

$$\mathbb{P}\left[(\mathrm{b}) \leq \frac{\alpha_t}{2}\right] \geq \mathbb{P}\left[\sup_{y\in\mathbb{R}}\left|\frac{1}{N_t}\sum_{i=1}^{N_t} 1(d_t^i \leq y) - \mathbb{P}[D_t \leq y]\right| \leq \frac{\alpha_t}{2\left(\sum_{s=t+1}^{T} h_s + b_s\right)}\right]$$
$$\geq 1 - \exp\left[-\frac{\alpha_t^2}{4\left(\sum_{s=t+1}^{T} h_s + b_s\right)^2} N_t\right] \geq 1 - \frac{\delta_t}{4},$$

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 $\left|\mathbb{E}\hat{V}_{t+1}^{\mathrm{r}}(y_t - D_t) - \mathbb{E}V_{t+1}^{\mathrm{r}}(y_t - D_t)\right| \leq \gamma_t$ with probability 1.

Altogether, we have the following guarantee for the right derivatives:

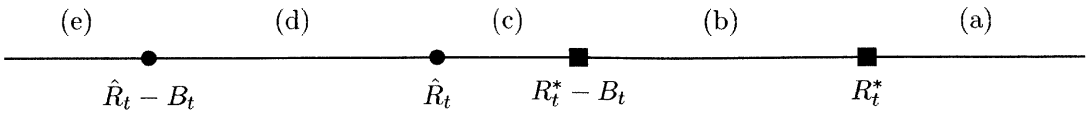
$$\mathbb{P}\left[\text{For all } y_t,\ \left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \leq \gamma_t + \alpha_t\right]$$
$$\geq \mathbb{P}\left[\text{For all } y_t,\ (\mathrm{a}) \leq \frac{\alpha_t}{2},\ (\mathrm{b}) \leq \frac{\alpha_t}{2},\ (\mathrm{c}) \leq \gamma_t\right] \geq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

Finally, by an almost identical analysis we also know that similar guarantee holds for the left derivatives, which proves the theorem. □

C.2 Proof of Theorem 4.5.3

Proof. Proof of Theorem 4.5.3 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 $U_t^{\mathrm{r}}(y_t) \geq 0$ for all $y_t \geq R_t^*$. 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 (4.8), (4.9). Consider the following 4 cases on $R_t^*, \hat{R}_t$:

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



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

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

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

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

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

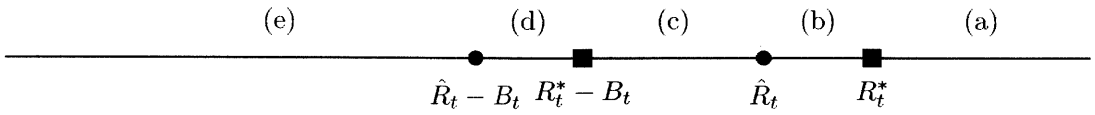
$$\begin{aligned}&\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| = \hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t) \leq \hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t)\\ \leq&\left|\hat{U}_t^{\mathrm{r}}(x_t + B_t) - U_t^{\mathrm{r}}(x_t + B_t)\right| \leq \eta_t.\end{aligned}$$

(e) We have $x_t \leq \hat{R}_t - B_t$. 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

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

2. **Case 2:** We have $R_t^* - B_t < \hat{R}_t \leq R_t^*$. 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 $\hat{R}_t \leq x_t < R_t^*$. Then $V_t^{\mathrm{r}}(x_t) = 0 \geq U_t^{\mathrm{r}}(x_t)$, and $\hat{V}_t^{\mathrm{r}}(x_t) = \hat{U}_t^{\mathrm{r}}(x_t) \geq 0$. Therefore

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

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

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

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

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

3. **Case 3:** We have $\hat{R}_t - B_t \leq R_t \leq \hat{R}_t$. 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. □

C.3 Proof of Corollary 4.5.4

Proof. Proof of Corollary 4.5.4 For a period t, we say that $\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t}^{T}$ is $\alpha-$good if for all $s \in \{t, \cdots, T\}$, the followings hold simultaneously:

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

Now, we have

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

By Theorem 4.5.1, conditioned on the event that $\left\{\hat{U}_s^{\mathrm{r}}, \hat{V}_s^{\mathrm{r}}\right\}_{s=t+1}^{T}$ is $\alpha-$good, we have

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

Next, by Theorem 4.5.3, we know that if $\left|\hat{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right| \leq \sum_{s=t}^{T} \alpha_s$ holds for all $y_t \in \mathbb{R}$, then $\left|\hat{V}_t^{\mathrm{r}}(x_t) - V_t^{\mathrm{r}}(x_t)\right| \leq \sum_{s=t}^{T} \alpha_s$ holds for all $x_t \in \mathbb{R}$ with probability 1. Altogether, this

shows that for all t

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

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

C.4 Proof of Claim 4.5.5

Proof. Proof of Claim 4.5.5 We would like to apply Lemma 4.5.6 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 4.5.6 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. □

C.5 Proof of Lemma 4.5.7

Proof. Proof of Lemma 4.5.7 First, by the definition of a modified base stock policy, the following equation holds:

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

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:

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

Now, suppose (C.2) holds for period $t+1$, and we will prove that (C.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 (C.1), the following inequality holds:

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

To prove the induction claim (C.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

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

where $(\dagger)$ 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 (C.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,

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

where in (Dagger) 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 (C.2) is still true.

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

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

C.6 Proof of Theorem 4.3.1

Proof. Proof of Theorem 4.3.1 For each $t = 1, \cdots, T$, let

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

By applying this value of α_t in Corollary 4.5.4, we know that the following holds:

$$\begin{aligned}
&\mathbb{P}\left[(\hat{R}_1, \cdots, \hat{R}_T) \text{ is a } (1+\epsilon)\text{-optimal modified base stock policy}\right] \\
\geq &\mathbb{P}\left[\text{For all } t,\, U_t(\hat{R}_t) \leq \left(1 + \frac{\epsilon}{2T}\right) U_t(R_t^*)\right] && \text{(C.4)} \\
\geq &\mathbb{P}\left[\text{For all } t,\, U_t(\hat{R}_t) \leq \left(1 + \frac{3\sum_{s=t}^{T} \alpha_s}{\min\{b_t, h_t\}}\right) U_t(R_t^*)\right] \\
\geq &\mathbb{P}\left[\text{For all } t \text{ and } y,\, \left|U_t^{\mathrm{r}}(y) - \hat{U}_t^{\mathrm{r}}(y)\right| \leq \sum_{s=t}^{T} \alpha_s\right] && \text{(C.5)} \\
\geq &1 - \delta. && \text{(C.6)}
\end{aligned}$$

Step (C.4) is by Lemma 4.5.7, step (C.5) is by Claim 4.5.5, and step (C.6) is by the derivation in the previous subsection. Finally, applying our chosen value of α_t in Corollary 4.5.4, we obtain the number of samples needed, as stated in Theorem 4.3.1. □

C.7 Proof of Lemma 4.6.1

Proof. Proof of 4.6.1 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 $\tilde{V}_{T+1}^{\mathrm{r}} = 0$. Now, suppose that it is true for $\tilde{V}_{t+1}^{\mathrm{r}}$, we will show that the induction claim is also true for $\tilde{U}_t^{\mathrm{r}}$.

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

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

where $\beta_1 < \cdots < \beta_{M-1}$ are the breakpoints of $\tilde{V}_{t+1}^{\mathrm{r}}(x)$. 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 $\tilde{U}_t^{\mathrm{r}}$ is defined as follows:

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

This implies that $\tilde{U}_t^{\mathrm{r}}(y_t)$ has breakpoints $\{d_i^t\}_{i=1}^{N_t} \cup \{d_i^t + \beta_j\}_{i=1,j=1}^{N_t,M-1}$. 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 $\tilde{U}_t^{\mathrm{r}}$.

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

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

where $p_0 = -\infty$ and $p_{N+1} = \infty$. Thus, to attain an explicit expression of $\tilde{U}_t^{\mathrm{r}}$, 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 $\{C_t^{\mathrm{r}}(q_j)\}_{j=1}^{N+1}$, as well as the computations of $\tilde{V}_{t+1}^{\mathrm{r}}(q_j - d_t^1), \cdots, \tilde{V}_{t+1}^{\mathrm{r}}(q_j - d_t^{N_t})$. The computations for C_t^{r} can be done by fist sorting $d_t^1, \cdots, d_t^{N_t}$, 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_t^{r}, which takes time

$$O(N \log N_t (\log d_{\max} + \log c^*)).$$

Thus, for C_t^{r} it takes time $O((N+N_t)\log N_t \log d_{\max} c^*)$ to compute. Next, for $\tilde{V}_{t+1}^{\mathrm{r}}$, it involves computing the function at $(N+1)N_t$ points, and each evaluation of $\tilde{V}_{t+1}^{\mathrm{r}}$ (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 $\tilde{U}^{\mathrm{r}}_t$ is dominated by the running time for the $N+1$ evaluations on $\tilde{U}_t(q_i)$s. Thus, given the explicit expression for $\tilde{V}^{\mathrm{r}}_{t+1}$, the run time for providing an explicit expression for $\tilde{U}^{\mathrm{r}}_t$ 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 $\tilde{U}^{\mathrm{r}}_t$, we prove the induction claims relevant to $\tilde{V}^{\mathrm{r}}_t$. First, observe that $\tilde{R}_t$ is the smallest break point p_j of $\tilde{U}^{\mathrm{r}}_t$ such that $\tilde{U}^{\mathrm{r}}_t(p_j) \geq 0$, which can be found efficiently by binary search. The search 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 $\tilde{U}^{\mathrm{r}}_t$ is non-decreasing. Now, given $\tilde{R}_t$ and the explicit expression of $\tilde{U}^{\mathrm{r}}_t$, by step 7 we can construct $\hat{\tilde{V}}\mathrm{r}_t$ 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 $\hat{\tilde{V}}^{\mathrm{r}}_t$ is equal to the number of break points of $\tilde{U}^{\mathrm{r}}_t$. Finally, to provide an explicit expression to $\tilde{V}^{\mathrm{r}}_t$ 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 sparsfication procedure, it suffices to perform it on the break points of $\tilde{U}^{\mathrm{r}}_t$.

Finally, since $\hat{V}^{\mathrm{r}}_t$ 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. □

C.8 Proof of Theorem 4.6.2

Proof. Proof of Theorem 4.6.2 First, we will show by a backward induction on t that for all y_t

$$\max\left\{\left|\tilde{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)\right|, \left|\tilde{V}_t^{\mathrm{r}}(y_t) - V_t^{\mathrm{r}}(y_t)\right|\right\} \leq \sum_{s=t}^{T}(\alpha_s + \eta)$$

where $\alpha_s = \frac{\epsilon \min_{t\in\{1,\cdots,T\}}\{\min\{h_t, b_t\}\}}{6T^2}$, just as was defined before.

We prove the inequality for the right derivatives by backward induction on t. First, for $t = T$, we see that $\tilde{U}_T = \hat{U}_T$, which implies $\tilde{U}_T^{\mathrm{r}} = \hat{U}_T^{\mathrm{r}}$. Thus, we have $|\tilde{U}_T^{\mathrm{r}}(x_T) - U_T^{\mathrm{r}}(x_T)| \leq \alpha_T$. In addition, by Theorem 4.5.3, we have $|\tilde{V}_T^{\mathrm{r}}(x_T) - V_T^{\mathrm{r}}(x_T)| \leq \alpha_T$. This establishes the induction claim for $t = T$. Now suppose the induction claim is true for $t+1$. Then by Theorem 4.5.1, we know that $|\tilde{U}_t^{\mathrm{r}}(y_t) - U_t^{\mathrm{r}}(y_t)| \leq \alpha_t + \sum_{s=t+1}^{T}(\alpha_s + \eta)$. By Theorem 4.5, this implies that

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

This is because $\tilde{R}_t$ is the smallest minimizer of $\tilde{U}_t$. Finally, we have

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

This establishes the induction claim.

Finally, we can conclude the $(1 + 2\epsilon)$ optimality of $(\tilde{R}_1, \cdots, \tilde{R}_T)$ by the argument in §5.2, and applying Lemma 4.5.6 and Lemma 4.5.7. □

C.9 Proof of Lemma 4.3.2

Proof. Proof of Lemma 4.3.2 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 order in periods $2, \cdots, T$, and 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.

The optimal modified base stock policy $(R_1^*, \cdots, R_T^*)$ has the following form:

$$\begin{aligned} R_1^* &= \min_{R \in \mathbb{R}} \left\{ R : \mathbb{P}\left[\sum_{t=1}^{T} D_t \leq R \right] \geq \frac{b_T}{h_T + b_T} \right\} \\ &= \min_{R \in \mathbb{R}} \left\{ R : \left| \left\{ S \subset \{1, \cdots, T\} : \sum_{t \in S} a_t \leq R \right\} \right| \geq \frac{b_T}{h_T + b_T} 2^T \right\}, \end{aligned}$$

and $R_2^*, \cdots R_T^*$ are arbitrary real numbers.

Suppose there exists an algorithm $\mathcal{A}$ that computes R_1^* in time polynomial in

$$(T, \log(h_T + b_T)).$$

Then, by letting b_T, h_T be positive integers such $b_T = \alpha \in \{1, \cdots, 2^T\}$, $h_T + b_T = 2^T$, one can use $\mathcal{A}$ as a subroutine to solve the **#KNAPSACK** problem by a binary search on α, and such a binary search algorithm runs in time polynomial in T. This concludes the proof. □ □

C.10 Proof of Theorem 4.3.4

Proof. Proof of Theorem 4.3.4 Recall h, b respectively denote the holding cost and backlog cost in the newsvendor problem. We first consider the case when $h \leq b$. Consider 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 $A = 2000 \max\left\{1, \frac{h}{b\epsilon}\right\}$. We claim that the following statements are true:

- **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: $\mathrm{KL}(D_1||D_2) \leq \frac{8h\epsilon^2}{b+h}$,

where we 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.

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. Let $\mathcal{A}$ be an algorithm which draws m samples and returns a $(1 + \frac{\epsilon}{20})$−optimal base stock with probability at least $1 - \delta$ under any latent distribution D. Given such $\mathcal{A}$, consider the following classification algorithm $\mathcal{B}$. The algorithm $\mathcal{B}$ has the input $d^m = (d_1, \cdots, d_m)$, where $d_1, \cdots, d_m$ be i.i.d. samples drawn from the distribution D. The distribution D is either D_1 or D_2, but its identity is unknown. The algorithm $\mathcal{B}$ outputs its decision on the identity of D.

Note that the "If" statement can be checked, since $\mathcal{B}$ has access to the CDFs of D_1, D_2. The construction above shows that given such an algorithm $\mathcal{A}$, the classifier $\mathcal{B}$ uses m

Algorithm 11 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.
5: **else**
6: Return 0.
7: **end if**

samples, and succeeds with probability $1-2\delta$.

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. That is, $\mathcal{B}(d^m) = \mathbf{1}(\mathcal{B}$ decides that d^m is drawn from $D_1)$. Also, for $i = 1, 2$, we use D_i^m to denote the product distributions of m i.i.d. random variables with common distribution D_i. Now, by our assumption on $\mathcal{A}$, we have the following bounds:

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

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

These result in the following bound:

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

Now, we relate the right hand side of the bound above to the KL divergence $\mathrm{KL}(D_1||D_2)$:

$$\begin{aligned}
\mathbb{E}_{d^m\sim D_1^m}\mathcal{B}(d^m) - \mathbb{E}_{d^m\sim D_2^m}\mathcal{B}(d^m) &= \sum_{d^m\in\{0,1,A\}^m} \mathcal{B}(d^m)\left(\mathbb{P}[D_1^m = d^m] - \mathbb{P}[D_2^m = d^m]\right) \\
&\leq \sum_{d^m\in\{0,1,A\}^m} \left|\mathbb{P}[D_1^m = d^m] - \mathbb{P}[D_2^m = d^m]\right| \qquad \text{(C.7)} \\
&\leq \sqrt{(2\log 2)\mathrm{KL}\left(D_1^m||D_2^m\right)} \qquad \text{(C.8)} \\
&\leq \sqrt{(2m\log 2)\mathrm{KL}\left(D_1||D_2\right)}
\end{aligned}$$

$$\leq \sqrt{\frac{(16m\log 2)h\epsilon^2}{h+b}} \tag{C.9}$$

The inequality (C.7) is true because $|\mathcal{B}(d^m)| \leq 1$ for all d^m. Note that the quantity in (C.7) is the total variation distance between D_1^m and D_2^m. The inequality (C.8) is an application of Pinsker Inequality (for example see [Cover and Thomas, 2006]) on D_1^m and D_2^m. The inequality (C.9) is by Claim 2.

Altogether, 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 4.3.4 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 4.3.4 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

[Levi et al., 2007], or deduce from §4.4) 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 4.3.4. □

Bibliography

[Aouad et al., 2015] Aouad, A., Farias, V. F., and Levi, R. (2015). Assortment optimization under consider-then-choose choice models. *Manuscript.*

[Arora et al., 2012] Arora, S., Hazan, E., and Kale, S. (2012). The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(6):121–164.

[Aviv and Federgruen, 1997] Aviv, Y. and Federgruen, A. (1997). Stochastic inventory models with limited production capacity and periodically varying parameters. *Probability in the Engineering and Informational Sciences*, 11:107–135.

[Aviv et al., 2009] Aviv, Y., Levin, Y., and Nediak, M. (2009). Counteracting strategic consumer behavior in dynamic pricing systems. In Tang, C. S. and Netessine, S., editors, *Consumer-Driven Demand and Operations Management Models*, volume 131 of *International Series in Operations Research and Management Science*, pages 323–352. Springer US.

[Babaioff et al., 2012] Babaioff, M., Dughmi, S., Kleinberg, R., and Slivkins, A. (2012). Dynamic pricing with limited supply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 74–91, New York, NY, USA. ACM.

[Badanidiyuru et al., 2013] Badanidiyuru, A., Kleinberg, R., and Slivkins, A. (2013). Bandits with knapsacks. In *FOCS.*

[Begen et al., 2012] Begen, M. A., Levi, R., and Queyranne, M. (2012). Technical noteâĂŤa sampling-based approach to appointment scheduling. *Operations Research*, 60(3):675–681.

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

[Bertsimas et al., 2011] Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Rev.*, 53(3):464–501.

[Besbes and Zeevi, 2009] Besbes, O. and Zeevi, A. (2009). Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Oper. Res.*, 57(6):1407–1420.

[Besbes and Zeevi, 2011] Besbes, O. and Zeevi, A. (2011). On the minimax complexity of pricing in a changing environment. *Operations Research*, 59(1):66–79.

[Besbes and Zeevi, 2012] Besbes, O. and Zeevi, A. (2012). Blind network revenue management. *Oper. Res.*, 60(6):1537–1550.

[Blanchet et al., 2013] Blanchet, J. H., Gallego, G., and Goyal, V. (2013). A markov chain approximation to choice modeling. In *EC*, pages 103–104.

[Broder and Rusmevichientong, 2012] Broder, J. and Rusmevichientong, P. (2012). Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980.

[Bront et al., 2009] Bront, J. J. M., Méndez-Díaz, I., and Vulcano, G. (2009). A column generation algorithm for choice-based network revenue management. *Oper. Res.*, 57(3):769–784.

[Bubeck and Cesa-Bianchi, 2012] Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122.

[Charikar et al., 2005] Charikar, M., Chekuri, C., and Pal, M. (2005). Sampling bounds for stochastic optimization. In *PROC. 9TH RANDOM*, pages 257–269. Springer.

[Charnes and Cooper, 1959] Charnes, A. and Cooper, W. W. (1959). Chance-constrained programming. *Management Science*, 6(1):73–79.

[Chen et al., 2015] Chen, X., Owen, Z., Pixton, C., and Simchi-Levi, D. (2015). A statistical learning approach to personalization in revenue management. *Manuscript.*

[Cheung et al., 2015] Cheung, W. C., Simchi-Levi, D., and Wang, H. (2015). Dynamic pricing and demand learning with limited price experimentation. *Available at SSRN 2457296.*

[Ciocan and Farias, 2012] Ciocan, D. F. and Farias, V. (2012). Model predictive control for dynamic resource allocation. *Math. of Oper. Res.*, 37(3).

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

[Dantzig, 1955] Dantzig, G. (1955). Linear programming under uncertainty. *Management Science*, 1(3-4):197–206.

[Davis et al., 2013] Davis, J., Gallego, G., and Topaloglu, H. (2013). Assortment planning under the multinomial logit model with totally unimodular constraint structures. *Manuscript.*

[den Boer, 2011] den Boer, A. V. (2011). Dynamic pricing with multiple products and partially specified demand distribution. *Mathematics of Operations Research*, page Forthcoming.

[den Boer, 2014] den Boer, A. V. (2014). Dynamic pricing and learning: Historical origins, current research, and new directions.

[Désir and Goyal, 2014] Désir, A. and Goyal, V. (2014). Near-optimal algorithms for capacity constrained assortment optimization. *Manuscript.*

[Désir et al., 2015] Désir, A., Goyal, V., Segev, D., and Ye, C. (2015). Capacity constrained assortment optimization under the markov chain based choice model. *Oper. Res., Forthcoming.*

[Feldman and Topaloglu, 2014] Feldman, J. B. and Topaloglu, H. (2014). Revenue management under the markov chain choice model. *Manuscript.*

[Feldman and Topaloglu, 2015] Feldman, J. B. and Topaloglu, H. (2015). Capacity constraints across nests in assortment optimization under the nested logit model. *Oper. Res.*, 63(4):812–822.

[Feng and Gallego, 1995] Feng, Y. and Gallego, G. (1995). Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, 41(8):1371–1391.

[Feng and Xiao, 2000] Feng, Y. and Xiao, B. (2000). Optimal policies of yield management with multiple predetermined prices. *Operations Research*, 48(2):332–343.

[Gallego et al., 2004] Gallego, G., Iyengar, G., Phillips, R., and Dubey, A. (2004). Managing flexible products on a network. *Manuscript.*

[Gallego et al., 2015a] Gallego, G., Li, A., Truong, V.-A., and Wang, X. (2015a). Online resource allocation with customer choice. *Manuscript.*

[Gallego et al., 2015b] Gallego, G., Ratliff, R., and Shebalov, S. (2015b). A general attraction model and sales-based linear program for network revenue management under customer choice. *Oper. Res.*, 63(1):212–232.

[Gallego and van Ryzin, 1994] Gallego, G. and van Ryzin, G. (1994). Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Manage. Sci.*, 40(8):999–1020.

[Gallego and Van Ryzin, 1997] Gallego, G. and Van Ryzin, G. (1997). A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research*, 45(1):24–41.

[Golrezaei et al., 2014] Golrezaei, N., Nazerzadeh, H., and Rusmevichientong, P. (2014). Real-time optimization of personalized assortments. *Mgmt. Sci.*, 60(6):1532–1551.

[Grötschel et al., 1993] Grötschel, M., Lovász, L., and Schrijver, A. (1993). *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer.

[Guang et al., 2015] Guang, L., Rusmevichientong, P., and Topaloglu, H. (2015). The d-level nested logit model: Assortment and price optimization problems. *Oper. Res.*, 63(2):325–342.

[Gupta et al., 2004] Gupta, A., Pál, M., Ravi, R., and Sinha, A. (2004). 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.

[Gupta et al., 2005] Gupta, A., PÃąl, M., Ravi, R., and Sinha, A. (2005). What about wednesday? approximation algorithms for multistage stochastic optimization. In *in APPROX*, pages 86–98.

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

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

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

[Halman et al., 2009] Halman, N., Klabjan, D., Mostagir, M., Orlin, J., and Simchi-Levi, D. (2009). A fully polynomial-time approximation scheme for single-item stochastic inventory control with discrete demand. *Mathematics of Operations Research*, 34(3):674–685.

[Harrison et al., 2012] Harrison, J. M., Keskin, N. B., and Zeevi, A. (2012). Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586.

[Immorlica et al., 2004] Immorlica, N., Karger, D., Minkoff, M., and Mirrokni, V. S. (2004). 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.

[INT, 2015] INT (2015). `https://www.internetretailer.com/2015/07/29/global-e-commerce-set-grow-25-2015`.

[Jasin and Kumar, 2012] Jasin, S. and Kumar, S. (2012). A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *MOR*, 37(2):313–345.

[Kallus and Udell, 2015] Kallus, N. and Udell, M. (2015). Learning preferences from assortment choices in a heterogeneous population. *Manuscript.*

[Kapuscinski and Tayur, 1998] Kapuscinski, R. and Tayur, S. (1998). A capacitated production-inventory model with periodic demand. *Operations Research*, 46:899–911.

[Khachiyan, 1980] Khachiyan, L. (1980). Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53 – 72.

[Kleinberg and Leighton, 2003] Kleinberg, R. and Leighton, T. (2003). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 594–605.

[Kleywegt et al., 2002] Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM J. on Optimization*, 12(2):479–502.

[Kök et al., 2015] Kök, A. G., Fisher, M. L., and Vaidyanathan, R. (2015). *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, chapter Assortment Planning: Review of Literature and Industry Practice. Springer.

[Kunnumkal and Talluri, 2016] Kunnumkal, S. and Talluri, K. (2016). On a piecewise-linear approximation for network revenue management. *Math. of Oper. Res.*, 41(1):72–91.

[Levi et al., 2014] Levi, R., Perakis, G., and Uichanco, J. (2014). The data-driven newsvendor problem: New bounds and insights. *Manuscript.*

[Levi et al., 2006] Levi, R., Roundy, R. O., and Shmoys, D. B. (2006). 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.

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

[Levi et al., 2008] Levi, R., Roundy, R. O., Shmoys, D. B., and Truong, V. A. (2008). Approximation algorithms for capacitated stochastic inventory control models. *Operations Research*, 56(5):1184–1199.

[Liu and van Ryzin, 2008] Liu, Q. and van Ryzin, G. (2008). On the choice-based linear programming model for network revenue management. *MSOM*, 10(2).

[Luce, 1959] Luce, R. D. (1959). *Individual Choice Behavior: A theoretical analysis*. Wiley.

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

[McFadden, 1974] McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, pages 105–142. Academic Press.

[Mittal and Schulz, 2013] Mittal, S. and Schulz, A. S. (2013). A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. *Oper. Res.*, 61(2):386–397.

[Myerson, 1981] Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

[Ozer and Philips, 2012] Ozer, O. and Philips, R. (2012). *The Oxford Handbook of Pricing Management*. Oxford University Press.

[Placket, 1975] Placket, R. L. (1975). The analysis of permutations. *Applied Statistics*, 24:193–202.

[Plotkin et al., 1995] Plotkin, S. A., Shmoys, D. B., and Tardos, É. (1995). Fast approximation algorithms for fractional packing and covering problems. *MOOR*, 20(2):257–301.

[Qin et al., 2016] Qin, H., Simchi-Levi, D., and Wang, L. (2016). Multi-period data-driven approximation to joint pricing and inventory control. *Manuscript*.

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

[Rusmevichientong et al., 2009] Rusmevichientong, P., Shen, Z.-J. M., and Shmoys, D. B. (2009). A ptas for capacitated sum-of-ratios optimization. *ORL*, 37(4):230–238.

[Rusmevichientong et al., 2010] Rusmevichientong, P., Shen, Z.-J. M., and Shmoys, D. B. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Oper. Res.*, 58(6):1666–1680.

[SAP, 2015] SAP (2015). `https://news.sap.com/`.

[Saure and Zeevi, 2013] Saure, D. and Zeevi, A. (2013). Optimal dynamic assortment planning with demand learning. *MSOM*, 15(3):387–404.

[Scarf, 2002] Scarf, H. (2002). Inventory theory. *Operations Research*, 50(1):186–191.

[Schrijver, 1986] Schrijver, A. (1986). *Theory of Linear and Integer Programming.* John Wiley & Sons, Inc., New York, NY, USA.

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

[Shapiro et al., 2009] Shapiro, A., Dentcheva, D., and RuszczyÅĎski, A. (2009). *Lectures on Stochastic Programming.* Society for Industrial and Applied Mathematics.

[Shapiro and Nemirovski, 2005] Shapiro, A. and Nemirovski, A. (2005). On complexity of stochastic programming problems. In Jeyakumar, V. and Rubinov, A., editors, *Continuous Optimization*, volume 99 of *Applied Optimization*, pages 111–146. Springer US.

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

[Simchi-Levi et al., 2013] Simchi-Levi, D., Bramel, J., and Chen, X. (2013). *The logic of logistics: theory, algorithms, and applications for logistics and supply chain management.* Springer Verlag.

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

[Tal and Nemirovski, 1998] Tal, A. B. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

[Talluri and Van Ryzin, 2004] Talluri, K. and Van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Mgmt. Sci.*, 50(1):15–33.

[Talluri and Van Ryzin, 2006] Talluri, K. T. and Van Ryzin, G. J. (2006). *The theory and practice of revenue management*, volume 68. Springer Science & Business Media.

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

[TCS, 2016] TCS (2016). `http://www.tcs.com/resources/white_papers/Pages/Online-assortment-game-plan.aspx`.

[Topaloglu, 2009] Topaloglu, H. (2009). Using lagrangian relaxation to compute capacity-dependent bid prices in network revenue management. *Oper. Res.*, 57(3):637–649.

[Topaloglu, 2013] Topaloglu, H. (2013). Joint stocking and product offer decisions under the multinomial logit model. *Production and Operations Management*, 22(5).

[Wang, 2016] Wang, H. (2016). *Dynamic Learning and Optimization for Operations Management Problems*. PhD thesis, Massachusetts Institute of Technology.

[Wang et al., 2011] Wang, Z., Deng, S., and Ye, Y. (2011). Close the gaps: A learning-while-doing algorithm for a class of single-product revenue management problems. *CoRR*, abs/1101.4681.

[Young, 1995] Young, N. E. (1995). Randomized rounding without solving the linear program. In *SODA*.

[Zhang and Adelman, 2009] Zhang, D. and Adelman, D. (2009). An approximate dynamic programming approach to network revenue management with customer choice. *Trans. Sci.*, 43(3):381–394.