Dynamic assortment optimization

From learning to earning

Peeters, Y.A.

Publication date
2022

Document Version
Final published version

Dynamic Assortment Optimization

From Learning to Earning

Yannik Peeters

Dynamic Assortment Optimization

From Learning to Earning

Yannik Peeters

Dynamic Assortment Optimization

From Learning to Earning

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 24 februari 2022, te 16.00 uur

door

Yannik Adriaan Peeters
geboren te Breda

Promotiecommissie

Promotores

Dr. A.V. den Boer	Universiteit van Amsterdam
Prof. dr. M.R.H. Mandjes	Universiteit van Amsterdam

Overige leden

Prof. dr. M. Salomon	Universiteit van Amsterdam
Prof. dr. ir. D. den Hertog	Universiteit van Amsterdam
Dr. T.A.L. van Erven	Universiteit van Amsterdam
Prof. dr. F.M. Spieksma	Universiteit Leiden
Dr. Y. Wang	University of Florida

Faculteit Economie en Bedrijfskunde

To my dearest Tara

Contents

Chapter 1

Introduction

As the infrastructure of information improves to better handle incoming real-time purchase data, the necessity for data-driven, automated assortment policies arises. In general, a seller of products has to consider multiple aspects, e.g., inventory management, demand management, budget constraints, pricing strategy and assortment planning. Proper management of these aspects can have a large impact on the success of his/her undertaking. In particular, sales data can be used effectively in order to maximize profit. This thesis considers *assortment optimization*, where an assortment is a collection or subset of all products – from which a customer chooses a product to purchase. The main question that we study is: how can a seller determine the optimal assortment of products – the subset which yields the highest expected profit – based on sales data.

To illustrate the potential benefits of adequate usage of sales data, we consider the airline industry. Before the COVID-19 pandemic, worldwide airline ancillary revenue – i.e., revenue made from non-ticket sources such as upgrades, baggage, seat selection, change or cancellation fees, etc. – increased by a staggering 485% over the course of nine years: from $22.6 billion in 2010 to $109.5 billion in 2019 (see IdeaWorksCompany, 2019; Xu & Wang, 2021). This demonstrates that a substantial amount of additional revenue can be gained by understanding customer's purchasing behavior. An airline designs combinations of travel class and booking options with the objective of maximizing its expected revenue. An understanding of consumers' preferences from real-world sales data can be exceedingly helpful in the design of such

options.

Analyzing sales data can potentially bolster gains in revenue. However, this data may not be available, e.g., due to a newly set up production line. Then, within a finite time frame, the acquisition of insightful sales data overlaps with the exploitation of those insights. These aspects are often described as *learning* and *earning*, or *exploration* and *exploitation*. First, we would like to learn customers' demand and preferences. Second, we would like to utilize that knowledge to make informed decisions. In this thesis, we focus on the decisions and strategies regarding assortment selection.

We explore customers' preference learning for adequate assortment planning in a rigorous mathematical and sequential setting. In this framework, the seller starts a selling period with no prior information regarding customers' preferences. However, as time goes on, more sales data becomes available and this data can be used by the seller to adjust the offered assortment accordingly. Closely related to this formulation is the *multi-armed bandit* (MAB) paradigm. In the MAB framework, we take on the role of a gambler with K slot machines at our disposal. Each of the K slot machines has a different expected pay-off, which is unknown to us, the gambler. We are given T coins to play T rounds on the slot machines and we are free to choose which machines and in which order we wish to play. The goal is to maximize our accumulated profit, which means we have to delicately balance the two concepts: learning and earning. With our T coins, we would like to learn which slot machine yields the highest expected pay-off. In addition, we would like to use this knowledge to earn as much as we can. By learning insufficiently, we are not sure which action is the most profitable. By emphasizing learning too much, there are lesser coins available to exploit our findings.

In our set-up, we consider the problem of assortment planning in which a seller can frequently alter the offered assortment to his/her customers. Starting with no prior knowledge regarding customers' preferences, the seller's intention is to offer assortments sequentially over a finite selling period. In this venture, it is key to balance the aforementioned concepts of learning and earning. To deal with this balance, we view the problem of a seller of products as a MAB problem, where we decide which assortment to offer instead of which slot machine to play. In this set-up, we are given

a time horizon T and for simplicity, we associate each time $t = 1, \ldots, T$ with a visiting customer. Before each visit, we can choose which assortment to offer to the visiting customer. Subsequently, the customer chooses a product from the offered assortment or does not purchase anything.

One can imagine that the seller's choice of the size of the assortments to offer is restricted. Such a constraint may arise due to limited shelf space in a physical store, finite inventory capacity or various logistic restrictions. Therefore, we include the possibility of a *capacity constraint*. The circumstance where we do consider a capacity constraint is referred to as the *capacitated* variant of the optimization problem. We refer to the setting where there is no capacity constraint as the *uncapacitated* variant.

We envision the total collection of products in two ways. First, we consider the classical problem of discrete assortment planning. In this framework, there are N distinct products and the seller selects an assortment S, which is a subset of the N products. Then, the customer chooses either a product from S or does not purchase anything. Second, we consider a continuous spectrum of products. Examples of such a continuous spectrum are the duration or amount of a mortgage and the amount of cellular data usage. Here, each product x lies on the continuous spectrum and the seller selects an assortment S to offer, which is a subset of that spectrum. Again, the customer either chooses a product from S or does not purchase anything.

1.1 Discrete Model

For discrete assortment optimization, there are N distinct products for the seller to consider offering. The seller – subject to a potential capacity constraint – chooses which assortment S, a subset of all products, to offer. The capacity constraint constitutes an integer $K \leqslant N$, meaning that each offered assortment may consist of only K products. A collection of at most K products is referred to as a feasible assortment. Being offered an assortment S, a customer then chooses a product from S, or does not purchase anything. For the seller, each product i yields a known marginal revenue per sold unit of product i and a no-purchase yields no marginal revenue.

Without a choice model for customers' behavior, the task of assortment optimization is nearly intractable. Indeed, without a choice model we cannot easily predict the

change of customers' behavior caused by adjusting the offered assortment and we can only observe the changing behavior by actually adjusting the assortment. Arguably the most-studied choice model in the extensive literature on assortment optimization with a finite number of products is the multinomial logit (MNL) model (see, e.g., Ben-Akiva & Lerman, 1985; Mahajan & van Ryzin, 2001, and the references therein). The MNL model arises as a special case of the *random utility model* (RUM). Within the RUM, the customer, upon arrival, takes notice of the offered assortment S and knows his/her utility for all products in S and the no-purchase option. Then, the customer chooses the product with the highest utility. Under the RUM, it is assumed that the utility value of each product follows some probability distribution. The MNL model is the special case where the utility is the sum of a deterministic part (depending on the product features and its overall appeal) and a random, probabilistic part (following a specific probability distribution). This set-up has compelling properties. Given an offered assortment, there exists a closed-form expression of the purchase probabilities of a particular product that is proportional to the attractiveness expressed by the utility parameter of that product. Consequently, we can explicitly consider the expected revenue of each assortment and measure the performance of a decision policy that sequentially outputs a feasible assortment using observed past purchase behavior. Regarding practical applications, various successful implementations of the MNL model in retail have been reported (see, e.g., Guadagni & Little, 1983; Ratliff et al., 2008; Newman et al., 2014; Feldman et al., 2019). For an excellent literature review of choice models – including MNL – as well as caveats and industry applications to assortment planning, we refer to Kök et al. (2015).

Recently, several authors have studied assortment optimization under the discrete MNL choice model where the seller has no prior knowledge regarding customers' preference (see, e.g., Rusmevichientong et al., 2010; Farias et al., 2013; Sauré & Zeevi, 2013; Agrawal et al., 2017; Cheung & Simchi-Levi, 2017; Chen & Wang, 2018; Ou et al., 2018; Agrawal et al., 2019; Kallus & Udell, 2020; Chen et al., 2021). A first step in assessing the profitability of a sequential decision policy is to analyze its *regret*. This metric is the accumulated expected loss due to offering suboptimal assortments. As it is not a straightforward task to directly assess how profitable a particular policy is, policy creators assess performance by mathematically determining upper bounds

on the regret. By providing such an upper bound in terms of the time horizon, the number of products and the capacity constraint, the asymptotic performance of proposed policies can be rigorously evaluate. Furthermore, by showing a matching lower bound on the regret that any policy must – in the worst case – endure, policy creators are able to guarantee that their proposed policy performs asymptotically optimal and gain insights regarding the learning problem.

Typically, regret rates for dynamic assortment optimization with capacity constraint grow as $\sqrt{NT}$ (see, e.g., Agrawal et al., 2017; Chen & Wang, 2018; Agrawal et al., 2019; Chen et al., 2021). In particular for constant revenue parameters and capacity constraint $K \leqslant N/4$, Chen & Wang (2018) show that the regret that any policy endures, is of the order $\sqrt{NT}$ in the worst case. This thesis builds upon this result by extending the $\sqrt{NT}$ lower bound to a more general setting. Interestingly, for the uncapacitated variant, Chen et al. (2021) provide a Trisection policy that achieves a regret of order $\sqrt{T}$, where the regret no longer depends on the number of products N. This indicates that the Trisection policy performs robustly for an increasing value of N. Moreover, Chen et al. (2021) show a regret lower bound of order $\sqrt{T}$ for specific values as revenue parameters. We continue the regret analysis by providing a $\sqrt{T}$ lower bound for arbitrary revenue parameters. In addition, we present a policy that achieves a $\sqrt{T}$ regret as well and we show its merits numerically.

Aside from contributing – see Section 1.3 – to dynamic assortment optimization under the discrete MNL model, we also consider a continuous counterpart of the MNL model.

1.2 Continuous Model

In the management science and operation research literature, assortments are traditionally thought of as being of a discrete nature. However, in several applications, attributes of products or services are adjusted in a *continuous* manner, leading to a spectrum of similar but distinct commodities, each with a possibly different selling price. In these situations, customers can be offered highly personalized, custom-made products – a phenomenon that the marketing literature refers to as *mass customization* (see, e.g., Pine, 1993; Fogliatto et al., 2012).

The idea of considering a continuous spectrum of products is a well-established concept in several branches of the literature. Within the economics literature, for example, this idea is studied in the context of *vertical product differentiation* (see, e.g., Mussa & Rosen, 1978; Moorthy, 1984). More recently, the idea of examining a continuum of products has been studied in the operations research literature (see, e.g., Dewan et al., 2003; Gaur & Honhon, 2006; Pan & Honhon, 2012; Fisher & Vaidyanathan, 2014; Keskin & Birge, 2019; den Boer et al., 2021). With the exception of Keskin & Birge (2019) and den Boer et al. (2021), the literature mentioned above assumes that the model primitives are known to the seller. Although the concept of a continuous spectrum of products is not novel, a sequential decision framework and learning the model primitives from sales data has – to the best of our knowledge – not yet been studied in the context of assortment optimization.

The utility value of a product in the discrete MNL choice model depends on the product's features. We consider a setting where these features can be adjusted continuously prior to purchasing and we propose a novel model that forms a continuous counterpart of the discrete MNL model. The novel continuous model arises from the discrete MNL model by letting the number of products N go to infinity. Under this continuous model, we consider a seller of a commodity or services with an attribute that can be infinitesimally adjusted to any value in the unit interval $[0, 1]$. Each value in $[0, 1]$ is referred to as a product and the seller has to decide which assortment of products, i.e., which subset of $[0, 1]$, to offer to each potential customer. Being offered an assortment S, a customer chooses a product from S or does not purchase anything. For each product x in $[0, 1]$, the marginal revenue obtained if x is purchased, is known to the seller; no revenue is obtained from a no-purchase.

Similar to the discrete model, we assess the performance of a sequential decision policy by its regret. As the continuous model that we propose has not been considered before, no prior knowledge exists regarding the growth rate of the regret expressed in terms of the time horizon T. Also alike the discrete setting, a seller may be restricted in the assortments he/she can offer. Aside from the aspect that the continuous model arises from the discrete model by taking the number of products N to infinity, one can imagine that in a truly continuous setting the seller's choice of the size of the offered assortment is limited due to various logistic constraints. Therefore, we consider both

the capacitated variant and uncapacitated variant of the optimization problem under the continuous model. The potential capacity constraint entails a constant c between 0 and 1 such that the seller may only offer assortments each with a combined length of at most c.

1.3 Outline and Scientific Contributions

This thesis contributes to multiple settings regarding dynamic assortment optimization. The state-of-the-art regret bounds – after this thesis – are summarized in Table 1.1 presented below. Thereafter, the specific contributions are discussed in detail.

	Uncapacitated		Capacitated	
Model	Upper bound	Lower bound	Upper bound	Lower bound
Continuous	$\log T$	$\log T$	$T^{2/3}$ (*)	$T^{2/3}$
Discrete	$\sqrt{T}$ (a)	$\sqrt{T}$ (b)	–	$\sqrt{NT}$ (c)

Table 1.1: Regret bounds (up to multiplicative constants) as provided in this thesis for several combination of settings. The continuous model is discussed in Chapter 2 and the discrete model in Chapter 3.
(*) Up to logarithmic terms.
(a) Matching the regret rate of Chen et al. (2021).
(b) Slightly generalizing Chen et al. (2021).
(c) Slightly generalizing Chen & Wang (2018).

A policy and a corresponding upper bound on its regret in the discrete setting with capacity constraint is not included in this thesis, as it has been covered extensively by several authors (see, e.g., Rusmevichientong et al., 2010; Agrawal et al., 2017, 2019; Chen et al., 2021).

The topics and contributions are presented in this thesis as follows. In Chapter 2, we consider assortment planning under the novel continuous model. Moreover, we consider the seller's problem under incomplete information. This topic is covered first – before we study the discrete model in Chapter 3 – as it is the main contribution of this thesis. In Chapter 2, we study the structure of the optimal assortment and we provide an algorithm to compute the optimal assortment when the model primitives are known. For the dynamic setting, where the model primitives are initially

unknown, we consider two distinct cases, one without capacity constraint and one with capacity constraint. For the setting without a capacity constraint, we propose a stochastic-approximation type of policy and we show that its regret grows at most as $\log T$ in the time horizon T. We complement this result by showing a matching lower bound on the regret of any policy, implying that our policy is asymptotically optimal. We then show that adding a capacity constraint significantly changes the structure of the problem: we construct a discretization policy and show that its regret after T time periods is bounded from above by a constant times $T^{2/3}$ (up to a logarithmic term). In addition, we show that the regret of any policy is bounded from below by a positive constant times $T^{2/3}$, so that also in the capacitated case we obtain asymptotic optimality. Additionally, we propose a kernel density estimation-based policy and show that – under an additional assumption – its regret is also bounded from above by a constant times $T^{2/3}$ (up to a logarithmic term). This chapter, with the exception of Section 2.5.5 and 2.5.6, corresponds to the paper titled "*Continuous Assortment Optimization with Logit Choice Probabilities and Incomplete Information*". This paper, Peeters et al. (2021), is a joint endeavor with dr. A.V. den Boer and prof. dr. M.R.H. Mandjes.

Chapter 3 is concerned with dynamic assortment optimization with incomplete information under the discrete MNL model. To recognize the relevance of our contribution, we briefly discuss some of the relevant literature. Chen et al. (2021) consider uncapacitated assortment optimization under the MNL model. They provide a Trisection policy and show an upper bound at the rate of $\sqrt{T}$ on its worst-case regret, which is independent of the number of products N. Furthermore, they show a lower bound at the rate of $\sqrt{T}$ on the worst-case regret of any policy. This lower bound is shown for a specific marginal revenue setting: the marginal revenue is 1 for odd-numbered products and 1/2 for even-numbered products. For capacitated assortment optimization under the MNL model, Agrawal et al. (2019) study an Upper Confidence Bound (UCB) policy. They show that its worst-case regret grows at the rate of $\sqrt{NT \log T}$. Moreover, they show a lower bound at the rate of $\sqrt{NT/K}$ for the worst-case regret of any policy. In addition, Agrawal et al. (2017) present a Thompson Sampling policy in the same setting and provide an constant times $\sqrt{NT} \log TK$ upper bound for its worst-case regret. The rate of the lower bound of Agrawal et al. (2019) is improved by

Chen & Wang (2018) to $\sqrt{NT}$, under the assumption that $K \leqslant N/4$. Both the lower bound of Agrawal et al. (2019) and Chen & Wang (2018) consider constant revenue parameters for all products.

Our contributions, as presented in Chapter 3, are the following. For no capacity constraint, we propose a stochastic-approximation type policy, which is the discrete counterpart of the policy provided in Chapter 2. Moreover, we show that its regret after T time periods is at most a constant times $\sqrt{T}$, which is independent of the number of products N. This matches the current theoretical standard up to a multiplicative constant. Numerical illustrations show the advantages of our policy over alternatives, especially the improved performance for moderately large N (see Chapter 4). This policy is horizon-free as it does not require T as input. Furthermore, we show that for arbitrary revenue parameters the regret of any policy is bounded from below by a constant times $\sqrt{T}$, generalizing the result from Chen et al. (2021) to a broader setting. In the discrete model, we also find that introducing a capacity constraint substantially alters the structure of the problem. We show for arbitrary revenue parameters that for any policy satisfying a capacity constraint $K < N/2$, the regret is bounded from below by a positive constant times $\sqrt{NT}$. This solidifies the result from Chen & Wang (2018) to a more general setting. Chapter 3 of this thesis consists of two papers. The first paper covers the stochastic-approximation type policy, its regret upper bound and the regret lower bound for uncapacitated assortment optimization. This paper, Peeters & den Boer (2021a), is titled "*Stochastic Approximation for Uncapacitated Assortment Optimization under the Multinomial Logit Model*". The second paper covers the regret lower bound for capacitated assortment optimization. This paper, Peeters & den Boer (2021b), is titled "*A Regret Lower Bound for Assortment Optimization under the Capacitated MNL Model with Arbitrary Revenue Parameters*".

Chapter 4 covers the numerical experiments supplementing Chapter 2 and 3. In Chapter 4, we compare the policies that we provide numerically with established, alternative policies. These experiments show that our policies outperform or are on par with alternatives. Moreover, we include numerical experiments to compare the predictive performance of the continuous model and the discrete model. Section 4.1 and 4.3 are contained in Peeters et al. (2021) and Section 4.2 is contained in Peeters

& den Boer (2021a).

The mathematical proofs of results stated in Chapter 2 and 3 are collected in Appendix A and B, respectively.

Chapter 2

Continuous Assortment Optimization

2.1 Introduction

2.1.1 Background and Motivation

In the management science and operations research literature, assortments are traditionally thought of as being of a discrete nature. However, in several applications, attributes of products or services are adjusted in a *continuous* manner, leading to a spectrum of similar but distinct commodities, each with a possibly different selling price. In these situations, customers can be offered highly personalized, custom-made products – a phenomenon that the marketing literature refers to as *mass customization* (see, e.g., Pine, 1993; Fogliatto et al., 2012). Examples of attributes that can be customized in such a continuous manner include the duration of renting a commodity, the duration or amount of a mortgage, the amount of cellular data usage, or the amount of (voluntary) deductible excess in insurances. A seller of such products or services faces, in particular in the product design phase, the concrete problem of having to decide which specific subset of the spectrum to offer to potential customers, so as to maximize expected profit.

The seller's problem can be translated into a mathematical optimization problem over an uncountable space of subsets of an interval. This type of problem can only be solved efficiently when some structure is imposed on how the consumers' purchase behavior and the seller's revenue depend on the assortment that is offered. In the

extensive literature on assortment optimization with a finite number of products, arguably the most-studied choice model is the so-called multinomial logit (MNL) model (see, e.g., Ben-Akiva & Lerman, 1985; Mahajan & van Ryzin, 2001, and the references therein). In this model, a nonnegative preference value is associated with each product (and also with the option of not purchasing a product), and the probability that a customer selects a particular product from an assortment of products is proportional to this preference value. To align our work with this rich strand of literature, we propose a choice model that is the continuous counterpart of the discrete MNL model, with the preference values replaced by a preference function.

Importantly, we study the seller's continuous assortment optimization problem in an *incomplete information* setting, meaning that the preference function is a priori unknown to the seller. To arrive at profitable assortment decisions, the seller thus has to learn the unknown preference function from accumulating sales data. This requires designing a policy that judiciously balances the two (sometimes conflicting) goals of *learning* and *earning*: on the one hand, the seller needs to offer assortments that support high-quality estimates of the unknown preference function; on the other hand, assortments need to be offered that yield a high profit given an available estimate of the preference function. This is an example of the well-known *exploration-exploitation* trade-off in multi-armed bandit (MAB) problems: a paradigm for sequential decision problems under uncertainty. Indeed, the problem studied in this chapter can be seen as a continuous, combinatorial MAB problem, where the objective is to dynamically learn which subset of the continuum maximizes the seller's expected revenue function. Designing and analyzing optimal decision policies for this novel and relevant question is the topic of this chapter.

2.1.2 Contributions

The contributions of this chapter are as follows.

- First, we propose a probabilistic choice model for the setting where customers select from assortments that are subsets of the unit interval. The choice model is the continuous counterpart of the widely studied MNL model, in the sense that the continuous model arises as a limit of discrete MNL models where the

number of products grows large, and, conversely, that discretizing the product space in the continuous model gives rise to a discrete MNL model.

◦ Next, assuming that products are labeled in increasing order of marginal profit, we show that the optimal assortment is an interval of the form $[y, 1]$, for some $y \in [0, 1]$, and that the corresponding optimal expected profit is the unique solution to a fixed point equation. Leveraging this property, we construct a stochastic-approximation type policy and show that its regret (the cumulative expected revenue loss compared with the optimal policy) after T time periods is $\mathcal{O}(\log T)$. In addition, relying on the Van Trees inequality (which can be seen as a Bayesian version of the well-known Cramér-Rao lower bound), we show that the worst-case regret for any policy grows as $\Omega(\log T)$, implying that our policy is asymptotically optimal.

◦ Inspired by analogous problems in the discrete setting, we then consider assortment optimization with a capacity constraint. We first show that the optimal assortment is not necessarily an interval anymore, but can have a much more complex structure. As a consequence it becomes necessary – in contrast to the uncapacitated case – to explore the whole product space in order to learn the optimal assortment. We propose a discretization policy and show that, up to a logarithmic term, its regret after T time periods is bounded from above by a constant times $T^{2/3}$. We then construct an instance in which the regret of any policy grows as $\Omega(T^{2/3})$, indicating that the capacitated setting indeed exhibits intrinsically different behavior than the uncapacitated case in which logarithmic regret is attainable. In addition, we propose a policy based on kernel density estimation, and show that under an additional assumption its regret after T time periods is bounded from above by a constant times $T^{2/3}$, up to logarithmic terms.

2.1.3 Outline

After providing an overview of relevant literature in Section 2.2, we introduce our model for continuous assortment optimization in Section 2.3. In Section 2.4 we study assortment optimization without capacity constraints: we propose a stochastic-approximation type policy, provide an upper bound on its regret, and prove a matching lower bound on the regret of any policy. The capacitated problem is discussed in Section 2.5: we propose a discretization policy, prove an upper bound on its regret, and prove a matching lower bound (up to a logarithmic term) on the regret of any policy. In addition, in Section 2.5.5 we discuss a policy based on kernel density estimation and in Section 2.5.6 we provide an upper bound on its regret. This policy and its regret analysis (relying on another assumption) are presented at the end of Section 2.5. For our numerical study regarding the performance of these policies, we refer to Section 4.1. The relation between the continuous and discrete logit choice model is discussed in Section 2.6. In Section 2.7, we present a bisection algorithm to compute the optimal continuous assortment. For our numerical study we refer to Chapter 4. Mathematical proofs are collected in Appendix A and an additional numerical study on the predictive performance of the continuous model is included in Section 4.3.

2.2 Literature

To put our work into the right perspective, we proceed by providing an account of the most relevant branches of the existing literature.

The idea of considering a continuous spectrum of products is a well-established concept in several branches of the literature. Within the economics literature, for example, this idea is studied in the context of vertical product differentiation and customer self-selection. The seminal work by Mussa & Rosen (1978) assumes a linear utility-based model in which a seller offers a continuous spectrum of quality levels and tries to optimally match customers of different types to prices and quality levels. Their model was generalized by Moorthy (1984) to include preferences that are nonlinear in the customer's type. More recently, Pan & Honhon (2012) considered vertical product differentiation in the context of assortment optimization, focusing on determining the

optimal positioning of products to offer and corresponding selling prices. Keskin & Birge (2019) consider a continuum of quality levels in a customer self-selection framework, and analyze dynamic learning of uncertain production costs. Den Boer et al. (2021) study the problem of optimally pricing and positioning a finite number of horizontally differentiated products represented by points on the unit interval, and design asymptotically optimal learning policies. Assortment optimization with product sets with a continuous structure have also been studied by Gaur & Honhon (2006) and Fisher & Vaidyanathan (2014), who both view products as entities in an attribute space and focus explicitly on modeling substitution for finding the optimal assortment. Another example is Dewan et al. (2003), which studies optimal product customization using the continuous, locational Salop model to determine an optimal (sub)spectrum of products to offer. With the exception of den Boer et al. (2021) and Keskin & Birge (2019), the literature mentioned above assumes that the model primitives are known to the seller.

The continuous choice model studied in the present chapter aligns well with the widely studied MNL choice model. Recently, several authors have studied assortment optimization under this choice model while assuming incomplete information: that is, the model parameters are unknown in advance and have to be learned from data. Rusmevichientong et al. (2010) focus on assortment optimization with a capacity constraint, and provide a bi-section algorithm to compute the optimal assortment under full information. Under incomplete information, they show under mild conditions that the expected loss (regret) of an explore-then-exploit type of algorithm after T time periods is bounded by a (instance-dependent) constant times $N^2 \log T$, where N denotes the number of products. Sauré & Zeevi (2013) consider a similar framework with a more general utility-based choice model, and implement procedures to quickly detect sub-optimal products. Agrawal et al. (2019) study an Upper Confidence Bound (UCB) algorithm for capacitated assortment optimization under the MNL model, and provide both a $\mathcal{O}(\sqrt{NT \log T})$ upper bound on the worst-case regret of their policy as well as an $\Omega(\sqrt{NT/K})$ lower bound for the regret of any policy, where N is the total number of products and K is the maximum number of products in the assortment. In addition, Agrawal et al. (2017) present a Thompson Sampling (TS) algorithm in the same setting, and provide an $\mathcal{O}(\sqrt{NT} \log TK)$ upper bound on the worst-case regret

of the policy.

The lower bound of Agrawal et al. (2019) is improved by Chen & Wang (2018) to $\Omega(\sqrt{NT})$, under the assumption that $K \leqslant N/4$. Without capacity constraint, Chen et al. (2021) provide an $\mathcal{O}(\sqrt{T})$ upper bound for the regret of their policy and an $\Omega(\sqrt{T})$ lower bound for the regret of any policy, under the assumption that only the first two products have positive marginal profit. A combination of a spatially structured product set and learning is studied by Ou et al. (2018). They present a learning algorithm for the assortment planning problem under the MNL model when the utility is a linear function of product attributes, as in the numerical study done by Rusmevichientong et al. (2010), and derive regret bounds.

The problem of learning the optimal assortment from accumulating data relates our work to *multi-armed bandit* (MAB) problems: a framework to study sequential learning-and-optimization problems. A central theme in these problems is to determine the optimal balance between exploration ('learning') and exploitation ('earning'). Classically, the number of arms is assumed to be finite (see, e.g., Robbins, 1952; Lai & Robbins, 1985; Agrawal, 1995; Auer et al., 2002). More recently, MAB problems have been studied where the action set is a continuum (see, e.g., Agrawal, 1995; Agarwal et al., 2011; Kleinberg, 2005; Auer et al., 2007; Kleinberg et al., 2008; Bubeck et al., 2009; Cope, 2009; Bubeck et al., 2011a,b; Flaxman et al., 2005; Shamir, 2013), or where the action set consists of a (typically large) number of combinatorial structures (see, e.g., Cesa-Bianchi & Lugosi, 2012; Chen et al., 2013; Combes et al., 2015). Our work is related to both these strands of literature: we study a MAB problem where the action sets consist of *subsets* of the unit interval, comprising a combinatorial MAB problem with uncountable action set. To the best of our knowledge, such a continuous, combinatorial MAB problem has not been considered before in the literature.

2.3 Model

We consider a seller of a commodity or service with an attribute that can be infinitesimally adjusted to any value in the interval $[0, 1]$. Each value in $[0, 1]$ is referred to as a product, and the seller has to decide which assortment of products, i.e., which subset

of $[0,1]$, to offer to each potential customer. Upon being offered an assortment, a customer either purchases a product from the assortment, or decides not to purchase – such a *no-purchase* is denoted by $\varnothing$. The *total collection of products* $\mathcal{X}$ is the union of the unit interval and the no-purchase option:

$$\mathcal{X} := [0,1] \cup \{\varnothing\}.$$

The goal of the seller is to identify an assortment that maximizes her expected revenue; as we shall see, this is not necessarily the entire interval $[0,1]$. We consider both capacitated and uncapacitated settings: in the former, the size of the assortment is bounded by a known constant $c < 1$, whereas in the latter case, this maximum size is $c = 1$. The set of feasible assortments is thus given by all (measurable) sets $S \subset [0,1]$ with volume at most c:

$$\mathcal{S} := \{S \in \mathcal{B}[0,1] : \mathrm{vol}(S) \leqslant c\},$$

where $\mathcal{B}[0,1]$ is the Borel sigma-algebra on $[0,1]$ and where

$$\mathrm{vol}(S) := \int_{x \in S} \mathrm{d}x.$$

For each product $x \in [0,1]$, the marginal revenue that the retailer obtains if x is purchased, is denoted by $w(x)$; no revenue is obtained from a no-purchase. We assume that w is a continuously differentiable function $[0,1] \to [0,1]$ with positive derivative bounded away from zero. It is worth observing that, in case x is a measure of quality, it is natural to assume that w is increasing.

For all $S \in \mathcal{S}$, we let X^S denote the random choice of an arbitrary customer who is offered assortment S. We assume the following structure on the distribution of X^S:

$$\mathbb{P}(X^S \in A) = \frac{\int_{x \in A} v(x)\mathrm{d}x}{1 + \int_{x \in S} v(x)\mathrm{d}x}, \tag{2.1}$$

for all (Borel measurable) $A \subseteq S$, and

$$\mathbb{P}(X^S = \varnothing) = \frac{1}{1 + \int_{x \in S} v(x)\mathrm{d}x},$$

where $v : [0,1] \to \mathbb{R}_+$ is an integrable function. The function v is referred to as the *preference function*, and is *unknown* to the seller. The expected revenue earned by

the seller after offering assortment $S \in \mathcal{S}$ to a customer is denoted by

$$r(S, v) := \frac{\int_{x \in S} v(x) w(x) \mathrm{d}x}{1 + \int_{x \in S} v(x) \mathrm{d}x}.$$

The aim of the seller is determining an assortment $S \in \mathcal{S}$ that maximizes $r(S, v)$. This is not directly possible, however, since the preference function is unknown. We, therefore, consider a sequential version of the problem that enables the seller to learn the optimal assortment from accumulating sales data. The seller offers assortments during $T \in \mathbb{N}$ consecutive time periods, indexed by $t = 1, \ldots, T$. Each time period t corresponds to a visit of a single customer. The assortment offered at time t is denoted by S_t, while $X_t \in \mathcal{X}$ denotes the (no-)purchase of the customer at time t. Conditionally on $S_t = S$, the purchase X_t is distributed as X^S, for all $S \in \mathcal{S}$ and all $t = 1, \ldots, T$.

The seller's decisions which assortments to offer are described by her *policy*: a sequence of mappings from available sales data (consisting of previously offered assortments and corresponding (no-)purchases) to a new assortment. Formally, a policy $\pi = (\pi_1, \ldots, \pi_T)$ is a vector of mappings $\pi_t : (\mathcal{S} \times \mathcal{X})^{t-1} \to \mathcal{S}$ such that

$$S_t = \pi_t(S_1, X_1, \ldots, S_{t-1}, X_{t-1}), \quad \text{for all } t = 1, \ldots, T; \tag{2.2}$$

here, we write $S_1 = \pi_1(\varnothing)$ for the initial assortment. Thus, a policy describes for each possible data-set of assortments and purchases how the seller selects the next assortment. The performance of a policy is measured by its *regret*: the cumulative expected loss caused by using sub-optimal assortments. Formally, the regret of a policy π is defined as

$$\Delta_\pi(T, v) := \sum_{t=1}^{T} \mathbb{E}_\pi \left[\max_{S \in \mathcal{S}} r(S, v) - r(S_t, v) \right], \tag{2.3}$$

where $S_1, \ldots, S_T$ satisfy (2.2), and where the subscript in the expectation operator indicates the dependence on the policy π. In the next sections we show that the maximum in (2.3) is attained. We also consider the *worst-case* regret over a class $\mathcal{V}$ of preference functions:

$$\Delta_\pi(T) := \sup_{v \in \mathcal{V}} \Delta_\pi(T, v).$$

The class of preference functions $\mathcal{V}$ under consideration consists of all functions v defined on the unit interval that satisfy the following assumptions.

Assumption 2.1. *(i) For all $v \in \mathcal{V}$ and $y \in [0,1]$,*

$$\underline{v} \leqslant v(y) \leqslant \overline{v},$$

for some $\overline{v} > \underline{v} > 0$ with $\overline{v} > w(0)/\int_0^1 (w(x) - w(0))\mathrm{d}x$.
(ii) All $v \in \mathcal{V}$ are differentiable on $(0,1)$ with uniformly bounded derivative, i.e.,

$$\sup_{y \in (0,1), v \in \mathcal{V}} |v'(y)| < \infty.$$

These assumptions are arguably mild, and allow us to obtain instance-independent regret upper bounds. If one is only interested in an instance-dependent bound of the form $\Delta_\pi(T, v) \leqslant C \log T$, where C may depend on v, then Assumption 1(ii) can be weakened; see Remark 2.4 for details. The assumption $\overline{v} > w(0)/\int_0^1 (w(x) - w(0))\mathrm{d}x$ is used in Section 2.4 to exclude trivialities; without this assumption, the unit interval $[0,1]$ is an optimal assortment for all $v \in \mathcal{V}$ (in case $c = 1$), and there is nothing to learn.

Remark 2.1. It is worth emphasizing that without assuming a particular structure of the choice probabilities $\mathbb{P}(X^S \in A)$, learning the optimal assortment from data is hopeless since the action space is uncountable. Our proposed model is motivated by its similarity to the well-known and frequently used *discrete* MNL choice model. In this model, the probability that a customer's choice lies in $A \subseteq S$ when being offered assortment S is equal to $\sum_{x \in A} v(x) / \big(1 + \sum_{x \in S} v(x)\big)$, for a function v defined on the product space and taking values in $(0, \infty)$. We essentially assume the same probabilistic structure, but with sums replaced by integrals.

Remark 2.2. The discrete MNL model can be derived from an assumed underlying random utility model in which a customer assigns utility $u(x) = \log(v(x)) + \varepsilon(x)$ to each product x and utility $\varepsilon(0)$ to the no-purchase option; here $\{\varepsilon(x)\}$ and $\varepsilon(0)$ are i.i.d. standard Gumbel distributed random variables. If the customer selects the

product (or no-purchase option) that maximizes her utility, then the probability that her choice lies in $A \subseteq S$ when being offered assortment S has a closed form and is equal to the above mentioned expression $\sum_{x \in A} v(x) / \big(1 + \sum_{x \in S} v(x)\big)$ (see Train, 2009, for a derivation). Whether a similar relation between choice probabilities and an underlying choice model exists when the product space is the continuum is not known. With uncountably many products, the arguments from the discrete case do not carry over, as one, e.g., would need to take a maximum over uncountably many random variables. Investigating the relation between choice probabilities and random utility models in case of a continuum of products is an interesting problem in its own right, but is outside the scope of this thesis. That said, our continuous model is closely connected to the discrete variant: it arises as a limit of discrete MNL models with the number of products N going to infinity, and, conversely, discretizing the continuum product space generates choice probabilities that are described by a discrete MNL model (see Section 2.6 for details). Furthermore, the policy that we propose in Section 2.5 to learn the optimal assortment with capacity constraint is effectively based on the fact that the continuous model can be approximated up to arbitrary precision by a discrete model.

Throughout this chapter, we use $[n]$ as a compact notation for the set $\{1, \dots, n\}$ (where $n \in \mathbb{N}$).

2.4 Uncapacitated Assortment Optimization

In this section, we investigate the uncapacitated case $c = 1$, in which the assortment can in principle cover the full interval $[0, 1]$. Our main finding is that the optimal asymptotic growth rate of regret is logarithmic in the time horizon. In what follows, we first show how to compute an optimal assortment. Next, we construct a policy and show that its regret is bounded from above by $\overline{C} \log T$ for some positive $\overline{C}$ independent of T. Then, we show that for *any* policy π the regret majorizes $\underline{C} \log T$ for some $\underline{C} > 0$ independent of T. This implies that our constructed policy achieves the smallest possible growth rate of regret, and is therefore asymptotically optimal.

The intuitive ideas underlying the mathematical statements in this section are given in the main text; the full proofs are contained in Section A.1.

2.4.1 Full Information Optimal Solution

It is known that the optimal assortment under the discrete MNL model without capacity constraints is of the form 'offer the k most profitable products' for some integer k (cf. Talluri & van Ryzin, 2004, Proposition 6). This result carries over to our model of continuous assortment optimization. Since we assume that products are labeled in such a way that w is increasing, the optimal assortment is of the form $[y, 1]$, for some $y \in [0, 1]$. The argument to show this is as follows (cf. Rusmevichientong et al., 2010, Section 2.1):

$$\begin{aligned}\max\{r(S, v) : S \in \mathcal{S}\} &= \max\{\varrho \in [0, 1] : \exists S \in \mathcal{S} : r(S, v) \geqslant \varrho\} \\ &= \max\left\{\varrho \in [0, 1] : \exists S \in \mathcal{S} : \int_S v(x)\big(w(x) - \varrho\big)\mathrm{d}x \geqslant \varrho\right\} \\ &= \max\left\{\varrho \in [0, 1] : \max_{S \in \mathcal{S}} \int_S v(x)\big(w(x) - \varrho\big)\mathrm{d}x \geqslant \varrho\right\}. \qquad (2.4)\end{aligned}$$

The inner maximization problem in (2.4) is maximized by $\{x \in [0, 1] : w(x) \geqslant \varrho\}$. Let $w^{-1}(\cdot)$ denote the generalized inverse of $w(\cdot)$, i.e.,

$$w^{-1}(\varrho) := \min\{x \in [0, 1] : w(x) \geqslant \varrho\}, \qquad \varrho \in [0, 1].$$

Since w is strictly increasing and continuous, the set $\{x \in [0, 1] : w(x) \geqslant \varrho\}$ is equal to the interval $[w^{-1}(\varrho), 1]$, and it follows that

$$\max\{r(S, v) : S \in \mathcal{S}\} = \max\{r([w^{-1}(\varrho), 1], v) : \varrho \in [0, 1]\}.$$

The fact that the optimal assortment is an interval of the form $[y, 1]$ has evident attractive computational implications, most notably that it reduces the original optimization problem over all subsets of the unit interval to an optimization problem in one variable $y \in [0, 1]$.

2.4.2 A Policy for Incomplete Information

We proceed by defining a data-driven policy that iteratively approximates the optimal assortment. The policy is parameterized by $\alpha \geqslant 1$ and $\beta \geqslant \alpha - 1$.

Stochastic Approximation Policy SAP(α, β)

1. Initialization. Let $\alpha \geqslant 1$, $\beta \geqslant \alpha - 1$ and $\varrho_1 \in [0,1]$. For all $t \in \mathbb{N}$ let $a_t := \alpha/(t+\beta)$. Put $t := 1$. Go to 2.
2. Assortment selection. Let

$$S_t := [w^{-1}(\varrho_t), 1], \quad R_t := w(X_t)\mathbf{1}\{X_t \in S_t\},$$

and

$$\varrho_{t+1} = \varrho_t + a_t\big(R_t - \varrho_t\big).$$

Put $t := t+1$. If $t \leqslant T$, then go to 2, else to 3.

3. Terminate.

The policy SAP(α, β) is a classic stochastic approximation policy (Robbins & Monro, 1951; Kushner & Yin, 1997) that aims at finding the value of $\varrho \in [0,1]$ such that $r([w^{-1}(\varrho), 1], v)$ equals ϱ. This condition uniquely defines the optimal ϱ that corresponds to the optimal assortment $[w^{-1}(\varrho), 1]$. Since only noisy observations R_t of the revenue function $r([w^{-1}(\varrho), 1], v)$ are available, the policy keeps changing ϱ_t based on observations of $R_t - \varrho_t$. The step sizes a_t decay roughly as $1/t$; this rate ensures that, on the one hand, ϱ_t does not converge 'too slowly' to the optimal value, while on the other hand, ϱ_t does not keep jumping 'over' the optimal ϱ which could potentially lead to a slow convergence rate.

2.4.3 Regret Upper Bound

We proceed by showing that the worst-case regret of SAP(α, β) grows at most logarithmically in T.

Theorem 2.1. *Let π correspond to* SAP(α, β) *with $\alpha \geqslant \overline{v}+1$ and $\beta \geqslant \alpha - 1$. Then, there is a $\overline{C} > 0$ such that, for all $T \geqslant 2$,*

$$\Delta_\pi(T) \leqslant \overline{C} \log T.$$

Write $g(y) := r([y,1], v)$ and $h(\varrho) := g(w^{-1}(\varrho))$, for $y, \varrho \in [0,1]$. The key idea

underlying the algorithm and the regret upper bound is the observation that the optimal expected revenue

$$\varrho^* := \max\{r(S, v) : S \in \mathcal{S}\},$$

solves the fixed-point equation

$$h(\varrho) = \varrho.$$

Because the noisy observation R_t has conditional expected value $h(\varrho_t)$, we can apply a Robbins-Monro scheme to find ϱ^* and the corresponding optimal assortment, without, e.g., having to estimate the gradient of the revenue function. This explains why we achieve a small regret rate of $\mathcal{O}(\log T)$ instead of, e.g., $\mathcal{O}(\sqrt{T})$ which is commonly seen in continuous multi-armed bandit problems.

Remark 2.3. The logarithmic growth rate of the regret in Theorem 2.1 holds for all choices of $\alpha \geqslant \overline{v} + 1$ and $\beta \geqslant \alpha - 1$. As the constant in front of the $\log T$ term may depend on these parameters, the finite-time performance of the policy may be fine-tuned by carefully selecting these α and β, for example based on initial simulations.

Remark 2.4. Theorem 2.1 presents a *worst-case* bound: the constant $\overline{C}$ is independent of $v \in \mathcal{V}$. To obtain this result we need to impose assumptions on uniform bounds on the derivative of $v \in \mathcal{V}$. If we are only interested in an *instance-dependent* upper bound $\Delta_\pi(T, v) \leqslant C_v \log T$, for all $v \in \mathcal{V}$ and some v-dependent constant $C_v > 0$, then Assumption 2.1.(ii) can be relaxed to v being continuously differentiable: this ensures inequality (A.2) in the proof of Lemma A.1.

2.4.4 Regret Lower Bound

Now that we have proven an upper bound on the regret of the policy SAP(α, β), we proceed by showing that this bound is, up to a multiplicative constant, asymptotically tight as T grows large. This implies that our policy is asymptotically optimal.

Theorem 2.2. *There is a $\underline{C} > 0$ such that, for all policies π and all $T \geqslant 2$,*

$$\Delta_\pi(T) \geqslant \underline{C} \log T.$$

To prove Theorem 2.2 we first define a collection of preference functions v_θ, indexed by a parameter θ that takes values in a closed interval Θ. Next, we show that the instantaneous regret incurred by offering assortment S instead of the optimal assortment $[y(\theta), 1]$ corresponding to θ, is bounded from below by a constant times the squared difference between the volumes of $[y(\theta), 1]$ and S, for any $S \in \mathcal{S}$ and $\theta \in \Theta$. This result is obtained by exploiting local quadratic behavior of the instantaneous regret for assortments close to the optimal one. Furthermore, this relation implies that it suffices to prove a lower bound on the mean squared error of any estimate of the *volume* of the optimal assortment: a reduction from subsets of $[0, 1]$ to one-dimensional variables in $[0, 1]$. To mitigate difficulties with the atom of the purchase distributions X^S on $\varnothing$, we define new, absolutely continuous random variables $Z_1, Z_2, \ldots$ and show that it suffices to prove a regret lower bound based on observations $Z_1, Z_2, \ldots$ instead of the purchases $X_1, X_2, \ldots$. Next, we bound the Fisher information corresponding to $Z_1, \ldots, Z_t$ from above by a positive constant times t, and define a probability measure λ on the support of θ. By the Van Trees inequality (Gill & Levit, 1995), we then conclude that the expected instantaneous regret in period $t + 1$, where the expectation is with respect to λ, is bounded from below by a constant times $1/t$, for all t. By summing over all $t = 1, \ldots, T$, the logarithmic lower bound follows.

2.5 Capacitated Assortment Optimization

In this section, we consider the setting in which the capacity c is strictly less than 1. We first characterize the optimal assortment under full information, and show that the optimal solutions in the capacitated case exhibit richer behavior than the intervals $[y, 1]$ observed in the uncapacitated case. Next, we show that this structural difference translates into a different complexity of the dynamic learning problem, finding that the optimal growth rate of the regret behaves as $T^{2/3}$ instead of $\log T$ as established in the previous section.

The intuitive ideas underlying the mathematical statements in this section are given in the main text; the full proofs are contained in Section A.2.

2.5.1 Full Information Optimal Solution

As shown in Section 2.4.1, the assortment optimization problem under full information can be written as

$$\max\{r(S,v) : S \in \mathcal{S}\} = \max\left\{\varrho \in [0,1] : \max_{S\in\mathcal{S}} \mathcal{I}(S,\varrho) \geqslant \varrho\right\}, \tag{2.5}$$

where

$$\mathcal{I}(S,\varrho) := \int_S v(x)\big(w(x)-\varrho\big)\mathrm{d}x,$$

for $S \in \mathcal{S}$ and $\varrho \in [0,1]$, and where $\mathcal{S}$ denotes the collection of all measurable subsets of the unit interval with volume at most c. Without a capacity constraint, $\mathcal{I}(S,\varrho)$ is maximized by the upper level set

$$W_\varrho := \{x \in [0,1] : w(x) \geqslant \varrho\},$$

for all $\varrho \in [0,1]$, since $v(x)(w(x)-\varrho)$ is nonnegative if and only if $x \in W_\varrho$. With capacity constraint, however, the optimization becomes slightly more subtle, because the set W_ϱ may have volume larger than c. We discuss how to solve the inner maximization problem in (2.5), i.e., how to construct an S_ϱ, for each $\varrho \in [0,1]$, such that

$$\mathcal{I}(S_\varrho,\varrho) = \max\{\mathcal{I}(S,\varrho) : S \in \mathcal{S}\}. \tag{2.6}$$

Next, we utilize this result to obtain an optimal solution for (2.5). To this end, let

$$h(x,\varrho) := v(x)\big(w(x)-\varrho\big), \qquad x \in [0,1], \varrho \in [0,1], \tag{2.7}$$

be the function that $\mathcal{I}$ integrates, let

$$L_\varrho(\ell) := \{x \in [0,1] : h(x,\varrho) \geqslant \ell\}, \qquad \varrho \in [0,1],\ \ell \in [0,\infty),$$

be the upper level sets of $h(\,\cdot\,,\varrho)$, and let

$$m_\varrho(\ell) := \mathrm{vol}(L_\varrho(\ell)), \qquad \varrho \in [0,1],\ \ell \in [0,\infty),$$

denote their volume. We first give an explicit characterization of the optimal solution(s) of (2.6).

Lemma 2.1. *Let* $\varrho \in [0,1]$.

(*i*) *If* $\mathrm{vol}(W_\varrho) \leqslant c$, *then the maximum of* $\mathcal{I}(S,\varrho)$ *over sets in* $\mathcal{S}$ *is attained by* $S = W_\varrho$.

(*ii*) *If* $\mathrm{vol}(W_\varrho) > c$, *then the maximum*

$$\ell_\varrho := \max\{\ell \geqslant 0 : m_\varrho(\ell) \geqslant c\}$$

exists, and the maximum of $\mathcal{I}(S,\varrho)$ *over sets in* $\mathcal{S}$ *is attained by* $S = L_\varrho^+ \cup L_\varrho^{\circlearrowleft}$, *where*

$$L_\varrho^+ := \{x \in [0,1] : h(x,\varrho) > \ell_\varrho\},$$

$$L_\varrho^= := \{x \in [0,1] : h(x,\varrho) = \ell_\varrho\},$$

and $L_\varrho^{\circlearrowleft}$ *is a subset of* $L_\varrho^=$ *such that* $\mathrm{vol}(S) = \mathrm{vol}(L_\varrho^+) + \mathrm{vol}(L_\varrho^{\circlearrowleft}) = c$.

As is intuitive, the upper level set W_ϱ maximizes $\mathcal{I}(S,\varrho)$ with respect to S if this does not result in a violation of the capacity constraint (case (*i*)). On the other hand, if the volume of W_ϱ exceeds the maximum capacity (case (*ii*)), then we construct an optimal assortment as follows. First, we 'fill' the assortment by the upper level set $\{x \in [0,1] : h(x,\varrho) > \ell\}$, where ℓ is as large as possible given the capacity constraint; this largest value of ℓ is denoted by ℓ_ϱ in Lemma 2.1. If the resulting assortment has size c then we are done; if not, then the function $h(x,\varrho)$ has 'flat' regions; that is, the level set $\{x \in [0,1] : h(x,\varrho) = \ell_\varrho\}$ has positive measure, and adding this set to the assortment would result in a violation of the capacity constraint. In that case, the optimal assortment S constructed in Lemma 2.1 consists of $\{x \in [0,1] : h(x,\varrho) > \ell_\varrho\}$ and a subset of $\{x \in [0,1] : h(x,\varrho) = \ell_\varrho\}$ such that the volume of the union of the two parts is exactly equal to c.

Based on the explicit solution of the inner maximization problem (2.6) given in Lemma 2.1, we now characterize an optimal solution to (2.5).

Proposition 2.1. *For each* $\varrho \in [0,1]$ *let* $S_\varrho \in \mathcal{S}$ *satisfy* (2.6). *Then, there is a unique solution* $\varrho^* \in [0,1]$ *to the fixed-point equation*

$$\mathcal{I}(S_\varrho,\varrho) = \varrho, \qquad \varrho \in [0,1],$$

and S_{ϱ^*} is an optimal assortment:

$$r(S_{\varrho^*}, v) = \max\{r(S, v) : S \in \mathcal{S}\}.$$

We prove the proposition by showing that $\mathcal{I}(S_\varrho, \varrho)$ is continuous and non-increasing as function of ϱ, with $\mathcal{I}(S_0, 0) \geqslant 0$ and $\mathcal{I}(S_1, 1) = 0$. By equality (2.5) and the observation

$$\mathcal{I}(S_\varrho, \varrho) = \varrho \quad \Longleftrightarrow \quad r(S_\varrho, v) = \varrho,$$

we conclude that if ϱ^* solves the fixed-point equation, then S_{ϱ^*} is an optimal assortment.

Remark 2.5. The optimal assortment can be efficiently computed up to any desired accuracy via a bisection method. In Section 2.7. we present an implementation of such a bisection algorithm.

Remark 2.6. In contrast to the setting discussed in Section 2.4, the optimal assortment in the presence of a capacity constraint does not have to be a connected interval. Consider, for example, the bi-modal preference function plotted in the left-hand panel of Figure 2.1, and let $c = 0.5$ and $w(x) = x$ for all $x \in [0, 1]$. The optimal assortment S^* in this instance consists of the union of two disjoint intervals:

$$S^* = [0.33, 0.48] \cup [0.63, 0.98],$$

with corresponding optimal expected profit $r(S^*, v) = 0.19$. In contrast, the largest expected profit that can be obtained from a single closed interval in this instance is equal to 0.13 (attained at the interval $[0.5, 1]$); a reduction in profit of more than thirty percent. This shows that restricting to single intervals can leave a significant amount of profit on the table.

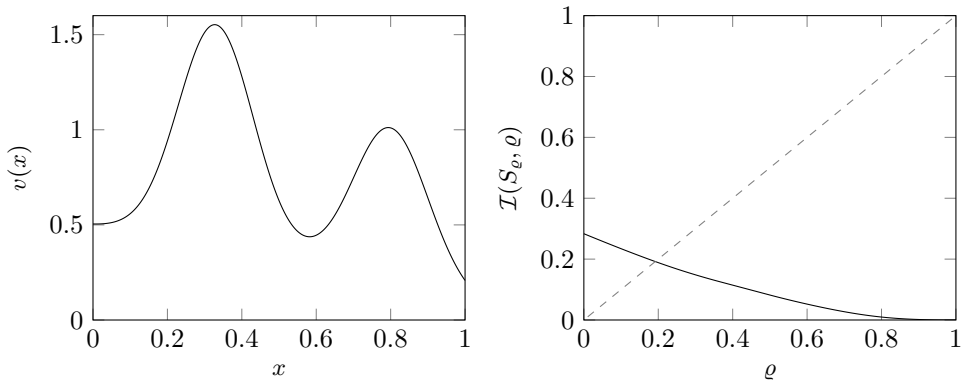

Figure 2.1: The left-hand panel shows the bi-modal preference function $v(x) = \frac{1}{10} + \frac{1}{5}(2 + x)(1-x) + \frac{2}{7}\varphi(x; 0.33, 0.1) + \frac{1}{5}\varphi(x; 0.8, 0.1)$, $x \in [0, 1]$, where $\varphi(\,\cdot\,; \mu, \sigma)$ is the normal probability density function with parameters μ and σ. The right-hand panel shows the corresponding function $\varrho \mapsto \mathcal{I}(S_\varrho, \varrho)$. The optimal $\varrho^* = 0.19$ is the unique ϱ such that $\varrho = \mathcal{I}(S_\varrho, \varrho)$.

The continuous model offers insight in the role of the capacity constraint in its discrete counterpart. To illustrate this, consider the instance of the discrete MNL assortment optimization problem discussed by Rusmevichientong et al. (2010) with $N = 4$ products, and preference values v_i and marginal revenues w_i given by

$$\boldsymbol{v} = (0.2, 0.6, 0.3, 5.2) \qquad \text{and} \qquad \boldsymbol{w} = (9.5, 9.0, 7.0, 4.5).$$

Rusmevichientong et al. (2010) shows that the optimal assortment, as function of the maximum assortment size C, is given by

C	1	2	3	4
Optimal assortment	$\{4\}$	$\{2, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 3, 4\}$

By defining

$$v(x) = N \sum_{i=1}^{N} v_i \mathbf{1}\left\{ \frac{i-1}{N} \leqslant x < \frac{i}{N} \right\},$$

and

$$w(x) = \sum_{i=1}^{N} w_i \mathbf{1}\left\{ \frac{i-1}{N} \leqslant x < \frac{i}{N} \right\},$$

for all $x \in [0, 1]$, we translate the problem into our continuous assortment optimization setting. For each fixed ϱ, the function $x \mapsto h(x, \varrho)$ defined in (2.7) is a piece-wise constant function that attains the values $N v_i(w_i - \varrho)$, for $i \in [N]$. The ordering of

the quantities $\{Nv_i(w_i - \varrho) : i \in [N]\}$ does not change when ϱ is slightly changed, except possibly if ϱ is of the form

$$\varrho_{i,j} := \frac{v_i w_i - v_j w_j}{v_i - v_j}, \qquad \text{for some } 1 \leqslant i < j \leqslant N.$$

If we consider the optimal revenue $\varrho^*(c)$ as function of the capacity constraint c, then it follows that the fraction of a product that is included in the optimal assortment might be discontinuous at points c such that $\varrho^*(c) = \varrho_{i,j}$, for some i, j. In our example, this happens at $c \approx 0.32$, $c \approx 0.61$, and $c \approx 0.66$. Figure 2.2 illustrates this behavior. The fraction of a particular product that is included in the optimal assortment is not monotone in c, and can in fact make jumps.

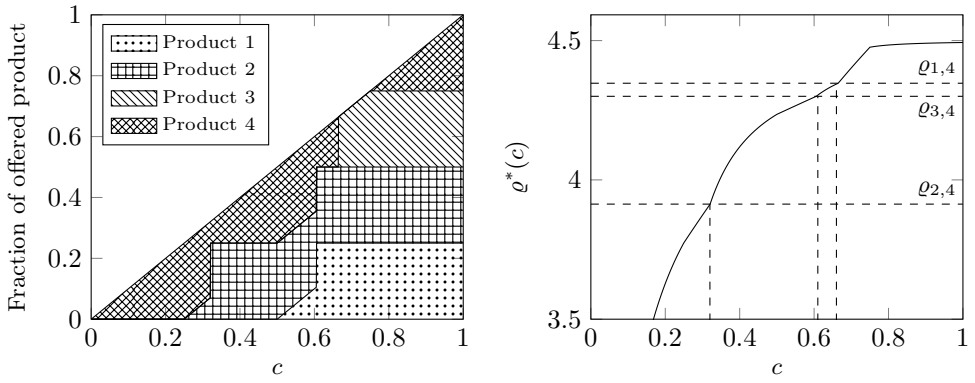

Figure 2.2: The left-hand panel shows the optimal amount of each products, as function of c. The right-hand panel shows the corresponding optimal expected profit $\varrho^*(c)$.

2.5.2 A Discretization Policy for Incomplete Information

We proceed by presenting a discretization policy for the continuous assortment optimization problem with capacity constraint and incomplete information. We first discuss the underlying intuition and the method of establishing upper confidence bounds, after which we formally present our policy Discretized Upper Confidence Bounds (DUCB).

The proposed policy is parameterized by an integer $N \in \mathbb{N}$. The policy DUCB(N) discretizes the set of products $[0, 1]$ into N bins of equal size, after which the policy exploits the similarity with the discrete MNL model. This is done by applying the

UCB policy from (Agrawal et al., 2019, (Algorithm 1)) to the bin structure. We regard a continuous purchase in the i-th bin as a purchase of product i in the discrete MNL model. The policy establishes upper confidence bounds on the preference parameters corresponding to the discrete MNL model. More specifically, define the bins as

$$B_i := \left[\frac{i-1}{N}, \frac{i}{N}\right) \tag{2.8}$$

for $i = 1, \dots, N-1$ and

$$B_N := \left[\frac{N-1}{N}, 1\right], \tag{2.9}$$

and define the parameters

$$v_i := \int_{B_i} v(x)\,\mathrm{d}x \qquad \text{and} \qquad w_i := N \int_{B_i} w(x)\,\mathrm{d}x, \qquad i \in [N].$$

Note that by our choice of v_i and w_i for $i \in [N]$, the expected profit of an assortment consisting of a collection of bins is the same for the continuous and discrete MNL model.

At each time t we observe a purchase $X_t \in S_t \cup \{\varnothing\}$, and translate this X_t to a discrete purchase Y_t by

$$Y_t := \sum_{i=1}^{N} i\,\mathbf{1}\{X_t \in B_i\}.$$

Observe that $X_t \in B_{Y_t}$ if $X_t \in S_t$ and $Y_t = 0$ if $X_t = \varnothing$. The policy at time t computes upper confidence parameters $v_{1,t}^{\mathrm{UCB}}, \dots, v_{N,t}^{\mathrm{UCB}}$ of the parameters $v_1, \dots, v_N$ using observed discrete purchases $Y_1, \dots, Y_t$. In the next step, at time $t+1$, the chosen assortment is the collection of bins $S_{t+1} = \bigcup_{i \in D_{t+1}} B_i$ where D_{t+1} is a subset of $[N]$ of size at most $\lfloor cN \rfloor$, which maximizes

$$D \mapsto \frac{\sum_{i \in D} v_{i,t}^{\mathrm{UCB}} w_i}{1 + \sum_{i \in D} v_{i,t}^{\mathrm{UCB}}}.$$

If such an optimal assortment is not unique, ties are broken by applying an arbitrary, fixed ordering of assortments.

The DUCB(N) policy starts by setting $v_{i,0}^{\mathrm{UCB}} = 1$ for all $i \in [N]$. To compute the upper confidence parameters $v_{1,t}^{\mathrm{UCB}}, \dots, v_{N,t}^{\mathrm{UCB}}$ for $t = 1, \dots, T$, the observed discrete purchases $Y_1, \dots, Y_t$ are used as follows. The time horizon is partitioned into epochs, where each epoch corresponds to a sequence of consecutive actual purchases. An

epoch ends when a no-purchase is observed, i.e., $X_t = \varnothing$ or, equivalently, $Y_t = 0$. Specifically, let $t_0 := 0$ and recursively define

$$t_\ell := \min\{t \in \{t_{\ell-1}+1, \dots, T\} : Y_t = 0\}, \qquad \ell \in \mathbb{N}_{\geqslant 1},$$

and $t_\ell := T$ if $\{t \in \{t_{\ell-1}+1, \dots, T\} : Y_t = 0\} = \varnothing$. Let L denote the first index such that $t_L = T$, that is,

$$L := \min\{\ell \in \mathbb{N}_{\geqslant 1} : t_\ell = T\}.$$

Then, the ℓ-th epoch $\mathcal{E}_\ell$ is defined as

$$\mathcal{E}_\ell := \{t_{\ell-1}+1, \dots, t_\ell\}, \qquad \ell \in [L].$$

Within each epoch $\mathcal{E}_\ell$ the upper confidence parameters remain unchanged, that is, $v_{i,t}^{\mathrm{UCB}} = v_{i,s}^{\mathrm{UCB}}$ for all $i \in [N]$ when $s, t \in \{t_{\ell-1}, \dots, t_\ell - 1\}$. As a result, and by the fixed tie-breaking rule, D_t remains the same within each epoch. Define $D^\ell := D_{t_{\ell-1}+1}$. At the end of an epoch, the upper confidence parameters are updated. Then, the upper confidence bounds $v_{1,t}^{\mathrm{UCB}}, \dots, v_{N,t}^{\mathrm{UCB}}$ become

$$v_{i,t}^{\mathrm{UCB}} := \begin{cases} \bar{v}_{i,\ell} + \sqrt{\bar{v}_{i,\ell} \dfrac{48\,\log(\sqrt{N}\ell+1)}{|\mathcal{T}_i(\ell)|}} + \dfrac{48\,\log(\sqrt{N}\ell+1)}{|\mathcal{T}_i(\ell)|}, & \text{if } t = t_\ell \text{ for some } \ell \in [L] \text{ and } i \in D^\ell, \\ v_{i,t-1}^{\mathrm{UCB}}, & \text{otherwise.} \end{cases} \tag{2.10}$$

Here, $\mathcal{T}_i(\ell)$ is the set of epochs up to ℓ in which product i is offered, that is,

$$\mathcal{T}_i(\ell) := \{\tau \in [\ell] : i \in D^\tau\}, \qquad i \in [N],$$

and $\bar{v}_{i,\ell}$ is the average of the number of times product i is purchased in epoch τ for epochs $\tau \in \mathcal{T}_i(\ell)$, that is,

$$\bar{v}_{i,\ell} := \frac{1}{|\mathcal{T}_i(\ell)|} \sum_{\tau \in \mathcal{T}_i(\ell)} \sum_{t \in \mathcal{E}_\tau} \mathbf{1}\{Y_t = i\}.$$

For all $i \in D^\ell$, $\bar{v}_{i,\ell}$ is an unbiased estimator of the discrete preference parameters v_i (see Corollary A.1 by Agrawal et al., 2019). Note that in (2.10) there exists an $\ell \in [L]$ such that $t = t_\ell$ if and only if $Y_t = 0$.

After the verbal description, we now formally present our DUCB(N) policy.

Discretized Upper Confidence Bounds DUCB(N)

1. Initialization. Let $N \in \mathbb{N}$ and put $K := \lfloor cN \rfloor$. Let B_i for $i \in [N]$ be as in (2.8) and (2.9). Let $w_i := N \int_{B_i} w(x)\mathrm{d}x$ and $v_{i,0}^{\mathrm{UCB}} := 1$ for $i \in [N]$ and $t := 1$. Go to 2.
2. Assortment selection. Let

$$D_t \in \underset{D \subseteq [N] : |D| \leqslant K}{\arg\max} \frac{\sum_{i \in D} v_{i,t-1}^{\mathrm{UCB}} w_i}{1 + \sum_{i \in D} v_{i,t-1}^{\mathrm{UCB}}}, \tag{2.11}$$

and

$$S_t := \bigcup_{i \in D_t} B_i.$$

Determine $v_{1,t}^{\mathrm{UCB}}, \ldots, v_{N,t}^{\mathrm{UCB}}$ as in (2.10), and let $t := t + 1$. If $t \leqslant T$, then go to 2, else to 3.

3. Terminate.

If the discrete assortment D_t as in (2.11) is not unique, ties are dealt with by applying an arbitrary fixed ordering of assortments.

2.5.3 Regret Upper Bound for Discretization

We proceed by showing that the worst-case regret of DUCB(N), with appropriately chosen N, grows at most as $T^{2/3}$ up to a logarithmic term.

Theorem 2.3. *Let π correspond to* DUCB(N) *with $N = \lfloor \gamma T^{1/3} \rfloor$ where $\gamma = \max\{\overline{v}, 1/c + 1\}$. Then, there is a $\overline{C} > 0$ such that, for all $T \geqslant 2$,*

$$\Delta_\pi(T) \leqslant \overline{C}\, T^{2/3} (\log T)^{1/2}.$$

To prove the theorem, we first establish a relation between the regret in our model and that of the discrete regret in the context of Agrawal et al. (2019). There is an obvious misalignment between those two notions: one deals with functions and the other with discrete parameters. However, we are able to bound the regret of DUCB(N) from above by the regret of UCB plus a discretization error of order T/N.

Since the regret of UCB is of order $\sqrt{NT}$ (up to a logarithmic term), the optimal value of N is proportional to $T^{1/3}$ which results in a $T^{2/3}$ upper bound for the regret of DUCB(N) (also up to a logarithmic term).

Then, it is observed that the discretization error consists of three sources. The first source is due to the fact that the discrete model approximates the actual preference function and marginal profit function by a piecewise constant function. The second source is caused by the fact that the true optimal assortment is not necessarily exactly equal to a collection of bins. The third source is the effect of the misalignment between the regret within our model with that of the regret of UCB as analyzed by Agrawal et al. (2019). When considering the regret of DUCB(N), we need to take this translation error into account.

To facilitate the analysis of the performance of DUCB(N), we define

$$\check{v}(x) := N \sum_{i=1}^{N} \mathbf{1}\{x \in B_i\} \int_{B_i} v(y)\,\mathrm{d}y, \qquad x \in [0,1], \tag{2.12}$$

$$\check{w}(x) := N \sum_{i=1}^{N} \mathbf{1}\{x \in B_i\} \int_{B_i} w(y)\,\mathrm{d}y, \qquad x \in [0,1]. \tag{2.13}$$

In addition, we introduce an adjustment of the currently used notation of the expected profit of an assortment $S \in \mathcal{S}$. We will explicitly denote that this expected profit depends on marginal profit function $w(x)$, as well as preference function $v(x)$:

$$r(S, v, w) := \frac{\int_S v(x)w(x)\,\mathrm{d}x}{1 + \int_S v(x)\,\mathrm{d}x}.$$

The effect of the first component of the discretization error is captured by Proposition 2.2 below.

Proposition 2.2. *Let $\check{v}$ and $\check{w}$ be as in* (2.12) *and* (2.13)*, respectively. Let S^* and $\check{S}$ in $\mathcal{S}$ be optimal assortments corresponding to v and w, and $\check{v}$ and $\check{w}$, respectively, that is,*

$$r(S^*, v, w) = \max_{S \in \mathcal{S}} r(S, v, w) \qquad \textit{and} \qquad r(\check{S}, \check{v}, \check{w}) = \max_{S \in \mathcal{S}} r(S, \check{v}, \check{w}). \tag{2.14}$$

Then, the difference between the expected revenue of S^* under v and w and the expected

revenue of $\check{S}$ under $\check{v}$ and $\check{w}$ is bounded from above by

$$r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) \leqslant \|v - \check{v}\|_1 + \overline{v}\,\|w - \check{w}\|_1, \tag{2.15}$$

where $\|\cdot\|_1 := \int_0^1 |\cdot|\,\mathrm{d}x$.

Note that the optimal assortment $\check{S}$ in the result stated above is the optimal assortment within $\mathcal{S}$. The UCB algorithm only considers discrete assortments, which translates to a collection of bins within our model. The effect of this is stated in Lemma 2.2 below.

Lemma 2.2. *Let $\check{v}$ and $\check{w}$ be as in* (2.12) *and* (2.13)*, respectively. Let $\mathcal{A}_K$ be the set of all collections of at most $K = \lfloor cN \rfloor$ bins B_i, that is,*

$$\mathcal{A}_K := \left\{ \bigcup_{i \in D} B_i : D \subset [N] \;\text{ and }\; |D| \leqslant K \right\}. \tag{2.16}$$

In addition, let $\check{S}$ in $\mathcal{S}$ and S^d in $\mathcal{A}_K$ be optimal assortments corresponding to $\check{v}$ and $\check{w}$, that is,

$$r(\check{S}, \check{v}, \check{w}) = \max_{S \in \mathcal{S}} r(S, \check{v}, \check{v}) \qquad \text{and} \qquad r(S^d, \check{v}, \check{w}) = \max_{S \in \mathcal{A}_K} r(S, \check{v}, \check{v}). \tag{2.17}$$

Then, the difference between the expected revenue under $\check{v}$ and $\check{w}$ of $\check{S}$ and S^d is bounded from above by

$$r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) \leqslant \frac{\overline{v}}{N}.$$

Recall that the first two components address the effect of the discretization error regarding the specifics of the optimal assortment. The third and last component concerns the translation error regarding the offered assortments $S_1, \ldots, S_T$. Since all these assortments lie in $\mathcal{A}_K$, as in (2.16), we present the result below for a general set in $\mathcal{A}_K$.

Lemma 2.3. *Let $\check{v}$ and $\check{w}$ be as in* (2.12) *and* (2.13)*, respectively. Let $\mathcal{A}_K$ be as in* (2.16) *and let $S \in \mathcal{A}_K$. Then, the difference between the expected profit of S under $\check{v}$*

and $\check{w}$, and v and w is bounded from above by

$$r(S, \check{v}, \check{w}) - r(S, v, w) \leqslant \|v - \check{v}\|_1 + \overline{v}\, \|w - \check{w}\|_1,$$

where $\|\cdot\|_1 := \int_0^1 |\cdot|\, \mathrm{d}x$.

The three components of the discretization error are combined as follows. Let S^*, $\check{S}$ and S^d be as in (2.14) and (2.17) and let $S_1, \ldots, S_T$ be the offered assortments. The instantaneous regret at time $t \in [T]$ can be split into four parts as

$$\begin{aligned} r(S^*, v, w) - r(S_t, v, w) = {} & r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) + && (2.18)\\ & r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) + && (2.19)\\ & r(S^d, \check{v}, \check{w}) - r(S_t, \check{v}, \check{w}) + && (2.20)\\ & r(S_t, \check{v}, \check{w}) - r(S_t, v, w). && (2.21) \end{aligned}$$

The idea is to apply the triangle inequality. For the right-hand side of (2.18), (2.19), and (2.21), we apply Proposition 2.2, Lemma 2.2 and Lemma 2.3, respectively. Note that the term in (2.20) corresponds to the instantaneous regret of UCB. The remainder of the proof of Theorem 2.3 consists of showing that both the L_1-differences $\|v - \check{v}\|_1$ and $\|w - \check{w}\|_1$ are of the order $1/N$ and applying Theorem 1 from Agrawal et al. (2019).

Remark 2.7. The analysis of the upper bound on the regret of DUCB extends to higher dimensional continuous assortment problems. In particular, if the dimension is $d \geqslant 2$, then one can discretize the set of products $[0, 1]^d$ into N^d bins. Under a smoothness assumption of the preference function and the marginal profit function, the order of the L_1-difference between the actual functions and the discretized functions remains $\mathcal{O}(1/N)$ as the difference can be bounded from above by a sum of N^d terms that each is of order $N^{-(d+1)}$, similar as in (A.12). As a result, the cumulative discretization error is of order T/N and the total regret in higher dimensions is of the order (up to a logarithmic factor)

$$\frac{T}{N} + \sqrt{N^d T}.$$

Hence, the optimal value of N is proportional to $T^{\frac{1}{d+2}}$ which results in a $T^{\frac{d+1}{d+2}}$ regret.

This corresponds to the regret rate for continuum-armed bandit in higher dimensions (see, e.g., Kleinberg et al., 2008; Bubeck et al., 2011a,b).

2.5.4 Regret Lower Bound

In this section, we construct an instance for the assortment optimization problem with capacity constraint, and we show that the regret of *any* policy after T time periods is at least a constant times $T^{2/3}$. This shows that the structural differences between optimal assortments with or without a capacity constraint under full information – Section 2.4.1 and 2.5.1 – translate into a different complexity of the corresponding data-driven optimization problem, characterized by the growth rate of the regret.

We consider the following instance. Let $\underline{v} \in (0, 0.04)$ and $\overline{v} \geqslant 9$. Moreover, let $c \in (0, 0.25]$, $s = 0.05c$ and $\delta = 0.2$, and consider the marginal profit function

$$w(x) = (1-s)\frac{1-\delta}{1-\delta x} + s, \qquad x \in [0,1].$$

To obtain a lower bound on the regret, we construct 'difficult instances' of preference functions that are hard to distinguish statistically, but that correspond to different optimal assortments. To this end, we first define a 'baseline' preference function v_0 by

$$v_0(x) := \frac{s}{c(w(x)-s)} = \frac{s(1-\delta x)}{c(1-s)(1-\delta)}, \qquad x \in [0,1].$$

This preference function has the property that $\varrho_0^* := \max_{S \in \mathcal{S}} r(S, v_0)$ is equal to s (see Section A.2, Lemma A.5), and that $v_0(x)(w(x) - \varrho_0^*)$ does not depend on x. As a result, *any* assortment of volume c is optimal for this preference function.

The next step is to perturb the baseline preference function with small, positive 'bumps' at different locations such that the corresponding optimal assortment will be a collection of intervals centered around these bumps. The perturbed preference functions are, in a sense, close to each other (measured, e.g., by the L_1 norm), but correspond to different and possibly even disjoint optimal assortments. In particular, let $K \geqslant 2$ be an integer and $N_K := \lfloor K/c \rfloor$, and define the *i-th bin* as the interval

$$B_i := \left[c\frac{i-1}{K}, c\frac{i}{K} \right), \qquad i \in [N_K].$$

Note that this definition differs from the bins presented in Section 2.5.2. The definition

here is convenient as the union of any K distinct bins has combined volume of precisely c. Let $\mathcal{D}_K$ denote the collection of all subsets of $[N_K]$ of size K, i.e.,

$$\mathcal{D}_K := \big\{ I \subseteq [N_K] : |I| = K \big\}.$$

For each collection of bins $I \in \mathcal{D}_K$ we now define a preference function v_I that, roughly speaking, consists of the baseline preference function with small, positive bumps added at all bins B_i, $i \in I$. In particular, define the bump function $b(x)$ as the normal probability density function with parameters $\mu = 0$ and $\sigma = 0.3$:

$$b(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \qquad x \in \mathbb{R}.$$

This function is shifted and re-scaled such that the probability mass on $[-1, 1)$ is mapped onto B_i, as follows. For $i \in [N_K]$ and $x \in \mathbb{R}$, let

$$\varphi_i(x) := \frac{2Kx}{c} - 2i + 1,$$

be a linear transformation that satisfies $\varphi_i(B_i) = [-1, 1)$, and define

$$\tau_i(x) := \frac{c}{K}\, b\big(\varphi_i(x)\big).$$

Finally, define the constant

$$\beta := \frac{c}{K} \frac{1}{\sigma\sqrt{2\pi}} \sum_{n \in \mathbb{Z}} \exp\left(-\frac{(2n-1)^2}{2\sigma^2} \right),$$

and, for each $I \in \mathcal{D}_K$, define the preference function

$$v_I(x) := v_0(x) \left(1 + \sum_{i \in I} \tau_i(x) - \beta \right), \qquad x \in [0, 1].$$

The subtraction of the (small) constant β ensures that $v_I(x) \leqslant v_0(x)$ for all $x \notin \bigcup_{i \in I} B_i$, i.e., the preference function dips just below the baseline function $v_0(x)$ for x outside the collection of bins in I. This ensures that the optimal assortment corresponding to v_I is approximately equal to the collection of intervals $\bigcup_{i \in I} B_i$ at which small bumps have been added.

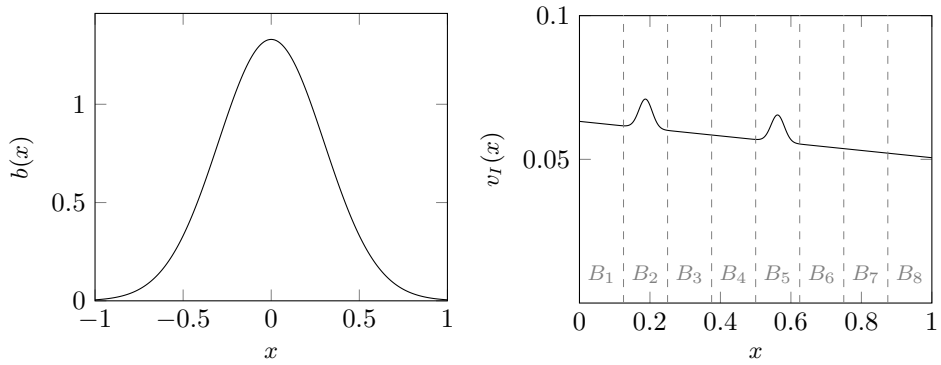

Figure 2.3: Left: bump function $b(x)$. Right: preference function $v_I(x)$ for $c = 0.25$, $K = 2$ and $I = \{2, 5\}$.

Having defined a collection of preference functions, we now proceed in proving a regret lower bound. First, for any policy and any $I \in \mathcal{D}_K$, we bound the regret corresponding two preference function v_I from below by an expression that counts how often products from the approximately optimal assortment $\bigcup_{i\in I} B_i$ were not offered. To state the result, let

$$\varepsilon_I(x) := \frac{v_I(x) - v_0(x)}{v_0(x)} = \sum_{i\in I} \tau_i(x) - \beta, \qquad I \in \mathcal{D}_K, x \in [0, 1],$$

and let

$$k(x) := \sum_{t=1}^{T} \mathbf{1}\{x \in S_t\}, \qquad x \in [0, 1],$$

count the number of times that $x \in [0, 1]$ is offered to consumers. Throughout the remainder of this section we fix an arbitrary policy π, and let $\mathbb{P}_I$ and $\mathbb{E}_I$ denote the probability law and the expectation operator under policy π and preference function v_I.

Proposition 2.3. *There are constants $C_1 > 0$, $C_2 > 0$, independent of π, such that, for any $T \in \mathbb{N}$ and $I \in \mathcal{D}_K$,*

$$\Delta_\pi(T, v_I) \geqslant C_1 \int_{\bigcup_{i\in I} B_i} \big(T - \mathbb{E}_I[k(x)]\big)\varepsilon_I(x)\mathrm{d}x - C_2 \frac{T}{K}.$$

The proposition is proven by exploiting the structure of the optimal assortment as outlined in Section 2.5.1 and the fact that the definition of v_I implies that the corresponding optimal assortment is approximately equal to $\bigcup_{i\in I} B_i$. The constants C_1, C_2 are given explicitly in the proof of Proposition 2.3.

The second step in the proof of the regret lower bound is the following result, which provides an upper bound on how the expected number of times that a product $x \in [0,1]$ is offered changes when the preference function is changed from v_I to $v_{I\setminus\{i\}}$, for some $i \in I$.

Proposition 2.4. *Let $x \in [0,1]$, $I \in \mathcal{D}_K$, and $J = I\setminus\{i\}$ for some $i \in I$. Then, there is a constant $C_c > 0$ independent of π, such that*

$$\left|\mathbb{E}_I[k(x)] - \mathbb{E}_J[k(x)]\right| \leqslant C_c \left(\frac{T}{K}\right)^{3/2}. \tag{2.22}$$

This bound is proven by relating the left-hand side of (2.22) to the Kullback-Leibler (KL) divergence of $\mathbb{P}_I$ and $\mathbb{P}_J$, using Pinsker's inequality, and subsequently bounding this expression from above by carefully analyzing its dependence on v_I and v_J. The constant C_c is given explicitly in the proof of Proposition 2.4.

With Propositions 2.3 and 2.4 at hand, we finally arrive at our regret lower bound.

Theorem 2.4. *There is a $\underline{C} > 0$, independent of π, such that, for $T \in \mathbb{N}$,*

$$\Delta_\pi(T) \geqslant \underline{C}\, T^{2/3}.$$

To prove the theorem, we first show that the preference functions $\{v_I : I \in \mathcal{D}_K, K \in \mathbb{N}\}$ satisfy Assumption 2.1. This implies that the worst-case regret is bounded from below by the expected regret when the preference function is chosen uniformly at random from $\{v_I : I \in \mathcal{D}_K\}$, for any fixed K. The regret corresponding to each v_I is then bounded from below by an expression that involves the expected number of times that products from the approximate optimal assortment $\bigcup_{i\in I} B_i$ are not offered, using Proposition 2.3. Proceeding in a similar fashion as in the proof of the regret lower bound obtained by Chen & Wang (2018) for discrete assortments, while dealing with all the intricacies of having a continuum product space, we connect

the expression in Proposition 2.3 to statement (2.22) of Proposition 2.4. By carefully selecting K, we arrive at the stated lower bound.

2.5.5 A Density Estimation Policy for Incomplete Information

We proceed by defining a density estimation policy for the continuous assortment optimization problem with capacity constraint and incomplete information. We first formally present our Kernel Density Estimation Policy (KDEP), after which we discuss the underlying intuition and our method of estimating the unknown preference function. The proposed policy is parameterized by an integer $M \in \mathbb{N}$.

Kernel Density Estimation Policy KDEP(M)

1. Initialization. Let $M \in \mathbb{N}$ and $J := \lceil 1/c \rceil$. For $i \in [J]$, put

$$S^i := \left[\frac{i-1}{J-1}(1-c), \frac{i-1}{J-1}(1-c) + c \right]. \tag{2.23}$$

   Put $t := 1$. Go to 2.

2. Exploration. Put $S_t := S^{\lceil t/M \rceil}$ and $t := t+1$. If $t \leqslant MJ$, then go to 2, else to 3.

3. Estimation. Put $\hat{v}(x)$ as in (2.25) and let

$$\hat{S} \in \arg\max_{S \in \mathcal{S}} r(S, \hat{v}). \tag{2.24}$$

   Go to 4.

4. Exploitation. Let

$$S_t = \hat{S}.$$

   Put $t := t+1$. If $t \leqslant T$, then go to 4, else to 5.

5. Terminate.

For how to compute $\hat{S}$ as in (2.24), we refer to Section 2.7. The policy presented above consists of two phases. In the initial *exploration* phase, a number of test assortments offered, each during M time periods. The observed purchases in this exploration phase are used to determine an estimate $\hat{v}$ of the true but unknown

preference function v. Our estimation method is based on kernel density estimation (KDE), and is explained in more detail below. Based on this estimate $\hat{v}$, we compute an optimal assortment $\hat{S}$. In the second, *exploitation* phase, the statistical knowledge obtained in the exploration phase is exploited by offering the estimated optimal assortment $\hat{S}$ during the remainder of the time horizon. The number of times M that each of the test assortments is offered during the exploration phase is a key tuning parameter that captures the well-known exploration-exploitation trade-off: choosing M large will improve the quality of the estimated optimal assortment $\hat{S}$, but choosing M small will decrease the total loss from the exploration phase. In Theorem 2.5 we will carefully choose M and derive a regret upper bound on KDEP(M).

The estimate $\hat{v}$ is established in the following manner. First, for all $x \in [0,1]$, let $k(x)$ denote the number of times that product x is contained in the test assortments $S^1, \ldots, S^J$:

$$k(x) := \sum_{i=1}^{J} \mathbf{1}\{x \in S^i\}, \qquad x \in [0,1].$$

Observe that the test assortments are constructed in such a way that $k(x) > 0$ for all $x \in [0,1]$. For each test assortment S^i we construct a corresponding estimate $\hat{v}_i(x)$ of $v(x)\mathbf{1}\{x \in S^i\}$, and then combine these into our estimate $\hat{v}$, as follows:

$$\hat{v}(x) := \frac{1}{k(x)} \sum_{i=1}^{J} \hat{v}_i(x), \qquad x \in [0,1]. \tag{2.25}$$

To define $\hat{v}_i$, define the Legendre polynomials

$$\varphi_0(x) := \frac{1}{\sqrt{2}}, \quad \varphi_j(x) := \sqrt{\frac{2j+1}{2}} \frac{1}{2^j j!} \frac{\mathrm{d}^j}{\mathrm{d}x^j} \left[(x^2-1)^j\right],$$

for $j \in \mathbb{N}$, which form an orthonormal basis in $L_2([-1,1])$. Write $S^i = [a_i, b_i]$ as in (2.23), for all $i \in [J]$, let $h \in (0, c/2]$ be a bandwidth parameter determined below (see Proposition 2.6), and for all $i \in [J]$ and $x \in \mathbb{R}$ define two shift coefficients γ_x^i and ζ_x^i as

$$\left(\gamma_x^i\,,\,\zeta_x^i\right) = \begin{cases} \left(\frac{2h}{h+x-a_i}\,,\quad -\frac{h-(x-a_i)}{h+x-a_i}\right), & \text{for } x \in [a_i, a_i+h), \\ (1,\quad 0), & \text{for } x \in [a_i+h, b_i-h], \\ \left(\frac{2h}{h+b_i-x}\,,\quad \frac{h-(b_i-x)}{h+b_i-x}\right), & \text{for } x \in (b_i-h, b_i]. \end{cases}$$

In addition, we define the shifted support I_x^i as

$$I_x^i = \left[-\min\left\{1, \tfrac{x-a_i}{h}\right\}, \min\left\{1, \tfrac{b_i-x}{h}\right\}\right].$$

and define the *Legendre kernel of order* ℓ for S^i by

$$K_x^i(u) := \gamma_x \sum_{j=0}^{\ell} \varphi_j\big(\zeta_x\big)\varphi_j\big(\gamma_x u + \zeta_x\big), \qquad x \in S^i,\ u \in I_x^i,$$

and $K_x^i(u) := 0$ for $x \in S^i$ and $u \notin I_x^i$.

Since $v(x)\mathbf{1}\{x \in S^i\}$ is not a proper density, we re-scale the kernel estimator based on the number of (no)-purchases corresponding to test assortment S^i, for all $i \in [J]$. To this end, let E_i denote the no-purchases observed when assortment S^i is offered:

$$E_i := \{X_t : X_t = \varnothing \text{ and } (i-1)M + 1 \leqslant t \leqslant iM\},$$

and let

$$A_i := \{X_t : X_t \neq \varnothing \text{ and } (i-1)M + 1 \leqslant t \leqslant iM\}$$

denote the actual purchases observed when S^i is offered. Then, $v(x)\mathbf{1}\{x \in S^i\}$ is estimated by

$$\hat{v}_i(x) := \frac{1}{(|E_i|+1)h} \sum_{X \in A_i} K_x^i\left(\frac{X-x}{h}\right), \qquad x \in S^i,$$

and set $\hat{v}_i(x) := 0$ for $x \notin S^i$. These estimates are combined into one estimate $\hat{v}$ of v, as given by (2.25).

Remark 2.8. Because traditional KDE does not perform well near endpoints of the support, we construct our estimate of v based on the so-called *boundary kernel method*, that locally adjusts the kernels near the edges of the support (see Müller, 1991; Zhang et al., 1999, for other demonstrations of this method). Also contrary to traditional KDE, we allow the order of the kernel to depend on the number of observations. To construct such a kernel of arbitrarily high order, it is natural to work with an orthonormal basis of polynomials. We specifically choose Legendre polynomials since this choice allows us to bound the convergence rate explicitly for kernels of flexible order.

2.5.6 Regret Upper Bound for Density Estimation

In this section, we provide an upper bound on the regret of KDEP(M), with appropriately chosen M. We conduct our analysis under an additional assumption on the derivatives of the unknown preference function:

Assumption 2.2. *There is a $C > 0$ such that*

$$\left| \frac{v^{(\ell)}(y)}{(\ell+1)!} \right| < C,$$

for all $y \in (0,1)$ and $\ell \in \mathbb{N}$.

We denote the class of functions that satisfy Assumption 2.2 as $\mathcal{V}_1$. It is worth emphasizing that Assumption 2.2 is satisfied by many commonly used functions (such as polynomials, sines, cosines and exponential functions, as well as sums, products and compositions of these). In addition, in the operations research and statistics literature such an assumption on the derivatives of the function of interest is often imposed (see, e.g., Wang et al., 2021; Tsybakov, 2008). Weakening the smoothness conditions on v leads to different convergence rates, as explained in more detail in Remark 2.10.

Theorem 2.5. *Let π correspond to* KDEP(M) *with $M = \lfloor T^{2/3}/J \rfloor$ where $J = \lceil 1/c \rceil$. Then, there is a $\overline{C} > 0$, for all v that satisfy Assumption 2.2, such that, for all $T \geqslant 2$,*

$$\Delta_\pi(T, v) \leqslant \overline{C}\, T^{2/3} (\log T)^{1/2}.$$

To prove the theorem, we first show that the instantaneous expected revenue loss caused by using an estimated optimal assortment $\hat{S}$, based on an estimate $\hat{v}$ of v, is bounded from above by the L_1-difference between v and $\hat{v}$. It is worth noting that this result closely resembles Proposition 2.2. Although, here we consider the optimal assortment under v and the optimal assortment under $\hat{v}$, both with respect to the same marginal profit function w, and we evaluate the difference in expected revenue between both assortments under specifically the actual preference function v and the marginal profit function w. We present the result below separately, as neither this result nor Proposition 2.2 is a simple corollary of the other.

Proposition 2.5. *Let $v, \hat{v} : [0,1] \to \mathbb{R}^+$. Let S^* and $\hat{S}$ in $\mathcal{S}$ be optimal assortments corresponding to v and $\hat{v}$, respectively, that is,*

$$r(S^*, v) = \max_{S \in \mathcal{S}} r(S, v) \qquad \text{and} \qquad r(\hat{S}, \hat{v}) = \max_{S \in \mathcal{S}} r(S, \hat{v}).$$

Then, the difference between the expected revenue under v of S^* and $\hat{S}$ is bounded from above by

$$r(S^*, v) - r(\hat{S}, v) \leqslant 2 \left\| v - \hat{v} \right\|_1,$$

where $\| \cdot \|_1 := \int_0^1 | \cdot | \, \mathrm{d}x$.

Now, let $\hat{v}$ specifically be the estimate of v obtained after the exploration phase. The next step is to bound the estimation error $\|v - \hat{v}\|_1$ in terms of M, the length of the exploration phase. Observe that we cannot directly use existing results on convergence rates of kernel density estimators, since $\hat{v}$ is not a conventional KDE: it is composed of the estimates $\hat{v}_i$, $i \in [J]$, which also depend on the number of no-purchases observed. In particular, each estimate $\hat{v}_i$, for $i \in [J]$, can be written as the product of two separate estimates $\hat{\alpha}_i$ and $\hat{f}_i$, where

$$\hat{\alpha}_i = \frac{|A_i|}{|E_i| + 1} \qquad \text{and} \qquad \hat{f}_i(x) = \frac{1}{|A_i| h} \sum_{X \in A_i} K_x^i \left(\frac{X - x}{h} \right) \tag{2.26}$$

are estimates of

$$\alpha_i := \int_{S^i} v(x) \mathrm{d}x \qquad \text{and} \qquad f_i(x) := \frac{v(x)}{\int_{S^i} v(y) \mathrm{d}y},$$

respectively. To analyze their convergence rates, we denote the no-purchase probability with respect to test assortment S^i as $p_i := \frac{1}{1+\alpha_i}$ and we let $\varepsilon_i := \frac{1}{2} \min\{p_i, 1 - p_i\}$, for $i \in [J]$. Morover, we define the *clean event* as

$$\mathcal{E} := \left\{ \forall i \in [J] : \frac{|E_i|}{M} \in (p_i - \varepsilon_i, p_i + \varepsilon_i) \right\}.$$

We show that the probability of the complement of $\mathcal{E}$ is $\mathcal{O}(M^{-1/2})$. On the event $\mathcal{E}$, we determine convergence rates of $\hat{\alpha}_i$, $\hat{f}_i$, and $\hat{v}$, in the following proposition.

Proposition 2.6. *Let $M \in \mathbb{N}$, $J = \lceil 1/c \rceil$ and let π correspond to* KDEP(M). *In addition, let $i \in [J]$ and $M_i = \max\{\frac{1}{p_i}, \frac{7}{1-p_i}\}$, and let $\hat{\alpha}_i$ and $\hat{f}_i$ be as in* (2.26). *Furthermore, set the bandwidth h^* and order $\ell^* = \lceil \beta^* \rceil$ such that*

$$h^* := \min\left\{\frac{c}{2}, \frac{1}{e}\right\} \qquad \text{and} \qquad \beta^* := \frac{1}{2}\log\big(-2|A_i|\log h^*\big) - \frac{1}{2}.$$

Then, for $M \geqslant M_i$,

$$\mathbb{E}_\pi\left[\, |\alpha_i - \hat{\alpha}_i| \;\middle|\; \mathcal{E}\,\right] \leqslant C_1 \frac{1}{\sqrt{M}} \tag{2.27}$$

and

$$\mathbb{E}_\pi\left[\int_{x\in S^i} \Big(f_i(x) - \hat{f}_i(x)\Big)^2 \mathrm{d}x \;\middle|\; \mathcal{E}\right] \leqslant C_2 \frac{\log M}{M}, \tag{2.28}$$

for universal constants C_1 and C_2. Moreover, let $\hat{v}$ be as in (2.25). *Then, the expected L_1-difference between v and $\hat{v}$, conditioned on the clean event, can be bounded for $M \geqslant \max_i M_i$ as*

$$\mathbb{E}_\pi\Big[\, \|v - \hat{v}\|_1 \;\Big|\; \mathcal{E}\,\Big] \leqslant C_3 \frac{(\log M)^{1/2}}{M^{1/2}}, \tag{2.29}$$

where C_3 is a universal constant.

Remark 2.9. Convergence rates on the (integrated) mean squared error of KDE can, for example, be found in Devroye & Györfi (1985) and Tsybakov (2008). As we are not aware of existing literature that provides convergence rates applicable to our specific context, we have included a full derivation of the proof of the desired convergence rates.

Putting everything together, we obtain that the regret on the complement of the event $\mathcal{E}$ is $\mathcal{O}(TM^{-1/2})$, while the regret on the event $\mathcal{E}$ is $\mathcal{O}(M + TM^{-1/2}(\log M)^{1/2})$; here, the first term corresponds to the duration of the exploration phase, while the second term corresponds to the expected loss during the exploitation phase. Choosing M proportional to $T^{2/3}$ yields a regret rate of $T^{2/3}(\log T)^{1/2}$.

Remark 2.10. If, instead of Assumption 2.2, we would only assume that $v(\cdot)$ is ℓ times continuously differentiable, for some given $\ell \in \mathbb{N}$, then analogously to Proposi-

tion 2.6 we can show that there exists a positive constant C such that

$$\mathbb{E}_\pi\Big[\, \|v - \hat{v}\|_1 \,\Big|\, \mathcal{E} \,\Big] \leqslant C M^{-\frac{\ell}{2\ell+1}}.$$

By repeating our analysis it follows that choosing M proportionally to $T^{\frac{2\ell+1}{3\ell+1}}$ then leads to an upper bound of the form

$$\Delta_\pi(T, v) \leqslant \overline{C}\, T^{\frac{2\ell+1}{3\ell+1}},$$

for some constant $\overline{C}$.

Remark 2.11. One could also try to use existing results for continuum-armed bandits to approach the capacitated assortment problem. However, there is in principle no upper bound on the number of (disjoint) subintervals of which the optimal assortment is composed. If we neglect this and assume that the optimal assortment consists of at most k disjoint subintervals, we essentially obtain a continuous multi-armed bandit problem with dimension $d = 2k$ (each subinterval contributing a beginning and an end point) and linear constraints (which guarantee that subintervals do not overlap and the capacity constraint is satisfied). A typical regret rate for continuum-armed bandit problems with not necessarily concave reward functions is $T^{\frac{d+1}{d+2}}$ (see, e.g., Kleinberg et al., 2008; Bubeck et al., 2011a,b). Thus, even if $k = 1$ and $d = 2$, this approach would endure a regret rate strictly higher than $T^{2/3}$.

2.6 Relation to Discrete MNL Choice Probabilities

The choice probabilities in our continuous assortment optimization model are closely connected to the discrete MNL model, in two regards.

First, our choice probabilities naturally arise as a limit of discrete models where the number of products grow large. To see this, consider a sequence of discrete MNL assortment optimization problems indexed by $n \in \mathbb{N}$, where the n-th problem corresponds to a setting with n products labeled $i = 1, \ldots, n$, each with associated location $i/(n+1)$ and valuation $v_i^{(n)} = v(i/(n+1))/(n+1)$, for all $i = 1, \ldots, n$ and some continuous function $v : [0, 1] \to \mathbb{R}^+$. Under the discrete MNL model, the

probability that a customer selects a product in a (measurable) set $A \in [0,1]$ when being offered assortment S is equal to

$$\frac{\sum_{i:\frac{i}{n+1}\in A} v_i^{(n)}}{1+\sum_{i:\frac{i}{n+1}\in S} v_i^{(n)}}.$$

It follows from classical results in integration theory (see, e.g., Stroock, 1994) that this expression converges to (2.1) as $n \to \infty$.

Second, when the product space is discretized into finitely many products, each corresponding to a subinterval in $[0,1]$, then our model translates into choice probabilities that are described by a discrete MNL model. To see this, suppose that $I_1, \ldots, I_n$ are mutually disjoint subsets of $[0,1]$, each corresponding to a 'discrete product' such that $\bigcup_{i=1}^n I_i = [0,1]$. Let $v_i := \int_{I_i} v(x)\mathrm{d}x$, for all i. Then, for each 'discrete assortment' $\tilde{S} \subseteq \{1,\ldots,n\}$ and for each $i \in \tilde{S}$, the probability $P(i \mid \tilde{S})$ that a customer selects from I_i when being offered assortment $\bigcup_{j\in\tilde{S}} I_j$, is equal to

$$\mathbb{P}(\,i\,|\,\tilde{S}\,) = \mathbb{P}\left(X^S \in I_i\right) = \frac{\int_{I_i} v(x)\mathrm{d}x}{1+\int_{\bigcup_{j\in\tilde{S}} I_j} v(x)\mathrm{d}x} = \frac{v_i}{1+\sum_{j\in\tilde{S}} v_j}.$$

This is precisely the structure of a discrete MNL choice model.

2.7 Bisection Algorithm for Section 2.5

According to Proposition 2.1, the optimal assortment can be computed up to any desired accuracy $\varepsilon > 0$. The algorithm COA(n) below shows how this is done, where $n := -\log\varepsilon$. Recall that

$$\mathcal{I}(S,\varrho) := \int_S v(x)\big(w(x)-\varrho\big)\mathrm{d}x. \tag{2.30}$$

The algorithm COA(n) uses bisection to find the fixed-point solution ϱ^* to the equation

$$\mathcal{I}(S_\varrho,\varrho) = \varrho.$$

The value of $\mathcal{I}(S_\varrho,\varrho)$ is computed by relying on the level ℓ_ϱ. This level value is calculated by an additional inner bisection using the algorithm IB(n,ϱ). This algorithm is also presented below.

Remark 2.12. As mentioned, the calculation of the level ℓ_ϱ for a single ϱ requires a bisection on its own. This means that the run time of $\mathrm{IB}(n, \varrho)$ is $\mathcal{O}(-\log \varepsilon)$, and hence the run time of $\mathrm{COA}(n)$ is $\mathcal{O}((\log \varepsilon)^2)$.

Capacitated Optimal Assortment $\mathtt{COA}(n)$

1. Initialization. Let $n \geqslant 1$. Put $a := 0$, $b := 1$, $\mathtt{piv} := (b-a)/2$ and $i := 1$. Go to 2.
2. Capacity check. Put

$$W_{\mathtt{piv}} := \{x \in [0,1] : w(x) \geqslant \mathtt{piv}\}.$$

(i) If $\mathrm{vol}(W_{\mathtt{piv}}) > c$, then go to 3.

(ii) If $\mathrm{vol}(W_{\mathtt{piv}}) \leqslant c$, then put $S_{\mathtt{piv}} := W_{\mathtt{piv}}$ and $I_{\mathtt{piv}} := \mathcal{I}(S_{\mathtt{piv}}, \mathtt{piv})$ as in (2.30) and go to 5.

3. Inner bisection. Compute $\ell_{\mathtt{piv}}$ according to $\mathrm{IB}(n, \mathtt{piv})$. Go to 4.
4. Level set. Put

$$L^+_{\mathtt{piv}} := \{x \in [0,1] : v(x)(w(x) - \mathtt{piv}) > \ell_{\mathtt{piv}}\},$$

$$L^=_{\mathtt{piv}} := \{x \in [0,1] : v(x)(w(x) - \mathtt{piv}) = \ell_{\mathtt{piv}}\}$$

and

$$x_{\mathtt{piv}} := \min\{x \in [0,1] : \mathrm{vol}(L^+_{\mathtt{piv}}) + \mathrm{vol}\big(L^=_{\mathtt{piv}} \cap [0,x]\big) = c\}$$

Put $S_{\mathtt{piv}} = L^+_{\mathtt{piv}} \cup \big(L^=_{\mathtt{piv}} \cup [0, x_{\mathtt{piv}}]\big)$ and $I_{\mathtt{piv}} := \mathcal{I}(S_{\mathtt{piv}}, \mathtt{piv})$ as in (2.30). Go to 5.

5. Pivot.

(i) If $I_{\mathtt{piv}} > \mathtt{piv}$, then put $a := \mathtt{piv}$.

(ii) If $I_{\mathtt{piv}} \leqslant \mathtt{piv}$, then put $b := \mathtt{piv}$.

Put $i := i + 1$. If $i \leqslant n$, then put $\mathtt{piv} := (b-a)/2$ and go to 2, else go to 6.

6. Optimization. Put $S^* := S_{\texttt{piv}}$. Go to 7.

7. Terminate.

Recall that there is a possible degree of freedom for picking S_ϱ if $\mathrm{vol}(W_\varrho) > c$. By the definition of $x_{\texttt{piv}}$ above, we explicitly choose the left-most version. The algorithm $\mathrm{IB}(n, \varrho)$ computes the level ℓ_ϱ for given ϱ. Recall by Lemma 2.1 that this level is defined as

$$\ell_\varrho := \max\{\ell \geqslant 0 : \mathrm{vol}\big(L(\varrho, \ell)\big) \geqslant c\}.$$

$\mathrm{IB}(n, \varrho)$ also uses the bisection method, which is facilitated by the fact that, as a function of $\ell \geqslant 0$, $\mathrm{vol}\big(L(\varrho, \ell)\big)$ is left-continuous and non-increasing by Lemma A.3.

Inner Bisection $\mathrm{IB}(n, \varrho)$

1. Initialization. Let $n \geqslant 1$ and $\varrho \in [0, 1]$. Put $a := 0$, $b := v_{\max}(w_{\max} - \varrho) + 1$, $\texttt{piv} := (b - a)/2$ and $i := 1$. Go to 2.

2. Level set. Put

$$L^{\texttt{piv}} := \{x \in [0, 1] : v(x)(w(x) - \varrho) \geqslant \texttt{piv}\}.$$

Go to 3.

3. Pivot.

(i) If $\mathrm{vol}(L^{\texttt{piv}}) > c$, then put $a := \texttt{piv}$.

(ii) If $\mathrm{vol}(L^{\texttt{piv}}) \leqslant c$, then put $b := \texttt{piv}$.

Put $i := i + 1$. If $i \leqslant n$, then put $\texttt{piv} := (b - a)/2$ and go to 2, else go to 4.

4. Optimization. Put $\ell_\varrho := \texttt{piv}$. Go to 5.

5. Terminate.

2.8 Concluding Remarks

In this chapter, we introduce the concept of continuous assortment optimization with demand learning. We distinguish between the capacitated and uncapacitated cases,

revealing intrinsically different regret behavior: we show that the asymptotically optimal regret rate in the absence of a capacity constraint grows logarithmically in the time horizon, whereas imposing a capacity constraint leads to $T^{2/3}$ regret. To our knowledge, this chapter is the first to extend discrete assortment optimization problems to the continuous realm.

Our work points to various directions for future research. First, the customer-purchase model used in this chapter is the natural continuous equivalent of the well-studied discrete multinomial logit choice model. It remains an open question how one constructs a random utility model that serves as a theoretical justification of the continuous choice model. Second, in line with the majority of the assortment optimization literature, our set-up assumes that product prices are exogenous. A question of practical interest is to consider price and assortment decisions simultaneously in our continuous model, potentially in a competitive setting. Third, we have constructed an example in which the optimal assortment is not an uninterrupted interval. It would be interesting to study under which conditions a single interval solution is optimal, and whether one can bound the maximum loss when the decision-maker is restricted to offering a single interval.

Chapter 3

Discrete Assortment Optimization

3.1 Introduction

Assortment optimization is vital for maximizing revenue. A seller of a large number of substitute products faces the challenge of determining the most profitable subset, i.e., assortment, of products to offer to consumers. In most practical situations, the seller does not know the specific demand distribution for all assortments, so that the optimal assortment optimization has to be studied in a sequential optimization framework with incomplete information. As the infrastructure of information improves to better handle incoming real-time purchase data, the necessity for computationally efficient and easily implementable data-driven algorithms arises.

Because the number of feasible assortments grows exponentially in the number of products, the problem is often studied under a particular choice model that describes how demand or choice probabilities depend on the offered assortment. Perhaps the most widely studied choice model is the multinomial logit (MNL) choice model, considering either the *uncapacitated* or the *capacitated* variant. In the latter setting, the seller is restricted in the size of the assortments that can be offered.

When the seller can offer assortments of *any* size, the optimal assortment under the MNL model is of the form 'offer the k most profitable products' – see, e.g., Talluri & van Ryzin (2004, Proposition 6). This structure greatly simplifies the problem as the seller merely has to learn the optimal value of the one-dimensional quantity k

instead of high-dimensional model parameters, suggesting that a conceptually simple stochastic-approximation type of policy might work well. In this chapter, we construct such a stochastic approximation policy for the dynamic assortment optimization problem without capacity constraint and with demand characterized by the MNL model with unknown parameters, and show that it is both asymptotically optimal and has excellent numerical performance.

Recently, for the dynamic assortment optimization problem *with* capacity constraint, Chen & Wang (2018) show that the regret – the cumulative expected revenue loss caused by offering suboptimal assortments – that any decision policy endures is bounded from below by a constant times $\sqrt{NT}$, where N denotes the number of products and T denotes the time horizon. This result is shown under the assumption that the product revenues are *constant*, and thus leaves the question open whether a lower regret rate can be achieved for non-constant revenue parameters. In this chapter, we show that this is not the case: we show that, for *any* vector of product revenues there is a positive constant such that the regret of any policy is bounded from below by this constant times $\sqrt{NT}$. Our result implies that policies that achieve $\mathcal{O}(\sqrt{NT})$ regret are asymptotically optimal for all product revenue parameters.

3.1.1 Literature

Within the management science and operations research community, the problem of dynamic assortment planning has recently received much attention (see, e.g., Rusmevichientong et al., 2010; Farias et al., 2013; Sauré & Zeevi, 2013; Agrawal et al., 2017; Cheung & Simchi-Levi, 2017; Chen & Wang, 2018; Ou et al., 2018; Agrawal et al., 2019; Kallus & Udell, 2020; Chen et al., 2021). All these papers study assortment optimization under the MNL model in a sequential decision framework. An important recent contribution closely related to this chapter is Chen et al. (2021). The authors construct a Trisection policy that exploits the structure of the optimal assortment under the uncapacitated MNL assortment optimization problem ('offer the k most profitable products'), and prove that its regret after T time periods is bounded by $\sqrt{T}$ times a constant *that does not depend on the number of products.* This improves earlier regret rates that do depend on the number of products – see, e.g., the bounds derived in Rusmevichientong et al. (2010, Theorem 3.4), Sauré &

Zeevi (2013, Theorem 4), Agrawal et al. (2017, Theorem 1) and Agrawal et al. (2019, Theorem 1, 3 and 4), applied to a maximum capacity of $K = N$ products. In addition, Chen et al. (2021) construct an instance where the marginal revenue is 1 for odd-numbered products and 1/2 for even-numbered products, and prove that, for this instance, the worst-case regret over a set of possible model parameters that any policy must endure is bounded from below by a constant times $\sqrt{T}$. This implies that the $\sqrt{T}$ regret rate can in general not be improved.

This chapter is also related to Chapter 2, where we study a stochastic approximation policy in the context of dynamic assortment optimization under a *continuous* logit choice model. Here, the set of feasible products is the unit interval. The stochastic approximation policy that we propose in Section 3.3 can been seen as a discrete counterpart of the policy constructed and analyzed in Section 2.4 but with different regret rates (and different proof techniques) caused by the structural differences between continuous and discrete assortment optimization. This is visible in the regret rates that we derive: in the continuous setting studied in Chapter 2 a regret growth rate of $\log T$ is optimal, whereas in the discrete setting studied in this chapter $\sqrt{T}$ is the best attainable rate.

Regarding *capacitated* assortment optimization, two notable contributions are from Agrawal et al. (2017) and Agrawal et al. (2019), who construct decision policies based on Thompson Sampling and Upper Confidence Bounds, respectively. They show that the regret of these policies is bounded by a constant times $\sqrt{NT}$ (up to logarithmic terms), where N denotes the number of products and $T \geqslant N$ denotes the length of the time horizon. These upper bounds are complemented by the recent work from Chen & Wang (2018), who show that the regret of any policy is bounded from below by a positive constant times $\sqrt{NT}$, implying that the policies by Agrawal et al. (2017) and Agrawal et al. (2019) are (up to logarithmic terms) asymptotically optimal.

The lower bound by Chen & Wang (2018) is proven under the assumption that the product revenues are *constant* – that is, each product generates the same amount of revenue when sold. In practice, it often happens that different products have *different* marginal revenues, and it is a priori not completely clear whether the policies by Agrawal et al. (2017) and Agrawal et al. (2019) are still asymptotically optimal or

that a lower regret can be achieved. In addition, Chen & Wang (2018) assume that K, the maximum number of products allowed in an assortment, is bounded by $\frac{1}{4} \cdot N$, but point out that this constant $\frac{1}{4}$ can probably be increased.

3.1.2 Contributions and Outline

In this chapter, we propose an easily implementable and asymptotically optimal data-driven policy for the uncapacitated assortment optimization problem under the MNL model with unknown parameters. Our policy is based on stochastic approximation and exploits structural properties of the optimal assortment so that not all unknown model parameters have to be learned from data. The policy does not require the time horizon as input. Under a mild positivity assumption on the no-purchase probability, we prove that the regret of our policy is bounded from above by $\sqrt{T}$ times a constant independent of the number of products. In addition, we prove a $\sqrt{T}$ regret lower bound that any policy must endure for *any* given vector of product revenues. This slightly generalizes the lower bound proven by Chen et al. (2021), and implies that policies with $\mathcal{O}(\sqrt{T})$ regret are asymptotically optimal *for any* product revenue parameters. Moreover, we conduct numerical experiments (see Section 4.2) that demonstrate that our policy has a robust performance in different instances, and can outperform alternative algorithms by a significant margin when T and the number of N are not too small. We emphasize that our policy is not the first to have been shown to be asymptotically optimal (that is achieved by Chen et al., 2021); the value of our policy lies in the fact that it is easy to understand and implement, and has superior numerical performance when N and T are not too small.

For dynamic assortment optimization *with* capacity constraint K, we prove a $\sqrt{NT}$ regret lower bound for *any* given vector of product revenues. This implies that policies with $\mathcal{O}(\sqrt{NT})$ regret are asymptotically optimal *regardless* of the product revenue parameters. Furthermore, our result is valid for all $K < \frac{1}{2}N$, thereby confirming the intuition of Chen & Wang (2018) that the constraint $K \leqslant \frac{1}{4}N$ is not tight.

This chapter is organized as follows. We introduce the model in Section 3.2. Section 3.3 regards uncapacitated assortment optimization. Here, we present our policy and the upper bound on its regret, as well as the lower bound result of the regret of any policy. In Section 3.4, we provide the regret lower bound and its mathematical

proof for capacitated assortment optimization. For our numerical study we refer to Section 4.2. Mathematical proofs from Section 3.3 are collected in the Appendix B.

3.2 Model

We consider a seller who has $N \in \mathbb{N}$ different products for sale during $T \in \mathbb{N}$ time periods, and who has to determine at the beginning of each time period which subset of products is available for purchase. We abbreviate the set of products $\{1 \dots, N\}$ as $[N]$ and the set of time instances $\{1, \dots, T\}$ as $[T]$. Each product $i \in [N]$ yields a known marginal revenue of $w_i > 0$. Without loss of generality due to scaling, we can assume that $w_i \leqslant 1$ for all $i \in [N]$. Each product $i \in [N]$ is associated with a preference parameter $v_i \geqslant 0$, unknown to the seller. Each offered assortment $S \subseteq [N]$ must satisfy a potential capacity constraint, i.e., $|S| \leqslant K$ for capacity constraint $K \in \mathbb{N}$, $K \leqslant N$. Note that, for uncapacitated assortment optimization; $K = N$. For notational convenience, we write

$$\mathcal{A}_K := \{S \subseteq [N] : |S| \leqslant K\}$$

for the collection of all feasible assortments of size at most K.

At the beginning of each time period $t \in [T]$ the seller selects an assortment $S_t \in \mathcal{A}_K$ based on the purchase information available up to and including time $t-1$. Thereafter, the seller observes a purchase $Y_t \in S_t \cup \{0\}$, where product 0 corresponds to a no-purchase. Clearly, $w_0 = 0$ and we set v_0 – the preference parameter for product 0 – equal to 1. The purchase probabilities under the MNL model are given by

$$\mathbb{P}(Y_t = i \,|\, S_t = S) = \frac{v_i}{1 + \sum_{j \in S} v_j}, \qquad \text{for all } i \in S \cup \{0\}.$$

The expected revenue earned by the seller from an assortment $S \in \mathcal{A}_K$ is denoted by

$$r(S, v) := \frac{\sum_{i \in S} v_i w_i}{1 + \sum_{i \in S} v_i}.$$

The assortment decisions of the seller are described by his/her policy: a collection of

probability distributions $\pi = (\pi(\cdot \mid h) : h \in H)$ on $\mathcal{A}_K$, where

$$H := \bigcup_{t \in [T]} \{(S, Y) : Y \in S \cup \{0\}, S \in \mathcal{A}_K\}^{t-1}$$

is the set of possible histories, and where, conditionally on $h = (S_1, Y_1, \ldots, S_{t-1}, Y_{t-1})$, assortment S_t has distribution $\pi(\cdot \mid h)$, for all $h \in H$ and all $t \in [T]$. Let $\mathbb{P}^{\pi}_{vj}$ denote the probability measure of $\{S_t, Y_t : t \in \mathbb{N}\}$ under policy π and preference vector v, and let $\mathbb{E}^{\pi}_{v}$ be the corresponding expectation operator. The objective for the seller is to find a policy π that maximizes the total accumulated revenue or, equivalently, minimizes the accumulated regret:

$$\Delta_\pi(T, v) := \sum_{t=1}^{T} \mathbb{E}^{\pi}_{v} \left[\max_{S \in \mathcal{A}_K} r(S, v) - r(S_t, v) \right].$$

In addition, we consider the worst-case regret over a class $\mathcal{V}$ of preference vectors:

$$\Delta_\pi(T) := \sup_{v \in \mathcal{V}} \Delta_\pi(T, v).$$

The class of preference vectors $\mathcal{V}$ under consideration consists of all vectors v that either satisfy Assumption 3.1 for Section 3.3 or Assumption 3.2 for Section 3.4.

3.3 Uncapacitated Assortment Optimization

In this section, we consider dynamic assortment optimization under the MNL model without capacity constraint. Here, we discuss the structure of the optimal assortment, propose a policy for incomplete information and show that its regret is bounded by $\sqrt{T}$ times a constant independent of N. Then, we show that, for arbitrary revenue parameters, the regret of any policy is bounded from below by a constant times $\sqrt{T}$, implying that our policy is asymptotically optimal in a general setting. The mathematical proofs of the results stated in this section are collected in the Appendix B.

Within this section, we consider the class of preference vectors $\mathcal{V}$ that satisfy the following assumption.

Assumption 3.1. *The $\mathcal{V}$ is the set of all components-wise non-negative vectors*

$(v_1, \ldots, v_N)$ *such that*

$$\frac{1}{1 + \sum_{i=1}^{N} v_i} \geqslant p_0,$$

for some $p_0 \in (0, 1)$ *known to the seller.*

This assumption is arguably mild as it simply puts a lower bound p_0 on the no-purchase probability that should be strictly positive but can otherwise be arbitrarily small. In addition, we assume that the products are ordered in strictly increasing order with respect to their marginal revenue: $0 < w_1 < \ldots < w_N \leqslant 1$. (Note that, in the uncapacitated MNL model, products that yield the same marginal profit to the seller can be considered as the same product.)

For uncapacitated assortment optimization under the MNL model, it is known that the optimal assortment is of the form 'offer the k most profitable products' for some integer $k \in [N]$, see, e.g., Talluri & van Ryzin (2004, Proposition 6).To facilitate our analysis, we define

$$S_\varrho := \{i \in [N] : w_i \geqslant \varrho\}, \qquad \varrho \in [0, 1].$$

Then, it is also known that the sequence $\mathcal{P} := \big(r(S_{w_1}, v), \ldots, r(S_{w_N}, v)\big)$ is unimodal. Consider the function $\varrho \mapsto r(S_\varrho, v)$. This function is piecewise constant and attains every value in $\mathcal{P}$. Moreover, the assortment $S^* = S_{\varrho^*}$ with

$$\varrho^* := \max_{S \subseteq [N]} r(S, v)$$

is optimal under preference vector v, see, e.g., Chen et al. (2021, Section 4). This structure has compelling computational implications, as we only need to approximate ϱ^* instead of the more straightforward approach of establishing estimates of $r(S_{w_i}, v)$ for all $i \in [N]$.

3.3.1 A Policy for Incomplete Information

In this section, we propose a policy to iteratively establish a sequence of assortments that converges to the optimal assortment. The policy follows the intuitive approach of offering products with marginal revenue above a threshold value $\varrho_t \in [0, 1]$ at each time $t \in [T]$. Based on the observed (no-)purchases, each next threshold value ϱ_{t+1} is selected. The policy is parameterized by $\alpha \geqslant 1$ and $\beta \geqslant \alpha - 1$.

Stochastic Approximation Policy SAP(α, β)

1. Initialization. Let $\varrho_1 \in [0,1]$. For all $t \in \mathbb{N}$ let $a_t := \alpha/(t+\beta)$, for some $\alpha \geqslant 1$ and $\beta \geqslant \alpha - 1$. Put $t := 1$. Go to 2.
2. Assortment selection. Let $S_t := S_{\varrho_t}$ and

$$\varrho_{t+1} := \varrho_t + a_t(w_{Y_t} - \varrho_t)$$

   Put $t := t+1$. If $t \leqslant T$, then go to 2, else to 3.
3. Terminate.

The policy SAP(α, β) is a classical stochastic approximation algorithm (Robbins & Monro (1951); Kushner & Yin (1997)) that relies on the observation that ϱ^* is the unique solution to the fixed point equation

$$\varrho = r(S_\varrho, v).$$

This is easily verified (see, e.g., Lemma 2 from Chen et al. (2021)). Note that we do not directly observe the value of $r(S_\varrho, v)$. However, we do have the unbiased, noisy observation w_Y given offered assortment S_ϱ at our disposal. As a result, the sign of $w_Y - \varrho$ approximately indicates the direction in which ϱ^* is situated in relation to ϱ. The step sizes a_t decays approximately as $1/t$, ensuring the correct convergence rate of ϱ_t to ϱ^*, as ϱ_t does not 'keep jumping over' ϱ^*, and ϱ_t does not converge 'too slow'.

3.3.2 Regret Upper Bound

We proceed by showing that the worst-case regret of SAP(α, β) is bounded from above by $\sqrt{T}$ times a constant independent of N.

Theorem 3.1. *Let π correspond to* SAP(α, β) *with $\alpha \geqslant 1/p_0$ and $\beta \geqslant \alpha - 1$. Then, there exists a $\overline{C} > 0$ such that, for all $T \geqslant 1$,*

$$\Delta_\pi(T) \leqslant \overline{C}\sqrt{T}.$$

The proof relies on two main steps. First, we show that the regret is bounded from above by the following expression:

$$\Delta_\pi(T) \leqslant c_1 \left(T\, \mathbb{E}_v^\pi \big[\, |\varrho^* - \varrho_{T+1}| \,\big] + \sum_{t=1}^{T} \mathbb{E}_v^\pi \big[\, |\varrho^* - \varrho_t| \,\big] \right), \tag{3.1}$$

for all $v \in \mathcal{V}$ and some instance-independent constant c_1. Second, we show a recursive relationship regarding the mean squared error of ϱ_t with respect to ϱ^*. Then, the convergence rate of ϱ_t to ϱ^* is a consequence of the recursive relation. This second step is summarized in the following lemma.

Lemma 3.1. *Let $v \in \mathcal{V}$ and let π correspond to* SAP(α, β) *with $\alpha \geqslant 1/p_0$ and $\beta \geqslant \alpha - 1$. For all $t \in [T]$ it holds that*

$$\mathbb{E}_v^\pi \big[(\varrho^* - \varrho_{t+1})^2\big] \leqslant \mathbb{E}_v^\pi \big[(\varrho^* - \varrho_t)^2\big]\,(1 - 2p_0 a_t) + a_t^2. \tag{3.2}$$

As a result, there exists an instance-independent constant c_2 such that, for all $t = 1, \ldots, T+1$,

$$\mathbb{E}_v^\pi \big[(\varrho^* - \varrho_t)^2\big] \leqslant \frac{c_2}{t + \beta}. \tag{3.3}$$

We then apply Jensen's inequality to the concave function $x \mapsto \sqrt{x}$ to conclude that the mean absolute error of ϱ_t with respect to ϱ^* converges as $1/\sqrt{t}$, that is,

$$\mathbb{E}_v^\pi \big[\, |\varrho^* - \varrho_t| \,\big] \leqslant \frac{\sqrt{c_2}}{\sqrt{t + \beta}},$$

for all $t = 1, \ldots, T+1$. With the convergence rate of the mean absolute error in combination with (3.1), the proof of Theorem 3.1 follows.

Remark 3.1. The requirement of $\alpha \geqslant 1/p_0$ in Theorem 3.1 is caused by the constant p_0 in (3.2), and ensures an $\mathcal{O}(\sqrt{T})$ regret for all preference vectors $v \in \mathcal{V}$, even with small corresponding no-purchase probability. However, as the performance of SAP depends on its parameters, a large value of α and β when p_0 is small might not be necessary in practice. The inequality (3.2) is also valid when p_0 is replaced by the instance-dependent constant $\gamma_v \in (p_0, 1]$, where

$$\gamma_v := \inf_{\varrho < \varrho_v^*} \frac{r(S_\varrho, v) - \varrho}{\varrho_v^* - \varrho}, \tag{3.4}$$

and where ϱ_v^* denotes the optimal revenue as function of the preference vector v. We still obtain an $\mathcal{O}(\sqrt{T})$ regret bound as long as $\alpha \geqslant 1/\gamma_v$, which is a weaker requirement on α since γ_v might be substantially larger than p_0.

To illustrate this possibility, consider the case where $N = 2$, $r_1 = 0.4$, and $r_2 = 1$. In addition, fix $p_0 \in (0, 1/2)$ and let $\mathcal{V}$ be as in Assumption 3.1. For all $v = (v_1, 1)$ such that $1/(1 + v_1 + 1) \geqslant p_0$ the optimal assortment is $\{2\}$ with $\varrho_v^* = 0.5$. In this example, it holds that the infimum in (3.4) is attained at $\varrho = r_1 = 0.4$ and as a result

$$\gamma_v = \frac{r\big(\{1,2\}, v\big) - 0.4}{0.1} = \frac{2}{2 + v_1}.$$

If $v_1 = 0$, then $\gamma_v = 1 > p_0$, and if $v_1 = 2/p_0 - 2$, then $\gamma_v = 2p_0$. This example shows that there exists instances within $\mathcal{V}$ where γ_v is of the order p_0 as well as instances where $\gamma_v \gg p_0$. Hence, the requirement that α is of the order $1/p_0$, although necessary for bounding the worst-case regret as done by Theorem 3.1, can for some preference vectors be mitigated while maintaining good case-specific performance in practice. For a numerical illustration we refer to Section 4.2.

3.3.3 Regret Lower Bound

Now that we have provided an upper bound on the regret of our SAP policy, we proceed by showing that this bound is asymptotically tight – up to a multiplicative constant – as T grows large. This implies that our policy performs asymptotically optimal. We prove our regret lower bound for values p_0 in Assumption 3.1 such that

$$p_0 \leqslant \max_{1 \leqslant k < \ell \leqslant N} \frac{2w_\ell - 2w_k}{5w_\ell + w_k}. \tag{3.5}$$

Observe that this condition can always be ensured to hold by choosing p_0 sufficiently small.

The regret lower bound is presented below.

Theorem 3.2. *There exists a $\underline{C} > 0$ such that, for all policies π and all $T \geqslant 1$,*

$$\Delta_\pi(T) \geqslant \underline{C}\sqrt{T}.$$

The proof of Theorem 3.2 is established in three steps. First, we construct two

preference vectors v^0 and v^1 which are statistically 'difficult to distinguish'. Second, we show that any estimator ψ that has the observed purchases $Y_1, \ldots, Y_T$ as inputs and outputs either 0 or 1 must satisfy

$$\max_{j=0,1} \mathbb{P}^{\pi}_{v^j}(\psi \neq j) \geqslant \frac{1}{4}. \tag{3.6}$$

Third, we define a specific estimator ψ and show that under the assumption that $\Delta_\pi(T) < \underline{C}\sqrt{T}$ it follows that

$$\mathbb{P}^{\pi}_{v^j}(\psi \neq j) < \frac{1}{4},$$

for both $j = 0, 1$. Having found a contradiction with (3.6), we thus conclude that the statement in Theorem 3.2 must hold.

The novelty of this proof is concentrated in the first step, which allows us to prove Theorem 3.2 for arbitrary revenue parameters. The second step and third step are conceptually the same as in Chen et al. (2021). We include the last two steps because of the slight deviation in our set-up as well as for the sake of completeness.

Starting with the first step, we define two quantities. Let $k, \ell \in [N]$ such that $k < \ell$ and

$$p_0 \leqslant \frac{2w_\ell - 2w_k}{5w_\ell + w_k}.$$

Note that such k, ℓ exist by equation (3.5). Furthermore, define

$$u_k := \frac{w_\ell}{w_\ell - w_k} \qquad \text{and} \qquad u_\ell := \frac{w_k}{w_\ell - w_k}.$$

The idea underlying this set-up is as follows. Let u denote the preference vector that has u_k as its k-th component, u_ℓ as its ℓ-th component and is zero everywhere else. Then, it holds that

$$r\big(\{k, \ell\}, u\big) = r\big(\{\ell\}, u\big) = \max_{S \subseteq [N]} r(S, u).$$

Then, by perturbing u in two ways we can construct two preference vectors v^0 and v^1 which are very close to u, but result in different solutions to the assortment optimization problem. Specifically, we define v^0 and v^1 as follows. Let $\varepsilon \in (0, 1/2]$ denote a small constant and let $v^0 = (v^0_1, \ldots, v^0_N)$ and $v^1 = (v^1_1, \ldots, v^1_N)$ denote the vectors

of preference parameters such that, for $i \in [N]$,

$$v_i^0 = \begin{cases} u_k, & \text{if } i = k, \\ (1-\varepsilon)u_\ell, & \text{if } i = \ell, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad v_i^1 = \begin{cases} u_k, & \text{if } i = k, \\ (1+\varepsilon)u_\ell, & \text{if } i = \ell, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

From this set-up it follows that $\{k, \ell\}$ is the optimal assortment under v^0 and $\{\ell\}$ is the optimal assortment under v^1 (up to inclusion of products with zero preference). That is,

$$\max_{S \subseteq [N]} r(S, v^0) < w_k \qquad \text{and} \qquad \{k, \ell\} \in \operatorname*{arg\,max}_{S \subseteq [N]} r(S, v^0),$$

and

$$\max_{S \subseteq [N]} r(S, v^1) > w_k \qquad \text{and} \qquad \{\ell\} \in \operatorname*{arg\,max}_{S \subseteq [N]} r(S, v^1).$$

In what follows we abbreviate the expectation value and probability $\mathbb{E}^\pi_{v^j}[\,\cdot\,]$ and $\mathbb{P}^\pi_{v^j}(\,\cdot\,)$ as $\mathbb{E}_j[\,\cdot\,]$ and $\mathbb{P}_j(\,\cdot\,)$ for $j = 0, 1$. We suppress the notation that these two notions depend on policy π, as Theorem 3.2 holds for any policy. Now, for the second step we use the following lemma, where we first bound the Kullback-Leibler (KL) divergence.

Lemma 3.2. *Let $S \subseteq [N]$. Then, there is constant $c_3 > 0$ such that*

$$\mathrm{KL}\Big(\mathbb{P}_0(\,\cdot \mid S) \,\Big\|\, \mathbb{P}_1(\,\cdot \mid S)\Big) \leqslant c_3\, \varepsilon^2.$$

In addition, let $\psi \in \{0, 1\}$ denote an arbitrary estimator which has random purchases $Y_1, \ldots, Y_T$ as inputs. Then

$$\max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geqslant \frac{1}{2}\left(1 - \sqrt{2c_3}\,\varepsilon\,\sqrt{T}\right).$$

The lemma above contains a similar statement as Lemma EC.1 and Lemma 12 from Chen et al. (2021). However, we look at different preference vectors than considered by Chen et al. (2021) because we allow arbitrary revenue parameters. The proof of the lemma above makes use of Pinsker's inequality as well as Le Cam's method.

From Lemma 3.2 and by setting ε equal to

$$\varepsilon := \min\left\{\frac{1}{2}, \left(2\sqrt{2c_3 T}\right)^{-1}\right\},$$

it follows that (3.6) holds.

As first part of the third step, we provide a lower bound of the regret under v_0 and v_1 in terms of T, ε, $\wp_0$ and $\wp_1$, where $\wp_0$ denotes how often k and ℓ are both contained in S_t for $t \in [T]$ and $\wp_1$ denotes how often k is excluded while ℓ is contained in S_t for $t \in [T]$. That is,

$$\wp_0 := \sum_{t=1}^{T} \mathbf{1}\{k, \ell \in S_t\} \qquad \text{and} \qquad \wp_1 := \sum_{t=1}^{T} \mathbf{1}\{k \notin S_t, \ell \in S_t\}.$$

This lower bound is stated in the lemma below.

Lemma 3.3. *There is a constant $c_4 > 0$ such that, for any policy π,*

$$\Delta_\pi(T, v^0) \geqslant c_4\, \mathbb{E}_0\Big[\varepsilon\wp_1 + T - \wp_0 - \wp_1\Big], \tag{3.8}$$

and

$$\Delta_\pi(T, v^1) \geqslant c_4\, \mathbb{E}_1\Big[\varepsilon\wp_0 + T - \wp_0 - \wp_1\Big]. \tag{3.9}$$

The lemma above covers similar statements shown by Chen et al. (2021) in the proof of Lemma 13. Here as well, however, we look at different preference vectors than considered by Chen et al. (2021), because we allow arbitrary revenue parameters.

The remainder of the third step is established by a contradiction. To this end, assume that

$$\Delta_\pi(T) < \underline{C}\sqrt{T}, \tag{3.10}$$

where

$$\underline{C} := \frac{c_4}{16\sqrt{2c_3}}.$$

In addition, in view of (3.8) and (3.9) we define L_0 and L_1 as

$$L_0 := c_4\Big(\varepsilon\wp_1 + T - \wp_0 - \wp_1\Big) \qquad \text{and} \qquad L_1 := c_4\Big(\varepsilon\wp_0 + T - \wp_0 - \wp_1\Big).$$

As a consequence of the assumption in (3.10), we conclude by Markov's inequality

and Lemma 3.3 that

$$\mathbb{P}_0\left(L_0 > 4\underline{C}\sqrt{T}\right) \leqslant \frac{\mathbb{E}_0 L_0}{4\underline{C}\sqrt{T}} \leqslant \frac{\Delta_\pi(T, v^0)}{4\underline{C}\sqrt{T}} < \frac{1}{4},$$

and likewise

$$\mathbb{P}_1\left(L_1 > 4\underline{C}\sqrt{T}\right) \leqslant \frac{\mathbb{E}_1 L_1}{4\underline{C}\sqrt{T}} \leqslant \frac{\Delta_\pi(T, v^1)}{4\underline{C}\sqrt{T}} < \frac{1}{4}.$$

Next, we define the estimator ψ as

$$\psi := \begin{cases} 0, & \text{if } \wp_0 > T/2, \\ 1, & \text{if } \wp_0 \leqslant T/2. \end{cases}$$

The remainder of the proof is to show that $\psi = 1$ implies $L_0 > 4\underline{C}\sqrt{T}$ and that $\psi = 0$ implies $L_1 > 4\underline{C}\sqrt{T}$. From this we conclude that

$$\mathbb{P}_0(\psi = 1) < \frac{1}{4} \qquad \text{as well as} \qquad \mathbb{P}_1(\psi = 0) < \frac{1}{4}.$$

This is a contradiction with the statement in (3.6), which holds by Lemma 3.2. Therefore, statement (3.10) cannot be true and we have thus proven Theorem 3.2.

3.4 Capacitated Assortment Optimization

Here, we consider dynamic assortment optimization under the MNL model *with* capacity constraint. Within this section, we consider the class of preference vectors $\mathcal{V}$ and capacity constraint K that satisfy the following assumption.

Assumption 3.2. *The $\mathcal{V}$ is the set of all components-wise non-negative vectors. In addition, the capacity constraint K is strictly less than $N/2$.*

For notational convenience we write

$$\mathcal{S}_K := \{S \subseteq [N] : |S| = K\}$$

for the collection of all assortments of exactly size K.

3.4.1 Regret Lower Bound and Mathematical Proof

The main result of this section – presented below – states that the regret of any policy can uniformly be bounded from below by a constant times $\sqrt{NT}$.

Theorem 3.3. *There exists a constant $\underline{C} > 0$ such that, for all $T \geqslant N$ and for all policies π,*

$$\Delta_\pi(T) \geqslant \underline{C}\sqrt{NT}.$$

The proof of Theorem 3.3 can be broken up into four steps. First, we define a baseline preference vector $v^0 \in \mathcal{V}$ and we show that under v^0 *any* assortment $S \in \mathcal{S}_K$ is optimal. Second, for each $S \in \mathcal{A}_K$ we define a preference vector $v^S \in \mathcal{V}$ by

$$v_i^S := \begin{cases} v_i^0(1+\varepsilon), & \text{if } i \in S, \\ v_i^0, & \text{otherwise,} \end{cases}$$

for some $\varepsilon \in (0, 1]$. For each such v^S, we show that the instantaneous regret from offering a sub-optimal assortment S_t is bounded from below by a constant times the number of products $|S \setminus S_t|$ not in S; cf. Lemma 3.4 below. This lower bound takes into account how much the assortments $S_1, \ldots, S_T$ overlap with S when the preference vector is v^S. Third, let N_i denote the number of times product $i \in [N]$ is contained $S_1, \ldots, S_T$, i.e.,

$$N_i := \sum_{t=1}^{T} \mathbf{1}\{i \in S_t\}.$$

Then, we use the KL divergence and Pinsker's inequality to upper bound the difference between the expected value of N_i under v^S and $v^{S\setminus\{i\}}$, see Lemma 3.5. Fourth, we apply a randomization argument over $\{v^S : S \in \mathcal{S}_K\}$, combine the previous steps, and set ε accordingly to conclude the proof.

The novelty of this section is concentrated in the first two steps. The third and fourth step closely follow the work of Chen & Wang (2018). These last steps are included (1) because of slight deviations in our set-up, (2) for the sake of completeness, and (3) since the proof techniques are extended to the case where $K/N < 1/2$. In the work of Chen & Wang (2018), the lower bound is shown for $K/N \leqslant 1/4$, but the authors already mention that this constraint can probably be relaxed. Our proof confirms that this is indeed the case.

Step 1: Construction of baseline preference vector

Let $\underline{w} := \min_{i\in[N]} w_i > 0$ and define the constant

$$s := \frac{\underline{w}^2}{3 + 2\underline{w}}.$$

Note that $s < \underline{w}$. The baseline preference vector is formally defined as

$$v_i^0 := \frac{s}{K(w_i - s)}, \qquad \text{for all } i \in [N].$$

Now, the expected revenue for any $S \in \mathcal{A}_K$ under v^0 can be rewritten as

$$r(S, v^0) = \frac{\sum_{i\in S} v_i^0 w_i}{1 + \sum_{i\in S} v_i^0} = \frac{s\sum_{i\in S} \frac{w_i}{w_i - s}}{K + \sum_{i\in S} \frac{s}{w_i - s}} = \frac{s\sum_{i\in S} \frac{w_i}{w_i - s}}{K - |S| + \sum_{i\in S} \frac{w_i}{w_i - s}}.$$

The expression on the right-hand side is only maximized by assortments S with maximal size $|S| = K$, in which case

$$r(S, v^0) = \max_{S'\in\mathcal{A}_K} r(S', v^0) = s.$$

It follows that *all* assortments S with size K are optimal.

Step 2: Lower bound on instantaneous regret of v^S

For the second step, we bound the instantaneous regret under v^S.

Lemma 3.4. *Let $S \in \mathcal{S}_K$. Then, there exists a constant $c_1 > 0$, only depending on $\underline{w}$ and s, such that, for all $t \in [T]$ and $S_t \in \mathcal{A}_K$,*

$$\max_{S'\in\mathcal{A}_K} r(S', v^S) - r(S_t, v^S) \geqslant c_1 \frac{\varepsilon \left|S \backslash S_t\right|}{K}.$$

As a consequence,

$$\sum_{t=1}^{T} \left(\max_{S'\in\mathcal{A}_K} r(S', v^S) - r(S_t, v^S) \right) \geqslant c_1\, \varepsilon \left(T - \frac{1}{K} \sum_{i\in S} N_i \right). \tag{3.11}$$

Proof. Fix $S \in \mathcal{S}_K$. First, note that since $\varepsilon \leqslant 1$, for any $S' \in \mathcal{A}_K$, it holds that

$$\sum_{i\in S'} v_i^S \leqslant \frac{2s}{\underline{w} - s}. \tag{3.12}$$

Second, let $S^* \in \arg\max_{S' \in \mathcal{A}_K} r(S', v^S)$ and $\varrho^* = r(S^*, v^S)$. By rewriting the inequality $\varrho^* \geqslant r(S', v^S)$ for all $S' \in \mathcal{A}_K$, we find that for all $S' \in \mathcal{A}_K$

$$\varrho^* \geqslant \sum_{i \in S'} v_i^S \big(w_i - \varrho^*\big). \tag{3.13}$$

Let $t \in [T]$ and $S_t \in \mathcal{A}_K$. Then, it holds that

$$\begin{aligned}
r(S^*, v^S) - r(S_t, v^S) &= \varrho^* - \frac{\sum_{i \in S_t} v_i^S w_i}{1 + \sum_{i \in S_t} v_i^S} \\
&= \frac{1}{1 + \sum_{i \in S_t} v_i^S} \left(\varrho^* + \sum_{i \in S_t} v_i^S \varrho^* - \sum_{i \in S_t} v_i^S w_i \right) \\
&\geqslant \frac{\underline{w} - s}{\underline{w} + s} \left(\varrho^* - \sum_{i \in S_t} v_i^S \big(w_i - \varrho^*\big) \right) \\
&\geqslant \frac{\underline{w} - s}{\underline{w} + s} \left(\sum_{i \in S} v_i^S \big(w_i - \varrho^*\big) - \sum_{i \in S_t} v_i^S \big(w_i - \varrho^*\big) \right) \\
&= \frac{\underline{w} - s}{\underline{w} + s} \Bigg(\underbrace{\sum_{i \in S} v_i^S \big(w_i - s\big) - \sum_{i \in S_t} v_i^S \big(w_i - s\big)}_{(a)} \\
&\qquad\qquad - \underbrace{(\varrho^* - s) \left(\sum_{i \in S} v_i^S - \sum_{i \in S_t} v_i^S \right)}_{(b)} \Bigg).
\end{aligned}$$

Here, the first inequality is due to (3.12) and the second inequality follows from (3.13) with $S' = S$. Next, note that since $|S_t| \leqslant K$ and $|S| = K$, we find that

$$\big|S_t \backslash S\big| \leqslant \big|S \backslash S_t\big|. \tag{3.14}$$

Now, term (a) can be bounded from below as

$$\begin{aligned}
(a) &= \sum_{i \in S \backslash S_t} v_i^S \big(w_i - s\big) - \sum_{i \in S_t \backslash S} v_i^S \big(w_i - s\big) \\
&= \frac{s}{K} \Big((1 + \varepsilon) \big|S \backslash S_t\big| - \big|S_t \backslash S\big| \Big) \\
&\geqslant s \, \frac{\varepsilon \, \big|S \backslash S_t\big|}{K}.
\end{aligned} \tag{3.15}$$

Here, at the final inequality, we used (3.14). Next, term (b) can be bounded from

above as

$$(b) \leqslant \underbrace{\left|\varrho^* - s\right|}_{(c)} \underbrace{\left|\sum_{i\in S} v_i^S - \sum_{i\in S_t} v_i^S\right|}_{(d)}.$$

Now, for term (c), we note that $v_i^S \geqslant v_i^0$ for all $i \in [N]$. In addition, since $r(S^*, v^0) \leqslant s$,

$$\begin{aligned}\varrho^* - s &\leqslant \frac{\sum_{i\in S^*} v_i^S w_i}{1+\sum_{i\in S^*} v_i^S} - \frac{\sum_{i\in S^*} v_i^0 w_i}{1+\sum_{i\in S^*} v_i^0} \\ &\leqslant \frac{1}{1+\sum_{i\in S^*} v_i^0} \sum_{i\in S^*} (v_i^S - v_i^0) w_i \\ &\leqslant \sum_{i=1}^{N} (v_i^S - v_i^0) = \varepsilon \sum_{i\in S} v_i^0 \leqslant \frac{s}{\underline{w}-s}\,\varepsilon.\end{aligned}$$

This entails an upper bound for (c). Term (d) is bounded from above as

$$\begin{aligned}(d) &\leqslant \sum_{i\in S\setminus S_t} v_i^S + \sum_{i\in S_t\setminus S} v_i^S \\ &\leqslant (1+\varepsilon) \sum_{i\in S\setminus S_t} v_i^0 + \sum_{i\in S_t\setminus S} v_i^0 \\ &\leqslant (1+\varepsilon) \frac{s}{K(\underline{w}-s)} \left|S\setminus S_t\right| + \frac{s}{K(\underline{w}-s)} \left|S_t\setminus S\right| \\ &\leqslant \frac{3s}{\underline{w}-s}\,\frac{\left|S\setminus S_t\right|}{K}.\end{aligned}$$

Here, at the final inequality, we used (3.14) and the fact that $\varepsilon \leqslant 1$. Now, we combine the upper bounds of (c) and (d) to find that

$$(b) \leqslant \frac{3s^2}{(\underline{w}-s)^2} \cdot \frac{\varepsilon \left|S\setminus S_t\right|}{K}. \tag{3.16}$$

It follows from (3.15) and (3.16) that

$$\begin{aligned}r(S^*, v^S) - r(S_t, v^S) &\geqslant \frac{\underline{w}-s}{\underline{w}+s}\left(s - \frac{3s^2}{(\underline{w}-s)^2}\right) \frac{\varepsilon \left|S\setminus S_t\right|}{K} \\ &\geqslant c_1 \frac{\varepsilon \left|S\setminus S_t\right|}{K},\end{aligned}$$

where

$$c_1 := \frac{\underline{w}-s}{\underline{w}+s}\left(s - \frac{3s^2}{(\underline{w}-s)^2}\right).$$

Note that the constant c_1 is positive if $(\underline{w}-s)^2 > 3s$. This follows from $s = \underline{w}^2/(3+2\underline{w})$

since

$$(\underline{w} - s)^2 - 3s > \underline{w}^2 - s(3 + 2\underline{w}).$$

Statement (3.11) follows from the additional observation

$$\sum_{t=1}^{T} \big|S \backslash S_t\big| = TK - \sum_{t=1}^{T} \big|S \cap S_t\big| = TK - \sum_{i \in S} N_i. \qquad \square$$

Step 3: KL divergence and Pinsker's inequality

We denote the dependence of the expected value and the probability on the preference vector v^S as $\mathbb{E}_S[\,\cdot\,]$ and $\mathbb{P}_S(\,\cdot\,)$ for $S \in \mathcal{A}_K$. In addition, we write $S \backslash i$ instead of $S \backslash \{i\}$. The lemma below states an upper bound on the KL divergence of $\mathbb{P}_S$ and $\mathbb{P}_{S\backslash i}$ and uses Pinsker's inequality to derive an upper bound on the absolute difference between the expected value of N_i under v^S and $v^{S\backslash i}$.

Lemma 3.5. *Let $S \in \mathcal{S}_K$, $S' \in \mathcal{A}_K$ and $i \in S$. Then, there exists a constant c_2, only depending on $\underline{w}$ and s, such that*

$$\mathrm{KL}\Big(\mathbb{P}_S(\,\cdot \,|\, S') \,\Big\|\, \mathbb{P}_{S\backslash i}(\,\cdot \,|\, S')\Big) \leqslant c_2 \, \frac{\varepsilon^2}{K}.$$

As a consequence,

$$\Big|\mathbb{E}_S[N_i] - \mathbb{E}_{S\backslash i}[N_i]\Big| \leqslant \sqrt{2c_2}\, \frac{\varepsilon\, T^{3/2}}{\sqrt{K}}. \tag{3.17}$$

Proof. Let $\mathbb{P}$ and $\mathbb{Q}$ be arbitrary probability measures on $S' \cup \{0\}$. It can be shown (see, e.g., Lemma 3 from Chen & Wang (2018)) that

$$\mathrm{KL}(\mathbb{P} \,\|\, \mathbb{Q}) \leqslant \sum_{j \in S' \cup \{0\}} \frac{(p_j - q_j)^2}{q_j},$$

where p_j and q_j are the probabilities of outcome j under $\mathbb{P}$ and $\mathbb{Q}$, respectively. We apply this result for p_j and q_j defined as

$$p_j := \frac{v_j^S}{1 + \sum_{\ell \in S'} v_\ell^S} \qquad \text{and} \qquad q_j := \frac{v_j^{S\backslash i}}{1 + \sum_{\ell \in S'} v_\ell^{S\backslash i}},$$

for $j \in S' \cup \{0\}$. First, note that by (3.12), for all $j \in S' \cup \{0\}$,

$$q_j \geqslant \frac{v_j^0}{1 + 2\frac{s}{\underline{w}-s}} = \frac{\underline{w}-s}{\underline{w}+s} v_j^0.$$

Now, we bound $|p_j - q_j|$ from above for $j \in S' \cup \{0\}$. Note that for $j = 0$ it holds that

$$\begin{aligned} |p_0 - q_0| &= \frac{\left|\sum_{\ell \in S'} v_\ell^S - \sum_{\ell \in S'} v_\ell^{S\setminus i}\right|}{\left(1 + \sum_{\ell \in S'} v_\ell^S\right)\left(1 + \sum_{\ell \in S'} v_\ell^{S\setminus i}\right)} \\ &\leqslant \left|(1+\varepsilon)v_i^0 - v_i^0\right| = v_i^0 \varepsilon. \end{aligned}$$

For $j \neq i$, since $\varepsilon \leqslant 1$, we find that

$$|p_j - q_j| = v_j^S |p_0 - q_0| \leqslant 2 v_j^0 v_i^0 \varepsilon.$$

For $j = i$, we find that

$$\begin{aligned} |p_i - q_i| &= v_i^0 \left|p_0 - q_0 + \varepsilon p_0\right| \\ &\leqslant v_i^0 \left(|p_0 - q_0| + \varepsilon p_0\right) \\ &\leqslant v_i^0 \left(v_i^0 + 1\right) \varepsilon \end{aligned}$$

Therefore, we conclude that

$$\begin{aligned} \mathrm{KL}\Big(\mathbb{P}_S(\,\cdot\,|S') \,\Big\|\, \mathbb{P}_{S\setminus i}(\,\cdot\,|S')\Big) &\leqslant \sum_{j \in S' \cup \{0\}} \frac{(p_j - q_j)^2}{q_j} \\ &\leqslant \frac{(p_0 - q_0)^2}{q_0} + \sum_{j \in S' : j \neq i} \frac{(p_j - q_j)^2}{q_j} + \frac{(p_i - q_i)^2}{q_i} \\ &\leqslant \frac{\underline{w}+s}{\underline{w}-s} \left((v_i^0 \varepsilon)^2 + 4(v_i^0 \varepsilon)^2 \sum_{j \in S' : j \neq i} v_j^0 + v_i^0 (v_i^0 + 1)^2 \varepsilon^2 \right) \\ &\leqslant \frac{s(\underline{w}+s)}{(\underline{w}-s)^2} \left(\frac{s}{\underline{w}-s} + \frac{4s^2}{(\underline{w}-s)^2} + \left(\frac{\underline{w}}{\underline{w}-s}\right)^2 \right) \frac{\varepsilon^2}{K} \\ &= c_2 \frac{\varepsilon^2}{K}, \end{aligned}$$

where

$$c_2 := \frac{s(\underline{w}+s)}{(\underline{w}-s)^2} \left(\frac{s}{\underline{w}-s} + \frac{4s^2 + \underline{w}^2}{(\underline{w}-s)^2} \right).$$

Next, note that the entire probability measures $\mathbb{P}_S$ and $\mathbb{P}_{S\setminus i}$ depend on T. Then, as

a consequence of the chain rule of the KL divergence, we find that

$$\mathrm{KL}\big(\mathbb{P}_S \,\big\|\, \mathbb{P}_{S\setminus i}\big) \leqslant c_2 \,\frac{\varepsilon^2\, T}{K}.$$

Now, statement (3.17) follows from

$$\begin{aligned}
\Big|\mathbb{E}_S[N_i] - \mathbb{E}_{S\setminus i}[N_i]\Big| &\leqslant \sum_{n=0}^{T} n\Big|\mathbb{P}_S(N_i = n) - \mathbb{P}_{S\setminus i}(N_i = n)\Big| \\
&\leqslant T \sum_{n=0}^{T} \Big|\mathbb{P}_S(N_i = n) - \mathbb{P}_{S\setminus i}(N_i = n)\Big| \\
&= 2T \max_{n=0,\dots,T} \Big|\mathbb{P}_S(N_i = n) - \mathbb{P}_{S\setminus i}(N_i = n)\Big| \qquad (3.18) \\
&\leqslant 2T \sup_{A} \Big|\mathbb{P}_S(A) - \mathbb{P}_{S\setminus i}(A)\Big| \\
&\leqslant T\sqrt{2\,\mathrm{KL}\big(\mathbb{P}_S \,\big\|\, \mathbb{P}_{S\setminus i}\big)},
\end{aligned}$$

where the step in (3.18) follows from e.g. Proposition 4.2 from Levin et al. (2017) and we used Pinsker's inequality at the final inequality. □

Step 4: Proving the main result

With all the established ingredients, we can finalize the proof of the lower bound on the regret.

Proof of Theorem 3.3. Since $v^S \in \mathcal{V}$ for all $S \in \mathcal{S}_K$ and by Lemma 3.4, we know that

$$\begin{aligned}
\Delta_\pi(T) &\geqslant \frac{1}{|\mathcal{S}_K|} \sum_{S\in\mathcal{S}_K} \Delta_\pi(T, v^S) \\
&\geqslant c_1\,\varepsilon \left(T - \underbrace{\frac{1}{|\mathcal{S}_K|} \sum_{S\in\mathcal{S}_K} \frac{1}{K} \sum_{i\in S} \mathbb{E}_S[N_i]}_{(a)} \right). \qquad (3.19)
\end{aligned}$$

We decompose (a) into two terms:

$$(a) = \underbrace{\frac{1}{|\mathcal{S}_K|} \sum_{S\in\mathcal{S}_K} \frac{1}{K} \sum_{i\in S} \mathbb{E}_{S\setminus i}[N_i]}_{(b)} + \underbrace{\frac{1}{|\mathcal{S}_K|} \sum_{S\in\mathcal{S}_K} \frac{1}{K} \sum_{i\in S} \Big(\mathbb{E}_S[N_i] - \mathbb{E}_{S\setminus i}[N_i]\Big)}_{(c)}.$$

Let $c := K/N \in (0, 1/2)$. By summing over $S' = S\setminus i$ instead of over S, we bound (b)

from above by

$$(b) = \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}} \frac{1}{K} \sum_{i \notin S'} \mathbb{E}_{S'}[N_i] \leqslant \frac{|\mathcal{S}_{K-1}|}{|\mathcal{S}_K|} T \leqslant \frac{c}{1-c} T,$$

where the first inequality follows from $\sum_{i \in [N]} \mathbb{E}_{S'}[N_i] \leqslant TK$, and the second inequality from

$$\frac{|\mathcal{S}_{K-1}|}{|\mathcal{S}_K|} = \frac{\binom{N}{K-1}}{\binom{N}{K}} = \frac{K}{N-K+1} \leqslant \frac{K/N}{1-K/N}.$$

Now, (c) can be bounded by applying Lemma 3.5:

$$(c) \leqslant \sqrt{2c_2}\, \frac{\varepsilon\, T^{3/2}}{\sqrt{K}} = \frac{\sqrt{2c_2}}{\sqrt{c}}\, \frac{\varepsilon\, T^{3/2}}{\sqrt{N}}.$$

By plugging the upper bounds on (b) and (c) in (3.19), we obtain

$$\begin{aligned} \Delta_\pi(T) &\geqslant c_1\, \varepsilon \left(T - \frac{c}{1-c}\, T - \frac{\sqrt{2c_2}}{\sqrt{c}}\, \frac{\varepsilon\, T^{3/2}}{\sqrt{N}} \right) \\ &= c_1\, \varepsilon \left(\frac{1-2c}{1-c}\, T - \frac{\sqrt{2c_2}}{\sqrt{c}}\, \frac{\varepsilon\, T^{3/2}}{\sqrt{N}} \right). \end{aligned}$$

Now, we set ε as

$$\varepsilon = \min \left\{ 1, \frac{(1-2c)\sqrt{c}}{2(1-c)\sqrt{2c_2}} \sqrt{N/T} \right\}.$$

This yields, for all $T \geqslant N$,

$$\Delta_\pi(T) \geqslant \min \left\{ \frac{c_1\sqrt{2c_2}}{\sqrt{c}}\, T, \frac{c_1(1-2c)^2 c}{8(1-c)\sqrt{2c_2}} \sqrt{NT} \right\}.$$

Finally, note that for $T \geqslant N$ it follows that $T \geqslant \sqrt{NT}$ and therefore

$$\Delta_\pi(T) \geqslant \underline{C}\, \sqrt{NT},$$

where

$$\underline{C} := \min \left\{ \frac{c_1\sqrt{2c_2}}{\sqrt{c}}, \frac{c_1(1-2c)^2 c}{8(1-c)\sqrt{2c_2}} \right\} > 0. \qquad \square$$

3.5 Concluding Remarks

In this chapter, we consider discrete assortment optimization with demand learning. We distinguish between the capacitated and uncapacitated cases, revealing intrinsically different regret behavior. We show that the asymptotically optimal regret rate

in the absence of a capacity constraint grows independently of the number of products N, as $\sqrt{T}$ in the time horizon T, whereas imposing a capacity constraint K leads to $\sqrt{NT}$ regret when $K < N/2$.

Regarding uncapacitated assortment optimization, we present a stochastic-approximation type policy that is easy to implement, and show its asymptotic optimality by providing an upper bound on the regret of our policy, as well as a matching lower bound on the regret that only policy – in the worst case – must endure. The resulting regret rate is $\sqrt{T}$, which interestingly differs from the $\log T$ rate as discussed in Section 2.4 due to structural differences in the continuous logit and the discrete MNL model.

Moreover, we prove an $\Omega(\sqrt{NT})$ regret lower bound for capacitated assortment optimization, where we notice the phase transition in regret rates for capacity constraint K from $K < N/2$ to $K = N$. As of now, the question remains open on how the regret behaves for capacity constraint K between $N/2$ and N.

Chapter 4

Numerical Experiments

In this chapter we present all numerical experiments. In Section 4.1, we compare the performance of the proposed policies for the continuous model from Chapter 2 – the stochastic approximation policy, the discretization policy and the density estimation policy – to alternative policies that are specifically designed for the discrete assortment problem. We find that our algorithms outperform or are on par with these alternatives

In Section 4.2, we compare the stochastic approximation policy for the discrete multinomial (MNL) model without capacity constraint – as presented in Chapter 3 – with established, alternative policies. We demonstrate that our policy has a robust performance in different instances, and outperforms alternative algorithms when the number of N is moderately large – sometimes by a substantial margin.

In Section 4.3, we compare the predictive performance of the continuous model with that of the discrete MNL model and we show that our continuous assortment model has good predictive properties compared to its discrete counterpart, even if the true data-generating model is discrete.

4.1 Continuous Model

Under the continuous model, we consider both the uncapacitated and the capacitated setting. In the uncapacitated case, we compare our algorithm SAP to (i) the Thompson Sampling-based algorithm by Agrawal et al. (2017), and (ii) the Trisection-based algorithm by Chen et al. (2021), both applied to discretized versions of the continuous problem. To have a fair comparison, we use in all our numerical experiments the same

discretization of the product space as in our DUCB algorithm. We refer to these two policies from the literature, applied to discretized versions of the continuous assortment problem, as Discretized Thompson Sampling (DTS) and Discretized Trisection (DTR).

In the capacitated case, we compare our algorithms DUCB and KDEP to DTS but not to DTR, since the Trisection-based algorithm of Chen et al. (2021) is not designed to handle capacity constraints. In addition, in the capacitated case we also evaluate the performance of an adjusted version of DUCB (called *ADUCB*) in which we replace the constant 48 in (2.10) by 1; our numerical results indicate that changing this constant significantly improves performance. Optimally tuning this constant is an interesting direction for future research but is outside the scope of this thesis. In this section, we report numerical results on the regret behavior for these different algorithms; Section 4.3 contains additional numerical experiments on the predictive performance of the continuous model.

We set the preference function v as the bi-modal function that is plotted in Figure 2.1. This function is defined as

$$v(x) = \frac{1}{10} + \frac{1}{5}(2+x)(1-x) + \frac{2}{7}\varphi(x; 0.33, 0.1) + \frac{1}{5}\varphi(x; 0.8, 0.1), \qquad x \in [0,1],$$

where $\varphi(\,\cdot\,; \mu, \sigma)$ denotes the normal probability density function with mean μ and standard deviation σ. In addition, we set $w(x) = x$, $x \in [0,1]$, as the marginal revenue function. We test our algorithms with $c = 1$ and $c = 0.5$, corresponding to capacity constraints $K = N$ and $K = \lfloor N/2 \rfloor$ in the discretized versions. In line with Theorem 2.3, we set the discretization parameter N as $\lfloor \gamma T^{1/3} \rfloor$ with $\gamma = \max\{\overline{v}, 1/c + 1\}$ and $\overline{v} = 2$. The parameters for KDEP are precisely in line with Theorem 2.5 and Proposition 2.6 (these parameters do not depend on $\underline{v}$ or $\overline{v}$). The parameters of SAP are set to $\alpha = 3$, $\beta = 2$, $\varrho_1 = 0$. The algorithms' average regrets over 100 simulations after T time periods, for $T \in \{1\,000, 2\,000, \ldots, 10\,000\}$, are recorded in Table 4.1 and 4.2.

4.1.1 Results

Table 4.1 shows that our algorithm SAP outperforms the alternatives DTR and DTS by a significant margin. The top row of Figure 4.1 plots the regret of SAP as function

of T, both on a linear (left-hand panel) and a logarithmic scale (right-hand panel). The linear growth rate of regret as function of $\log T$ in Figure 4.1 confirms our theoretical result on the regret behavior of SAP. Fitting the curve $\mathcal{R}(T) = \gamma_1 + \gamma_2 \log T$ using linear regression, we find that $\gamma_1 = 0.00171$ and $\gamma_2 = 0.0545$.

Table 4.2 records the regret of DTS, DUCB, ADUCB and KDEP; the results are visualized in the second, third and bottom row of Figure 4.1. The figure illustrates that the regrets of both DUCB, ADUCB and KDEP grow sublinearly. The adjusted policy ADUCB and KDEP perform comparably and on par with DTS, while ADUCB, KDEP and DTS outperform DUCB. This suggests that fine-tuning the constants in the updating formula for the upper confidence bounds can lead to less regret. Fitting the curve $\log \mathcal{R}(T) = \gamma_1 + \gamma_2 \log T$ using linear regression, we find that $\gamma_2 = 0.70$ for DUCB, $\gamma_2 = 0.67$ for ADUCB and $\gamma_2 = 0.72$ for KDEP. This confirms, particularly for ADUCB, our theoretical regret bounds of $T^{2/3}$ (up to a logarithmic term). It is worth observing and illustrated by Figure 4.1 that the regret for our policies is not necessarily monotone in T; this is a result of the discretization to an integer number of products.

	Time horizon T									
Policy	1 000	2 000	3 000	4 000	5 000	6 000	7 000	8 000	9 000	10 000
DTR	8.67	18.2	25.5	32.1	35.9	40.9	48.3	55.8	63.7	70.0
DTS	1.46	1.94	2.10	2.25	2.56	2.45	2.87	3.06	2.85	3.34
SAP	0.380	0.417	0.439	0.452	0.463	0.474	0.483	0.492	0.500	0.507
N	19	25	28	31	34	36	38	39	41	43

Table 4.1: Simulated average regret of the policies with $c = 1$ based on 100 simulations.

	Time horizon T									
Policy	1 000	2 000	3 000	4 000	5 000	6 000	7 000	8 000	9 000	10 000
DTS	12.8	19.8	26.8	32.0	38.8	33.2	46.8	52.6	43.8	48.0
DUCB	89.6	153	206	252	295	334	371	403	440	470
ADUCB	9.81	16.9	23.6	29.8	35.3	32.6	46.7	52.1	45.8	50.1
KDEP	9.00	15.6	19.6	25.0	29.8	33.4	35.7	41.6	43.6	48.9
N	29	37	43	47	51	54	57	59	62	64

Table 4.2: Simulated average regret of the policies with $c = 0.5$ based on 100 simulations.

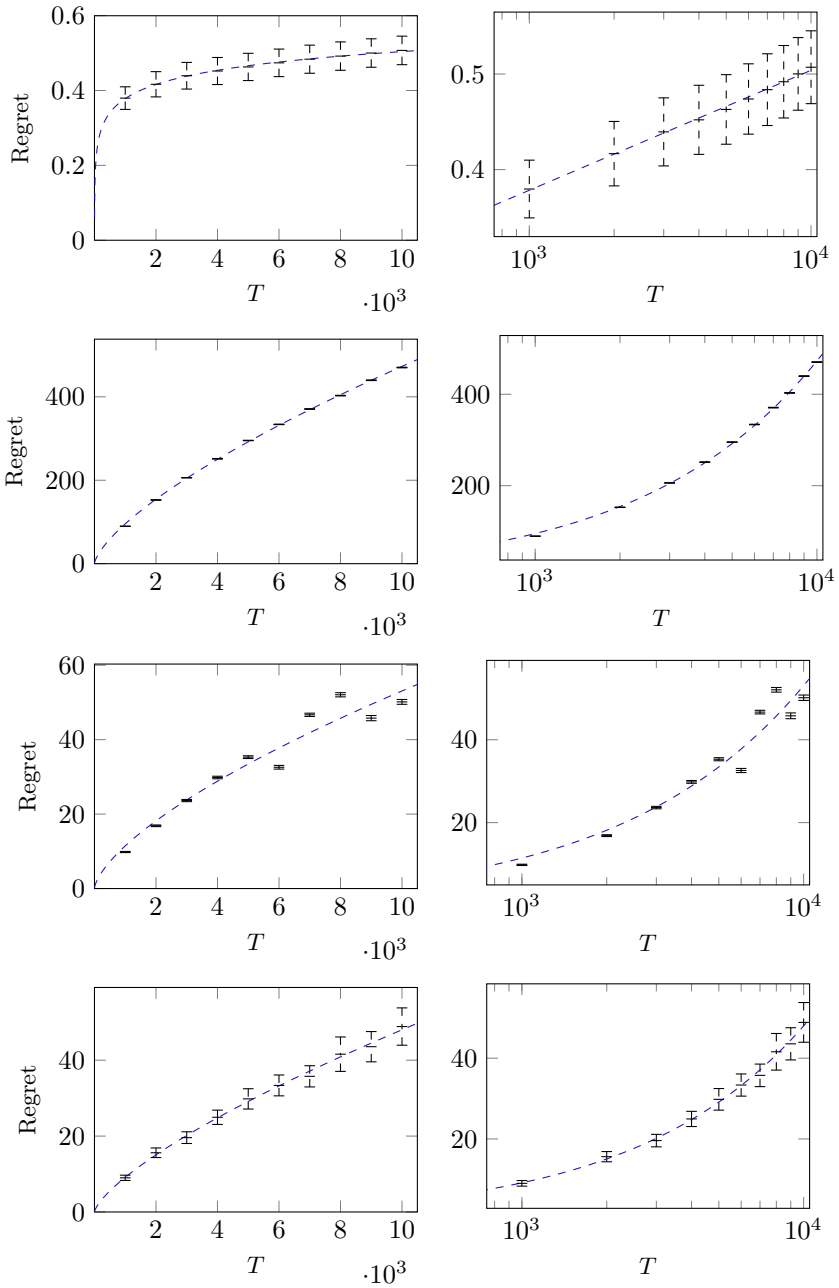

Figure 4.1: The whiskers show the 99% confidence interval of the mean cumulative regret for SAP (top row), DUCB (second row), ADUCB (third row) and KDEP (bottom row) with regular axes (left panels) and a logarithmic axis for T (right panels), based on 100 simulations. The dashed line shows the fitted curves $\gamma_1 + \gamma_2 \log T$ with $\gamma_1 = 0.00171$ and $\gamma_2 = 0.0545$ (top row), $\gamma_1 T^{\gamma_2}$ with $\gamma_1 = 0.783$ and $\gamma_2 = 0.700$ (second row), $\gamma_1 T^{\gamma_2}$ with $\gamma_1 = 0.115$ and $\gamma_2 = 0.666$ (third row) and $\gamma_1 T^{\gamma_2}$ with $\gamma_1 = 0.0628$ and $\gamma_2 = 0.721$ (bottom row).

4.2 Discrete Model

In this section, we compare the performance of our proposed policy from Chapter 3 with alternatives for dynamic assortment optimization under the discrete MNL model. In particular, we compare our policy SAP with the Thompson sampling (TS) based policy from Agrawal et al. (2017) and the Trisection (TR) policy from Chen et al. (2021) using simulated data. It is worth observing that only for SAP and TR an $\mathcal{O}(\sqrt{T})$ regret bound has been proven. We simulate purchase data in two scenarios.

Scenario 1: For different values of N and T we draw N values uniformly at random from $[0.4, 0.5]$, order the randomly drawn values in increasing order, and set the ordered values as the revenue parameters $w_1, \ldots, w_N$. We draw the preference parameters $v_1, \ldots, v_N$ uniformly at random from $[10/N, 20/N]$.

Scenario 2: As in scenario 1, but now we draw the unsorted revenue parameters uniformly at random from $[0, 1]$, order the randomly drawn values in increasing order, and set the ordered values as the revenue parameters $w_1, \ldots, w_N$; in addition, we draw the preference parameters $v_1, \ldots, v_N$ uniformly at random from $[1/N, 100/N]$.

The experimental setup described by scenario 1 follows the set-up from Chen et al. (2021). This setup ensures that finding the optimal solution is non-trivial as the optimal assortment is equal to $\{i : w_i \geqslant x\}$ for some x that lies approximately within $[0.4, 0.5]$. We include the simulated results from scenario 2 to investigate how the policies perform for a broader range of parameters: the range of possible no-purchase probabilities when offering the entire set of products is $[1/101, 1/2]$ for scenario 2, whereas for scenario 1 the range is $[1/21, 1/11]$.

In both simulation scenarios we set N to values ranging from 25 to 5000, roughly doubling each step, and we simulate the regret of the policies a total of 1000 times. For each N, we draw the preference and revenue parameters and, for the Trisection policy, let the policy run for $T = 500, 1000, 2500, 5000$. We let our policy SAP and the Thompson Sampling policy run for $T = 5000$ and record the intermediate regret at $t = 500, 1000, 2500$.

For the Trisection policy, we use the version with adaptive confidence levels, see, e.g., Chen et al. (2021, Section 6). In addition, in line with the simulations of Chen et al. (2021), we replace their $\sqrt{\frac{2\log(8/(\delta\ell))}{\ell}}$ confidence interval configuration with

$\sqrt{\frac{0.1\log(8/(\delta\ell))}{\ell}}$. For our policy SAP, we notice that the preference parameters satisfy Assumption 3.1 for $p_0 = 1/21$ and $p_0 = 1/101$ for scenario 1 and 2, respectively. Theorem 3.1 indicates that we should set the parameter of SAP as $\alpha = 21$ and $\beta = 20$, and $\alpha = 101$ and $\beta = 100$. Following Remark 3.1, we can obtain a better performance with smaller parameters. Hence, we include the performance of SAP with $\alpha = 1$ and $\beta = 0$. This policy is referred to as Adjusted SAP (ASAP). In addition, in all instances we start SAP with $\varrho_1 = w_1$.

4.2.1 Results

Table 4.3 reports the average regret of the policies in scenario 1 and 2 over 1000 simulation runs. All the standard errors are within 3%. From Table 4.3 we see that in both scenarios, the regret of SAP and ASAP does not appear to grow in N, as we would expect. In addition, we find that, in scenario 1, SAP performs on par with TR for $T = 500, 1000$, and substantially outperforms TR for larger values of T. In scenario 2, SAP outperforms TR significantly for all values of N and T. We also find that the adjusted policy ASAP performs improves performance even further by several magnitudes. When we compare SAP to TS, we see that TS performs well for small values of N whereas SAP is better for large values of N. The fact that the regret of TS appears to be growing in N suggests that TS is not an asymptotically optimal policy (i.e. with regret bounded by a factor that is independent of N).

Overall, we find that SAP and ASAP perform well and robustly for increasing values for N. In addition, we see that adequate parameter adjustment for SAP can yield a significant benefit over established alternatives, especially for large values of N. This suggests that designing an algorithm where α, β are adaptively tuned may be an interesting direction for future research.

For illustrative purposes we include visual graphs in Figure 4.2 for the average regret of the policies for N fixed at 5000.

		Scenario 1				Scenario 2			
N	T	TS	TR	SAP	ASAP	TS	TR	SAP	ASAP
25	500	0.447	3.10	3.09	0.457	0.683	101	22.0	1.14
50	500	0.660	2.82	3.22	0.407	1.15	95.0	20.6	2.07
100	500	1.43	2.52	2.99	0.426	0.758	122	25.6	1.50
250	500	2.72	3.18	3.13	0.492	1.64	134	25.1	1.76
500	500	4.74	3.22	3.11	0.493	3.27	121	24.6	1.88
1000	500	8.40	3.05	3.15	0.485	5.69	122	25.8	1.91
2500	500	15.2	3.13	3.16	0.495	13.0	123	26.1	1.76
5000	500	22.6	3.14	3.15	0.481	20.0	126	25.7	1.84
25	1000	0.496	3.78	3.77	0.499	0.758	175	27.5	1.12
50	1000	0.700	3.48	3.71	0.442	1.43	109	27.0	2.33
100	1000	1.56	4.83	3.58	0.499	0.867	157	31.6	1.58
250	1000	2.91	4.52	3.80	0.568	1.79	177	31.2	1.89
500	1000	5.20	4.66	3.78	0.573	3.61	145	30.8	2.05
1000	1000	9.80	4.81	3.83	0.554	6.05	157	32.5	2.05
2500	1000	19.0	4.68	3.84	0.569	14.4	153	32.8	1.88
5000	1000	29.9	4.64	3.82	0.554	23.5	161	32.3	1.96
25	2500	0.556	9.71	4.70	0.567	0.802	316	36.0	1.22
50	2500	0.756	8.87	4.29	0.495	1.90	161	37.2	2.79
100	2500	1.71	10.2	4.45	0.609	1.07	263	38.9	1.74
250	2500	3.11	11.2	4.77	0.666	1.97	288	38.6	2.07
500	2500	5.58	11.6	4.73	0.684	4.01	230	39.6	2.28
1000	2500	11.2	11.6	4.79	0.646	6.40	265	41.3	2.22
2500	2500	23.7	11.6	4.83	0.666	15.7	252	41.5	2.04
5000	2500	39.6	11.5	4.78	0.647	27.1	267	41.1	2.14
25	5000	0.618	15.8	5.31	0.634	0.807	354	42.8	1.22
50	5000	0.808	13.6	4.69	0.547	2.43	288	46.3	3.23
100	5000	1.81	13.5	5.23	0.689	1.27	386	44.0	1.91
250	5000	3.23	17.3	5.52	0.730	2.10	378	44.0	2.21
500	5000	5.80	17.8	5.48	0.768	4.29	369	46.7	2.46
1000	5000	11.8	17.1	5.54	0.717	6.64	399	47.8	2.32
2500	5000	26.1	17.3	5.58	0.742	16.3	398	47.8	2.16
5000	5000	46.1	17.3	5.52	0.713	28.8	395	47.5	2.28

Table 4.3: Average (mean) of the simulated regret in scenario 1 and 2 for the policies TS, TR, SAP (with $\alpha = 1/p_0$ and $\beta = \alpha - 1$) and ASAP (with $\alpha = 1$ and $\beta = 0$) based on 1000 runs.

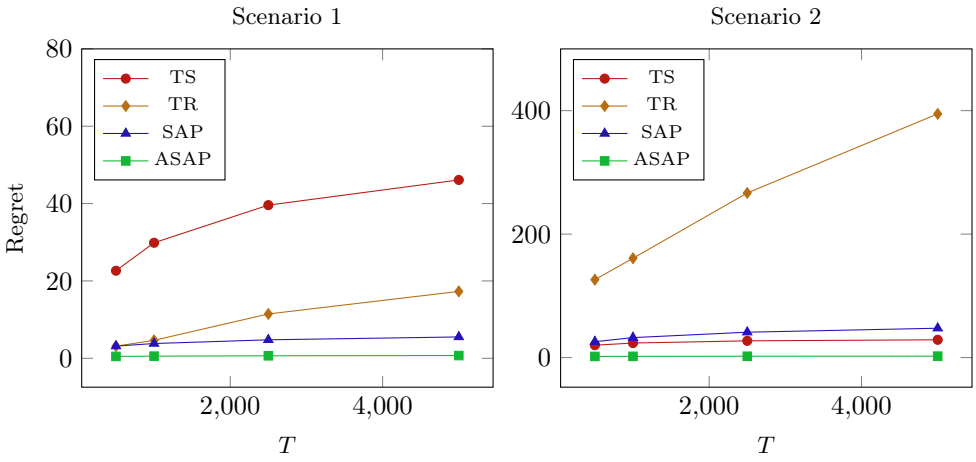

Figure 4.2: The average (mean) of the simulated regret in scenario 1 and 2 with $N = 5000$ for the policies TS, TR, SAP (with $\alpha = 1/p_0$ and $\beta = \alpha - 1$) and ASAP (with $\alpha = 1$ and $\beta = 0$) based on 1000 runs.

4.3 Continuous versus Discrete

In this section, we report the results of additional numerical experiments in which we compare the predictive performance of the continuous logit model with that of the discrete MNL model.

4.3.1 Experimental Set-up

The goal of these additional numerical experiments is to compare the predictive performance of the continuous and the discrete logit choice model. To make such a comparison, we need to define an estimator of the model parameters, for both the continuous and the discrete choice model. For the discrete choice model we use the well-known maximum-likelihood estimator (MLE) to estimate the model parameters. To estimate the preference function of the continuous model, we use the kernel density estimator (KDE) as discussed in Section 2.5.5. Throughout this section we use the same notations and concepts as in Section 2.5.2 and 2.5.3.

We compare the predictive performance of the two models in different scenarios. For each scenario we randomly generate transaction data according to a true 'ground truth model', which is either the discrete or the continuous model. Based on this data we estimate the preference values $v_1, \ldots, v_N$ of the discrete model and the preference

function v of the continuous model, using the MLE and KDE, respectively. We then evaluate the predictive performance of both models using three performance measures: (1) the relative revenue loss of the estimated optimal assortment compared to the true optimal revenue, (2) the L_1-difference between the estimated and true model parameters, and (3), following Berbeglia et al. (2021), the absolute error of the estimated no-purchase probability.

In what follows, we describe in detail the different scenarios, the MLE and KDE, and the three performance measures that we consider.

Scenarios. We consider three different scenarios. In the first scenario the discrete model is the ground truth, with parameters $v_1^{(1)}, \ldots, v_N^{(1)}$ drawn uniformly at random from $[\frac{1}{10N}, \frac{1}{2N}]$, for $N \in \{10, 30, 50\}$. This grossly violates our assumption imposed in the continuous model that the preference values are Lipschitz continuous. In the second scenario the discrete model is again the ground truth; however, the preference values $v_1^{(2)}, \ldots, v_N^{(2)}$ are set to $v_i^{(2)} := f(i/(N+1))/N$, for $i = 1, \ldots, N$, where $N \in \{10, 30, 50\}$,

$$f(x) = \frac{1}{10} + \varphi(x; \mu, \sigma), \qquad x \in [0, 1],$$

and where $\varphi(\,\cdot\,; \mu, \sigma)$ is the normal probability density function with μ drawn uniformly at random from $[0, 1]$, and σ drawn uniformly at random from $[0.1, 0.2]$. Thus, in this second scenario, the continuous model might provide a relatively accurate description of the choice probabilities, despite being a misspecified model. Finally, in the third and last scenario we assume that the continuous model is the ground truth, and we test up to what extent the discrete model is able to produce accurate predictions of consumer's choice behavior. The preference function is set to

$$v^{(3)}(x) = \frac{1}{10} + \frac{1}{5}(2 + x)(1 - x) + \frac{2}{7}\varphi(x; 0.33, 0.1) + \frac{1}{5}\varphi(x; 0.8, 0.1), \qquad x \in [0, 1].$$

The discrete model is estimated for $N \in \{10, 30, 50\}$ products. In all scenarios we set $w(x) := x$ for all $x \in [0, 1]$. For each scenario, for each $c \in \{\frac{1}{2}, 1\}$, and for each $N \in \{10, 30, 50\}$, we randomly generate 1 000 transaction data sets of size $T \in \{50, 100, 200, 500, 1\,000, 2\,000, 5\,000\}$. In these transaction data sets, the assortments are set to the unit interval for $c = 1$. For $c = 1/2$ we let the assortments be $[0, 0.5]$ in

the first $T/2$ time periods, and $[0.5, 1]$ in the second $T/2$ time periods. In the third scenario, in which the continuous model is the ground truth, the observed purchases for the discrete model are of the form $Y_t = \sum_{i=1}^{N} i\, \mathbf{1}\{X_t \in B_i\}$.

We refer to a specific vector of preference parameters as an *instance of the discrete model*, and to a specific preference function as an *instance of the continuous model*. Each instance $\boldsymbol{v} = (v_1, \dots, v_N)$ of the discrete model corresponds to an instance of the continuous model, by letting the discrete purchase Y_t coincide with the continuous purchase $X_t \in B_{Y_t}$ (and $X_t = \varnothing$ if $Y_t = 0$) and by setting the preference function $v(x)$ equal to

$$v(x) := N \sum_{i=1}^{N} v_i \mathbf{1}\{x \in B_i\}.$$

Conversely, each instance of the continuous model with preference function v that is constants on bins $B_1, \dots, B_N$ corresponds to an instance of the discrete model by setting $v_i = \int_{B_i} v(x)\mathrm{d}x$, for all $i = 1, \dots, N$. Concretely, we let $v^{(1)}(\cdot)$ and $v^{(2)}(\cdot)$ be the preference functions of the continuous model that correspond to the (discrete) instance in scenario 1 and 2, and we let $(v_1^{(3)}, \dots, v_N^{(3)})$ be the vector of preference values that correspond to the (continuous) instance in scenario 3.

Estimators. For $j = 1, 2, 3$, let $\hat{v}^{(j),\mathrm{KDE}}(x)$ denote the kernel density estimator of $v^{(j)}(x)$ as presented in Section 2.5.5 and let $\hat{v}^{(j),\mathrm{MLE}}(x)$ denote the stepwise constant function

$$\hat{v}^{(j),\mathrm{MLE}}(x) := \sum_{i=1}^{N} \hat{v}_i^{(j),\mathrm{MLE}}\, \mathbf{1}\{x \in B_i\},$$

where $\hat{v}_i^{(j),\mathrm{MLE}}$ denotes the MLE of $v_i^{(j)}$ for $i \in [N]$. That is,

$$\hat{v}_i^{(j),\mathrm{MLE}} := \frac{\sum_{t=1}^{T} \mathbf{1}\{Y_t = i\}}{\sum_{t=1}^{T} \mathbf{1}\{Y_t = 0\}},$$

for $c = 1$ and

$$\hat{v}_i^{(j),\mathrm{MLE}} := \frac{\sum_{t=(k-1)T/2+1}^{kT/2} \mathbf{1}\{Y_t = i\}}{\sum_{t=(k-1)T/2+1}^{kT/2} \mathbf{1}\{Y_t = 0\}}, \qquad i \in \{(k-1)N/2 + 1, \dots, kN/2\},\ k = 1, 2,$$

for $c = 0.5$, where Y_t are simulated from scenario j. We set the assumed upper bound of $v(x)$ in all scenarios to $\overline{v} = 5$. For $c = 1$, we let $\hat{v}_i^{(j),\mathrm{MLE}}$ be the fixed constant $\overline{v}/N$ if $\sum_{t=1}^{T} \mathbf{1}\{Y_t = 0\} = 0$ and for $c = 0.5$, we let $\hat{v}_{i,k}^{(j),\mathrm{MLE}} = \overline{v}/N$ if

$\sum_{t=(k-1)T/2+1}^{kT/2} \mathbf{1}\{Y_t = 0\} = 0$ with $k = 1, 2$. For the derivation of the MLE we refer to Section 4.3.3.

Performance measures. Given a simulated data sample of size T, the predictive performance is measured in three ways: (1) the instantaneous relative regret of the estimated optimal assortment, (2) the L_1 error of the estimated preference vector/function, and (3), in the same spirit as Berbeglia et al. (2021), the relative absolute difference between the estimated no-purchase probability and the actual no-purchase probability.

To ensure a fair comparison for the first performance measure, the optimal assortment in the first two scenarios is computed over $\mathcal{A}_K$, the collection of all unions of at most $K = cN$ bins. This is because, if the discrete model is the ground truth, then partial products can not be offered. In addition, in these first two scenarios, the estimated optimal assortment under the continuous model is computed with the function w replaced by $\breve{w}$, in line with (2.13). The instantaneous relative regret (IRR) is thus computed as

$$\mathrm{IRR}^{(j),\mathrm{E}} := \frac{r(S^{(j)}, v^{(j)}, w) - r(\hat{S}^{(j),\mathrm{E}}, v^{(j)}, w)}{r(S^{(j)}, v^{(j)}, w)}, \qquad j = 1, 2, 3, \quad \mathrm{E} \in \{\mathrm{KDE}, \mathrm{MLE}\},$$

where $S^{(j)}$ is the optimal assortment in scenario j and $\hat{S}^{(j),\mathrm{E}}$ the estimated optimal assortment, for both estimators $\mathrm{E} \in \{\mathrm{KDE}, \mathrm{MLE}\}$. The second performance measure is defined as

$$\mathrm{L}_1^{(j),\mathrm{E}} := \int_0^1 \left| v^{(j)}(x) - \hat{v}^{(j),\mathrm{E}}(x) \right| \mathrm{d}x, \qquad j = 1, 2, 3, \quad \mathrm{E} \in \{\mathrm{KDE}, \mathrm{MLE}\},$$

where $\hat{v}^{(j),\mathrm{MLE}}$ and $\hat{v}^{(j),\mathrm{KDE}}$ are the MLE and KDE estimator for scenario j, respectively. Finally, our third performance measure is the relative absolute difference of the actual no-purchase probability and the estimated no-purchase probability, where for $c = 1/2$ we average the relative absolute difference of the no-purchase probabilities for assortment $[0, 0.5]$ and $[0.5, 1]$. Thus, defining

$$Q^{(j)} := \frac{1}{1 + \int_0^1 v^{(j)}(x)\mathrm{d}x} \quad \text{and} \quad Q_k^{(j)} := \frac{1}{1 + \int_{S^k} v^{(j)}(x)\mathrm{d}x}, \qquad j = 1, 2, 3,\ k = 1, 2,$$

and

$$\hat{Q}^{(j),\mathrm{E}} := \frac{1}{1+\int_0^1 \hat{v}^{(j),\mathrm{E}}(x)\mathrm{d}x} \quad \text{and} \quad \hat{Q}_k^{(j),\mathrm{E}} := \frac{1}{1+\int_{S^k} \hat{v}^{(j),\mathrm{E}}(x)\mathrm{d}x}, \quad \begin{matrix} j=1,2,3, \\ \mathrm{E}\in\{\mathrm{KDE},\mathrm{MLE}\}, \\ k=1,2, \end{matrix}$$

then our third performance measure is equal to

$$\mathrm{RAD}^{(j),\mathrm{E}} := \frac{\left|Q^{(j)}-\hat{Q}^{(j),\mathrm{E}}\right|}{Q^{(j)}}, \qquad j=1,2,3, \quad \mathrm{E}\in\{\mathrm{KDE},\mathrm{MLE}\}.$$

for $c=1$, and

$$\mathrm{MRAD}^{(j),\mathrm{E}} := \frac{\left|Q_1^{(j)}-\hat{Q}_1^{(j),\mathrm{E}}\right|}{2Q_1^{(j)}} + \frac{\left|Q_2^{(j)}-\hat{Q}_2^{(j),\mathrm{E}}\right|}{2Q_2^{(j)}} \qquad j=1,2,3, \quad \mathrm{E}\in\{\mathrm{KDE},\mathrm{MLE}\},$$

for $c=0.5$.

4.3.2 Results

A priori one would expect that, in scenario 1, the predictive performance of the discrete model outperforms that of the continuous model, and that in scenario 3 it is the other way around. What happens in scenario 2 might be less predictable. The performance metrics in the three different scenarios are displayed in Figures 4.3 through 4.8.

Regarding the third performance measure, there is hardly any difference between the continuous and discrete model. For the other two performance measures, however, we observe marked differences. In scenario 1 the continuous model outperforms the discrete model in several instances, especially for small values of T, both when $c=1$ and when $c=0.5$. Similar behavior is seen in scenario 2: the continuous model outperforms the discrete model under the first two performance measures, except for $N=10$ and sufficiently large T. In scenario 3, the continuous model outperforms the discrete model when measured by the first or second performance measure when $c=0.5$; when $c=1$, the first and second performance measure are either approximately equal, or the continuous model outperforms the discrete model.

These observations demonstrate that there is value in using the continuous model for predictive purposes, also in situations where this model is misspecified.

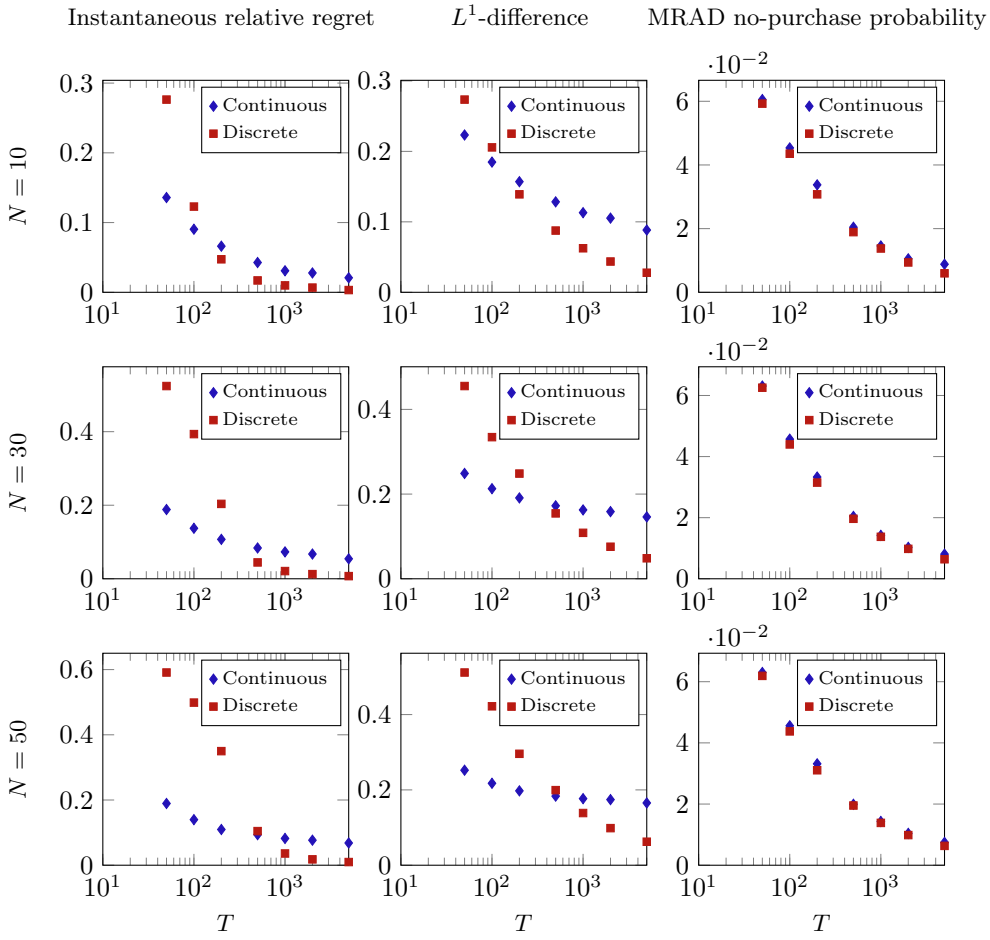

Figure 4.3: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 1 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

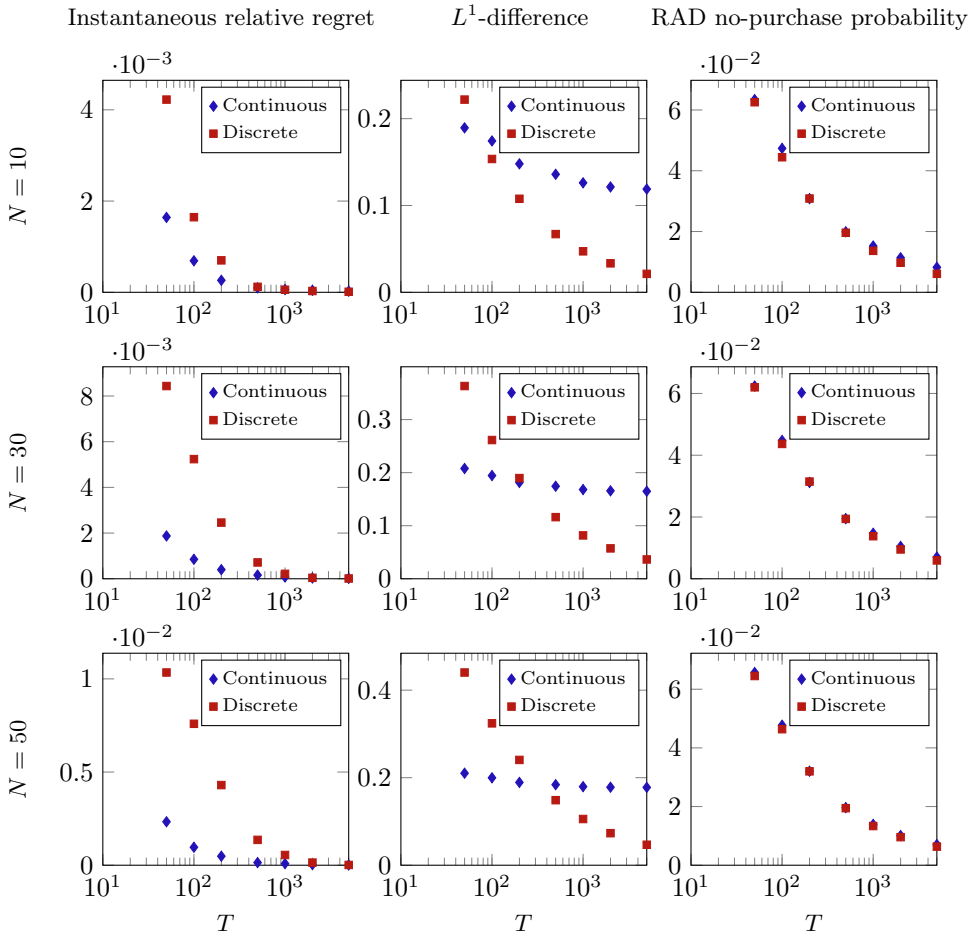

Figure 4.4: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 1 with $c = 1$ and $K = N$ based on 1 000 simulations.

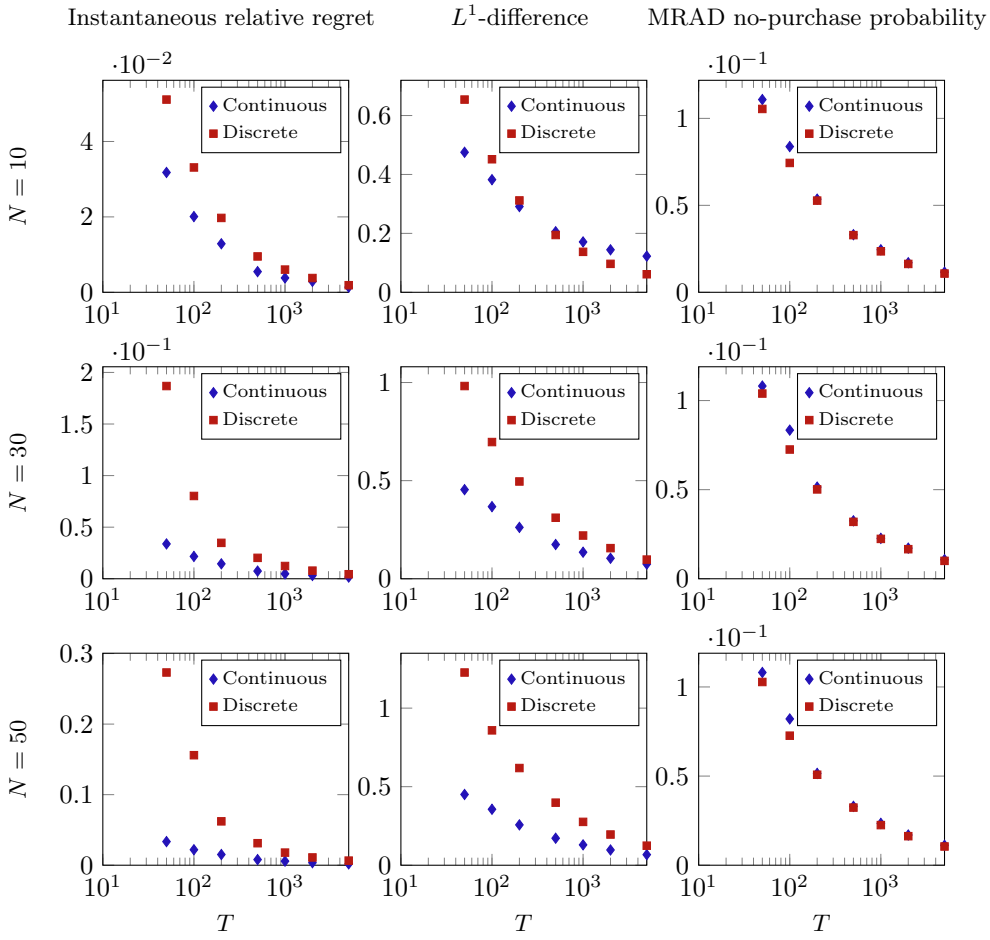

Figure 4.5: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 2 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

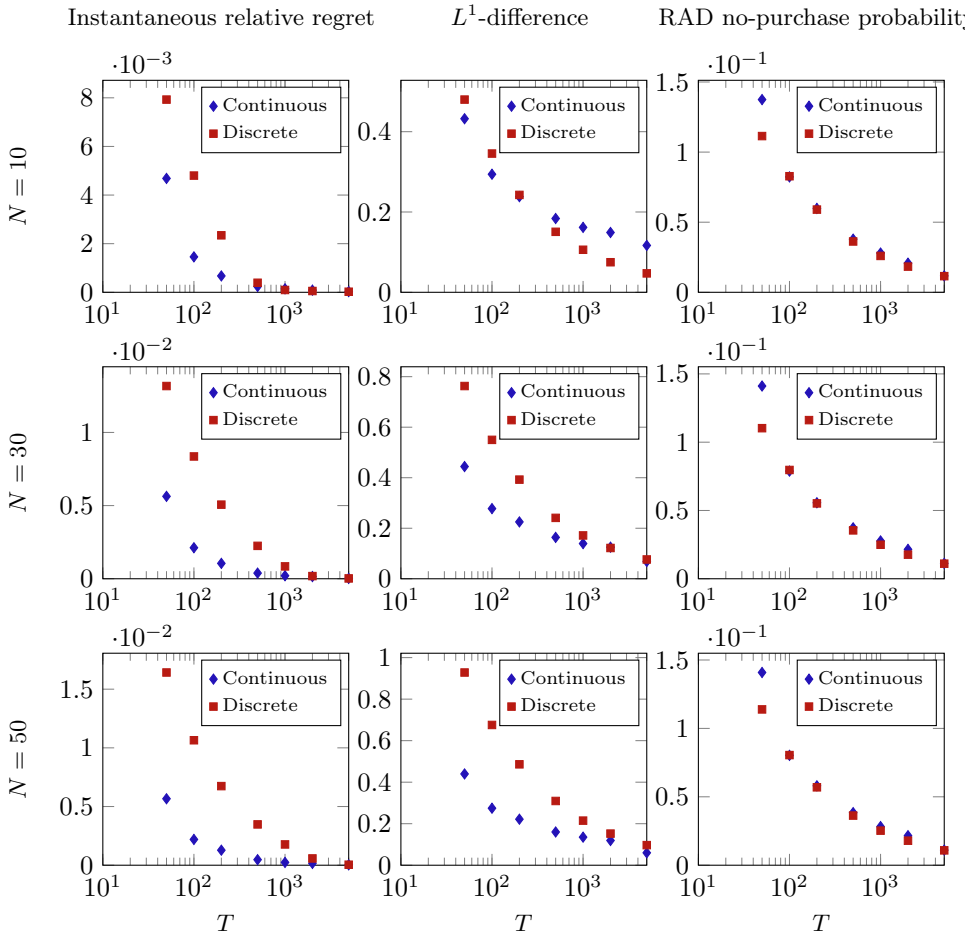

Figure 4.6: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 2 with $c = 1$ and $K = N$ based on 1 000 simulations.

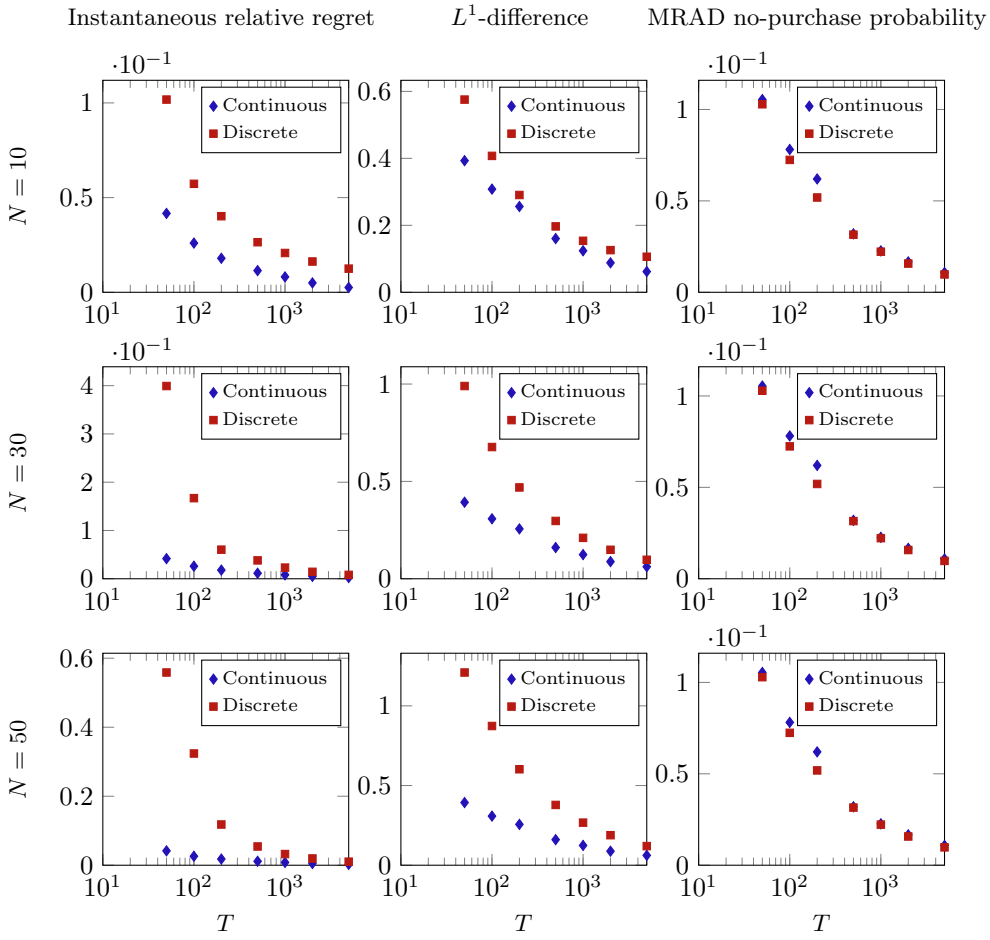

Figure 4.7: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 3 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

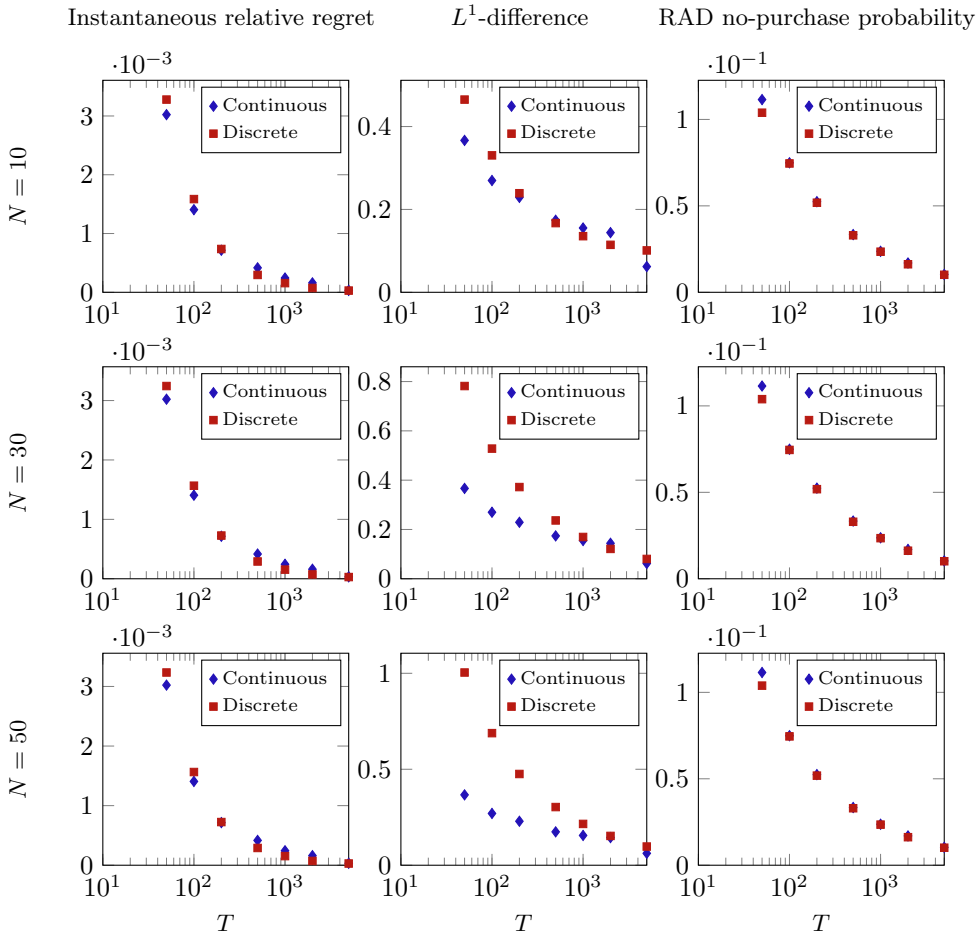

Figure 4.8: The performance metrics comparing the predictive performance of the continuous and the discrete logit choice model for scenario 3 with $c = 1$ and $K = N$ based on 1 000 simulations.

4.3.3 Derivation of the Maximum Likelihood Estimator

Here we derive the maximum likelihood estimator for the preference parameters in the discrete MNL model. We denote as the estimators as $\hat{v}_1, \dots, \hat{v}_N$. Following Section 4.3.1, we consider (i) $K = N$ and offer the entire set of products $[N]$ at all time instances, as well as (ii) $K = N/2$ and offer the assortments $\{1, \dots, N/2\}$ and $\{N/2+1, \dots, N\}$ (each in half of all time instances, that is).

First we consider that $K = N$ and $D_t = [N]$ for all $t \in [T]$. Let Y_t denote the discrete purchase observed at time t when offering $D_t \subseteq [N]$. Then, the log likelihood is

$$L(v_1, \dots, v_N) = \sum_{t=1}^{T} \log \left(\frac{v_{Y_t}}{1 + \sum_{i=1}^{N} v_i} \right) = \sum_{t=1}^{T} \log v_{Y_t} - \sum_{t=1}^{T} \log \left(1 + \sum_{i=1}^{N} v_i \right).$$

Taking the derivative of the log likelihood with respect to v_j for $j \in [N]$ yields

$$\frac{\partial}{\partial v_j} L(v_1, \dots, v_N) = \frac{1}{v_j} \sum_{t=1}^{T} \mathbf{1}\{Y_t = j\} - \sum_{t=1}^{T} \frac{1}{1 + \sum_{i=1}^{N} v_i}.$$

These partial derivatives are equal to zero, so as to obtain $\hat{v}_j$ for $j \in [N]$; we obtain

$$\sum_{t=1}^{T} \mathbf{1}\{Y_t = j\} = \sum_{t=1}^{T} \frac{\hat{v}_j}{1 + \sum_{i=1}^{N} \hat{v}_i}. \tag{4.1}$$

Summing all these equations for $j \in [N]$ yields

$$\sum_{t=1}^{T} \mathbf{1}\{Y_t \neq 0\} = \sum_{t=1}^{T} \frac{\sum_{j=1}^{N} \hat{v}_j}{1 + \sum_{i=1}^{N} \hat{v}_i},$$

or, equivalently,

$$\sum_{t=1}^{T} \mathbf{1}\{Y_t = 0\} = \sum_{t=1}^{T} \frac{1}{1 + \sum_{i=1}^{N} \hat{v}_i}. \tag{4.2}$$

Combining (4.1) and (4.2), we obtain

$$\hat{v}_j := \frac{\sum_{t=1}^{T} \mathbf{1}\{Y_t = j\}}{\sum_{t=1}^{T} \mathbf{1}\{Y_t = 0\}}, \qquad j \in D,$$

where we set $\hat{v}_j := \overline{v}/N$ if $\sum_{t=1}^{T} \mathbf{1}\{Y_t = 0\} = 0$.

Next, we consider that $K = N/2$. Denote $D^1 = \{1, \dots, N/2\}$ and $D^2 = \{N/2 + 1, \dots, N\}$, as well as $\mathcal{T}^1 = \{1, \dots, T/2\}$ and $\mathcal{T}^2 = \{T/2 + 1, \dots, T\}$. Then, $D_t = D^1$

for $t \in \mathcal{T}^1$ and $D_t = D^2$ for $t \in \mathcal{T}^2$. Let $i_1, \dots, Y_t$ denote the discrete purchases observed at time t when offering $D_t \subseteq [N]$. Then, the log likelihood is

$$L(v_1, \dots, v_N) = \sum_{t=1}^{T} \log \left(\frac{v_{Y_t}}{1 + \sum_{i \in D_t} v_i} \right) = \sum_{t=1}^{T} \log v_{Y_t} - \sum_{t=1}^{T} \log \left(1 + \sum_{i \in D_t} v_i \right).$$

Taking the derivative of the log likelihood with respect to v_j for $j \in [N]$ yields

$$\frac{\partial}{\partial v_j} L(v_1, \dots, v_N) = \begin{cases} \dfrac{1}{v_j} \displaystyle\sum_{t \in \mathcal{T}^1} \mathbf{1}\{Y_t = j\} - \sum_{t \in \mathcal{T}^1} \frac{1}{1 + \sum_{i \in D^1} v_i}, & \text{for } j \in D^1, \\ \dfrac{1}{v_j} \displaystyle\sum_{t \in \mathcal{T}^2} \mathbf{1}\{Y_t = j\} - \sum_{t \in \mathcal{T}^2} \frac{1}{1 + \sum_{i \in D^2} v_i}, & \text{for } j \in D^2. \end{cases}$$

These partial derivatives are set equal to zero, to obtain $\hat{v}_j$ for $j \in D^k$ and $k = 1, 2$. We thus obtain

$$\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t = j\} = \sum_{t \in \mathcal{T}^k} \frac{\hat{v}_j}{1 + \sum_{i = D^k} \hat{v}_i}. \tag{4.3}$$

Summing all these equations over $j \in D^k$ yields

$$\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t \neq 0\} = \sum_{t \in \mathcal{T}^k} \frac{\sum_{j \in D^k} \hat{v}_j}{1 + \sum_{i \in D^k} \hat{v}_i},$$

or, equivalently,

$$\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t = 0\} = \sum_{t \in \mathcal{T}^k} \frac{1}{1 + \sum_{i \in D^k} \hat{v}_i}. \tag{4.4}$$

Combining (4.3) and (4.4), we obtain

$$\hat{v}_j := \frac{\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t = j\}}{\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t = 0\}}, \qquad j \in D^k,\ k = 1, 2,$$

where we set $\hat{v}_j := \overline{v}/N$ if $\sum_{t \in \mathcal{T}^k} \mathbf{1}\{Y_t = 0\} = 0$.

Appendix A

A.1 Mathematical Proofs for Section 2.4

A.1.1 Proofs of the Results in Section 2.4.3

Proof of Theorem 2.1.

Define $g(y) := r([y,1],v)$ for $y \in [0,1]$ and $h(\varrho) := g(w^{-1}(\varrho))$ for $\varrho \in [0,1]$. Also, let ϱ^* denote the optimal expected profit, i.e.,

$$\varrho^* := \max\{r(S,v) : S \in \mathcal{S}\}.$$

The following auxiliary results turn out to be useful; the proof of Lemma A.1 follows after the proof of Theorem 2.1.

Lemma A.1. *It holds that $h(\varrho^*) = \varrho^*$. Moreover, for $\varrho \in [0,1]$, the following properties hold:*

$$(\varrho - \varrho^*)(h(\varrho) - \varrho) \leqslant -\frac{1}{1+\overline{v}}(\varrho - \varrho^*)^2, \tag{A.1}$$

$$h(\varrho^*) - h(\varrho) \leqslant C(\varrho - \varrho^*)^2, \tag{A.2}$$

for a universal constant $C > 0$.

Note that by our choice of $\alpha \geqslant 1 + \overline{v}$ and $\beta \geqslant \alpha - 1$ it follows that $\varrho_t \in [0,1]$ for all $t = 1, \ldots, T$. With these properties at our disposal, we continue the proof of the worst-case bound for Case 1, which closely follows the analysis of Broadie et al. (2011) on stochastic approximation schemes. For the policy $\pi = \mathrm{SAP}(\alpha, \beta)$, it holds for all $t = 1, \ldots, T$ that

$$\mathbb{E}_\pi[(\varrho_{t+1} - \varrho^*)^2 \,|\, \varrho_t]$$

$$
\begin{aligned}
&= \mathbb{E}_\pi \left[\left(\varrho_t + a_t(R_t - \varrho) - \varrho^* \right)^2 \mid \varrho_t \right] \\
&= \mathbb{E}_\pi \left[(\varrho_t - \varrho^*)^2 + 2(\varrho_t - \varrho^*) a_t (R_t - \varrho_t) + a_t^2 (R_t - \varrho_t)^2 \mid \varrho_t \right] \\
&\leqslant (\varrho_t - \varrho^*)^2 + 2(\varrho_t - \varrho^*) a_t (h(\varrho_t) - \varrho_t) + a_t^2 \\
&\leqslant (\varrho_t - \varrho^*)^2 \left(1 - \frac{2a_t}{1 + \overline{v}} \right) + a_t^2
\end{aligned}
$$

where the first inequality follows from $R_t - \varrho_t \in [-1, 1]$ and the second inequality from Lemma A.1, i.e., inequality (A.1). Recalling the definition of a_t, an immediate consequence of the above bound is that we have, with $\delta_t := \mathbb{E}_\pi[(\varrho_t - \varrho^*)^2]$, for any $t = 1, \ldots, T$,

$$
\delta_{t+1} \leqslant \delta_t \left(1 - \frac{2}{1 + \overline{v}} \cdot \frac{\alpha}{t + \beta} \right) + \frac{\alpha^2}{(t + \beta)^2}. \tag{A.3}
$$

From inequality (A.3) one can derive the following lemma in a relatively straightforward way. Its (inductive) proof follows after the proof of Theorem 2.1.

Lemma A.2. *There exists a constant κ such that for all $t = 1, \ldots, T$,*

$$
\delta_t \leqslant \frac{\kappa}{t + \beta}. \tag{A.4}
$$

We proceed by deriving an upper bound on the regret of the policy $\pi = \mathrm{SAP}(\alpha, \beta)$, relying on the upper bound on δ_t stated in Lemma A.2. Let C denote the constant as in Lemma A.1. The regret can be majorized as follows:

$$
\begin{aligned}
\Delta_\pi(T, v) = \sum_{t=1}^{T} \mathbb{E}_\pi[h(\varrho^*) - h(\varrho_t)] &\leqslant C \sum_{t=1}^{T} \delta_t \\
&\leqslant C \sum_{t=1}^{T} \frac{\kappa}{t + \beta} \leqslant 3C\kappa \log T,
\end{aligned}
$$

for all $T \geqslant 2$, where the first inequality follows by (A.2), the second inequality by (A.4), and the third inequality by $\sum_{t=1}^{T} (t + \beta)^{-1} \leqslant 3 \log T$ for all $T \geqslant 2$. We have proven the stated with $\overline{C} := 3C\kappa$. □

Proof of Lemma A.1.

We prove the three claims separately.

▷ Following the reasoning at (2.4), we find that

$$\begin{aligned}\varrho^* &= \max\left\{\varrho \in [0,1] : \max_{S\in\mathcal{S}} \int_S v(x)\big(w(x)-\varrho\big)\mathrm{d}x \geqslant \varrho\right\} \\ &= \max\left\{\varrho \in [0,1] : \int_{w^{-1}(\varrho)}^1 v(x)\big(w(x)-\varrho\big)\mathrm{d}x \geqslant \varrho\right\}.\end{aligned}$$

Since $w^{-1}(\cdot)$ is continuous, we know that, with $\varrho \in [0,1]$,

$$\mathcal{I}(\varrho) := \int_{w^{-1}(\varrho)}^1 v(x)\big(w(x)-\varrho\big)\mathrm{d}x$$

is continuous. Also, since $w^{-1}(\cdot)$ is non-decreasing and $\varrho \mapsto v(x)\big(w(x)-\varrho\big)$ is decreasing, we know that $\mathcal{I}(\cdot)$ is non-increasing. Moreover, note that $\mathcal{I}(0) > 0$ and $\mathcal{I}(1) = 0$. As a result, there exists a unique solution to $\mathcal{I}(\varrho) = \varrho$, and that this equation is precisely solved by ϱ^*. The proof is completed by observing that the equation $\mathcal{I}(\varrho) = \varrho$ is equivalent to $h(\varrho) = \varrho$.

▷ For $\varrho = \varrho^*$, (A.1) immediately holds. Now, assume that $\varrho \in [0, \varrho^*)$, then

$$\begin{aligned}h(\varrho) - \varrho &= \frac{\int_{w^{-1}(\varrho)}^1 v(x)w(x)\mathrm{d}x}{1+\int_{w^{-1}(\varrho)}^1 v(x)\mathrm{d}x} - \varrho = \frac{\mathcal{I}(\varrho)-\varrho}{1+\int_{w^{-1}(\varrho)}^1 v(x)\mathrm{d}x} \\ &\geqslant \frac{\mathcal{I}(\varrho^*)-\varrho}{1+\int_{w^{-1}(\varrho)}^1 v(x)\mathrm{d}x} = \frac{\varrho^*-\varrho}{1+\int_{w^{-1}(\varrho)}^1 v(x)\mathrm{d}x} \\ &\geqslant -\frac{1}{1+\overline{v}}(\varrho-\varrho^*).\end{aligned}$$

where the first inequality holds by the non-increasingness of $\mathcal{I}(\cdot)$. As a result,

$$(\varrho-\varrho^*)(h(\varrho)-\varrho) \leqslant -\frac{1}{1+\overline{v}}(\varrho-\varrho^*)^2.$$

Next, assume that $\varrho \in (\varrho^*, 1]$. It holds that $h(\varrho) \leqslant h(\varrho^*) = \varrho^*$ which implies $h(\varrho)-\varrho \leqslant -(\varrho-\varrho^*)$ and therefore

$$(\varrho-\varrho^*)(h(\varrho)-\varrho) \leqslant -(\varrho-\varrho^*)^2 \leqslant -\frac{1}{1+\overline{v}}(\varrho-\varrho^*)^2.$$

Hence, for all $\varrho \in [0,1]$ it holds that

$$(\varrho-\varrho^*)(h(\varrho)-\varrho) \leqslant -\frac{1}{1+\overline{v}}(\varrho-\varrho^*)^2.$$

▷ First, note that

$$\begin{aligned} g'(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\frac{\int_y^1 v(x)w(x)\mathrm{d}x}{1+\int_y^1 v(x)\mathrm{d}x} \\ &= \big(r([y,1],v) - w(y)\big)\cdot\frac{v(y)}{1+\int_y^1 v(x)\mathrm{d}x} \\ &= \big(g(y)-w(y)\big)\cdot\xi(y), \end{aligned}$$

where, for $y\in[0,1]$,

$$\xi(y) := \frac{v(y)}{1+\int_y^1 v(x)\mathrm{d}x}.$$

Second, we show that there exists a universal constant C_0 such that

$$\sup_{y\in(0,1)}\{-g''(y)\} \leqslant C_0. \tag{A.5}$$

To prove (A.5) observe that $g''(y) = \big(g(y)-w(y)\big)\big(\xi'(y)+\xi(y)^2\big) - w'(y)\xi(y)$, and

$$\xi'(y) = \frac{v'(y)}{1+\int_y^1 v(x)\mathrm{d}x} + \xi(y)^2.$$

Since $g(y)-w(y)\in[-1,1]$ for all $y\in(0,1)$, we obtain

$$\begin{aligned} -g''(y) &= -\big(g(y)-w(y)\big)\big(\xi'(y)+\xi(y)^2\big) + w'(y)\xi(y) \\ &\leqslant \sup_{y\in(0,1)}\{|\xi'(y)|+\xi(y)^2\} + \sup_{y\in[0,1]} w'(y)\overline{v} \\ &\leqslant \sup_{y\in(0,1),v\in\mathcal{V}}\{|v'(y)|+2\overline{v}^2\} + \sup_{y\in[0,1]} w'(y)\overline{v} =: C_0. \end{aligned}$$

Now, let $\varrho\in[0,1]$ and denote $y = w^{-1}(\varrho)$ and $y^* = w^{-1}(\varrho^*)$. We distinguish two cases. First, assume that $\varrho^* \geqslant w(0)$ or, equivalently, $g'(y^*)=0$. Then, there is a $\tilde{y}\in(0,1)$ such that $g(y) = g(y^*) + \frac{1}{2}g''(\tilde{y})(y-y^*)^2$. Therefore, we can apply (A.5) to obtain, with

$$k_w := \inf_{x\in(0,1)} w'(x)$$

that

$$\begin{aligned} h(\varrho^*) - h(\varrho) = g(y^*) - g(y) &= -\tfrac{1}{2}g''(\tilde{y})(y-y^*)^2 \\ &\leqslant \tfrac{1}{2}C_0(y-y^*)^2 \leqslant \frac{C_0}{2(k_w)^2}(\varrho-\varrho^*)^2, \end{aligned}$$

where at the final inequality we used that $w^{-1}(\cdot)$ is $(k_w)^{-1}$-Lipschitz continuous on $[0,1]$; note that k_w is strictly positive due to the assumptions imposed on w. Now we consider the second case: assume that $\varrho^* < w(0)$ or, equivalently, $g'(y^*) < 0$. In this case, $\varrho^* = g(0)$ and $w^{-1}(\varrho^*) = 0$. For $\varrho \in [0, w(0))$, $w^{-1}(\varrho) = w^{-1}(\varrho^*)$, and statement (A.2) holds for any constant $C \geqslant 0$. Now, let $\varrho \in [w(0), 1]$. Then, note that by (A.5)

$$g(0) - g(y) \leqslant -g'(0)y + \tfrac{1}{2}C_0 y^2.$$

Next, note that since $w^{-1}(\cdot)$ is non-decreasing and $(k_w)^{-1}$-Lipschitz continuous

$$y = w^{-1}(\varrho) - w^{-1}(\varrho^*) \leqslant \frac{1}{k_w}(\varrho - \varrho^*)$$

and note that

$$0 \leqslant -g'(0) = \big(w(0) - g(0)\big)\xi(0) \leqslant \xi(0)(\varrho - \varrho^*).$$

We conclude that

$$\begin{aligned} h(\varrho^*) - h(\varrho) &= g(0) - g(y) \\ &\leqslant \left(\frac{\xi(0)}{k_w} + \frac{C_0}{2(k_w)^2}\right)(\varrho - \varrho^*)^2 \leqslant \left(\frac{\overline{v}}{k_w} + \frac{C_0}{2(k_w)^2}\right)(\varrho - \varrho^*)^2. \end{aligned}$$

This proves (A.2) for all $\varrho \in [0,1]$ with

$$C = \frac{\overline{v}}{k_w} + \frac{C_0}{2(k_w)^2}.$$ □

Proof of Lemma A.2.

We show, by induction, that inequality (A.3) implies that, for some $\kappa > 0$, for all $t = 1, \ldots, T$ it holds that $\delta_t \leqslant \kappa/(t+\beta)$. To this end, let $K_0 := (1+\overline{v})^{-1}$ and

$$\kappa := \max\left\{1 + \beta, \alpha(1+\overline{v})\right\}.$$

For $t = 1$, we note that

$$\delta_1 \leqslant 1 \leqslant \frac{\kappa}{1+\beta}.$$

Now, suppose $\delta_t \leqslant \kappa/(t+\beta)$ for $t \leqslant t_0$ for some t_0. Then, for $t > t_0$, it follows that

$$\frac{t+\beta}{t+\beta+1} - 2\alpha K_0 < 1 - 2\alpha K_0 \leqslant -\alpha K_0,$$

since $\alpha \geqslant K_0^{-1}$ and therefore

$$\kappa \left(\frac{t+\beta}{t+\beta+1} - 2\alpha K_0 \right) + \alpha^2 < -\kappa \alpha K_0 + \alpha^2 \leqslant 0,$$

by definition of κ. This implies that

$$\kappa \left((t+\beta) - 2\alpha K_0 - \frac{(t+\beta)^2}{t+\beta+1} \right) + \alpha^2 \leqslant 0,$$

and thus

$$\frac{\kappa}{t+\beta} \left(1 - 2\frac{\alpha K_0}{t+\beta} \right) + \frac{\alpha^2}{(t+\beta)^2} \leqslant \frac{\kappa}{t+\beta+1}.$$

This, by (A.3) in combination with the induction hypothesis, yields

$$\delta_{t+1} \leqslant \frac{\kappa}{t+1+\beta},$$

so that we have proven the lemma. □

A.1.2 Proofs of the Results in Section 2.4.4

Proof of Theorem 2.2.

This proof relies on the Van Trees inequality, which can be seen as a Bayesian counterpart of the Cramér-Rao lower bound. Let $\Theta := [\theta_{\min}, \theta_{\max}]$, with $\theta_{\max} = \overline{v}$, $\theta_{\min} = c_0 + (\overline{v} - c_0)/2$, and

$$c_0 := \max \left\{ \underline{v}, \frac{w(0)}{\int_0^1 (w(x) - w(0)) \mathrm{d}x} \right\}.$$

Observe that $\underline{v} < \theta_{\min} < \theta_{\max} = \overline{v}$, because of the assumption

$$\overline{v} > \frac{w(0)}{\int_0^1 (w(x) - w(0)) \mathrm{d}x}.$$

For later reference, we introduce the probability density function $\lambda(\cdot)$ on Θ by

$$\lambda(\theta) := \frac{2}{\theta_{\max} - \theta_{\min}} \cos^2 \left(\pi \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}} - \pi/2 \right).$$

Observe that $\lambda(\cdot)$ is zero on the boundary of Θ. Later, when applying the Van Trees inequality, we work with a random θ, sampled from a distribution with density $\lambda(\cdot)$.

We start the proof with a number of definitions and preliminary observations. Let

$v_\theta(x) := \theta$ for all $x \in [0,1]$ and all $\theta \in \Theta$. Also, define $g(y,\theta) := r([y,1], v_\theta)$, for $y \in [0,1]$ and $\theta \in \Theta$. Let $g'(y,\theta)$ denote the partial derivative of $g(y,\theta)$ with respect to y, for $y \in (0,1)$. As in the proof of Theorem 2.1,

$$g'(y,\theta) = (g(y,\theta) - w(y)) \cdot \xi(y,\theta), \qquad \text{where} \quad \xi(y,\theta) := \frac{v_\theta(y)}{1 + \int_y^1 v_\theta(x)\mathrm{d}x}.$$

In addition, all $y \in (0,1)$ such that $g'(y,\theta) = 0$ satisfy $g''(y,\theta) < 0$, where $g''(y,\theta)$ is the second derivative of $g(y,\theta)$ to y. Observe that $g(0,\theta) - w(0) > 0$ for all $\theta \in \Theta$, since $\theta_{\min} > c_0$. It follows that for all $\theta \in \Theta$ there is a unique maximizer $y(\theta) \in (0,1)$ of $g(y,\theta)$ with respect to y; this maximizer is the unique solution $y \in [0,1]$ to the equation $g(y,\theta) = w(y)$. Moreover, observe that $g(y,\theta)$ is strictly increasing in θ, for all $y \in (0,1)$, and therefore

$$0 = g(y(\theta),\theta) - w(y(\theta)) < g(y(\theta),\theta') - w(y(\theta))$$

for all $\theta_{\min} \leqslant \theta < \theta' \leqslant \theta_{\max}$, which implies that $y(\theta') > y(\theta)$. Thus, $y(\theta)$ is increasing in θ, for $\theta \in \Theta$.

A complication in the proof is that in principle we can optimize over all sets $S \in \mathcal{S}$, which we would like to somehow convert into an optimization over intervals. This explains the relevance of the following objects: for $\theta \in \Theta$ and $S \in \mathcal{S}$, we define

$$\psi(\theta) := \mathrm{vol}([y(\theta),1]) = 1 - y(\theta), \quad \psi^S := \mathrm{vol}(S).$$

$\triangleright$ Step 1. We first show that $r([y(\theta),1], v_\theta)$ and $r(S, v_\theta)$ can only be close if $[y(\theta),1]$ and S are close (a necessary condition for which is that $\psi(\theta)$ and ψ^S are close). More concretely, for all $\theta \in \Theta$ and all $S \in \mathcal{S}$,

$$r([y(\theta),1], v_\theta) - r(S, v_\theta) \geqslant \kappa_0(\psi(\theta) - \psi^S)^2, \qquad \text{where} \quad \kappa_0 := \frac{\theta_{\min} k_w/2}{1 + \theta_{\max}}.$$

To this end, for $v \in \mathcal{V}$ let $\varrho_v^* = \max_{S \in \mathcal{S}} r(S, v)$, and let $S^*(v)$ be a corresponding maximizer. From

$$\varrho_v^* = \frac{\int_{S^*(v)} v(x)w(x)\mathrm{d}x}{1 + \int_{S^*(v)} v(x)\mathrm{d}x},$$

it follows $\varrho_v^* = \int_{S^*(v)} v(x)(w(x) - \varrho_v^*)\mathrm{d}x$, and thus, for all $S \in \mathcal{S}$,

$$\begin{aligned} r(S^*(v), v) - r(S, v) &= \varrho_v^* \frac{1 + \int_S v(x)\mathrm{d}x}{1 + \int_S v(x)\mathrm{d}x} - \frac{\int_S v(x)w(x)\mathrm{d}x}{1 + \int_S v(x)} \\ &= \frac{1}{1 + \int_S v(x)\mathrm{d}x} \left(\varrho_v^* + \int_S v(x)(\varrho_v^* - w(x))\mathrm{d}x \right) \\ &= \frac{1}{1 + \int_S v(x)\mathrm{d}x} \left(\int_{S^*(v)} v(x)(w(x) - \varrho_v^*)\mathrm{d}x - \int_S v(x)(w(x) - \varrho_v^*)\mathrm{d}x \right) \\ &= \frac{1}{1 + \int_S v(x)\mathrm{d}x} \quad \left(\int_{S^*(v)\setminus S} v(x)(w(x) - \varrho_v^*)\mathrm{d}x \right. \\ &\qquad \left. + \int_{S\setminus S^*(v)} v(x)(\varrho_v^* - w(x))\mathrm{d}x \right). \end{aligned}$$

Let $\theta \in \Theta$ and $S \in \mathcal{S}$. If $x \in S^*(v_\theta)\setminus S$, then $x \in S^*(v_\theta) = [y(\theta), 1]$, which implies that $w(x) - \varrho_{v_\theta}^* \geqslant w(y(\theta)) - \varrho_{v_\theta}^* = w(y(\theta)) - g(y(\theta), \theta) = 0$. Similarly, if $x \in S\setminus S^*(v_\theta)$, then $x \in [0, y(\theta))$ and consequently $\varrho_{v_\theta}^* - w(x) \geqslant \varrho_{v_\theta}^* - w(y(\theta)) = g(y(\theta), \theta) - w(y(\theta)) = 0$. It follows that

$$\begin{aligned} r(S^*(v_\theta), v_\theta) - r(S, v_\theta) \geqslant \frac{\theta_{\min}}{1 + \theta_{\max}} \left(\int_{[y(\theta),1]\setminus S} (w(x) - \varrho_{v_\theta}^*)\mathrm{d}x \right. \\ \left. + \int_{S\setminus[y(\theta),1]} (\varrho_{v_\theta}^* - w(x))\mathrm{d}x \right). \end{aligned}$$

Recall that $k_w = \inf_{y\in(0,1)} w'(y) > 0$. Since $\varrho_{v_\theta}^* = w(y(\theta))$, we have by the mean value theorem

$$w(x) - \varrho_{v_\theta}^* = w(x) - w(y(\theta)) \geqslant k_w(x - y(\theta)),$$

for all $x \in [y(\theta), 1]$, and

$$\varrho_{v_\theta}^* - w(x) = w(y(\theta)) - w(x) \geqslant k_w(y(\theta) - x),$$

for all $x \in [0, y(\theta))$. Upon combining the above, we arrive at the lower bound

$$\begin{aligned} r(S^*(v_\theta), v_\theta) - r(S, v_\theta) \geqslant \frac{\theta_{\min} k_w}{1 + \theta_{\max}} \left(\int_{[y(\theta),1]\setminus S} (x - y(\theta))\mathrm{d}x \right. \\ \left. + \int_{S\setminus[y(\theta),1]} (y(\theta) - x)\mathrm{d}x \right). \end{aligned}$$

Let $m_1 := \mathrm{vol}([y(\theta),1] \cap S^c)$ and $m_2 := \mathrm{vol}([0,y(\theta)) \cap S)$. Observe that

$$\int_{[y(\theta),1]\setminus S} (x - y(\theta))\mathrm{d}x \geqslant \int_{y(\theta)}^{y(\theta)+m_1} (x - y(\theta))\mathrm{d}x = \tfrac{1}{2}m_1^2,$$
$$\int_{S\setminus[y(\theta),1]} (y(\theta) - x)\mathrm{d}x \geqslant \int_{y(\theta)-m_2}^{y(\theta)} (y(\theta) - x)\mathrm{d}x = \tfrac{1}{2}m_2^2.$$

In addition,

$$\begin{aligned}\psi^S - \psi(\theta) &= \mathrm{vol}\big(S \cap [0,y(\theta))\big) + \mathrm{vol}\big(S \cap [y(\theta),1]\big) \\ &\quad - \mathrm{vol}\big(S \cap [y(\theta),1]\big) - \mathrm{vol}\big(S^c \cap [y(\theta),1]\big) \\ &= m_2 - m_1, \\ m_1^2 + m_2^2 &\geqslant m_1^2 + m_2^2 - 2m_1m_2 = (m_1 - m_2)^2 = (\psi^S - \psi(\theta))^2.\end{aligned}$$

From the above we conclude that our claim applies: for all $\theta \in \Theta$ and $S \in \mathcal{S}$,

$$r(S^*(v_\theta), v_\theta) - r(S, v_\theta) \geqslant \frac{\theta_{\min} k_w/2}{1 + \theta_{\max}} (\psi^S - \psi(\theta))^2.$$

▷ Step 2. For $S \in \mathcal{S}$ and $\theta \in \Theta$, let Z_θ^S be the random variable with support $[0,2]$ and probability density function

$$f_S(\, z \,|\, \theta\,) := \begin{cases} \dfrac{v_\theta(z)}{1 + \int_S v_\theta(\xi)\mathrm{d}\xi}, & \text{if } z \in S, \\[2ex] \dfrac{\big|[0,2]\setminus S\big|^{-1}}{1 + \int_S v_\theta(\xi)\mathrm{d}\xi}, & \text{if } z \in [0,2]\setminus S. \end{cases}$$

Observe that, when $v = v_\theta$, X^S is in distribution equal to the random variable that equals Z_θ^S if $Z_\theta^S \in S$ and equals $\varnothing$ if $Z_\theta^S \in [0,2]\setminus S$. Hence, for each $t \in \{1,\dots,T\}$ there is a function $\pi_t : [0,2]^{t-1} \to \mathcal{S}$ such that $S_t = \pi_t(Z_1,\dots,Z_t)$ a.s., where $Z_t \overset{\mathrm{d}}{=} Z_\theta^{S_t}$ for all $t = 1,\dots,T$, and where we write $\pi_1(\varnothing) := S_1$. In other words: to prove the regret lower bound we may assume that assortments are a function of the observations $Z_1, Z_2, \dots$ instead of the purchase observations $X_1, X_2, \dots$.

Let $t \in \{1,\dots,T\}$ and let $\mathcal{Z} := [0,2]^t$. The probability density function of $(Z_1,\dots,Z_t)$ is equal to

$$f(\, \boldsymbol{z}_t \,|\, \theta\,) = \prod_{i=1}^{t} f_{\pi_i(\boldsymbol{z}_{i-1})}(\, z_i \,|\, \theta\,),$$

for all $\boldsymbol{z}_t = (z_1, \ldots, z_t) \in \mathcal{Z}$, where we write $\boldsymbol{z}_{i-1} = (z_1, \ldots, z_{i-1})$ for the first $i-1$ components of $\boldsymbol{z}_t$, for all $i = 1, \ldots, t$, and $\boldsymbol{z}_0 := \varnothing$. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(\, \boldsymbol{z}_t \,|\, \theta\,) &= \sum_{i=1}^{t} \frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\pi_i(\boldsymbol{z}_{i-1})}(\, z_i \,|\, \theta\,) \\ &= \sum_{i=1}^{t} \frac{\mathrm{d}}{\mathrm{d}\theta} \left\{ \log\theta \cdot \mathbf{1}\{z_i \in \pi_i(\boldsymbol{z}_{i-1})\} - \log\left(1 + \theta \int_{\pi_i(\boldsymbol{z}_{i-1})} \mathrm{d}\xi\right)\right\} \\ &= \sum_{i=1}^{t} \theta^{-1} \mathbf{1}\{z_i \in \pi_i(\boldsymbol{z}_{i-1})\} - \frac{\mathrm{vol}(\pi_i(\boldsymbol{z}_{i-1}))}{1 + \theta\,\mathrm{vol}(\pi_i(\boldsymbol{z}_{i-1}))},\end{aligned}$$

and

$$-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f(\boldsymbol{z}_t \,|\, \theta) = \sum_{i=1}^{t} \theta^{-2} \mathbf{1}\{z_i \in \pi_i(\boldsymbol{z}_{i-1})\} - \frac{\mathrm{vol}(\pi_i(\boldsymbol{z}_{i-1}))^2}{\big(1 + \theta\,\mathrm{vol}(\pi_i(\boldsymbol{z}_{i-1}))\big)^2} \leqslant \frac{t}{\underline{v}^2},$$

since $\theta_{\min} \geqslant \underline{v}$. By taking expectation, it follows that the Fisher information corresponding to $\boldsymbol{Z}_t = (Z_1, \ldots, Z_t)$ satisfies

$$\mathcal{I}_t(\theta) = \mathbb{E}\left[-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f(\, \boldsymbol{Z}_t \,|\, \theta\,)\right] \leqslant \frac{t}{\underline{v}^2}.$$

The Fisher information $\mathcal{I}(\lambda)$ corresponding to the density $\lambda(\cdot)$ equals

$$\int_{\theta_{\min}}^{\theta_{\max}} \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log\lambda(\theta)\right)^2 \lambda(\theta)\mathrm{d}\theta = \frac{4\pi^2}{(\theta_{\max} - \theta_{\min})^2} = \frac{\pi^2}{(\overline{v} - c_0)^2}.$$

For each $\theta \in \Theta$, $y(\theta)$ is the unique solution to $g(y, \theta) - w(y) = 0$. By the Implicit Function Theorem, the derivative $\psi'(\theta)$ of $\psi(\theta)$ exists and is equal to

$$\begin{aligned}\psi'(\theta) &= -\frac{\mathrm{d}}{\mathrm{d}\theta} y(\theta) = \frac{\frac{\mathrm{d}g}{\mathrm{d}\theta}(y(\theta), \theta)}{\frac{\mathrm{d}g}{\mathrm{d}y}(y(\theta), \theta) - \frac{\mathrm{d}w}{\mathrm{d}y}(y(\theta))} \\ &= -\frac{(1 + \theta(1 - y(\theta)))^{-2}}{w'(y(\theta))} \leqslant -\frac{1}{\max\{w'(y) : y \in (0,1)\}} =: \kappa_1;\end{aligned}$$

for the last step, observe that w being continuously differentiable implies that $\max\{w'(y) : y \in (0,1)\}$ is finite. Now, let θ be a random variable with probability density function $\lambda(\cdot)$; we denote by $\mathbb{E}_\lambda[\,\cdot\,]$ expectation with respect to this density. Let $\psi_t := \psi^{S_{t+1}}$. Now, we are in a position to apply the Van Trees inequality, in particular the form featuring in Gill & Levit (1995). Using the notation used there, their Equation (4)

directly yields (realizing that $\psi'(\theta) \leqslant \kappa_1 < 0$ uniformly in θ)

$$\mathbb{E}_\lambda[(\psi_t - \psi(\theta))^2] \geqslant \frac{\mathbb{E}_\lambda[\psi'(\theta)]^2}{\mathbb{E}_\lambda[\mathcal{I}_t(\theta)] + \mathcal{I}(\lambda)} \geqslant \frac{\kappa_1^2}{t/\underline{v}^2 + \pi^2/(\overline{v} - c_0)^2}.$$

With this lower bound essentially behaving as t^{-1}, the corresponding partial sums (up to the T-th term) grow as $\log T$, as desired. More formally, summing over all $t = 1, \ldots, T-1$, we obtain, applying the lower bound established in Step 1,

$$\begin{aligned}\Delta_\pi(T) &= \sup_{v \in \mathcal{V}} \Delta_\pi(T, v) \geqslant \mathbb{E}_\lambda[\Delta_\pi(T, v_\theta)] \\ &\geqslant \kappa_0 \sum_{t=1}^{T-1} \mathbb{E}_\lambda[(\psi_t - \psi(\theta))^2] \geqslant \kappa_0 \sum_{t=1}^{T-1} \frac{\kappa_1^2 \underline{v}^2}{t + \pi^2 \underline{v}^2/(\overline{v} - c_0)^2} \geqslant \underline{C} \log T,\end{aligned}$$

where $\underline{C} := \kappa_0 \kappa_1^2 \underline{v}^2/(1 + \pi^2 \underline{v}^2/(\overline{v} - c_0)^2) > 0$, and where we used that

$$\sum_{t=1}^{T-1} (t + a)^{-1} \geqslant (1 + a)^{-1} \sum_{t=1}^{T-1} t^{-1} \geqslant (1 + a)^{-1} \log T$$

for all $T \geqslant 2$ and $a \geqslant 0$. □

A.2 Mathematical Proofs for Section 2.5

A.2.1 Proofs of the Results in Section 2.5.1

Proof of Lemma 2.1.

We start the proof by the general remark that it is clear that the optimizing S should only contain x such that $h(x, \varrho) \geqslant 0$, i.e., $x \in W_\varrho$.

First consider case (i), i.e., $\mathrm{vol}(W_\varrho) \leqslant c$. Including in S *all* $x \in W_\varrho$ thus leads to a set in $\mathcal{S}$. Since $h(x, \varrho) < 0$ for $x \notin W_\varrho$, we conclude that the maximum of $\mathcal{I}(S, \varrho)$ over sets in $\mathcal{S}$ is attained by $S = W_\varrho$.

Now, we consider case (ii), i.e., $\mathrm{vol}(W_\varrho) > c$; this means that we should select the subset of W_ϱ that maximizes $\mathcal{I}(S, \varrho)$. Our construction makes use of the following technical properties of $m_\varrho(l)$; their proofs will be given below.

Lemma A.3. *Let $\varrho \in [0, 1]$. Then, $m_\varrho(\ell)$ is non-increasing and left-continuous in ℓ, as well as $m_\varrho(\ell) \to 0$ as $\ell \to \infty$.*

We first concentrate on claim (1). To this end, observe that $m_\varrho(0) = \mathrm{vol}(L_\varrho(0)) =$

$\mathrm{vol}(W_\varrho) \geqslant c > 0$. In addition, by virtue of Lemma A.3, $m_\varrho(\ell) \to 0$ as $\ell \to \infty$. Hence, the set of $\ell \geqslant 0$ such that $m_\varrho(\ell) \geqslant c$ is nonempty and bounded, so that its supremum exists; because of the left-continuity that has been established in Lemma A.3 the supremum is actually attained (and hence is a maximum). This proves the first claim of (ii).

We now consider the second claim of (ii). The intuitive idea is that we start with $S = \varnothing$, and that we keep adding x from W_ϱ to S that have the highest value of $h(x, \varrho)$, until $\mathrm{vol}(S) = c$; at that point S consists of x such that $h(x, \varrho) \geqslant \ell_\varrho$. Bearing in mind, though, that the set of $x \in [0, 1]$ such that $h(x, \varrho)$ equals some given value may have positive Lebesgue measure, there may be still a degree of freedom, which is reflected in the way the set $L_\varrho^{\circlearrowleft}$ has been defined.

The formal argumentation is as follows. First we prove that $\mathrm{vol}(L_\varrho^+) \leqslant c$: as a consequence of the continuity of the Lebesgue measure and the fact that $m_\varrho(\ell)$ is non-increasing in ℓ,

$$\begin{aligned}\mathrm{vol}(L_\varrho^+) &= \mathrm{vol}\left(\bigcup_{k=1}^{\infty} L_\varrho(\ell_\varrho + 1/k)\right) = \mathrm{vol}\left(\lim_{n\to\infty} \bigcup_{k=1}^{n} L_\varrho(\ell_\varrho + 1/k)\right) \\ &= \lim_{n\to\infty} \mathrm{vol}\left(\bigcup_{k=1}^{n} L_\varrho(\ell_\varrho + 1/k)\right) = \lim_{n\to\infty} \mathrm{vol}\left(L_\varrho(\ell_\varrho + 1/n)\right) \\ &= \lim_{n\to\infty} m_\varrho(\ell_\varrho + 1/n) \leqslant c.\end{aligned}$$

Hence, there exists a set $L_\varrho^{\circlearrowleft}$ that is a (possibly empty) subset of $L_\varrho^{=}$ and that is such that $\mathrm{vol}(S) = \mathrm{vol}(L_\varrho^+) + \mathrm{vol}(L_\varrho^{\circlearrowleft}) = c$.

The next objective is to prove that $S = L_\varrho^+ \cup L_\varrho^{\circlearrowleft}$ maximizes $\mathcal{I}(\,\cdot\,, \varrho)$ over sets in $\mathcal{S}$. Take an arbitrary $R \in \mathcal{S}$. Since $\mathrm{vol}(S) = c$, we know that

$$c = \mathrm{vol}(S) = \mathrm{vol}(S \cap R) + \mathrm{vol}(S\backslash R) = \mathrm{vol}(R) - \mathrm{vol}(R\backslash S) + \mathrm{vol}(S\backslash R)$$

and since $\mathrm{vol}(R) \leqslant c$, we obtain $\mathrm{vol}(S\backslash R) \geqslant \mathrm{vol}(R\backslash S)$. Now, since $x \in S$ implies $h(x, \varrho) \geqslant \ell_\varrho$ and $x \in R\backslash S$ implies $h(x, \varrho) \leqslant \ell_\varrho$ we conclude

$$\mathcal{I}(S, \varrho) - \mathcal{I}(R, \varrho) = \mathcal{I}(S\backslash R, \varrho) - \mathcal{I}(R\backslash S, \varrho) \geqslant \ell_\varrho\big(\mathrm{vol}(S\backslash R) - \mathrm{vol}(R\backslash S)\big) \geqslant 0.$$

This proves the second claim of (ii). □

Proof of Lemma A.3.

The set $L_\varrho(\ell)$ is non-increasing in ℓ, hence so is the function $m_\varrho(\ell)$. The next step is to prove that $m_\varrho(\ell)$ is left-continuous. To this end, let ℓ_n be a strictly increasing sequence converging to $\ell < \infty$ as $n \to \infty$. As we have seen, $L_\varrho(\ell_n) \supseteq L_\varrho(\ell)$, and therefore

$$\begin{aligned} m_\varrho(\ell) - m_\varrho(\ell_n) &= \mathrm{vol}\big(\{x \in [0,1] : h(x,\varrho) \in [\ell_n, \ell)\}\big) \\ &= \sum_{k=n}^{\infty} \mathrm{vol}\big(\{x \in [0,1] : h(x,\varrho) \in [\ell_n, \ell_{n+1})\}\big). \end{aligned}$$

From the fact that the left-hand side is finite, it follows that the right-hand side is finite as well, implying left-continuity.

Along the same lines,

$$1 = \mathrm{vol}\big([0,1]\big) = \sum_{k=-\infty}^{\infty} \mathrm{vol}\big(\{x \in [0,1] : h(x,\varrho) \in [k, k+1)\}\big).$$

This entails that, with $n \to \infty$ along the integers,

$$\lim_{n\to\infty} m_\varrho(n) = \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathrm{vol}(\{x \in [0,1] : h(x,\varrho) \in [k, k+1)\}) = 0.$$

From the monotonicity of $m_\varrho(\ell)$, we also have that $m_\varrho(\ell) \to 0$ as $\ell \to \infty$ along the reals. □

Proof of Proposition 2.1.

First, we show that there exists a unique solution to the fixed-point equation

$$g(\varrho) = \varrho, \tag{A.6}$$

where $g(\varrho) := \mathcal{I}(S_\varrho, \varrho)$ for $\varrho \in [0,1]$. As the right-hand side of (A.6) is strictly increasing in ϱ, it suffices to prove that $g(\cdot)$ is continuous and non-increasing in ϱ, and that $g(0) \geqslant 0$ and $g(1) = 0$. To this end, consider $0 \leqslant \varrho_1 \leqslant \varrho_2 \leqslant 1$. Then, indeed, as $\mathcal{I}(S, \varrho)$ is non-increasing in ϱ for any fixed $S \in \mathcal{S}$, and recalling that S_{ϱ_1} maximizes $\mathcal{I}(S, \varrho_1)$,

$$g(\varrho_1) = \mathcal{I}(S_{\varrho_1}, \varrho_1) \geqslant \mathcal{I}(S_{\varrho_2}, \varrho_1) \geqslant \mathcal{I}(S_{\varrho_2}, \varrho_2) = g(\varrho_2).$$

The next step is to prove that $g(\cdot)$ is continuous. Let $\varrho_1, \varrho_2 \in [0,1]$. Then

$$\begin{aligned}\mathcal{I}(S_{\varrho_1}, \varrho_1) - \mathcal{I}(S_{\varrho_2}, \varrho_2) &\leqslant \mathcal{I}(S_{\varrho_1}, \varrho_1) - \mathcal{I}(S_{\varrho_1}, \varrho_2) \\ &= (\varrho_2 - \varrho_1) \int_{S_{\varrho_1}} v(x)\mathrm{d}x \\ &\leqslant |\varrho_1 - \varrho_2| \int_{[0,1]} v(x)\mathrm{d}x,\end{aligned}$$

where the first inequality is due to the fact that S_{ϱ_2} maximizes $\mathcal{I}(\,\cdot\,, \varrho_2)$. With the same token, the same upper bound applies when the roles of the ϱ_1 and ϱ_2 in the left-hand side are interchanged. It thus follows that $g(\cdot)$ is continuous; it is actually even Lipschitz continuous.

Obviously, $g(0) \geqslant 0$. Using that $\sup_{x\in[0,1]} w(x) \leqslant 1$, we also obtain

$$g(1) = \max_{S\in\mathcal{S}} \int_S v(x)(w(x) - 1)\mathrm{d}x = 0.$$

Second, we show that S_{ϱ^*} has the maximum expected revenue over all sets in $\mathcal{S}$. Note that, since $g(\varrho^*) = \varrho^*$, it follows that $r(S_{\varrho^*}) = \varrho^*$. Hence, as we proceed from (2.4) by invoking Lemma 2.1, we obtain

$$\max\left\{\varrho \in [0,1] : \max_{S\in\mathcal{S}} \mathcal{I}(S_\varrho, \varrho) \geqslant \varrho\right\} = \max\left\{\varrho \in [0,1] : g(\varrho) \geqslant \varrho\right\} = \varrho^* = r(S_{\varrho^*}).$$

□

A.2.2 Proofs of the Results in Section 2.5.3

Proof of Proposition 2.2.

In addition to optimal assortments S^* and $\check{S}$ as in (2.14), we define S^p as the optimal assortment under $\check{v}$ and w, that is,

$$r(S^p, \check{v}, w) = \max_{S\in\mathcal{S}} r(S, \check{v}, w).$$

This assortment S^p plays a pivotal role as we break up the left-hand side of (2.15) as follows:

$$r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) = r(S^*, v, w) - r(S^p, \check{v}, w) + \tag{A.7}$$

$$r(S^p, \check{v}, w) - r(\check{S}, \check{v}, \check{w}). \tag{A.8}$$

We start by bounding the right-hand side of (A.7) from above. Define

$$\mathcal{I}(S,\varrho)=\int_S v(x)(w(x)-\varrho)\mathrm{d}x \qquad \text{and} \qquad \mathcal{I}^p(S,\varrho):=\int_S \check{v}(x)(w(x)-\varrho)\mathrm{d}x$$

for $S\in\mathcal{S}$ and $\varrho\in[-\overline{v},1]$. Note that these definitions allow for negative values of ϱ (as opposed to (2.6)). Next, denote the L_1-difference between v and $\check{v}$ as $\delta:=\|v-\check{v}\|_1$. For $\varrho\in[-\overline{v},1]$, let S_ϱ be the maximizer of $\mathcal{I}(\,\cdot\,,\varrho)$ over $\mathcal{S}$ and let S^p_ϱ be the maximizer of $\mathcal{I}^p(\,\cdot\,,\varrho)$ over $\mathcal{S}$, that is,

$$\mathcal{I}(S_\varrho,\varrho)=\max_{S\in\mathcal{S}}\mathcal{I}(S,\varrho) \qquad \text{and} \qquad \mathcal{I}^p(S^p_\varrho,\varrho)=\max_{S\in\mathcal{S}}\mathcal{I}^p(S,\varrho).$$

Then, let ϱ^* and ϱ^p solve the fixed-point equations

$$\varrho=\mathcal{I}(S_\varrho,\varrho) \qquad \text{and} \qquad \varrho=\mathcal{I}^p(S^p_\varrho,\varrho),$$

respectively. Note that $S^p_{\varrho^p}$ is an optimal assortment under $\check{v}$ and w by Proposition 2.1. Hence, we may assume that $S^p=S^p_{\varrho^p}$. Also, we have $0\leqslant w(x)-\varrho^*\leqslant 1$ for all $x\in S^*$ and therefore,

$$\begin{aligned}\mathcal{I}^p(S^*,\varrho^*)-\mathcal{I}(S^*,\varrho^*)&=\int_{S^*}\check{v}(x)(w(x)-\varrho^*)\mathrm{d}x-\int_{S^*}v(x)(w(x)-\varrho^*)\mathrm{d}x\\&\leqslant\int_{S^*}|v(x)-\check{v}(x)|\mathrm{d}x\leqslant\delta.\end{aligned}$$

Now, we find that

$$\mathcal{I}^p(S^*,\varrho^*-\delta)\geqslant\mathcal{I}^p(S^*,\varrho^*)\geqslant\mathcal{I}(S^*,\varrho^*)-\delta=\varrho^*-\delta.$$

Hence, there exists an $S\in\mathcal{S}$ such that $\mathcal{I}^p(S,\varrho^*-\delta)\geqslant\varrho^*-\delta$, which by (2.4) entails $\varrho^p\geqslant\varrho^*-\delta$. Thus, (A.7) is bounded from above as

$$r(S^*,v,w)-r(S^p,\check{v},w)\leqslant\|v-\check{v}\|_1.$$

Bounding (A.8) from above follows in almost an identical manner, but instead of $0\leqslant w(x)-\varrho^*\leqslant 1$ we now use $0\leqslant\check{v}(x)\leqslant\overline{v}$. As a result, we conclude that

$$r(S^p,\check{v},w)-r(\check{S},\check{v},\check{w})\leqslant\overline{v}\,\|w-\check{w}\|_1.$$

Combining the above concludes the proof. □

Proof of Lemma 2.2.

First, let $\varrho^d = r(S^d, \check{v}, \check{w})$ and define the sets $\check{\mathcal{M}}$ and $\mathcal{M}^d$ as arguments of maxima as

$$\check{\mathcal{M}} := \operatorname*{arg\,max}_{S\in\mathcal{S}} \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \quad \text{and} \quad \mathcal{M}^d := \operatorname*{arg\,max}_{S\in\mathcal{A}_K} \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x.$$

Note that since $\mathcal{A}_K \subset \mathcal{S}$, we know for any $S_1 \in \check{\mathcal{M}}$ and $S_2 \in \mathcal{M}^d$ that

$$\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \geqslant \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \geqslant 0. \tag{A.9}$$

Since $\varrho^d \geqslant r(S, \check{v}, \check{w})$ for any $S \in \mathcal{A}_K$, it also holds for $S \in \mathcal{M}^d$ that

$$\varrho^d \geqslant \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x. \tag{A.10}$$

Then, for any $S_1 \in \check{\mathcal{M}}$ and $S_2 \in \mathcal{M}^d$, it follows that

$$\begin{aligned}
r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) &= \frac{\int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} - \varrho^d \\
&= \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{\check{S}} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \varrho^d \right) \\
&\leqslant^{(*)} \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \varrho^d \right) \\
&\leqslant^{(**)} \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \right) \\
&\leqslant^{(***)} \int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x.
\end{aligned} \tag{A.11}$$

Here, at $(*)$ we use that $S_1 \in \check{\mathcal{M}}$, at $(**)$ we use (A.10) and $(***)$ holds because of (A.9).

Now, we claim there exist assortments $S_1 \in \check{\mathcal{M}}$ and $S_2 \in \mathcal{M}^d$ such that $S_2 \subseteq S_1$ and $\mathrm{vol}(S_1 \backslash S_2) \leqslant 1/N$. To this end, let $y_i \in B_i$ for $i \in [N]$ and define h_i as

$$h_i := \check{v}(y_i)\big(\check{w}(y_i) - \varrho^d\big), \qquad i \in [N].$$

In addition, let $\sigma : [N] \to [N]$ be an ordering, such that,

$$h_{\sigma(1)} \geqslant \ldots \geqslant h_{\sigma(N)},$$

where we break ties arbitrarily. As in Lemma 2.1, we first consider the case that $\mathrm{vol}(W_{\varrho^d}) \leqslant c$. Then, we know by Lemma 2.1 that $W_{\varrho^d} \in \check{\mathcal{M}}$. Since $\check{w}$ is constant

on each bin, there exists an integer n such that $\mathrm{vol}(W_{\varrho^d}) = n/N$. If $n/N \leqslant c$, then $n \leqslant K$ and hence $W_\varrho \in \mathcal{M}^d$ as well. This concludes the claim for $\mathrm{vol}(W_{\varrho^d}) \leqslant c$. Next, we consider the case that $\mathrm{vol}(W_{\varrho^d}) > c \geqslant K/N$. Then, $h_{\sigma(K)} \geqslant 0$ and

$$S_1 := \bigcup_{i=1}^{K} B_{\sigma(i)} \in \mathcal{M}^d.$$

In addition, note that $h_{\sigma(K+1)} \geqslant 0$ as well as $K < N$ since $c < 1$ and define

$$R := \left[\frac{\sigma(K+1)-1}{N}, \frac{\sigma(K+1)-1}{N} + c - \frac{K}{N}\right) \subset B_{\sigma(K+1)}.$$

Recall the definitions from Lemma 2.1 and note that, as $\check{v}$ and $\check{w}$ are constant on each bin,

$$m_{\varrho^d}(\ell) = \frac{i}{N}, \qquad \ell \in \big(h_{\sigma(i+1)}, h_{\sigma(i)}\big] \cap [0,\infty), \;\; i = 1, \dots, N-1.$$

As a result, $c = K/N$ implies $\ell_{\varrho^d} = h_{\sigma(K)}$ and $R = \varnothing$, and $c > K/N$ implies $\ell_{\varrho^d} = h_{\sigma(K+1)}$. Either way, it follows that

$$L^+_{\varrho^d} \subseteq S_1 \subseteq S_1 \cup R \subseteq L^+_{\varrho^d} \cup L^=_{\varrho^d}.$$

Since $\mathrm{vol}(S_1 \cup R) = c$, it follows from Lemma 2.1 that $S_2 := S_1 \cup R \in \check{\mathcal{M}}$. This concludes the claim for $\mathrm{vol}(W_{\varrho^d}) > c$.

From (A.11), the shown claim and the fact that $\check{w}(x) - \varrho^d \leqslant 1$, it follows that

$$r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) \leqslant \frac{\overline{v}}{N}. \qquad \square$$

Proof of Lemma 2.3.

Since $S \in \mathcal{A}_K$, we know that

$$\int_S v(x)\mathrm{d}x = \int_S \check{v}(x)\mathrm{d}x.$$

Therefore,

$$\begin{aligned} r(S, \check{v}, \check{w}) - r(S, v, w) &= \frac{1}{1 + \int_S v(x)\mathrm{d}x} \int_S \big(\check{v}(x)\check{w}(x) - v(x)w(x)\big)\mathrm{d}x \\ &\leqslant \|vw - \check{v}\check{w}\|_1 = \|vw - \check{v}w + \check{v}w - \check{v}\check{w}\|_1 \\ &\leqslant \|v - \check{v}\|_1 + \overline{v}\,\|w - \check{w}\|_1, \end{aligned}$$

where we have used that $w(x) \leqslant 1$ and $\check{v}(x) \leqslant \overline{v}$ for all $x \in [0,1]$. □

Proof of Theorem 2.3.

We start by showing that $\|v - \check{v}\|_1$ and $\|w - \check{w}\|_1$ are of order $1/N$. For $i \in [N]$, denote the constant $b_i = \check{v}(x)$ for some $x \in B_i$. Note that $b_i = \check{v}(x)$ for all $x \in B_i$ and that

$$\|v - \check{v}\|_1 = \int_0^1 |v(x) - \check{v}(x)| \mathrm{d}x = \sum_{i=1}^N \int_{B_i} |v(x) - b_i| \mathrm{d}x.$$

By the Mean Value Theorem, for every $i \in [N]$, there exists a c_i in the closure of B_i such that $v(c_i) = b_i$. Hence,

$$\|v - \check{v}\|_1 = \sum_{i=1}^N \int_{B_i} |v(x) - v(c_i)| \mathrm{d}x \leqslant L \sum_{i=1}^N \int_{B_i} |x - c_i| \mathrm{d}x \leqslant L \sum_{i=1}^N \frac{1}{2N^2} \leqslant \frac{L}{2N}, \quad \text{(A.12)}$$

where $L := \sup_{x \in [0,1]} |v'(x)|$. Likewise,

$$\|w - \check{w}\|_1 \leqslant \frac{Q}{2N},$$

where $Q := \sup_{x \in [0,1]} |w'(x)|$.

Now, let $\Delta_{\mathrm{UCB}}(T)$ denote the cumulative regret of UCB within the discrete MNL model. Recall that the preference parameters $v_1, \ldots, v_N$ satisfy

$$v_i = \int_{B_i} v(x) \mathrm{d}x, \qquad i \in [0,1],$$

and the parameters $w_1, \ldots, w_N$ satisfy

$$w_i = N \int_{B_i} v(x) \mathrm{d}x, \qquad i \in [0,1].$$

Let $S = \bigcup_{i \in D} B_i \in \mathcal{A}_K$ for some $D \subset [N]$. Then, the probability under v, as well as under $\check{v}$, that a purchase from assortment S lies in B_i is

$$\mathbb{P}(X^S \in B_i) = \frac{v_i}{1 + \sum_{i \in D} v_i},$$

In addition, the expected profit of assortment $S \in \mathcal{A}_K$ under $\check{v}$ and $\check{w}$ is

$$r(S, \check{v}, \check{w}) = \frac{\sum_{i \in D} v_i w_i}{1 + \sum_{i \in D} v_i}.$$

As a result, if $S_1, \ldots, S_T$ denote the offered assortment under DUCB(N) and S^d as

in (2.17), then

$$\sum_{t=1}^{T} \mathbb{E}_\pi \left[r(S^d, \check{v}, \check{w}) - r(S_t, \check{v}, \check{w}) \right] = \Delta_{\mathrm{UCB}}(T).$$

Following the steps of (2.18)–(2.21), in combination with the above and Proposition 2.2, Lemma 2.2 and Lemma 2.3, we find that, with $C_1 := L + \overline{v}(Q+1)$,

$$\Delta_\pi(T) \leqslant C_1 \frac{T}{N} + \Delta_{\mathrm{UCB}}(T).$$

By our choice of γ, we know that $\lfloor \gamma \rfloor \geqslant 1/c$. Hence, $N \geqslant 1/c \geqslant 1$ and $K \geqslant 1$. Second, γ is chosen such that $\overline{v} \leqslant N$ and therefore $v_i \leqslant 1$ for all $i \in [N]$. By Theorem 1 from Agrawal et al. (2019), there exists constants C_2 and C_3 such that

$$\Delta_{\mathrm{UCB}}(T) \leqslant C_2 \sqrt{NT \log NT} + C_3 N \log^2 NT.$$

Since $N \leqslant \gamma T^{1/3}$, it follows that

$$\log NT \leqslant \log \gamma T^{4/3} = \frac{4}{3} \log T + \log \gamma \leqslant C_4 \log T,$$

where $C_4 := \frac{4}{3} + \log \gamma / \log 2$. Hence,

$$\Delta_{\mathrm{UCB}}(T) \leqslant C_2 \sqrt{\gamma C_4} \sqrt{T^{4/3} \log T} + \gamma C_3 C_4^2 T^{1/3} \log^2 T.$$

Now we note that

$$\log T \leqslant \frac{9}{2e} T^{2/9}$$

and therefore

$$T^{1/3} \log^2 T \leqslant \left(\frac{9}{2e} \right)^{3/2} T^{2/3} (\log T)^{1/2}.$$

Thus we obtain that $\Delta_{\mathrm{UCB}}(T) \leqslant C_5 T^{2/3} (\log T)^{1/2}$, where

$$C_5 := C_2 \sqrt{\gamma C_4} + \left(\frac{9}{2e} \right)^{3/2} \gamma C_3 C_4^2.$$

Next, we point out that $N \geqslant (\gamma - 1) T^{1/3}$ with $\gamma \geqslant 2$. Thus,

$$\frac{T}{N} \leqslant \frac{1}{\gamma - 1} T^{2/3} \leqslant \frac{1}{(\gamma - 1)(\log 2)^{1/2}} T^{2/3} (\log T)^{1/2}.$$

From this we conclude that

$$\Delta_\pi(T) \leqslant C_1 \frac{T}{N} + C_5 T^{2/3} (\log T)^{1/2} \leqslant \overline{C}\, T^{2/3} (\log T)^{1/2},$$

where

$$\overline{C} := \frac{C_1}{(\gamma-1)(\log 2)^{1/2}} + C_5. \qquad \square$$

A.2.3 Proofs of the Results in Section 2.5.4

Before stating the proofs of the results in Section 2.5.4, we recollect the notations and concepts introduced in that section. Let $c \in (0, 0.25]$, $s = 0.05c$, $\delta = 0.2$ and $\sigma = 0.3$. Let $K \geqslant 2$ be an integer, chosen at the end of the proof of Theorem 2.4. Furthermore, for all $x \in [0,1]$, $i \in \{1, \dots, N_K\}$, and $I \subseteq \{1, \dots, N_K\}$, let

$$\begin{aligned}
N_K &= \lfloor K/c \rfloor, & [N_K] &= \{1, \dots, N_K\},\\
\mathcal{D}_K &= \{I \subseteq [N_K] : |I| = K\}, & B_i &= \left[c\frac{i-1}{K}, c\frac{i}{K}\right),\\
w(x) &= (1-s)\frac{1-\delta}{1-\delta x} + s, & v_0(x) &= \frac{s}{c(w(x)-s)},\\
b(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, & \varphi_i(x) &= \frac{2Kx}{c} - 2i + 1,\\
\tau_i(x) &= \frac{c}{K} b\big(\varphi_i(x)\big), & \beta &= \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} e^{-(2n-1)^2/2\sigma^2},\\
\varepsilon_I(x) &= \sum_{i\in I} \tau_i(x) - \beta, & v_I(x) &= v_0(x)\big(1 + \varepsilon_I(x)\big).
\end{aligned}$$

In addition, we use the following notation throughout this section. For $I \in \mathcal{D}_K$ we write

$$I^\dagger := \bigcup_{i\in I} B_i.$$

Furthermore, we define the following quantities.

$$\begin{aligned}
H &:= \frac{1}{\sigma\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} e^{-2n^2/\sigma^2},\\
L &:= \frac{1}{\sigma\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} e^{-(2n-1)^2/2\sigma^2} \qquad \text{and}\\
P &:= \mathbb{P}(-1/\sigma \leqslant Z \leqslant 1/\sigma),
\end{aligned}$$

where $Z \sim N(0,1)$. Observe that $\beta = Lc/K$.

We proceed by stating two preliminary lemmas that will be used throughout the proofs. Lemma A.4 contains a number of inequalities related to the quantities defined above, and Lemma A.5 shows that the optimal expected profit under v_0 is precisely

equal to s. The proof of these lemmas is given below.

Lemma A.4. *Let $I \subseteq [N_K]$. Then*

(i) for any $x \in [0,1]$, it holds that $\sum_{i \in I} \tau_i(x) \leqslant H\frac{c}{K}$.

(ii) for any $S \in \mathcal{S}$ and $\beta > 0$, it holds that

1. $\int_S v_I(x)\mathrm{d}x \leqslant \frac{s}{(1-s)(1-\delta)}(1+H)$ *and*
2. $\int_S (v_I(x))^2\mathrm{d}x \leqslant \frac{s^2}{c(1-s)^2(1-\delta)^2}(1+H)^2$,

(iii) for $x \notin I^\dagger$, it holds that $\sum_{i \in I} \tau_i(x) \leqslant \beta$,

(iv) if $|I| = K$ and $S \in \mathcal{S}$, it holds that $\mathrm{vol}(I^\dagger \backslash S) \geqslant \mathrm{vol}(S \backslash I^\dagger)$,

(v) for all $i \in [N]$;

1. $\frac{c^2}{2K^2}P = \int_{B_i} \tau_i(x)\mathrm{d}x \leqslant \int_0^1 \tau_i(x)\mathrm{d}x \leqslant \frac{c^2}{2K^2}$ *and*
2. $\int_0^1 (\tau_i(x))^2\mathrm{d}x \leqslant \frac{c^3}{4\sigma\sqrt{\pi}K^3}$,

(vi) for any $i \in I$, $x \in B_i$ and $\beta' \geqslant \beta$, it holds that $|\varepsilon_I(x;\beta')| \leqslant \tau_i(x) + \beta'$.

Lemma A.5. *The optimal expected revenue under preference function v_0 equals s:*

$$\max_{S \in \mathcal{S}} r(S, v_0) = s.$$

Proof of Proposition 2.3.

Let

$$C_1 := \frac{s(1-s)(1-\delta)}{c(1-s)(1-\delta)+cs} \qquad \text{and} \qquad C_2 := \frac{s^2(c+2L)(4+Hc)}{4\big((1-s)(1-\delta)+s\big)(1-s)(1-\delta)}.$$

Let π be a policy, $T \in \mathbb{N}$, and let $I \in \mathcal{D}_K$. Write $v(x) := v_I(x)$, and let S^* denote an optimal assortment under v. Recall that S^* also maximizes the inner maximization problem (2.6) for $\varrho = \varrho^* = \max_{S \in \mathcal{S}} r(S, v)$. Therefore,

$$\int_{S^*} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \geqslant \int_{I^\dagger} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x. \tag{A.13}$$

In addition, observe that for $x \in [0,1]$

$$v_0(x) \leqslant \frac{s}{c(1-s)(1-\delta)}. \tag{A.14}$$

It now follows that, for all $S \in \mathcal{S}$,

$$\begin{aligned}
r(S^*, v) - r(S, v) &= \varrho^* - \frac{\int_S v(x)w(x)\mathrm{d}x}{1 + \int_S v(x)\mathrm{d}x} \\
&= \frac{1}{1 + \int_S v(x)\mathrm{d}x} \left(\varrho^* - \int_S v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \right) \\
&=^{(*)} \frac{1}{1 + \int_S v(x)\mathrm{d}x} \left(\int_{S^*} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \right) \\
&\geqslant^{(**)} \frac{(1-s)(1-\delta)}{(1-s)(1-\delta) + s} \left(\int_{S^*} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \right) \\
&\geqslant^{(***)} \frac{(1-s)(1-\delta)}{(1-s)(1-\delta) + s} \left(\int_{I^\dagger} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \right).
\end{aligned} \tag{A.15}$$

Here, $(*)$ follows from Proposition 2.1, $(**)$ follows by (A.14), and $(***)$ follows by (A.13). The terms within the large parentheses in (A.15) can be bounded from below as

$$\begin{aligned}
\int_{I^\dagger} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x & \\
&= \int_{I^\dagger} v(x)\big(w(x) - s\big)\mathrm{d}x - \int_S v(x)\big(w(x) - s\big)\mathrm{d}x \\
&\quad - (\varrho^* - s) \left(\int_{I^\dagger} v(x)\mathrm{d}x - \int_S v(x)\mathrm{d}x \right) \\
&\geqslant^{(*)} \frac{s}{c} \int_{I^\dagger \setminus S} \big(1 + \varepsilon_I(x)\big)\mathrm{d}x - \frac{s}{c} \int_{S \setminus I^\dagger} \big(1 + \varepsilon_I(x)\big)\mathrm{d}x \qquad \text{(A.16)} \\
&\quad - |\varrho^* - s| \left(\int_{I^\dagger \setminus S} v(x)\mathrm{d}x + \int_{S \setminus I^\dagger} v(x)\mathrm{d}x \right), \qquad \text{(A.17)}
\end{aligned}$$

where at $(*)$ we use that by design $v(x)\big(w(x) - s\big) = \frac{s}{c}(1 + \varepsilon_I(x))$. We proceed by bounding the two terms in (A.17) from above. The absolute difference between ϱ^* and s can be bounded from above by the L_1-difference between v and v_0, denoted as $\|v - v_0\|_1$, as follows. For $S \in \mathcal{S}$ and $\varrho \in [0,1]$, let

$$\mathcal{I}_0(S, \varrho) = \int_S v_0\big(w(x) - \varrho\big) \qquad \text{and} \qquad \mathcal{I}(S, \varrho) = \int_S v\big(w(x) - \varrho\big).$$

Since $w(x) - \varrho^* \in [0,1]$ for all $x \in S^*$, we therefore know that

$$\begin{aligned}\mathcal{I}(S^*, \varrho^*) - \mathcal{I}_0(S^*, s) &\leqslant \int_{S^*} \big|v(x) - v_0(x)\big| \mathrm{d}x \\ &\leqslant \int_0^1 \big|v(x) - v_0(x)\big| \mathrm{d}x = \|v - v_0\|_1.\end{aligned}$$

Furthermore,

$$\mathcal{I}_0(S^*, \varrho^* - \|v - v_0\|_1) \geqslant \mathcal{I}_0(S^*, \varrho^*) \geqslant \mathcal{I}(S^*, \varrho^*) - \|v - v_0\|_1 = \varrho^* - \|v - v_0\|_1.$$

Hence, there exists an $S \in \mathcal{S}$ such that $\mathcal{I}(S, \varrho^* - \|v - v_0\|_1) \geqslant \varrho^* - \|v - v_0\|_1$ and by (2.4) this entails $s \geqslant \varrho^* - \|v - v_0\|_1$. Likewise, we derive $\varrho^* \geqslant s - \|v - v_0\|_1$ and so $|\varrho^* - s| \leqslant \|v - v_0\|_1$. We proceed by developing an upper bound on the L_1-difference between v and v_0:

$$\begin{aligned}\int_0^1 \big|v(x) - v_0(x)\big| \mathrm{d}x = \int_0^1 v_0(x) \big|\varepsilon_I(x)\big| \mathrm{d}x &\leqslant \frac{s}{c(1-s)(1-\delta)} \int_0^1 \big|\varepsilon_I(x)\big| \mathrm{d}x \\ &\leqslant \frac{s}{c(1-s)(1-\delta)} \left(\sum_{i \in I} \int_0^1 \tau_i(x) \mathrm{d}x + \beta \right) \\ &\leqslant^{(*)} \frac{s(c+2L)}{2(1-s)(1-\delta)} \frac{1}{K}.\end{aligned}$$

Here, $(*)$ is justified by Lemma A.4.(v).1. The remaining term from (A.17) is bounded from above as

$$\begin{aligned}\int_{I^\dagger \setminus S} v(x) \mathrm{d}x + \int_{S \setminus I^\dagger} v(x) \mathrm{d}x &\leqslant^{(*)} \frac{s}{c(1-s)(1-\delta)} \left(\int_{I^\dagger \setminus S} (1 + \varepsilon_I(x)) \mathrm{d}x + \mathrm{vol}(S \setminus I^\dagger) \right) \\ &\leqslant^{(**)} \frac{s}{c(1-s)(1-\delta)} \int_{I^\dagger \setminus S} (2 + \varepsilon_I(x)) \mathrm{d}x \\ &\leqslant \frac{s}{c(1-s)(1-\delta)} \int_{I^\dagger \setminus S} \left(2 + \sum_{i \in I} \tau_i(x) \right) \mathrm{d}x \\ &\leqslant^{(***)} \frac{s(4 + Hc)}{2(1-s)(1-\delta)}\end{aligned}$$

Here, at $(*)$ we apply (A.14), $(**)$ holds since $\mathrm{vol}(S \setminus I^\dagger) \leqslant \mathrm{vol}(I^\dagger \setminus S)$ by Lemma A.4.(iv) and $(***)$ follows from Lemma A.4.(i) and $K \geqslant 2$. Next, consider the expression in (A.16). Since $\varepsilon_I(x) \leqslant 0$ for $x \notin I^\dagger$ by Lemma A.4.(iii) and $\mathrm{vol}(S \setminus I^\dagger) \leqslant \mathrm{vol}(I^\dagger \setminus S)$ by

Lemma A.4.(iv), we conclude that

$$\begin{aligned}\int_{I^\dagger\setminus S}(1+\varepsilon_I(x))\,\mathrm{d}x - \int_{S\setminus I^\dagger}(1+\varepsilon_I(x))\,\mathrm{d}x \\ &\geqslant \int_{I^\dagger\setminus S}(1+\varepsilon_I(x))\,\mathrm{d}x - \mathrm{vol}(S\setminus I^\dagger) \\ &\geqslant \int_{I^\dagger\setminus S}(1+\varepsilon_I(x))\,\mathrm{d}x - \mathrm{vol}(I^\dagger\setminus S) = \int_{I^\dagger\setminus S}\varepsilon_I(x)\mathrm{d}x.\end{aligned}$$

Hence,

$$\begin{aligned}r(S^*,v) - r(S,v) \geqslant &\frac{s(1-s)(1-\delta)}{c(1-s)(1-\delta)+cs}\int_{I^\dagger\setminus S}\varepsilon_I(x)\mathrm{d}x \\ &- \frac{s^2(c+2L)(4+Hc)}{4\big((1-s)(1-\delta)+s\big)(1-s)(1-\delta)}\,\frac{1}{K}.\end{aligned}$$

Applying the latter inequality to $S = S_t$, for $t = 1,\dots,T$, and taking the expectation of the sum of these terms yields the desired result, since

$$\begin{aligned}\mathbb{E}_I\left[\sum_{t=1}^{T}\int_{I^\dagger\setminus S_t}\varepsilon_I(x)\mathrm{d}x\right] &= \int_{I^\dagger}\mathbb{E}_I\left[\sum_{t=1}^{T}\big(1-\mathbf{1}\{x\in S_t\}\big)\varepsilon_I(x)\mathrm{d}x\right] \\ &= \int_{I^\dagger}\big(T-\mathbb{E}_I[k(x)]\big)\varepsilon_I(x)\mathrm{d}x. \qquad \square\end{aligned}$$

Proof of Proposition 2.4.

Let $x\in[0,1]$, $I\in\mathcal{D}_K$, $i\in I$, and $J = I\setminus\{i\}$. It suffices to show that there is a $C_c>0$ such that

$$\Big|\mathbb{E}_I[k(x)] - \mathbb{E}_J[k(x)]\Big| \leqslant T\sqrt{2\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J)}, \tag{A.18}$$

and

$$\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J) \leqslant \tfrac{1}{2}C_c^2\frac{T}{K^3}. \tag{A.19}$$

We first prove (A.18), using Pinsker's inequality that states that for any probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined on the same probability space $(\Omega,\mathcal{F})$,

$$2\sup_{A\in\mathcal{F}}\Big(\mathbb{P}(A)-\mathbb{Q}(A)\Big)^2 \leqslant \mathrm{KL}(\mathbb{P}\|\mathbb{Q}),$$

or, equivalently,

$$\sup_{A\in\mathcal{F}}\Big|\mathbb{P}(A)-\mathbb{Q}(A)\Big| \leqslant \sqrt{\tfrac{1}{2}\mathrm{KL}(\mathbb{P}\|\mathbb{Q})}. \tag{A.20}$$

Consider the probability measures p and q on $\{0,\dots,T\}$, defined by

$$p(n) := \mathbb{P}_I(k(x)=n) \qquad \text{and} \qquad q(n) := \mathbb{P}_J(k(x)=n), \quad n \in \{0,\dots,T\}.$$

From the equality

$$\sup_{n=0,\dots,T} |p(n)-q(n)| = \frac{1}{2}\sum_{n=0}^{T} |p(n)-q(n)|. \tag{A.21}$$

we obtain

$$\begin{aligned}\Big|\mathbb{E}_I[k(x)] - \mathbb{E}_J[k(x)]\Big| &= \left|\sum_{n=0}^{T} n(p(n)-q(n))\right| \\ &\leqslant \sum_{n=0}^{T} n\,|p(n)-q(n)| \leqslant T\sum_{n=0}^{T}|p(n)-q(n)| \\ &=^{(*)} 2T \sup_{n=0,\dots,T}|p(n)-q(n)| \leqslant^{(**)} T\sqrt{2\,\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J)},\end{aligned}$$

where $(*)$ follows by (A.21), and $(**)$ follows by (A.20). This proves (A.18).

We now prove (A.19). Write $v(x) = v_I(x)$ and $u(x) = v_J(x)$, for $x \in [0,1]$. We denote the no-purchase probabilities at time t as

$$p_t := \frac{1}{1+\int_{S_t} v(x)\mathrm{d}x} \qquad \text{and} \qquad q_t := \frac{1}{1+\int_{S_t} u(x)\mathrm{d}x}.$$

Note by Lemma A.4.(ii).1 that $p_t, q_t \in [p_0, 1]$, where

$$p_0 := \frac{(1-s)(1-\delta)}{(1-s)(1-\delta)+s(1+H)}.$$

The Kullback-Leibler (KL) divergence $\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J)$ can be written as

$$\begin{aligned}\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J) &= \mathbb{E}_I\sum_{t=1}^{T}\left(p_t\log\frac{p_t}{q_t} + \int_{S_t}\log\left(\frac{p_t v(x)}{q_t u(x)}\right)p_t v(x)\mathrm{d}x\right) \\ &= \mathbb{E}_I\sum_{t=1}^{T}\left(p_t\log\left(1+\frac{p_t-q_t}{q_t}\right) + \int_{S_t}\log\left(1+\frac{p_t v(x)-q_t u(x)}{q_t u(x)}\right)p_t v(x)\mathrm{d}x\right).\end{aligned}$$

Since $\log(1+x) \leqslant x$ for all $x > -1$, we find the following upper bound:

$$\begin{aligned}\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J) = \mathbb{E}_I\sum_{t=1}^{T}\Bigg(&p_t\log\left(1+\frac{p_t-q_t}{q_t}\right) \\ &+ \int_{S_t}\log\left(1+\frac{p_t v(x)-q_t u(x)}{q_t u(x)}\right)p_t v(x)\mathrm{d}x\Bigg)\end{aligned}$$

$$
\begin{aligned}
&\leqslant \mathbb{E}_I \sum_{t=1}^{T} \left(p_t \frac{p_t - q_t}{q_t} + \int_{S_t} \frac{p_t v(x) - q_t u(x)}{q_t u(x)} p_t v(x) \mathrm{d}x \right) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\quad + \mathbb{E}_I \sum_{t=1}^{T} \left(p_t - q_t + \int_{S_t} \big(p_t v(x) - q_t u(x)\big) \mathrm{d}x \right) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\quad + \mathbb{E}_I \sum_{t=1}^{T} \big(p_t - q_t + (1 - p_t) - (1 - q_t)\big) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right).
\end{aligned}
$$

Note that $q_t \geqslant p_0$ and $u(x) \geqslant 1/C_c^a$ for all $x \in [0,1]$, where

$$
C_c^a := \frac{c(1-s)}{s(1-\beta)} > 0.
$$

Hence, we can bound the KL divergence further as

$$
\begin{aligned}
\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) &\leqslant \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\leqslant \frac{1}{p_0} \mathbb{E}_I \sum_{t=1}^{T} \left(\underbrace{(p_t - q_t)^2}_{(a)} + C_c^a \underbrace{\int_{S_t} \big(p_t v(x) - q_t u(x)\big)^2 \mathrm{d}x}_{(b)} \right). \qquad \text{(A.22)}
\end{aligned}
$$

We bound both (a) and (b) in (A.22) from above. Let $t \in \{1, \dots, T\}$. For (a), observe that

$$
\begin{aligned}
(p_t - q_t)^2 &=^{(*)} \frac{\left(\int_{S_t} (v(x) - u(x)) \mathrm{d}x \right)^2}{\left(1 + \int_{S_t} v(x) \mathrm{d}x \right)^2 \left(1 + \int_{S_t} u(x) \mathrm{d}x \right)^2} \\
&\leqslant \left(\int_{S_t} (v(x) - u(x)) \mathrm{d}x \right)^2 \\
&\leqslant \left(\frac{s}{c(1-s)(1-\delta)} \int_{S_t} \tau_i(x) \mathrm{d}x \right)^2 \leqslant^{(**)} C_c^b \frac{c^2}{4K^4}, \qquad \text{(A.23)}
\end{aligned}
$$

where

$$C_c^b := \frac{s^2}{c^2(1-s)^2(1-\delta)^2},$$

and where $(*)$ holds since the cross terms cancel out and $(**)$ follows by Lemma A.4.(v).1.

We now bound (b) in (A.22) from above. Observe that

$$\begin{aligned}\int_{S_t} \big(p_t v(x) - q_t u(x)\big)^2 \mathrm{d}x &= \int_{S_t} \big(p_t v(x) - q_t v(x) + q_t v(x) - q_t u(x)\big)^2 \mathrm{d}x \\ &= (p_t - q_t)^2 \int_{S_t} v(x)^2 \mathrm{d}x \qquad \text{(A.24)} \\ &\quad + 2q_t(p_t - q_t) \int_{S_t} v(x)\tau_i(x)\mathrm{d}x \qquad \text{(A.25)} \\ &\quad + q_t^2 \int_{S_t} (\tau_i(x))^2 \,\mathrm{d}x. \qquad \text{(A.26)}\end{aligned}$$

The integral in (A.24) can be bounded by applying Lemma A.4.(v).2. Combining that with the bound for $(p_t - q_t)^2$ from (A.23), gives

$$(p_t - q_t)^2 \int_{S_t} v(x)^2 \mathrm{d}x \leqslant (C_c^b)^2(1+H)\frac{c^3}{4K^4}.$$

For the term (A.25), Lemma A.4.(i) shows that $\tau_i(x) \leqslant Hc/K$. Together with (A.23) and Lemma A.4.(ii).1 we find

$$\begin{aligned}2q_t(p_t - q_t) \int_{S_t} v(x)\tau_i(x)\mathrm{d}x &\leqslant 2|p_t - q_t| \int_{S_t} v(x)\tau_i(x)\mathrm{d}x \\ &\leqslant 2|p_t - q_t| \left(\max_{x\in[0,1]} \tau_i(x)\right) \int_S v(x)\mathrm{d}x \\ &\leqslant 2\sqrt{C_c^b}\frac{c}{2K^2}\left(H\frac{c}{K}\cdot c\sqrt{C_c^b}(1+H)\right) \\ &= C_c^b H(1+H)\frac{c^3}{K^3}\end{aligned}$$

Finally, we bound the term (A.26). As a consequence of Lemma A.4.(v).2, we have

$$q_t^2 \int_{S_t} (\tau_i(y))^2 \,\mathrm{d}y \leqslant \int_{S_t} (\tau_i(y))^2 \,\mathrm{d}y \leqslant \frac{c^3}{4\sigma\sqrt{\pi}K^3}.$$

Inserting the derived upper bounds on (A.24), (A.25), (A.26) in (A.22), we obtain

$$\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) \leqslant \frac{1}{p_0}\mathbb{E}_I \sum_{t=1}^{T} \left((p_t - q_t)^2 + C_c^a \int_{S_t} \big(p_t v(y) - q_t u(y)\big)^2 \mathrm{d}y\right)$$

$$\leqslant \frac{1}{p_0}\mathbb{E}_I \sum_{t=1}^{T}\left(C_c^b\frac{c^2}{4K^4}+C_c^a\left((C_c^b)^2(1+H)\frac{c^3}{4K^4}+C_c^bH(1+H)\frac{c^3}{K^3}+\frac{c^3}{4\sigma\sqrt{\pi}K^3}\right)\right)$$

$$\leqslant \frac{1}{4p_0}\left(c^2C_c^b+C_c^a\left(c^3(C_c^b)^2(1+H)+4c^3C_c^bH(1+H)+\frac{c^3}{\sigma\sqrt{\pi}}\right)\right)\frac{T}{K^3}.$$

This implies (A.19). □

Proof of Theorem 2.4.

We first show that the preference functions v_0 and $\{v_I : I \in \mathcal{D}_K, K \geqslant 2\}$ satisfy Assumption 2.1. To see this observe that, for all $x \in [0,1]$ and all $c \in (0, 0.25]$, the choice $s = 0.05c$ and $\delta = 0.2$ implies

$$v_0(x) \in \left[\frac{0.05}{1-s}, \frac{0.0625}{1-s}\right] \subseteq [0.05, 0.07]\,.$$

Moreover, for all $K \geqslant 2$ and $I \in \mathcal{D}_K$ we have $\beta \leqslant L/8 \leqslant 0.0013$, and therefore

$$v_I(x) \geqslant v_0(x)(1-\beta) \geqslant 0.04,$$

and Lemma A.4.(i) implies

$$v_I(x) \leqslant v_0(x)\left(1+\frac{H}{8}\right) \leqslant 0.09,$$

for all $c \in (0, 0.25]$. Moreover, we note that

$$\frac{w(0)}{\int_0^1 (w(x)-w(0))\mathrm{d}x} < 9,$$

for all $c \in (0, 0.25]$. This shows that Assumption 2.1.(i) is satisfied with $\overline{v} = 0.04$ and $\underline{v} = 9$.

We now show that $v_I'(\cdot)$ is uniformly bounded and hence Assumption 2.1.(ii) is satisfied as well. To this end, observe that

$$|v_I'(x)| = |v_0'(x)|\left|1+\sum_{i\in I}\tau_i(x)\right| + |v_0(x)|\left|\sum_{i\in I}\tau_i'(x)\right|,$$

for all $x \in [0,1]$. Therefore, by Lemma A.4.(i) it suffices to show that $\sum_{i\in I}\tau_i'(x)$ is

uniformly bounded. Note that

$$\tau_i'(x) = -\frac{2}{\sigma^2}\varphi_i(x)b\big(\varphi_i(x)\big).$$

For all $x \in [0,1]$, let $i_x := \lfloor Kx/c \rfloor$. Then, $x \in B_{i_x}$ for all $x \in [0,1]$, where $B_{N_K+1} := [0,1]\backslash\bigcup_{i\in[N_K]} B_i$, and $\varphi_i(B_{i_x}) = \big[2(i_x-i)-1, 2(i_x-i)+1\big)$. Since $|yb(y)|$ is decreasing for $y \geqslant 1$ and increasing for $y \leqslant -1$, we obtain that, for all $i < i_x$,

$$0 < \varphi_i(x)b\big(\varphi_i(x)\big) \leqslant \big(2(i_x-i)-1\big)b\big(2(i_x-i)-1\big),$$

and for all $i > i_x$,

$$0 < -\varphi_i(x)b\big(\varphi_i(x)\big) \leqslant \big(2(i_x-i)+1\big)b\big(2(i_x-i)+1\big).$$

From this we conclude that

$$\begin{aligned}
\left|\sum_{i\in I}\tau_i'(x)\right| &= \left|\sum_{i\in I}\frac{2}{\sigma^2}\varphi_i(x)b\big(\varphi_i(x)\big)\right| \\
&\leqslant \frac{2}{\sigma^2}\left(\big|\varphi_{i_x}(x)b\big(\varphi_{i_x}(x)\big)\big| + \sum_{i=1}^{i_x-1}\varphi_i(x)b\big(\varphi_i(x)\big) - \sum_{i=i_x+1}^{N_K}\varphi_i(x)b\big(\varphi_i(x)\big)\right) \\
&\leqslant \frac{2}{\sigma^2}\left(\big|\varphi_{i_x}(x)b\big(\varphi_{i_x}(x)\big)\big| + \sum_{i=1}^{i_x-1}\big(2(i_x-i)-1\big)b\big(2(i_x-i)-1\big)\right. \\
&\qquad \left. - \sum_{i=i_x+1}^{N_K}\big(2(i_x-i)+1\big)b\big(2(i_x-i)+1\big)\right) \\
&\leqslant \frac{2}{\sigma^2}\left(\frac{\sigma}{\sqrt{e}} + 2\sum_{n=1}^{\infty}(2n-1)b\big(2n-1\big)\right) < \infty.
\end{aligned}$$

As a result, v_0 and $\{v_I : I \in \mathcal{D}_K, K \geqslant 2\}$ satisfy Assumption 2.1. This implies

$$\begin{aligned}
\Delta_\pi(T) &= \sup_{v\in\mathcal{V}}\Delta_\pi(T,v) \\
&\geqslant \frac{1}{|\mathcal{D}_K|}\sum_{I\in\mathcal{D}_K}\Delta_\pi(T,v_I) \\
&\geqslant \frac{1}{|\mathcal{D}_K|}\sum_{I\in\mathcal{D}_K}\left(C_1\int_{I^\dagger}(T-\mathbb{E}_I[k(x)])\varepsilon_I(x)\mathrm{d}x - C_2\frac{T}{K}\right), \qquad \text{(A.27)}
\end{aligned}$$

where C_1 and C_2 are as in Proposition 2.3.

The integral $\int_{I^\dagger} \varepsilon_I(x)\mathrm{d}x$ can be bounded from below as

$$\begin{aligned}\int_{I^\dagger} \varepsilon_I(x)\mathrm{d}x &= \sum_{i\in I}\int_{I^\dagger} \tau_i(x)\mathrm{d}x - \beta c = \sum_{i\in I}\int_{B_i} \tau_i(x)\mathrm{d}x - \beta c \\ &\geqslant^{(*)} P\frac{c^2}{2K} - L\frac{c^2}{K} = \frac{c^2(P-2L)}{2K},\end{aligned}$$

where at $(*)$ we used Lemma A.4.(v).1. We use this lower bound to analyze (A.27). To this end, let $C_3 := c^2 C_1(P-2L)/2$. Then

$$\Delta_\pi(T) \geqslant (C_3 - C_2)\frac{T}{K} - \frac{C_1}{|\mathcal{D}_K|}\underbrace{\sum_{I\in\mathcal{D}_K}\int_{I^\dagger} \mathbb{E}_I[k(x)]\varepsilon_I(x)\mathrm{d}x}_{(a)}. \tag{A.28}$$

We now bound the term (a) in (A.28) from above, using Proposition 2.4. Let C_c denote the constant from Proposition 2.4, and let $I \in \mathcal{D}_K$ and $J = I\backslash\{i\}$ for some $i \in I$. Then, for $x \in B_i$,

$$\mathbb{E}_I[k(x)]\varepsilon_I(x) \leqslant \left(\mathbb{E}_J[k(x)] + C_c\left(\frac{T}{K}\right)^{3/2}\right)|\varepsilon_I(x)|. \tag{A.29}$$

To apply (A.29) in order to bound (a) in (A.28), we change the order of summation and integration and rewrite the summation itself. Let $U = \bigcup_{i=1}^{N_K} B_i$ denote the union of all bins, and for all $x \in U$, let $i_x = \lfloor Kx/c \rfloor$ again denote the index of the bin B_{i_x} such that $x \in B_{i_x}$, for all $x \in [0,1]$. Note that for each $x \in U$ that the mapping $I \mapsto I\backslash\{i_x\}$ between

$$E_K^x := \{I \in \mathcal{D}_K : x \in I^\dagger\} \qquad \text{and} \qquad F_{K-1}^x := \{J \in \mathcal{D}_{K-1} : x \notin J^\dagger\}$$

is a bijection. Hence,

$$\begin{aligned}\sum_{I\in\mathcal{D}_K}\int_{I^\dagger} \mathbb{E}_I[k(x)]\varepsilon_I(x)\mathrm{d}x &= \int_{x\in U}\sum_{I\in E_K^x} \mathbb{E}_I[k(x)]\varepsilon_I(x)\mathrm{d}x \\ &= \int_{x\in U}\sum_{J\in F_{K-1}^x} \mathbb{E}_{J\cup\{i_x\}}[k(x)]\varepsilon_{J\cup\{i_x\}}(x)\mathrm{d}x \\ &\leqslant^{(*)} \int_{x\in U}\sum_{J\in F_{K-1}^x} \mathbb{E}_J[k(x)]\big|\varepsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x \qquad (A.30) \\ &\quad + C_c\left(\frac{T}{K}\right)^{3/2}\int_{x\in U}\sum_{J\in F_{K-1}^x}\big|\varepsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x, \qquad (A.31)\end{aligned}$$

where at $(*)$ we apply (A.29). We now bound (A.30) and (A.31) from above. For (A.30), $|\varepsilon_I(x)|$ is bounded uniformly in x by Lemma A.4.(i):

$$\begin{aligned}\int_{x\in U}\sum_{J\in F^x_{K-1}}\mathbb{E}_J[k(x)]\big|\varepsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x &\leqslant (H+L)\,\frac{c}{K}\int_{x\in U}\sum_{J\in F^x_{K-1}}\mathbb{E}_J[k(x)]\mathrm{d}x\\ &= (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\int_{x\in U\setminus J^\dagger}\mathbb{E}_J[k(x)]\mathrm{d}x\\ &\leqslant (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\int_0^1\mathbb{E}_J[k(x)]\mathrm{d}x\\ &\leqslant (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\sum_{t=1}^{T}\mathbb{E}_J[\mathrm{vol}(S_t)]\\ &\leqslant (H+L)\,\frac{c^2}{K}|\mathcal{D}_{K-1}|T.\end{aligned}$$

We now consider (A.31). Observe that $|\varepsilon_I(x)|$ is bounded locally on B_i:

$$\begin{aligned}&\int_{x\in U}\sum_{J\in F^x_{K-1}}\big|\varepsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x=\sum_{J\in\mathcal{D}_{K-1}}\int_{x\in U\setminus J^\dagger}\big|\varepsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x\\ &=\sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\int_{B_i}|\varepsilon_{J\cup\{i\}}(x)|\mathrm{d}x\leqslant^{(*)}\sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\int_{B_i}\big(\tau_i(x)+\beta\big)\mathrm{d}x\\ &\leqslant^{(**)}\sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\frac{c^2(1+2L)}{2K^2}=\frac{c^2(1+2L)}{2K^2}|\mathcal{D}_{K-1}|(N_K-K+1),\end{aligned}$$

where we apply Lemma A.4.(vi) at $(*)$ and Lemma A.4.(v).1 at $(**)$. After inserting these upper bounds for (A.30) and (A.31) into (A.28), we conclude

$$\begin{aligned}\Delta_\pi(T)\geqslant{}&(C_3-C_2)\frac{T}{K}\\ &-\frac{C_1}{|\mathcal{D}_K|}\left((H+L)\,\frac{c^2}{K}|\mathcal{D}_{K-1}|T+\frac{c^2C_c(1+2L)}{2}|\mathcal{D}_{K-1}|(N-K+1)\frac{T^{3/2}}{K^{7/2}}\right).\end{aligned}$$

Rewriting the expression above yields

$$\begin{aligned}\Delta_\pi(T)\geqslant{}&\left(C_3-C_2-\frac{c^2C_1\,(H+L)\,|\mathcal{D}_{K-1}|}{|\mathcal{D}_K|}\right)\frac{T}{K}\\ &-\frac{c^2C_1C_c(1+2L)|\mathcal{D}_{K-1}|}{2|\mathcal{D}_K|}(N-K+1)\frac{T^{3/2}}{K^{7/2}}.\end{aligned}$$

Next, note that

$$\frac{|\mathcal{D}_{K-1}|}{|\mathcal{D}_K|}=\frac{K}{N_K-K+1},$$

and therefore

$$\Delta_\pi(T) \geqslant \underbrace{\left(C_3 - C_2 - \frac{(H+L)\,c^2 C_1 K}{N_K - K + 1}\right)}_{(b)} \frac{T}{K} - \frac{c^2 C_1 C_c (1+2L)}{2} \frac{T^{3/2}}{K^{5/2}}. \tag{A.32}$$

We abbreviate the constant $C_4 := c^2 C_1 C_c (1+2L)/2$. The factor (b) in front of the T/K term above can be bounded further from below. To this end, note that

$$N_K - K + 1 \geqslant \left(\frac{1}{c} - 1\right) K,$$

and therefore (A.32) implies

$$\Delta_\pi(T) \geqslant \left(C_3 - C_2 - \frac{(H+L)\,c^3 C_1}{1-c}\right) \frac{T}{K} - C_4 \frac{T^{3/2}}{K^{5/2}}.$$

Let

$$C_6 := \frac{P - 2L}{2} - \frac{(H+L)c}{1-c},$$

and

$$C_5 := C_3 - C_2 - \frac{(H+L)c^3 C_1}{1-c} = c^2 C_1 C_6 - C_2.$$

By computation and the assumption $c \in (0, 0.25]$ we obtain $C_6 \geqslant (P-2L)/2 - (H+L)/3 \geqslant 0.042 > 0$. In addition, our choice of $s = 0.05c$ implies

$$4c(1-s)^2(1-\delta)^2 C_6 > s(c+2L)(4+Hc),$$

for $c \in (0, 0.25]$ and therefore $C_5 = c^2 C_1 C_6 - C_2 > 0$. Now, choose

$$\gamma = \left(\frac{5C_4}{C_5}\right)^{2/3} \qquad \text{and} \qquad K = \max\left\{2, \left\lceil \gamma T^{1/3} \right\rceil\right\}.$$

For $T > 1/\gamma^3$, we know that $K = \lceil \gamma T^{1/3} \rceil$ as well as $K < \gamma T^{1/3} + 1 < 2\gamma T^{1/3}$ and $K \geqslant \gamma T^{1/3}$. Therefore, for $T > 1/\gamma^3$

$$\begin{aligned}\Delta_\pi(T) &\geqslant \frac{C_5}{2\gamma}\, T^{1/3} - \frac{C_4}{\gamma^{5/2}}\, T^{1/3} \\ &= \left(\tfrac{1}{2}\left(\tfrac{1}{5}\right)^{2/3} - \left(\tfrac{1}{5}\right)^{5/3}\right) \frac{C_5^{5/3}}{C_4^{2/3}}\, T^{2/3}.\end{aligned}$$

For T such that $1 \leqslant T \leqslant 1/\gamma^3$, we know that $K = 2$ as well as $\sqrt{T} \leqslant C_5/5C_4$ and

thus

$$\begin{aligned}\Delta_\pi(T) &\geqslant \frac{C_5}{2}\,T - \frac{\sqrt{2}C_4}{8}\,T^{3/2}\\ &= \left(\frac{C_5}{2} - \frac{\sqrt{2}C_4}{8}\,\sqrt{T}\right)T\\ &\geqslant \left(\tfrac{1}{2} - \tfrac{\sqrt{2}}{40}\right)C_5\,T \geqslant \left(\tfrac{1}{2} - \tfrac{\sqrt{2}}{40}\right)C_5\,T^{2/3}.\end{aligned}$$

Therefore, we have shown the desired result for

$$\underline{C} = \min\left\{\left(\tfrac{1}{2} - \tfrac{\sqrt{2}}{40}\right)C_5, \left(\tfrac{1}{2}\left(\tfrac{1}{5}\right)^{2/3} - \left(\tfrac{1}{5}\right)^{5/3}\right)\frac{C_5^{5/3}}{C_4^{2/3}}\right\} > 0. \qquad \square$$

Proof of Lemma A.4.

For $x \in [0,1]$, let $i_0 \in [N_K]$ $y = 2Kx/c - 2i_0 + 1$. Then, we find that (i) holds due to

$$\begin{aligned}\sum_{i\in I}\tau_i(x) &= \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{i\in I}\exp\big(-\tfrac{1}{2\sigma^2}(y + 2i_0 - 2i)^2\big)\\ &\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\exp\big(-\tfrac{1}{2\sigma^2}(y - 2n)^2\big)\\ &\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{2n^2}{\sigma^2}\right) = H\frac{c}{K}.\end{aligned}$$

Observe that (ii) is a corollary of (i), since $v_0(x) \leqslant \frac{s}{c(1-s)(1-\delta)}$ for all $x \in [0,1]$, and therefore

$$v_I(x) \leqslant \frac{s}{c(1-s)(1-\delta)}\left(1 + \sum_{i\in I}\tau_i(x)\right).$$

For (iii), let $x \notin I^\dagger$ and $i_x := \lfloor Kx/c \rfloor$ such that $x \in B_{i_x}$, where $B_{N_K+1} := [0,1]\setminus\bigcup_{i\in[N_K]} B_i$. Note that τ_i is either increasing or decreasing on B_{i_x} for $i \neq i_x$. Then,

$$\tau_i(x) \leqslant \max\left\{\tau_i\left(c\frac{i_x - 1}{K}\right), \tau_i\left(c\frac{i_x}{K}\right)\right\} = \frac{c}{K}\max_{\ell\in\{-1,1\}}\Big\{b\big(2(i_x - i) + \ell\big)\Big\},$$

for $i \neq i_x$. From this, we derive for any $x \notin I^\dagger$,

$$\begin{aligned}\sum_{i\in I}\tau_i(x) &\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{i\in I}\max_{\ell\in\{-1,1\}}\left\{\exp\left(-\frac{1}{2\sigma^2}(2(i_x - i) + \ell)^2\right)\right\}\\ &\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{1}{2\sigma^2}(2n - 1)^2\right),\end{aligned}$$

which implies (iii). For (iv), we observe that by $\mathrm{vol}(I^\dagger) = c$,

$$c = \mathrm{vol}(I^\dagger) = \mathrm{vol}(I^\dagger \cap S_t) + \mathrm{vol}(I^\dagger \backslash S_t) = \mathrm{vol}(S_t) - \mathrm{vol}(S_t \backslash I^\dagger) + \mathrm{vol}(I^\dagger \backslash S_t),$$

Since $\mathrm{vol}(S_t) \leqslant c$, (iv) follows. Item (v) is derived by straightforward computation: for both results $(v).1$ and $(v).2$ we apply the variable substitution $y = 2Kx/c - 2i + 1$ to obtain

$$\begin{aligned}\int_0^1 \tau_i(x)\mathrm{d}x &= \frac{c}{K}\int_0^1 b\left(\frac{2Kx}{c} - 2i + 1\right)\mathrm{d}x \\ &\leqslant \frac{c}{K}\int_{\mathbb{R}} b\left(\frac{2Kx}{c} - 2i + 1\right)\mathrm{d}x = \frac{c^2}{2K^2}\int_{\mathbb{R}} b(y)\mathrm{d}y = \frac{c^2}{2K^2}.\end{aligned}$$

For the equality in $(v).1$, we find that

$$\begin{aligned}\int_{B_i} \tau_i(x)\mathrm{d}x &= \frac{c}{K}\int_{B_i} b\left(\frac{2Kx}{c} - 2i + 1\right)\mathrm{d}x \\ &= \frac{c^2}{2K^2}\int_{[-1,1]} b(y)\mathrm{d}y = \frac{c^2}{2K^2}P.\end{aligned}$$

For the integral in $(v).2$, we derive

$$\begin{aligned}\int_{[0,1]} (\tau_i(x))^2\mathrm{d}x &= \frac{c^2}{K^2}\int_{B_i}\left(b\left(\frac{2Kx}{c} - 2i + 1\right)\right)^2\mathrm{d}x \\ &\leqslant \frac{c^2}{K^2}\int_{\mathbb{R}}\left(b\left(\frac{2Kx}{c} - 2i + 1\right)\right)^2\mathrm{d}x \\ &= \frac{c^3}{2K^3}\int_{\mathbb{R}} \left(b\left(y\right)\right)^2\mathrm{d}x = \frac{c^3}{4\sigma\sqrt{\pi}K^3}.\end{aligned}$$

Finally, for (vi) we point out that as a corollary of (iii), for $i \in I$, $\beta' \geqslant \beta$, and $x \in B_i$,

$$-\beta' \leqslant \varepsilon_{I\backslash\{i\}}(x;\beta') \leqslant 0,$$

since $x \notin (I\backslash\{i\})^\dagger$. Hence,

$$|\varepsilon_I(x;\beta')| = \left|\tau_i(x) + \varepsilon_{I\backslash\{i\}}(x;\beta')\right| \leqslant \tau_i(x) + \left|\varepsilon_{I\backslash\{i\}}(x;\beta')\right| \leqslant \tau_i(x) + \beta'. \qquad \square$$

Proof of Lemma A.5.

For any $\varrho \in [0, 1-\delta]$ and any $x \in [0,1]$ it holds that $w(x) \geqslant 1-\delta \geqslant \varrho$ and therefore $\mathrm{vol}(W_\varrho) = \mathrm{vol}(\{x \in [0,1] : w(x) \geqslant \varrho\}) = 1$. In particular this implies that $\mathrm{vol}(S_\varrho) = c$, for all $\varrho \in [0, 1-\delta]$, where S_ϱ is a maximizer of (2.6). Now, let $\varrho = s$. Since

$s \in [0, 1-\delta]$, it follows that

$$\mathcal{I}(S_\varrho, \varrho) = \int_{S_\varrho} v_0(x)\big(w(x) - \varrho\big)\mathrm{d}x = \frac{s}{c}\int_{S_\varrho} \frac{w(x)-\varrho}{w(x)-s}\mathrm{d}x = s\frac{\mathrm{vol}(S_\varrho)}{c} = \varrho,$$

and therefore $\varrho^* = s$ by Proposition 2.1. □

A.2.4 Proofs of the Results in Section 2.5.6

In this section, abbreviate the expectation value and probability $\mathbb{E}_\pi[\,\cdot\,]$ and $\mathbb{P}_\pi(\,\cdot\,)$ as $\mathbb{E}[\,\cdot\,]$ and $\mathbb{P}(\,\cdot\,)$, where we suppress the notation that these two notions depend on policy $\pi = \mathrm{KDEP}(M)$.

Proof of Proposition 2.5.

Define

$$\mathcal{I}(S, \varrho) = \int_S v(x)(w(x) - \varrho)\mathrm{d}x \qquad \text{and} \qquad \hat{\mathcal{I}}(S, \varrho) := \int_S \hat{v}(x)(w(x) - \varrho)\mathrm{d}x$$

for $S \in \mathcal{S}$ and $\varrho \in \mathbb{R}$. Note that these definitions allow for negative values of ϱ (as opposed to (2.6)). Next, denote the L_1-difference between v and $\hat{v}$ as $\delta := \|v - \tilde{v}\|_1$. For $\varrho \in \mathbb{R}$, let $\hat{S}_\varrho$ be the maximizer of $\hat{\mathcal{I}}(\cdot, \varrho)$ over $\mathcal{S}$, that is,

$$\hat{\mathcal{I}}(\hat{S}_\varrho, \varrho) = \max_{S \in \mathcal{S}} \hat{\mathcal{I}}(S, \varrho),$$

Then, let ϱ^* and $\hat{\varrho}$ solve the fixed-point equations

$$\varrho = \mathcal{I}(S_\varrho, \varrho) \qquad \text{and} \qquad \varrho = \hat{\mathcal{I}}(\hat{S}_\varrho, \varrho)$$

respectively. Note that $\hat{S}_{\hat{\varrho}}$ is an optimal assortment under $\hat{v}$ by Proposition 2.1. Hence, we may assume that $\hat{S} = \hat{S}_{\hat{\varrho}}$. Also, we have $0 \leqslant w(x) - \hat{\varrho} \leqslant 1$ for all $x \in \hat{S}$ and therefore,

$$\begin{aligned}\mathcal{I}(\hat{S}, \hat{\varrho}) - \hat{\mathcal{I}}(\hat{S}, \hat{\varrho}) &= \int_{\hat{S}} v(x)(w(x) - \hat{\varrho})\mathrm{d}x - \int_{\hat{S}} \hat{v}(x)(w(x) - \hat{\varrho})\mathrm{d}x \\ &\leqslant \int_{\hat{S}} |v(x) - \hat{v}(x)|\mathrm{d}x \leqslant \delta.\end{aligned}$$

Now, we find that

$$\mathcal{I}(\hat{S}, \hat{\varrho} - \delta) \geqslant \mathcal{I}(\hat{S}, \hat{\varrho}) \geqslant \hat{\mathcal{I}}(\hat{S}, \hat{\varrho}) - \delta = \hat{\varrho} - \delta. \tag{A.33}$$

Hence, there exists an $S \in \mathcal{S}$ such that $\mathcal{I}(S, \hat{\varrho} - \delta) \geqslant \hat{\varrho} - \delta$, which by (2.4) entails $\varrho^* \geqslant \hat{\varrho} - \delta$. Likewise, we derive $\hat{\varrho} \geqslant \varrho^* - \delta = r(S^*, v) - \delta$. By rewriting (A.33) we obtain

$$\int_{\hat{S}} v(x)w(x)\mathrm{d}x \geqslant (\hat{\varrho} - \delta)\left(1 + \int_{\hat{S}} v(x)\mathrm{d}x\right).$$

Hence, we may conclude

$$r(\hat{S}, v) = \frac{\int_{\hat{S}} v(x)w(x)\mathrm{d}x}{1 + \int_{\hat{S}} v(x)\mathrm{d}x} \geqslant \hat{\varrho} - \delta \geqslant r(S^*, v) - 2\delta. \qquad \square$$

Proof of Proposition 2.6.

We start by showing the rate of convergence of $\hat{\alpha}_i$ as in (2.27). Let $p = p_i$ and $\hat{p} = |E_i|/M$. Define $p_- := p - \frac{1-p}{M}$ and

$$g(x) := \frac{1-x}{x + 1/M}, \qquad x \in [0, 1].$$

Note that $\alpha_i = g(p_-)$ and $\hat{\alpha}_i = g(\hat{p})$. Let $\delta = \frac{3}{4}\min\{p, 1-p\}$. Claim: $g'(x)$ is bounded for x such that $|p_- - x| < \delta$. Since g is differentiable, convex and decreasing on $I_\delta := [p_- - \delta, p_- + \delta]$, the maximum value of g' is at attained at the left edge of I_δ;

$$|g'(x)| \leqslant |g'(p_- - \delta)| = \frac{1 + 1/M}{(p-\delta)^2} \leqslant \frac{32}{p^2}.$$

Here in the final inequality above, we used that $p - \delta \geqslant \frac{1}{4}p$. Next, note that $[p - \varepsilon, p + \varepsilon] \subset I_\delta$ if $M \geqslant \frac{1-p}{\delta - \varepsilon}$. This is guaranteed since

$$\frac{1-p}{\delta - \varepsilon} \leqslant \frac{1}{4p} \leqslant M_i \leqslant M.$$

As a result, if the event $\mathcal{E}$ applies, then $\hat{p} \in I_\delta$ and

$$|\alpha_i - \hat{\alpha}_i| = |g(p_-) - g(\hat{p})| \leqslant \frac{32}{p^2}|p_- - \hat{p}| \leqslant \frac{32}{p^2}\left(|p - \hat{p}| + \frac{1}{M}\right). \tag{A.34}$$

The expected value of $|p - \hat{p}|$ can be bounded from above by Hoeffding's inequality;

$$\begin{aligned}\mathbb{E}\big[\,|p - \hat{p}| \,\big|\, \mathcal{E}\,\big] \leqslant \mathbb{E}\big[\,|p - \hat{p}|\,\big] &= \int_0^\infty \mathbb{P}(|p - \hat{p}| \geqslant x)\mathrm{d}x \\ &\leqslant \int_0^\infty 2e^{-2Mx^2}\mathrm{d}x = \frac{\sqrt{\pi}}{\sqrt{2M}},\end{aligned}$$

which concludes (2.27) for

$$C_1 = 32(1 + c\overline{v})^2 \left(\frac{\sqrt{\pi}}{\sqrt{2}} + 1 \right)$$

since $p \geqslant 1/(1 + c\overline{v})$.

Regarding the convergence rate of $\hat{f}_i(\cdot)$ as in (2.28), we will first verify the claim that for some constant C_4 and $n \geqslant 4$ it holds that

$$\mathbb{E}\left[\int_{x \in S^i} \Big(f_i(x) - \hat{f}_i(x) \Big)^2 \mathrm{d}x \;\Big|\; |A_i| = n \right] \leqslant C_4 \frac{\log(n)}{n}. \tag{A.35}$$

We introduce the compact notation $\mathbb{E}_{i,n}[\,\cdot\,] = \mathbb{E}\big[\;\cdot\;\big|\;|A_i| = n\;\big]$. Note that the (conditional) mean integrated squared error can be written as two components as

$$\mathbb{E}_{i,n}\left[\int_{x \in S} \Big(f_i(x) - \hat{f}_i(x) \Big)^2 \mathrm{d}x \right] = \int_{x \in S} b^2(x) \mathrm{d}x + \int_{x \in S} \sigma^2(x) \mathrm{d}x, \tag{A.36}$$

where

$$b_i(x) := \mathbb{E}_{i,n}\left[\hat{f}_i(x) \right] - f_i(x) \qquad \text{and} \qquad \sigma_i^2(x) := \mathbb{E}_{i,n}\left[\Big(\hat{f}_i(x) - \mathbb{E}_{i,n} \hat{f}_i(x) \Big)^2 \right].$$

Also $x \in S^i$ we define the following recurring constant

$$C_x := \frac{8}{\pi} \cdot \begin{cases} \sqrt{\frac{h}{x-a}}, & \text{for } x \in (a, a+h), \\ 1, & \text{for } x \in [a+h, b-h], \\ \sqrt{\frac{h}{b-x}}, & \text{for } x \in (b-h, b), \end{cases}$$

Now, showing (A.35) relies on an auxiliary result, which is given in the lemma below.

Lemma A.6. *Let $h \in (0, \frac{c}{2}]$ and $\beta \geqslant \frac{1}{2}$. Let $K_x^i(\cdot)$ be a Legendre kernel for S^i of order $\ell = [\beta]$. Then, there exist uniform constants C_b and C_σ such that for all $x \in (a_i, b_i)$*

$$b_i^2(x) \leqslant C_b C_x \, h^{2\beta} \qquad \text{and} \qquad \sigma_i^2(x) \leqslant C_\sigma C_x \, \frac{\beta}{nh}.$$

Using the lemma above, note that C_x is integrable on (a, b) with respect to x:

$$\int_a^b C_x \mathrm{d}x = \frac{8}{\pi} \Big(2h + b - a - 2h + 2h \Big) \leqslant \frac{16c}{\pi}.$$

Therefore, we can use that the local bounds from Lemma A.6:

$$\begin{aligned}\mathbb{E}_{i,n}\left[\int_{x\in S^i}\Big(f_i(x)-\hat{f}_i(x)\Big)^2\,\mathrm{d}x\right] &= \int_{x\in S^i} b_i^2(x)\mathrm{d}x + \int_{x\in S^i}\sigma_i^2(x)\mathrm{d}x \\ &\leqslant \left(C_b h^{2\beta} + C_\sigma\frac{\beta}{nh}\right)\int_a^b C_x\mathrm{d}x \\ &\leqslant \frac{16c}{\pi}\left(C_b h^{2\beta}+C_\sigma\frac{\beta}{nh}\right) \\ &\leqslant \frac{16c}{\pi}\max\{C_b,C_\sigma\}\left(h^{2\beta}+\frac{\beta}{nh}\right), \end{aligned} \tag{A.37}$$

Now, note that $h^* \in \left[\frac{c}{2e},\frac{1}{e}\right]$ and since $n \geqslant 4$

$$\beta^* = \frac{1}{2}\log(-2n\log h^*) - \frac{1}{2} \geqslant \frac{1}{2}\log 8 - \frac{1}{2} = 0.54 > \frac{1}{2}.$$

Filling in $h = h^*$ and $\beta = \beta^*$ in the expression in (A.37) we find that

$$\begin{aligned}(h^*)^{2\beta^*} + \frac{\beta^*}{nh^*} &= \frac{1}{nh^*}\Big(\log n + \log\big(-2\log h^*\big) - 4\log h^* - 1\Big) \\ &\leqslant \frac{2e}{c}\,\frac{1}{n}\Big(\log n + C_5\Big) \\ &\leqslant \frac{2e}{c}\,(1+C_5)\,\frac{\log n}{n},\end{aligned}$$

where

$$C_5 = \log\left(-2\log\left(\frac{c}{2e}\right)\right) - 4\log\frac{c}{2e} - 1.$$

So, we have shown (A.35) with

$$C_4 = \frac{32e}{\pi}(1+C_5)\max\{C_b,C_\sigma\}.$$

Given (A.35), we note that the occurrence of the clean event implies the event

$$\mathcal{B}_i := \big\{(1-p_i-\varepsilon_i)M < |A_i| < (1-p_i+\varepsilon_i)M\big\}.$$

Let $n_0 := \lceil(1-p_i-\varepsilon)M\rceil$ and $n_1 := \lfloor(1-p_i+\varepsilon)M\rfloor$. Note that $n_0 \geqslant 4$ since

$$(1-p_i+\varepsilon)M \geqslant \frac{1}{2}(1-p_i)M \geqslant \frac{1}{2}(1-p_i)M_i \geqslant \frac{7}{2}.$$

As a consequence,

$$\mathbb{E}\left[\int_{x\in S^i}\Big(f_i(x)-\hat{f}_i(x)\Big)^2\,\mathrm{d}x \;\middle|\; \mathcal{B}_i\right]$$

$$
\begin{aligned}
&= \sum_{n=n_0}^{n_1} \mathbb{E}\left[\int_{x\in S^i} \left(f_i(x) - \hat{f}_i(x)\right)^2 \mathrm{d}x \;\middle|\; |A_i| = n \right] \cdot \mathbb{P}(|A_i| = n) \\
&\leqslant \max_{n=n_0,\dots,n_1} \mathbb{E}\left[\int_{x\in S^i} \left(f_i(x) - \hat{f}_i(x)\right)^2 \mathrm{d}x \;\middle|\; |A_i| = n \right] \\
&\leqslant C_4 \frac{\log n_0}{n_0} \leqslant \frac{2C_4}{1-p_i} \cdot \frac{\log M}{M}.
\end{aligned}
$$

Thus, we conclude that (2.28) holds since

$$
\begin{aligned}
\mathbb{E}\left[\int_{x\in S^i} \left(f_i(x) - \hat{f}_i(x)\right)^2 \mathrm{d}x \;\middle|\; \mathcal{E} \right] = \mathbb{E}\left[\mathbb{E}\left[\int_{x\in S^i} \left(f_i(x) - \hat{f}_i(x)\right)^2 \mathrm{d}x \;\middle|\; \mathcal{B}_i \right] \middle|\; \mathcal{E}\, s \right] \\
\leqslant C_2 \frac{\log M}{M},
\end{aligned}
$$

with

$$
C_2 := 2C_4 \frac{1 + c\overline{v}}{c\underline{v}}.
$$

For showing (2.29), note that, for all $x \in [0,1]$, we can write $v(x)$ as a weighted sum over the test assortments, i.e.,

$$
v(x) = \frac{1}{k(x)} \sum_{i=1}^{J} \alpha_i f_i(x),
$$

Then,

$$
\begin{aligned}
|v(x) - \hat{v}(x)| &\leqslant \frac{1}{k(x)} \sum_{i=1}^{J} \left| \alpha_i f_i(x) - \hat{\alpha}_i \hat{f}_i(x) \right| \\
&= \frac{1}{k(x)} \sum_{i=1}^{J} \left| \alpha_i f_i(x) - \hat{\alpha}_i f_i(x) + \hat{\alpha}_i f_i(x) - \hat{\alpha}_i \hat{f}_i(x) \right| \\
&\leqslant \sum_{i=1}^{J} |\alpha_i - \hat{\alpha}_i|\, f_i(x) + \sum_{i=1}^{J} \hat{\alpha}_i \left| f_i(x) - \hat{f}_i(x) \right|. \qquad \text{(A.38)}
\end{aligned}
$$

Note that by (A.34) we know that on $\mathcal{E}$

$$
\hat{\alpha}_i \leqslant \alpha_i + \frac{64}{p_i^2} \leqslant c\overline{v} + \frac{64}{(1 + c\underline{v})^2} =: C_6.
$$

Then, by integrating (A.38) and applying the Cauchy-Schwarz inequality, we find that on $\mathcal{E}$

$$
\int_0^1 |v(x) - \hat{v}(x)| \mathrm{d}x \leqslant \sum_{i=1}^{J} |\alpha_i - \hat{\alpha}_i| + C_6 \sum_{i=1}^{J} \int_{x\in S^i} \left| f_i(x) - \hat{f}_i(x) \right| \mathrm{d}x
$$

$$\leqslant \sum_{i=1}^{J} |\alpha_i - \hat{\alpha}_i| + \sqrt{c} C_6 \sum_{i=1}^{J} \left(\int_{x \in S^i} \left(f_i(x) - \hat{f}_i(x) \right)^2 \mathrm{d}x \right)^{1/2}. \tag{A.39}$$

Let $\mathbb{E}_{\mathrm{cl}}[\,\cdot\,]$ denote the conditional expectation given the clean event. By taking the expectation of (A.39), conditioned on the clean event, and applying Jensen's inequality for concave functions, we find the following upper bound:

$$\begin{aligned}
\mathbb{E}_{\mathrm{cl}}\Big[\, \|v - \hat{v}\|_1 \,\Big] &\leqslant \sum_{i=1}^{J} \mathbb{E}_{\mathrm{cl}}\Big[\, |\alpha_i - \hat{\alpha}_i| \,\Big] + \sqrt{c} C_6 \sum_{i=1}^{J} \left(\mathbb{E}_{\mathrm{cl}}\Big[\int_{x \in S^i} \left(f_i(x) - \hat{f}_i(x) \right)^2 \mathrm{d}x \,\Big] \right)^{1/2} \\
&\leqslant C_1 J \frac{1}{\sqrt{M}} + \sqrt{cC_2}\, C_6 J \frac{(\log M)^{1/2}}{M^{1/2}} \\
&\leqslant J \left(C_1 + \sqrt{cC_2}\, C_6 \right) \frac{(\log M)^{1/2}}{M^{1/2}}.
\end{aligned}$$

Since $J \leqslant 1 + 1/c$ we conclude the proof for

$$C_3 := \left(1 + \frac{1}{c} \right) \left(C_1 + \sqrt{cC_2}\, C_6 \right). \qquad \square$$

Proof of Theorem 2.5.

Let $M_0 := \max_i \{ \frac{1}{p_i}, \frac{7}{1-p_i} \}$. Note that $M = \lfloor T^{2/3}/J \rfloor$ for the number test assortments J, and that $JM \leqslant T^{2/3} \leqslant T$. For now, assume that $T \geqslant (J(M_0+1))^{3/2}$. Then, it holds that

$$M \geqslant \frac{T^{2/3}}{J} - 1 \geqslant \left(1 - \frac{1}{M_0 + 1} \right) \frac{T^{2/3}}{J} \geqslant M_0.$$

Now, note that by Hoeffding's inequality we know that for each $i \in \{1 \ldots, J\}$ that

$$\mathbb{P}(\mathcal{B}_i) \geqslant 1 - 2e^{-2\varepsilon_i^2 M}.$$

From this the probability that the clean event does not occur can be bounded from above by Boole's inequality:

$$\mathbb{P}(\mathcal{E}^c) = \mathbb{P}\left(\bigcup_{i=1}^{J} \mathcal{B}_i^c \right) \leqslant \sum_{i=1}^{J} \mathbb{P}\left(\mathcal{B}_i^c \right) \leqslant 2 \sum_{i=1}^{J} e^{-2\varepsilon_i^2 M} \leqslant \frac{1}{\sqrt{M}} \sum_{i=1}^{J} \frac{1}{\sqrt{2}\varepsilon_i}.$$

Next we note that

$$p_i \geqslant \frac{1}{1 + c\overline{v}} \qquad 1 - p_i \geqslant \frac{c\underline{v}}{1 + c\overline{v}}. \tag{A.40}$$

Therefore, $1/\varepsilon_i \leqslant C_7$, where

$$C_7 := 2(1 + c\overline{v}) \max\left\{1, \frac{1}{c\underline{v}}\right\}.$$

Hence,

$$\mathbb{P}(\mathcal{E}^c) \leqslant \sqrt{2}JC_7 \frac{1}{\sqrt{M}}. \tag{A.41}$$

Now, recall that $r(S) \leqslant 1$ for any $S \in \mathcal{S}$. We split $\mathbb{E}\big[r(S^*) - r(\hat{S})\big]$ into the contributions due to two complementary events;

$$\begin{aligned}\mathbb{E}\left[r(S^*) - r(\hat{S})\right] &= \mathbb{E}\left[r(S^*) - r(\hat{S})\middle|\, \mathcal{E}\right] \cdot \mathbb{P}(\mathcal{E}) + \mathbb{E}\left[r(S^*) - r(\hat{S})\middle|\, \mathcal{E}^c\right] \cdot \mathbb{P}(\mathcal{E}^c) \\ &\leqslant \mathbb{E}\left[r(S^*) - r(\hat{S})\middle|\, \mathcal{E}\right] + \mathbb{P}(\mathcal{E}^c).\end{aligned}$$

Now, we can apply (A.41) and Propositions 2.5 and 2.6. Let C_3 denote the constant as in Proposition 2.6. Then,

$$\begin{aligned}\mathbb{E}\left[r(S^*) - r(\hat{S})\right] &\leqslant \mathbb{E}\left[r(S^*) - r(\hat{S})\middle|\, \mathcal{E}\right] + \sqrt{2}JC_7 \frac{1}{\sqrt{M}} \\ &\leqslant 2\mathbb{E}\Big[\, \|v - \hat{v}\|_1 \,\Big|\, \mathcal{E}\Big] + \sqrt{2}JC_7 \frac{1}{M^{1/2}} \\ &\leqslant \Big(2C_3 + \sqrt{2}JC_7\Big) \frac{(\log M)^{1/2}}{M^{1/2}},\end{aligned}$$

Therefore, the exploitation regret can be bounded from above as

$$(T - JM)\mathbb{E}\left[r(S^*) - r(\hat{S})\right] \leqslant (2C_3 + \sqrt{2}JC_7) \frac{T(\log M)^{1/2}}{M^{1/2}}.$$

Also, the exploration regret can be bounded as

$$M \sum_{i=1}^{J} \mathbb{E}\Big[r(S^*) - r(S^i)\Big] \leqslant JM.$$

Hence, for $T \geqslant (J(M_0 + 1))^{3/2}$

$$\begin{aligned}\Delta_\pi(T, v) &\leqslant JM + (2C_3 + \sqrt{2}JC_7) \frac{T(\log M)^{1/2}}{M^{1/2}} \\ &\leqslant T^{2/3} + (2C_3 + \sqrt{2}JC_7) J^{1/2}\, T^{2/3} \left(\tfrac{2}{3}\log T - \log J\right)^{1/2} \\ &\leqslant \Big(1 + \tfrac{2}{3}(2C_3 + \sqrt{2}JC_7) J^{1/2}\Big)\, T^{2/3} (\log T)^{1/2}.\end{aligned}$$

On the other hand, if $2 \leqslant T < (J(M_0+1))^{3/2}$ we conclude that

$$\begin{aligned}\Delta_\pi(T,v) &\leqslant T \leqslant (J(M_0+1))^{3/2} \\ &\leqslant \frac{(J(M_0+1))^{3/2}}{2^{2/3}(\log 2)^{1/2}} T^{2/3}(\log T)^{1/2}.\end{aligned}$$

Finally, note that J is bounded from above by $1+1/c$ and by (A.40)

$$M_0 \leqslant (1+c\overline{v}) \max\left\{1, \frac{7}{c\underline{v}}\right\} =: C_8.$$

Therefore, we have shown the upper bound with

$$\overline{C} = \max\left\{1 + \tfrac{2}{3}\Big(2C_3 + \sqrt{2}(1+\tfrac{1}{c})C_7\Big)(1+\tfrac{1}{c})^{1/2}, \frac{\big((1+\frac{1}{c})(C_8+1)\big)^{3/2}}{2^{2/3}(\log 2)^{1/2}}\right\}. \qquad \square$$

Proof of Lemma A.6.

We start by showing that the first ℓ moments of $K_x^i(\cdot)$ disappear, that is,

$$\int_{u\in I_x^i} u^j K_x^i(u)\mathrm{d}u = \begin{cases} 1 & \text{for } j=0, \\ 0 & \text{for } j=1,\dots,\ell. \end{cases} \tag{A.42}$$

Note that $T_x^i(u) := \gamma_x^i u + \zeta_x^i$ maps I_x^i into $[-1,1]$. Since $\varphi_q\big(T_x^i(\cdot)\big)$ is a polynomial of degree q, there exist coefficients b_{qj} for $j \leqslant \ell$ and $q = 0,\dots,j$ such that, for all $u \in I_x^i$,

$$u^j = \sum_{q=0}^{j} b_{qj}\varphi_q\big(T_x^i(u)\big).$$

By setting $v = T_x^i(u)$, we obtain $\mathrm{d}v = \gamma_x^i \mathrm{d}u$ and, since $\zeta_x^i = T_x^i(0)$,

$$\begin{aligned}\int_{u\in I_x^i} u^j K_x^i(u)\mathrm{d}u &= \sum_{q=0}^{j}\sum_{k=0}^{\ell} \gamma_x^i \int_{u\in I_x^i} b_{qj}\varphi_q\big(T_x^i(u)\big)\varphi_k(\zeta_x^i)\varphi_k\big(T_x^i(u)\big)\mathrm{d}u \\ &= \sum_{q=0}^{j}\sum_{k=0}^{\ell} b_{qj}\varphi_k(\zeta_x^i)\int_{v\in[-1,1]} \varphi_q(v)\varphi_k(v)\mathrm{d}v \\ &=^{(*)} \sum_{q=0}^{j} b_{qj}\varphi_q\big(T_x^i(0)\big) = 0^j.\end{aligned}$$

In $(*)$ we use the fact that the Legendre polynomials form an orthonormal basis in $L_2([-1,1])$.

Next, we show two local upper bounds, which are used for bounding the bias and

variance component as in (A.36). These upper bounds are

$$\int_{u\in I_x^i} K_x^i(u)^2 \mathrm{d}u \leqslant C_x \ell, \tag{A.43}$$

and

$$\int_{u\in I_x^i} |u|^\beta |K_x^i(u)| \mathrm{d}u \leqslant \sqrt{C_x}. \tag{A.44}$$

By the orthonormality of the Legendre polynomials $\{\varphi_j\}_{j\geqslant 0}$ we obtain

$$\int_{u\in I_x^i} K_x^i(u)^2 \mathrm{d}u = \sum_{j=0}^{\ell} \gamma_x^i \big(\varphi_j(\zeta_x^i)\big)^2$$

for every $x \in [a_i, b_i]$. We can bound the $\gamma_x^i \big(\varphi_j(\zeta_x^i)\big)^2$ term as follows. By Theorem 7.3.3 from Szegö (1939) we know that for all $j \geqslant 1$ and $u \in (-1, 1)$

$$\big(1-u^2\big)^{1/4} \big|\varphi_j(u)\big| \leqslant \sqrt{\frac{2}{\pi}} \cdot \sqrt{\frac{j+1}{j}},$$

and, as a consequence,

$$\big(\varphi_j(u)\big)^2 \leqslant \frac{2}{\pi\sqrt{1-u^2}} \cdot \frac{j+1}{j} \leqslant \frac{4}{\pi\sqrt{1-u^2}}.$$

Therefore, for all $j \geqslant 0$ and $x \in (a, b)$,

$$\gamma_x^i \big(\varphi_j(\zeta_x^i)\big)^2 \leqslant \frac{1}{2} C_x.$$

Now, we obtain (A.43) by

$$\int_{u\in I_x^i} K_x^i(u)^2 \mathrm{d}u \leqslant \frac{\ell+1}{2} C_x \leqslant C_x \ell.$$

Now, (A.44) follows from (A.43) and the Cauchy-Schwarz inequality:

$$\begin{aligned}\left(\int_{u\in I_x^i} |u|^\beta |K_x^i(u)| \mathrm{d}u\right)^2 &\leqslant \int_{u\in I_x^i} |u|^{2\beta} \mathrm{d}u \int_{u\in I_x^i} K_x^i(u)^2 \mathrm{d}u \\ &\leqslant 2 \int_{u\in[0,1]} u^{2\beta} \mathrm{d}u \, C_x \ell = \frac{2\ell}{2\beta+1} C_x \leqslant C_x.\end{aligned}$$

The upper bound for bias component is now obtained as follows. Denote $J_x^i :=$

$\left[-\frac{x-a_i}{h}, \frac{b_i-x}{h}\right]$. Then, $J_x^i \supseteq I_x^i$ and by a change of variable $u = \frac{z-x}{h}$ we find

$$\begin{aligned} b_i(x) &= \frac{1}{h}\int_{z\in S^i} K_x^i\left(\frac{z-x}{h}\right) f_i(z)\mathrm{d}z - f_i(x) \\ &= \int_{u\in J_x^i} K_x^i(u)\, f_i(x+uh)\mathrm{d}u - \int_{u\in I_x^i} K_x^i(u) f_i(x)\mathrm{d}u \\ &= \int_{u\in I_x^i} K_x^i(u)\, f_i(x+uh)\mathrm{d}u - \int_{u\in I_x^i} K_x^i(u) f_i(x)\mathrm{d}u \\ &= \int_{u\in I_x^i} K_x^i(u)\Big(f_i(x+uh) - f_i(x)\Big)\mathrm{d}u. \end{aligned}$$

Now, we point out that

$$f_i(x+uh) = f_i(x) + uhf_i^{(1)}(x) + \ldots + \frac{(uh)^\ell}{\ell!} f_i^{(\ell)}(x+\tau uh)$$

for some $\tau \in [0,1]$. Since $K_x^i(\cdot)$ is of order ℓ, we obtain by (A.42)

$$\begin{aligned} b_i(x) &= \int_{u\in I_x^i} K_x^i(u)\frac{(uh)^\ell}{\ell!} f_i^{(\ell)}(x+\tau uh)\mathrm{d}u \\ &= \int_{u\in I_x^i} K_x^i(u)\frac{(uh)^\ell}{\ell!} f_i^{(\ell)}(x+\tau uh)\mathrm{d}u - \frac{h^\ell f_i^{(\ell)}(x)}{\ell!}\int_{u\in I_x^i} u^\ell K_x^i(u)\mathrm{d}u \\ &= \int_{u\in I_x^i} K_x^i(u)\frac{(uh)^\ell}{\ell!}\Big(f_i^{(\ell)}(x+\tau uh) - f_i^{(\ell)}(x)\Big)\mathrm{d}u. \end{aligned}$$

Now, we denote

$$C_9 := \sup_{v\in\mathcal{V}_1, \ell\in\mathbb{N}, y\in(0,1)} \left|\frac{v^{(\ell)}(y)}{(\ell+1)!}\right| \qquad \text{and} \qquad C_b := \frac{C_9^2}{(c\underline{v})^2},$$

where we note that the constant C_9 is the same constant as in Assumption 2.2. Moreover, note that $\alpha_i \geqslant c\underline{v}$ and thus the ℓ-th derivative of f_i is Lipschitz continuous with constant $\sqrt{C_b}\ell!$ by Assumption 2.2. Hence, by (A.44),

$$\begin{aligned} |b_i(x)| &\leqslant \int_{u\in I_x^i} |K_x^i(u)|\frac{|uh|^\ell}{\ell!}\Big|f_i^{(\ell)}(x+\tau uh) - f_i^\ell(x)\Big|\mathrm{d}u \\ &\leqslant \sqrt{C_b}\int_{u\in I_x^i} |K_x^i(u)||uh|^\ell|\tau uh|\mathrm{d}u \\ &\leqslant \sqrt{C_b}\int_{u\in I_x^i} |K_x^i(u)||uh|^\ell|\tau uh|^{\beta-\ell}\mathrm{d}u \leqslant \sqrt{C_bC_x}h^\beta. \end{aligned}$$

Squaring both sides of this final inequality, we obtain the result. Now, we consider the bound of the variance component. Let $Z_1, \ldots, Z_n$ be independent f_i-distributed

random variables. Then, for $k = 1, \ldots, n$, we define

$$\eta_k(x) = K_x^i\left(\frac{Z_k - x}{h}\right) - \mathbb{E}_{i,n}\left[K_x^i\left(\frac{Z_k - x}{h}\right)\right]$$

and find that these random variables are *iid* with mean zero. Now, we denote the constant

$$C_\sigma := \frac{2\overline{v}}{c\underline{v}}.$$

Next, we apply the same change of variables as before; $u = \frac{z-x}{h}$. Then, since $f_i(x) \leqslant C_\sigma$, we can bound the expected squared value of η_k, for $k = 1, \ldots, n$, locally in $x \in (a_i, b_i)$ by (A.43) as

$$\begin{aligned}\mathbb{E}_{i,n}\left[\eta_k(x)^2\right] = \mathbb{E}_{i,n}\left[K_x^i\left(\frac{Z_1 - x}{h}\right)^2\right] &= \int_{z\in S^i} K_x^i\left(\frac{z-x}{h}\right)^2 f_i(z)\mathrm{d}z \\ &\leqslant \frac{C_\sigma h}{2}\int_{u\in I_x^i} K_x^i(u)^2\mathrm{d}u \leqslant \frac{C_\sigma\, C_x}{2}\, h\,\ell.\end{aligned}$$

Hence,

$$\sigma_i^2(x) = \mathbb{E}_{i,n}\left[\left(\frac{1}{nh}\sum_{k=1}^n \eta_k(x)\right)^2\right] = \frac{1}{nh^2}\mathbb{E}_{i,n}\left[\eta_1(x)^2\right] \leqslant \frac{C_\sigma\, C_x}{2}\cdot\frac{\ell}{nh}, \leqslant C_\sigma\, C_x\,\frac{\beta}{nh},$$

since $\ell \leqslant \beta + \frac{1}{2} \leqslant 2\beta$. □

Appendix B

B.1 Mathematical Proofs for Section 3.3

B.1.1 Proofs of the Results in Section 3.3.2

Throughout this section we abbreviate the expectation $\mathbb{E}_v^\pi[\,\cdot\,]$ depending on $v \in \mathcal{V}$ and $\pi = \mathrm{SAP}(\alpha, \beta)$ as $\mathbb{E}[\,\cdot\,]$. In addition, recall that $S_\varrho = \{i \in [N] : w_i \geqslant \varrho\}$ with $\varrho \in [0,1]$.

Proof of Lemma 3.1

First, observe that

$$\begin{aligned}
&\mathbb{E}\big[(\varrho^* - \varrho_{t+1})^2 \,\big|\, \varrho_t\,\big] \\
&\quad= \mathbb{E}\big[\,(\varrho^* - \varrho_t - a_t(w_{Y_t} - \varrho_t))^2 \,\big|\, \varrho_t\,\big] \\
&\quad= \mathbb{E}\big[(\varrho^* - \varrho_t)^2 - 2a_t(\varrho^* - \varrho_t)(w_{Y_t} - \varrho_t) + a_t^2(w_{Y_t} - \varrho_t)^2 \,\big|\, \varrho_t\,\big] \\
&\quad\leqslant (\varrho^* - \varrho_t)^2 - 2a_t(\varrho^* - \varrho_t)(h(\varrho_t) - \varrho_t) + a_t^2. \qquad \text{(B.1)}
\end{aligned}$$

For bounding the cross term in (B.1), first suppose that $\varrho_t < \varrho^*$. Then, we define $\mathcal{I}(\varrho)$ as

$$\mathcal{I}(\varrho) = \sum_{i \in S_\varrho} v_i(w_i - \varrho), \qquad \varrho \in [0,1].$$

Note that $\varrho^* = \mathcal{I}(\varrho^*)$ since $\varrho^* = r(S_{\varrho^*}, v)$ and note that $S_{\varrho^*} \subseteq S_{\varrho_t}$. Therefore,

$$\begin{aligned}
\mathcal{I}(\varrho_t) &= \sum_{S_{\varrho^*}} v_i(w_i - \varrho_t) + \sum_{S_{\varrho_t} \setminus S_{\varrho^*}} v_i(w_i - \varrho_t) \\
&= \sum_{S_{\varrho^*}} v_i(w_i - \varrho^*) + (\varrho^* - \varrho_t) \sum_{S_{\varrho^*}} v_i + \sum_{S_{\varrho_t} \setminus S_{\varrho^*}} v_i(w_i - \varrho_t)
\end{aligned}$$

$$= \varrho^* + (\varrho^* - \varrho_t) \sum_{S_{\varrho^*}} v_i + \sum_{S_{\varrho_t} \setminus S_{\varrho^*}} v_i (w_i - \varrho_t)$$
$$\geqslant \varrho^* + (\varrho^* - \varrho_t) \sum_{S_{\varrho^*}} v_i.$$

Therefore,

$$h(\varrho_t) - \varrho_t = \frac{\sum_{i \in S_{\varrho_t}} v_i w_i}{1 + \sum_{i \in S_{\varrho_t}} v_i} - \varrho_t = \frac{\mathcal{I}(\varrho_t) - \varrho_t}{1 + \sum_{i \in S_{\varrho_t}} v_i}$$
$$\geqslant \frac{1 + \sum_{i \in S_{\varrho^*}} v_i}{1 + \sum_{i \in S_{\varrho_t}} v_i} (\varrho^* - \varrho_t) \geqslant \frac{1}{1 + \sum_{i=1}^N v_i} (\varrho^* - \varrho_t)$$
$$\geqslant p_0 (\varrho^* - \varrho_t).$$

Next, consider the case that $\varrho_t \geqslant \varrho^*$. Then, it holds that $h(\varrho_t) \leqslant h(\varrho^*) = \varrho^*$ and since $p_0 \leqslant 1$

$$h(\varrho_t) - \varrho_t \leqslant \varrho^* - \varrho_t \leqslant p_0 (\varrho^* - \varrho_t).$$

Either way, the cross term in (B.1) is bounded from above as

$$-2a_t (\varrho^* - \varrho_t)(h(\varrho_t) - \varrho_t) \leqslant -2p_0 a_t (\varrho^* - \varrho_t)^2.$$

As a result, we conclude that

$$\mathbb{E}\big[(\varrho^* - \varrho_{t+1})^2 \,\big|\, \varrho_t \big] \leqslant (\varrho^* - \varrho_t)^2 (1 - 2p_0 a_t) + a_t^2.$$

Taking the expectation on both sides yields (3.2).

For $t = 1, \ldots, T+1$, denote the expected squared difference of ϱ^* and ϱ_t as

$$\delta_t := \mathbb{E}\big[(\varrho^* - \varrho_t)^2 \big].$$

By induction, we show that the inequality (3.2) implies $\delta_t \leqslant c_2/(t + \beta)$, for all $t = 1, \ldots, T+1$ and some $c_2 > 0$. We note that the arguments given to show this result resemble the arguments in the proof of Lemma A.2. We include all arguments to show (3.3) because the constant c_2 is different from the constant κ as in (A.4) as well as for the sake of completeness. Specifically, let c_2 be the constant

$$c_2 = \max\{1 + \beta, \alpha^2\}.$$

For $t = 1$, we note that

$$\delta_1 \leqslant 1 \leqslant \frac{c_2}{1+\beta}.$$

Now, suppose $\delta_t \leqslant c_2/(t+\beta)$ for $t \leqslant t_0$ for some t_0. Then, for $t > t_0$, it follows that

$$\frac{t+\beta}{t+\beta+1} - 2\alpha p_0 < 1 - 2\alpha p_0 \leqslant -\alpha p_0 \leqslant -1,$$

since $\alpha \geqslant 1/p_0$ and therefore

$$c_2 \left(\frac{t+\beta}{t+\beta+1} - 2\alpha p_0 \right) + \alpha^2 < -c_2 + \alpha^2 \leqslant 0,$$

by definition of c_2. This implies that

$$c_2 \left((t+\beta) - 2\alpha p_0 - \frac{(t+\beta)^2}{t+\beta+1} \right) + \alpha^2 \leqslant 0,$$

and thus

$$\frac{c_2}{t+\beta} \left(1 - 2\frac{\alpha p_0}{t+\beta} \right) + \frac{\alpha^2}{(t+\beta)^2} \leqslant \frac{c_2}{(t+\beta+1)}.$$

This, by (3.2) in combination with the induction hypothesis, yields

$$\delta_{t+1} \leqslant \frac{c_2}{t+\beta+1},$$

so that we have proven the lemma. □

Proof of Theorem 3.1

First, define $h(\varrho) := r(S_\varrho, v)$ and

$$\varrho^* = \max_{S \subseteq [N]} r(S, v).$$

Recall that $\varrho^* = h(\varrho^*)$; see, e.g., Chen et al. (2021, Section 4). Next, note that

$$T = \sum_{t=1}^{T} \frac{1}{a_t} \big(1 - (1 - a_t)\big). \tag{B.2}$$

Also, note that, for $t \in [T]$, w_{Y_t} can be written in terms of a_t, ϱ_t and ϱ_{t+1} as

$$w_{Y_t} = \frac{1}{a_t} \big(\varrho_{t+1} - (1 - a_t)\varrho_t\big). \tag{B.3}$$

As a consequence of (B.2) and (B.3), it holds that

$$\begin{aligned}
T\varrho^* - \sum_{t=1}^{T} w_{Y_t} &= \sum_{t=1}^{T} \frac{1}{a_t}\big((\varrho^* - \varrho_{t+1}) - (1-a_t)(\varrho^* - \varrho_t)\big) \\
&= \sum_{t=1}^{T} \frac{1}{a_t}(\varrho^* - \varrho_{t+1}) - \sum_{t=1}^{T} \frac{1-a_t}{a_t}(\varrho^* - \varrho_t) \\
&= \frac{1}{a_T}(\varrho^* - \varrho_{T+1}) + \sum_{t=1}^{T-1} \frac{1}{a_t}(\varrho^* - \varrho_{t+1}) \\
&\qquad - \sum_{t=2}^{T} \frac{1-a_t}{a_t}(\varrho^* - \varrho_t) - \frac{1-a_1}{a_1}(\varrho^* - \varrho_1) \\
&= \frac{1}{a_T}(\varrho^* - \varrho_{T+1}) + \sum_{t=2}^{T} \frac{1}{a_{t+1}}(\varrho^* - \varrho_t) \\
&\qquad - \sum_{t=2}^{T} \frac{1-a_t}{a_t}(\varrho^* - \varrho_t) - \frac{1-a_1}{a_1}(\varrho^* - \varrho_1) \\
&= \frac{1}{a_T}(\varrho^* - \varrho_{T+1}) + \sum_{t=2}^{T} \left(\frac{1}{a_{t+1}} - \frac{1-a_t}{a_t}\right)(\varrho^* - \varrho_t) - \frac{1-a_1}{a_1}(\varrho^* - \varrho_1).
\end{aligned}$$

Given that $|\varrho^* - \varrho_1| \leqslant 1$, it follows that

$$T\varrho^* - \sum_{t=1}^{T} w_{Y_t} \leqslant \frac{T+\beta}{\alpha}(\varrho^* - \varrho_{T+1}) + \frac{\alpha+1}{\alpha}\sum_{t=2}^{T}(\varrho^* - \varrho_t) + \frac{\beta+1-\alpha}{\alpha}.$$

From Lemma 3.1 and Jensen's inequality for the concave function $x \mapsto \sqrt{x}$ it follows that

$$\mathbb{E}[\varrho^* - \varrho_t] \leqslant \mathbb{E}\big[\,|\varrho^* - \varrho_t|\,\big] \leqslant \big(\mathbb{E}\big[(\varrho^* - \varrho_t)^2\big]\big)^{1/2} \leqslant \frac{\sqrt{c_2}}{\sqrt{t+\beta}}.$$

From this, we conclude that the regret is bounded from above as

$$\begin{aligned}
\Delta_\pi(T) &\leqslant \frac{T+\beta}{\alpha}\mathbb{E}[\varrho^* - \varrho_{T+1}] + \frac{\alpha+1}{\alpha}\sum_{t=2}^{T}\mathbb{E}[\varrho^* - \varrho_t] + \frac{\beta+1-\alpha}{\alpha} \\
&\leqslant \sqrt{c_2}\frac{T+\beta}{\alpha\sqrt{T+\beta+1}} + \sqrt{c_2}\frac{\alpha+1}{\alpha}\sum_{t=2}^{T}\frac{1}{\sqrt{t+\beta}} + \frac{\beta+1-\alpha}{\alpha} \\
&\leqslant \sqrt{c_2}\frac{\sqrt{\beta+1}}{\alpha}\sqrt{T} + \sqrt{c_2}\frac{2\alpha+2}{\alpha}\sqrt{T} + \frac{\beta+1-\alpha}{\alpha} \\
&\leqslant \overline{C}\sqrt{T},
\end{aligned}$$

where

$$\overline{C} = \sqrt{c_2}\frac{\sqrt{\beta+1}+2\alpha+2}{\alpha} + \frac{\beta+1-\alpha}{\alpha}. \qquad \square$$

B.1.2 Proofs of the Results in Section 3.3.3

In this section, we abbreviate the expectation value and probability $\mathbb{E}^{\pi}_{v^j}[\,\cdot\,]$ and $\mathbb{P}^{\pi}_{v^j}(\,\cdot\,)$ as $\mathbb{E}_j[\,\cdot\,]$ and $\mathbb{P}_j(\,\cdot\,)$ for $j = 0, 1$ with v^0 and v^1 as in (3.7).

Proof of Lemma 3.2

First, note that if $\ell \notin S$, then $\mathbb{P}_0(\,\cdot\,|\,S) = \mathbb{P}_1(\,\cdot\,|\,S)$. Hence, we may assume that $\ell \in S$. In addition, note that $\mathbb{P}_j(\,\cdot\,|\,S) = \mathbb{P}_j(\,\cdot\,|\,S \cap \{k,\ell\})$ for $j = 0, 1$. Therefore, it suffices to check $S = \{k,\ell\}$ and $S = \{\ell\}$. To this end, define

$$p_i = \mathbb{P}_0(Y = i\,|\,S) \qquad \text{and} \qquad q_i = \mathbb{P}_1(Y = i\,|\,S), \qquad \text{for } i = 0, k, \ell,$$

where Y denotes a random purchase from S. First, consider the case that $S = \{k,\ell\}$. Then, for $i = 0, k, \ell$ it holds that

$$q_i = \frac{v^1_i}{1 + u_k + (1+\varepsilon)u_\ell} > \frac{\min\{1, u_\ell\}}{4u_k},$$

since $v^1_\ell > u_\ell$ and $\varepsilon < 1$ and $u_\ell < u_k$ and $1 < u_k$. Next, note that

$$\big(1 + u_k + (1-\varepsilon)u_\ell\big)\big(1 + u_k + (1+\varepsilon)u_\ell\big) \geqslant 2(1+u_k)u_\ell,$$

and as a result

$$|p_0 - q_0| \leqslant \frac{2u_\ell\varepsilon}{2(1+u_k)u_\ell} = \frac{\varepsilon}{1+u_k} < \varepsilon,$$

and

$$|p_k - q_k| = u_k|p_0 - q_0| \leqslant \frac{u_k\varepsilon}{1+u_k} < \varepsilon,$$

and

$$|p_\ell - q_\ell| = \frac{2(1+u_k)u_l\varepsilon}{\big(1 + u_k + (1-\varepsilon)u_\ell\big)\big(1 + u_k + (1+\varepsilon)u_\ell\big)} \leqslant \varepsilon.$$

Consequently,

$$\mathrm{KL}\Big(\mathbb{P}_0(\,\cdot\,|\,S)\,\Big\|\,\mathbb{P}_1(\,\cdot\,|\,S)\Big) \leqslant \sum_{i=0,k,\ell}\frac{(p_i - q_i)^2}{q_i} \leqslant \frac{12u_k}{\min\{1,u_\ell\}}\,\varepsilon^2 = c_3\,\varepsilon,$$

where the first inequality is easily verified (see, e.g., Lemma 3 from Chen & Wang (2018)) and

$$c_3 := \frac{12u_k}{\min\{1, u_\ell\}}.$$

Now, consider the case that $S = \{\ell\}$. Similarly, we derive for $i = 0, \ell$ that

$$q_i = \frac{v_i^1}{1 + (1+\varepsilon)u_\ell} > \frac{\min\{1, u_\ell\}}{1 + 2u_\ell} > \frac{\min\{1, u_\ell\}}{3u_k},$$

since $v_\ell^1 > u_\ell$ and $\varepsilon < 1$ and $u_\ell < u_k$ and $1 < u_k$. Next, note that

$$\big(1 + (1-\varepsilon)u_\ell\big)\big(1 + (1+\varepsilon)u_\ell\big) \geqslant 2u_\ell$$

and as a result

$$|p_0 - q_0| \leqslant \frac{2u_\ell\varepsilon}{2u_\ell} = \varepsilon,$$

and

$$|p_\ell - q_\ell| = \frac{2u_l\varepsilon}{\big(1 + (1-\varepsilon)u_\ell\big)\big(1 + (1+\varepsilon)u_\ell\big)} \leqslant \varepsilon.$$

Consequently,

$$\mathrm{KL}\Big(\mathbb{P}_0(\,\cdot \,|\, S) \,\Big\|\, \mathbb{P}_1(\,\cdot \,|\, S)\Big) \leqslant \sum_{i=0,\ell} \frac{(p_i - q_i)^2}{q_i} \leqslant \frac{6u_k}{\min\{1, u_\ell\}}\,\varepsilon^2 \leqslant c_3\,\varepsilon^2,$$

where the first inequality is again easily verified (see, e.g., Lemma 3 from Chen & Wang (2018)). The final statement of the lemma follows by applying Pinsker's inequality and Le Cam's method as follows. First, consider the entire probability measures $\mathbb{P}_0$ and $\mathbb{P}_1$. By the chain rule of the KL divergence it follows that

$$\mathrm{KL}\big(\mathbb{P}_0 \,\big\|\, \mathbb{P}_1\big) \leqslant c_3\,\varepsilon^2\,T.$$

From Pinsker's inequality it follows that the total variation (TV) norm between $\mathbb{P}_0$ and $\mathbb{P}_1$ is bounded from above as

$$\big\|\mathbb{P}_0 - \mathbb{P}_1\big\|_{\mathrm{TV}} = \sup_A \big|\mathbb{P}_0(A) - \mathbb{P}_1(A)\big| \leqslant \sqrt{2\mathrm{KL}\big(\mathbb{P}_0 \,\big\|\, \mathbb{P}_1\big)} \leqslant \sqrt{2c_3}\,\varepsilon\,\sqrt{T}.$$

Next, Le Cam's method entails to consider $B := \{\psi = 0\}$. Then, it follows that

$$\begin{aligned}\max_{j=0,1} \mathbb{P}_j(\psi \neq j) &\geqslant \frac{1}{2}\Big(\mathbb{P}_0(\psi = 1) + \mathbb{P}_1(\psi = 0)\Big) \geqslant \frac{1}{2}\Big(1 - \big(\mathbb{P}_0(B) - \mathbb{P}_1(B)\big)\Big) \\ &\geqslant \frac{1}{2}\Big(1 - \big\|\mathbb{P}_0 - \mathbb{P}_1\big\|_{\mathrm{TV}}\Big) \geqslant \frac{1}{2}\left(1 - \sqrt{2c_3}\,\varepsilon\,\sqrt{T}\right),\end{aligned}$$

which concludes our proof. □

Proof of Lemma 3.3

First note that, for $j = 0, 1$, the expected profit under v^j is independent of the inclusion of products $i \notin \{k, \ell\}$. That is, for all $S \subseteq [N]$,

$$r(S, v^j) = r(S \cap \{k, \ell\}, v^j). \tag{B.4}$$

We start with bounding the regret under v^0 from below. By (B.4) we only need to consider $S_t = \{k\}, \{\ell\}, \varnothing$. Then, as $u_k w_k = u_\ell w_\ell$ and $u_k = u_\ell + 1$ and $\varepsilon \in (0, 1/2]$, it follows that

$$r\big(\{k,\ell\}, v^0\big) - r\big(\{k\}, v^0\big) = u_\ell w_\ell \frac{2(1-\varepsilon)}{\big(2 + (2-\varepsilon)u_\ell\big)(2+u_\ell)} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2},$$

as well as,

$$r\big(\{k,\ell\}, v^0\big) - r\big(\{\ell\}, v^0\big) = u_\ell w_\ell \frac{\varepsilon}{\big(2 + (2-\varepsilon)u_\ell\big)\big(1 + (1-\varepsilon)u_\ell\big)} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}\,\varepsilon,$$

and, in addition,

$$r\big(\{k,\ell\}, v^0\big) - r(\varnothing, v^0) = u_\ell w_\ell \frac{2-\varepsilon}{2 + (2-\varepsilon)u_\ell} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}.$$

In general, we know that

$$r\big(\{k,\ell\}, v^0\big) - r(S_t, v^0) \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}\Big(\varepsilon \mathbf{1}\{k \notin S_t, \ell \in S_t\} + 1 - \mathbf{1}\{k, \ell \in S_t\} - \mathbf{1}\{k \notin S_t, \ell \in S_t\}\Big),$$

and by taking the expectation with respect to v^0 and summing over $t \in [T]$ we have shown (3.8) for

$$c_4 = \frac{u_\ell w_\ell}{(2+3u_\ell)^2}.$$

Next, we bound the regret under v^1 from below. Again by (B.4) we only need to consider $S_t = \{k\}, \{k, \ell\}, \varnothing$. Then, as $u_k w_k = u_\ell w_\ell$ and $u_k = u_\ell + 1$ and $\varepsilon \in (0, 1/2]$, it follows that

$$r\big(\{\ell\}, v^1\big) - r\big(\{k\}, v^1\big) = u_\ell w_\ell \frac{1+2\varepsilon}{\big(1 + (1+\varepsilon)u_\ell\big)(2+u_\ell)} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2},$$

as well as,

$$r\big(\{\ell\}, v^1\big) - r\big(\{k,\ell\}, v^1\big) = u_\ell w_\ell \frac{\varepsilon}{\big(1+(1+\varepsilon)u_\ell\big)\big(2+(2+\varepsilon)u_\ell\big)} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}\,\varepsilon,$$

and, in addition,

$$r\big(\{\ell\}, v^1\big) - r(\varnothing, v^1) = u_\ell w_\ell \frac{1+\varepsilon}{1+(1+\varepsilon)u_\ell} \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}.$$

In general, we know that

$$r\big(\{\ell\}, v^1\big) - r(S_t, v^1) \geqslant \frac{u_\ell w_\ell}{(2+3u_\ell)^2}\Big(\varepsilon \mathbf{1}\{k,\ell \in S_t\} + 1 - \mathbf{1}\{k,\ell \in S_t\} - \mathbf{1}\{k \notin S_t, \ell \in S_t\}\Big),$$

and by taking the expectation with respect to v^1 and summing over $t \in [T]$ we have shown (3.8) with

$$c_4 = \frac{u_\ell w_\ell}{(2+3u_\ell)^2}$$

as well. □

Proof of Theorem 3.2

As discussed in Section 3.3.3, we have established the first two steps: constructing two preference vectors v_0 and v_1 and showing that – as a consequence of Lemma 3.2 – for any estimator ψ that has as input the observed purchases $Y_1, \ldots, Y_T$ and outputs either 0 or 1 it holds that

$$\max_{j=0,1} \mathbb{P}^\pi_{v^j}(\psi \neq j) \geqslant \frac{1}{4}. \tag{B.5}$$

It remains to show that $v^0, v^1 \in \mathcal{V}$ and finish the third step by establishing a contradiction for a specific estimator ψ. First, note that for both $j = 0, 1$ we have that

$$\sum_{i=1}^N v_i^j \leqslant \frac{5}{2}u_k - \frac{3}{2}$$

as $u_\ell = u_k - 1$ and $\varepsilon \leqslant 1/2$. Consequently, v^0 and v^1 lie in the class $\mathcal{V}$ since

$$p_0 \leqslant \frac{2w_\ell - 2w_k}{5w_\ell + w_k} = \frac{2}{5u_k - 1}.$$

We continue our proof by establishing a contradiction. To this end, recall that $\underline{C} = c_4/(16\sqrt{2c_3})$ and recall the definitions of $\wp_0$, $\wp_1$, L_0 and L_1:

$$\wp_0 := \sum_{t=1}^{T} \mathbf{1}\{k, \ell \in S_t\} \qquad \text{and} \qquad \wp_1 := \sum_{t=1}^{T} \mathbf{1}\{k \notin S_t, \ell \in S_t\},$$
$$L_0 := c_4\Big(\varepsilon\wp_1 + T - \wp_0 - \wp_1\Big) \qquad \text{and} \qquad L_1 := c_4\Big(\varepsilon\wp_0 + T - \wp_0 - \wp_1\Big).$$

Now, we assume that

$$\Delta_\pi(T) < \underline{C}\sqrt{T}. \tag{B.6}$$

As a consequence of the assumption above, we conclude by Markov's inequality and Lemma 3.3 that

$$\mathbb{P}_0\left(L_0 > 4\underline{C}\sqrt{T}\right) \leqslant \frac{\mathbb{E}_0 L_0}{4\underline{C}\sqrt{T}} \leqslant \frac{\Delta_\pi(T, v^0)}{4\underline{C}\sqrt{T}} < \frac{1}{4},$$

and likewise

$$\mathbb{P}_1\left(L_1 > 4\underline{C}\sqrt{T}\right) \leqslant \frac{\mathbb{E}_1 L_1}{4\underline{C}\sqrt{T}} \leqslant \frac{\Delta_\pi(T, v^1)}{4\underline{C}\sqrt{T}} < \frac{1}{4}.$$

Next, define the estimator ψ as

$$\psi := \begin{cases} 0 & \text{if } \wp_0 > T/2, \\ 1 & \text{if } \wp_0 \leqslant T/2. \end{cases}$$

From $\psi = 1$ it follows that

$$\varepsilon\wp_1 + T - \wp_0 - \wp_1 > \varepsilon(T - \wp_0) \geqslant \frac{\varepsilon T}{2},$$

as $\wp_0 + \wp_1 \leqslant T$ and $\varepsilon < 1$. Now note that if $\varepsilon = 1/2$, then

$$\frac{\varepsilon T}{2} = \frac{T}{4} > \frac{T}{4\sqrt{2c_3}} \geqslant \frac{\sqrt{T}}{4\sqrt{2c_3}},$$

since $c_3 > 1$. Also, if $\varepsilon = (2\sqrt{2c_3 T})^{-1}$, then

$$\frac{\varepsilon T}{2} = \frac{\sqrt{T}}{4\sqrt{2c_3}}.$$

From this we conclude that $\psi = 1$ implies $L_0 > 4\overline{C}\sqrt{T}$ and therefore

$$\mathbb{P}_0(\psi = 1) \leqslant \mathbb{P}_0\left(L_0 > 4\underline{C}\sqrt{T}\right) < \frac{1}{4}. \tag{B.7}$$

Now consider $\psi = 0$. This implies

$$\varepsilon\wp_0 + T - \wp_0 - \wp_1 \geqslant \varepsilon\wp_0 > \frac{\varepsilon T}{2} \geqslant \frac{\sqrt{T}}{4\sqrt{2c_3}},$$

as $\wp_0 + \wp_1 \leqslant T$ and where the last inequality is established as before. From this we conclude that $\psi = 0$ implies $L_1 > 4\overline{C}\sqrt{T}$ and therefore

$$\mathbb{P}_1(\psi = 0) \leqslant \mathbb{P}_1\left(L_1 > 4\underline{C}\sqrt{T}\right) < \frac{1}{4}. \tag{B.8}$$

We conclude that (B.7) and (B.8) contradict (B.5). Hence, the assumption in (B.6) cannot be true. □

Bibliography

Agarwal A, Foster DP, Hsu DJ, Kakade SM & Rakhlin A (2011). Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems (NIPS)*, 1035–1043.

Agrawal R (1995). Sample mean based index policies with O(log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, **27**, 1054–1078.

Agrawal S, Avadhanula V, Goyal V & Zeevi A (2017). Thompson sampling for the MNL-bandit. *Conference on Learning Theory (COLT)*, 76–78.

Agrawal S, Avadhanula V, Goyal V & Zeevi A (2019). MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, **67**, 1453–1485.

Auer P, Cesa-Bianchi N, Freund Y & Schapire RE (2002). The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, **32**, 48–77.

Auer P, Ortner R & Szepesvári C (2007). Improved rates for the stochastic continuum-armed bandit problem. *Conference on Learning Theory (COLT)*, 454–468.

Ben-Akiva M & Lerman SR (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, USA.

Berbeglia G, Garassino A & Vulcano G (2021). A comparative empirical study of discrete choice models in retail operations. *SSRN e-print*.
URL `https://ssrn.com/abstract=3136816`

Den Boer AV, Chen B & Wang Y (2021). Pricing and positioning of horizontally differentiated products with incomplete demand information. *SSRN e-print*.
URL `https://ssrn.com/abstract=3682921`

Broadie M, Cicek D & Zeevi A (2011). General bounds and finite-time improvement for the Kiefer-Wolfowitz stochastic approximation algorithm. *Operations Research*, **59**, 1211–1224.

Bubeck S, Munos R, Stoltz G & Szepesvári C (2011a). X-armed bandits. *Journal of Machine Learning Research*, **12**, 1655–1695.

Bubeck S, Stoltz G, Szepesvári C & Munos R (2009). Online optimization in x-armed bandits. *Advances in Neural Information Processing Systems (NIPS)*, 201–208.

Bubeck S, Stoltz G & Yu JY (2011b). Lipschitz bandits without the Lipschitz constant. *International Conference on Algorithmic Learning Theory (ALT)*, 144–158.

Cesa-Bianchi N & Lugosi G (2012). Combinatorial bandits. *Journal of Computer and System Sciences*, **78**, 1404–1422.

Chen W, Wang Y & Yuan Y (2013). Combinatorial multi-armed bandit: General framework and applications. *International Conference on Machine Learning (ICML)*, 151–159.

Chen X & Wang Y (2018). A note on a tight lower bound for MNL-bandit assortment selection models. *Operations Research Letters*, **46**, 534–537.

Chen X, Wang Y & Zhou Y (2021). Optimal policy for dynamic assortment planning under multinomial logit models. *Mathematics of Operations Research. Published online in Articles in Advance 13 May 2021.*

Cheung WC & Simchi-Levi D (2017). Assortment optimization under unknown multinomial logit choice models. *ArXiv e-print.*
URL `https://arxiv.org/abs/1704.00108`

Combes R, Sadegh Talebi M, Proutiere A & Lelarge M (2015). Combinatorial bandits revisited. *Advances in Neural Information Processing Systems (NIPS)*, 2116–2124.

Cope EW (2009). Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Transactions on Automatic Control*, **54**, 1243–1253.

Devroye L & Györfi L (1985). *Nonparametric Density Estimation: The L1 View*. Wiley, New York, NY, USA.

Dewan R, Jing B & Seidmann A (2003). Product customization and price competition on the internet. *Management Science*, **49**, 1055–1070.

Farias VF, Jagabathula S & Shah D (2013). A nonparametric approach to modeling choice with limited data. *Management Science*, **59**, 305–322.

Feldman J, Zhang D, Liu X & Zhang N (2019). Customer choice models versus machine learning: Finding optimal product displays on Alibaba. *SSRN e-print*. URL `https://ssrn.com/abstract=3232059`

Fisher M & Vaidyanathan R (2014). A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, **60**, 2401–2415.

Flaxman AD, Kalai AT & McMahan HB (2005). Online convex optimization in the bandit setting: Gradient descent without a gradient. *Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 385–394.

Fogliatto FS, Da Silveira GJ & Borenstein D (2012). The mass customization decade: An updated review of the literature. *International Journal of Production Economics*, **138**, 14–25.

Gaur V & Honhon D (2006). Assortment planning and inventory decisions under a locational choice model. *Management Science*, **52**, 1528–1543.

Gill R & Levit B (1995). Applications of the van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli*, **1**, 59–79.

Guadagni PM & Little JD (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, **2**, 203–238.

IdeaWorksCompany (2019). CarTrawler worldwide estimate of ancillary revenue for 2019.
URL `https://www.cartrawler.com/ct/ancillary-revenue/worldwide-ancillary-revenue-2019`

Kallus N & Udell M (2020). Dynamic assortment personalization in high dimensions. *Operations Research*, **68**, 1020–1037.

Keskin NB & Birge JR (2019). Dynamic selling mechanisms for product differentiation and learning. *Operations Research*, **67**, 1069–1089.

Kleinberg R (2005). Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems (NIPS)*, 697–704.

Kleinberg R, Slivkins A & Upfal E (2008). Multi-armed bandits in metric spaces. *Fourtieth Annual ACM Symposium on Theory of Computing (STOC)*, 681–690.

Kök AG, Fisher ML & Vaidyanathan R (2015). Assortment planning: Review of literature and industry practice. N. Agrawal, & S. A. Smith (Eds.) *Retail Supply Chain Management*, 175–236. Springer Science & Business Media, New York, NY.

Kushner HJ & Yin GG (1997). *Stochastic Approximation and Recursive Algorithms and Applications*. Springer-Verlag, New York, NY, USA.

Lai TL & Robbins H (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, **6**, 4–22.

Levin DA, Peres Y & Wilmer EL (2017). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, USA.

Mahajan S & van Ryzin G (2001). Inventory competition under dynamic consumer choice. *Operations Research*, **49**, 646–657.

Moorthy KS (1984). Market segmentation, self-selection, and product line design. *Marketing Science*, **3**, 288–307.

Müller HG (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521–530.

Mussa M & Rosen S (1978). Monopoly and product quality. *Journal of Economic Theory*, **18**, 301–317.

Newman JP, Ferguson ME, Garrow LA & Jacobs TL (2014). Estimation of choice-based models using sales data from a single firm. *Manufacturing & Service Operations Management*, **16**, 184–197.

Ou M, Li N, Zhu S & Jin R (2018). Multinomial logit bandit with linear utility functions. *Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, 2602–2608.

Pan XA & Honhon D (2012). Assortment planning for vertically differentiated products. *Production and Operations Management*, **21**, 253–275.

Peeters Y & den Boer AV (2021a). Stochastic approximation for uncapacitated assortment optimization under the multinomial logit model. *Submitted*.

Peeters Y & den Boer AV (2021b). A regret lower bound for assortment optimization under the capacitated MNL model with arbitrary revenue parameters. *Probability in the Engineering and Informational Sciences, forthcoming*.

Peeters Y, den Boer AV & Mandjes M (2021). Continuous assortment optimization with logit choice probabilities and incomplete information. *Operations Research, forthcoming*.

Pine BJ (1993). *Mass Customization*. Harvard Business School Press, Boston, MA, USA.

Ratliff RM, Rao BV, Narayan CP & Yellepeddi K (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, **7**, 153–171.

Robbins H (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, **58**, 527–535.

Robbins H & Monro S (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, **22**, 400–407.

Rusmevichientong P, Shen ZJM & Shmoys DB (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, **58**, 1666–1680.

Sauré D & Zeevi A (2013). Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, **15**, 387–404.

Shamir O (2013). On the complexity of bandit and derivative-free stochastic convex optimization. *Conference on Learning Theory (COLT)*, 3–24.

Stroock DW (1994). *A Concise Introduction to the Theory of Integration*. Birkhäuser, Boston, MA, USA.

Szegö G (1939). *Orthonormal Polynomials*. American Mathematical Society, Providence, RI, USA.

Talluri K & van Ryzin G (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, **50**, 15–33.

Train KE (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.

Tsybakov AB (2008). *Introduction to Nonparametric Estimation*. Springer Science & Business Media, New York, NY, USA.

Wang Y, Chen B & Simchi-Levi D (2021). Multi-modal dynamic pricing. *Management Science*. *Published online in Articles in Advance 27 Jan 2021.*

Xu Y & Wang Z (2021). Assortment optimization for a multi-stage choice model. *SSRN e-print.*
URL `https://ssrn.com/abstract=3243742`

Zhang S, Karunamuni RJ & Jones MC (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, **94**, 1231–1241.

Summary

This thesis considers assortment optimization – where an assortment is a collection or subset of products offered to customers. The main question that we study is: how can a seller determine the optimal assortment of products – the subset which yields the highest expected profit – based on sales data. In particular, we consider dynamic assortment optimization over a finite time horizon in which we can adjust the offered assortment. To focus on the aspect of learning customers' preferences we consider a sequential decision framework. Then, the sequential decisions in a finite time window are based on past purchase behavior and are described by a policy.

We envision the total collection of products in two ways. In Chapter 2, we consider a continuous spectrum of products. Here, each product lies on the continuous spectrum and the seller selects an assortment to offer, which is a subset of that spectrum. Additionally, Chapter 3 regards the classical problem of discrete assortment optimization. In this framework, there are N distinct products and the seller selects an assortment to offer, which is a subset the N products.

For both continuous and discrete assortment optimization, we provide sequential decision policies and analyze their performance. In this analysis, the performance metric of interest is the *regret*, i.e., the accumulated expected loss due to offering suboptimal assortments. In general, however, it is not a straightforward task to directly determine the regret of a particular policy. As a result, when designing a policy, the performance is initially assessed by mathematically determining an upper bound on the regret. By providing such an upper bound in terms of the time horizon (and, if applicable, the finite number of products), we can rigorously evaluate the asymptotic performance of the policy that we provide. In addition, by showing a matching lower bound on the regret that any policy must endure, we are able to show

the asymptotic optimality of our proposed policies. Moreover, the classification of regret rate serves as an indication of how difficult the learning problem in different settings is. For example, given a finite time horizon T, a $\log T$ regret in one setting indicates that the learning problem is easier than the learning problem in another setting with a $T^{2/3}$ regret rate.

In Chapter 2 of this thesis, we consider dynamic assortment optimization over a continuous spectrum of products represented by the unit interval, where the seller's problem consists of determining the optimal subset of products to offer to potential customers. To describe the relation between assortment and customer choice, we propose a probabilistic choice model that forms the continuous counterpart of the widely studied discrete multinomial logit (MNL) model. We consider the seller's problem under incomplete information, propose a stochastic-approximation type of policy and show that its regret is only logarithmic in the time horizon. We complement this result by showing a matching lower bound on the regret of any policy, implying that our policy is asymptotically optimal. We then show that adding a capacity constraint significantly changes the structure of the problem. Here, the capacity constraint refers to the maximum size of the offered assortments. We construct a discretization policy and show that its regret after T time periods is bounded above by a constant times $T^{2/3}$ (up to a logarithmic term); in addition, we show that the regret of any policy is bounded from below by a positive constant times $T^{2/3}$, so that also in the capacitated case we obtain asymptotic optimality. Moreover, we provide a density estimation policy and we show that – under an additional assumption – its regret is also bounded above by a constant times $T^{2/3}$ (up to a logarithmic term).

Chapter 3 concerns dynamic assortment optimization under the discrete MNL choice model. For the setting without capacity constraint on the offered assortments, we propose a stochastic approximation policy – a discrete version of the stochastic approximation policy from Chapter 2 – and prove that the regret after T time periods is bounded by $\sqrt{T}$ times a constant that is independent of the number of products N. In addition, we prove a matching lower bound on the regret that is valid for arbitrary model parameters – slightly generalizing a similar recent regret lower bound derived for specific revenue parameters. Note that the regret rate of $\sqrt{T}$ differs from the logarithmical regret rate as discussed in Section 2.4. This is caused by the structural

difference between continuous and discrete assortment optimization. We continue Chapter 3 by considering the setting with capacity constraint $K < N/2$. In this setting we show that, for any vector of product revenues, there is a positive constant such that the regret of any policy is bounded from below by this constant times $\sqrt{NT}$. This result implies that policies that achieve a regret rate of $\sqrt{NT}$ here are asymptotically optimal for all product revenue parameters.

We present numerical experiments in Chapter 4, where we compare the performance of our policies with the performance of alternative policies. These numerical experiments show that our policies from Chapter 2 outperform or are on par with alternatives. Moreover, the experiments suggest that our policy from Chapter 3 outperforms alternatives by a significant margin when the number of products N is moderately large. In addition, in Chapter 4, we provide a numerical experiment to compare the predictive performance of the continuous logit model with that of the discrete MNL model. This experiment shows that our continuous assortment model has good predictive properties compared to its discrete counterpart, even if the true data-generating model is discrete.

Samenvatting

Deze dissertatie behandelt assortimentsoptimalisatie, waarbij een assortiment een collectie of deelverzameling van producten is die aan klanten worden aangeboden. De belangrijkste vraag die we onderzoeken is: hoe kan een verkoper het optimale assortiment van producten bepalen – de deelverzameling die de hoogste verwachte winst oplevert – op basis van verkoopgegevens. In het bijzonder beschouwen we dynamische assortimentsoptimalisatie over een eindige tijdshorizon waarin we het aangeboden assortiment kunnen aanpassen. Om ons voornamelijk te richten op het aspect van het leren van de voorkeuren van klanten, beschouwen we een sequentieel beslissingskader. Vervolgens zijn de sequentiële beslissingen gebaseerd op geobserveerd aankoopgedrag en worden beschreven door een algoritme.

De totale collectie van producten beschouwen we op twee verschillende manieren. In Hoofdstuk 2 beschouwen we een continu spectrum van producten. Hier ligt elk product op het continue spectrum en selecteert de verkoper een assortiment om aan te bieden. Het assortiment is op deze manier een deelverzameling van dat spectrum. Daarnaast behandelt Hoofdstuk 3 het klassieke probleem van discrete assortimentsoptimalisatie. In dit kader zijn er N verschillende producten en selecteert de verkoper een assortiment om aan te bieden, wat een deelverzameling is van de N producten.

Voor zowel continue als discrete assortimentsoptimalisatie presenteren we sequentiële algoritmes en analyseren we hun prestaties. De prestatiemaatstaf van belang in deze analyse is de *regret*, dit is het geaccumuleerde verwachte verlies als gevolg van het aanbieden van suboptimale assortimenten. In het algemeen is het echter geen eenvoudige taak om direct de *regret* van een bepaald algoritme vast te stellen. Als gevolg hiervan wordt bij het ontwerpen van een algoritme de prestatie in eerste instantie beoordeeld door wiskundig een bovengrens voor de *regret* te bepalen. Door

een zodanige bovengrens af te leiden in termen van de tijdshorizon (en, indien van toepassing, het eindig aantal producten), kunnen we de asymptotische prestaties van het algoritme dat we voorstellen, rigoureus evalueren. Daarnaast kunnen we, door een bijpassende ondergrens te tonen voor de *regret* die elk algoritme moet ondergaan, de asymptotische optimaliteit van onze voorgestelde algoritmes aantonen. Bovendien dient de classificatie van de *regret rate* (op welke manier de *regret* verandert ten opzichte van de tijdshorizon en eventueel het aantal producten) als een indicatie van hoe moeilijk het leerprobleem in verschillende situaties is. Bijvoorbeeld, gegeven een eindige tijdshorizon T, geeft een *regret* van $\log T$ in de ene setting aan dat het leerprobleem gemakkelijker is dan het leerprobleem in een andere setting met een *regret rate* van $T^{2/3}$.

In Hoofdstuk 2 van deze dissertatie beschouwen we dynamische assortimentsoptimalisatie over een continu spectrum van producten vertegenwoordigd door het eenheidsinterval. Hierbij bestaat het probleem van de verkoper uit het bepalen van de optimale deelverzamaling van producten die aan potentiële klanten kan worden aangeboden. Om de relatie tussen assortiment en klantkeuze te beschrijven, stellen we een stochastisch keuzemodel voor dat de continue tegenhanger vormt van het veel bestudeerde discrete multinomiale logit (MNL) model. We beschouwen het probleem van de verkoper onder onvolledige informatie, stellen een algoritme gebasseerd op stochastische benadering voor en laten zien dat zijn *regret* slechts logaritmisch is in de tijdshorizon. We vullen dit resultaat aan door een bijpassende ondergrens te tonen voor de *regret* van welk algoritme dan ook, wat impliceert dat ons algoritme asymptotisch optimaal is. Vervolgens laten we zien dat het toevoegen van een capaciteitsbeperking de structuur van het probleem aanzienlijk verandert. De capaciteitsbeperking verwijst hier naar de maximale grootte van de aangeboden assortimenten. We construeren een algoritme gebaseerd op discretisatie en laten zien dat zijn *regret* na T tijdsperioden van boven wordt begrensd door een constante maal $T^{2/3}$ (op een logaritmische term na); bovendien laten we zien dat de *regret* van elk algoritme van onderaf wordt begrensd door een positieve constante maal $T^{2/3}$, zodat we ook in het geval met capaciteitsbeperking asymptotische optimaliteit verkrijgen. Daarnaast presenteren we een algoritme voor het schatten van de dichtheid en laten we zien dat – onder een additionele aanname – zijn *regret* ook van boven wordt begrensd door een

constante maal $T^{2/3}$ (op een logaritmische term na).

Hoofdstuk 3 betreft dynamische assortimentsoptimalisatie onder het discrete MNL-keuzemodel. Voor de setting zonder capaciteitsbeperking op de aangeboden assortimenten, stellen we een algoritme gebaseerd op stochastische benadering voor – een discrete versie van het algoritme gebaseerd op stochastische benadering uit Hoofdstuk 2 – en bewijzen we dat de *regret* na T tijdsperioden wordt begrensd door $\sqrt{T}$ maal een constante die onafhankelijk is van het aantal producten N. Bovendien bewijzen we een overeenkomende ondergrens voor de *regret* die geldig is voor willekeurige modelparameters – een lichte generalisatie van een vergelijkbare recente *regret* ondergrens die is afgeleid voor specifieke winstparameters voor productopbrengsten. Opvallend is dat de *regret rate* van $\sqrt{T}$ verschilt met de logaritmische *regret rate* zoals besproken in Sectie 2.4. Dit wordt veroorzaakt door het structurele verschil tussen continue en discrete assortimentsoptimalisatie. We vervolgen Hoofdstuk 3 door het geval te beschouwen met capaciteitsbeperking $K < N/2$. In deze setting laten we zien dat er voor elke vector van winstparameters voor productopbrengsten een positieve constante is, zodat de *regret* van elk algoritme van onderaf wordt begrensd door deze constante maal $\sqrt{NT}$. Dit resultaat houdt in dat algoritmes die een *regret rate* van $\sqrt{NT}$ behalen, asymptotisch optimaal zijn voor alle winstparameters voor productopbrengsten.

We presenteren onze numerieke experimenten in Hoofdstuk 4, waarin we de prestaties van onze algoritmes vergelijken met de prestaties van alternatieve algoritmes. Deze numerieke experimenten laten zien dat onze algoritmes uit Hoofdstuk 2 beter of vergelijkbaar presteren ten opzichte van alternatieven. Bovendien suggereren de experimenten dat ons algoritme uit Hoofdstuk 3 met een aanzienlijke marge beter presteert dan alternatieven wanneer het aantal producten N tamelijk hoog is. Daarnaast presenteren we in Hoofdstuk 4 een numeriek experiment om de voorspellende prestaties van het continue keuzemodel te vergelijken met die van het discrete MNL-model. Dit experiment laat zien dat ons continue assortimentsmodel goede voorspellende eigenschappen heeft in vergelijking met zijn discrete tegenhanger, zelfs als het daadwerkelijke datagenererende model discreet is.

Acknowledgments

"*If we knew what we were doing, it would not be called research, would it?*"
- Albert Einstein

First, I would like to thank my supervisors Arnoud den Boer and Michel Mandjes for their patience, insights and guidance throughout this academic endeavor. Arnoud has been my steady co-author for all three papers collected in this thesis. Arnoud, thank you for your expertise on the related research and your guidance in academic writing. Michel was the one to recruit me for the PhD position at the Amsterdam Business School (ABS) and provided precious overarching feedback during all stages of research and writing. Michel, thank you for your challenging my point of view and substantially increasing the quality of the academic output. Additionally, I would like to thank my BSc thesis supervisor, Marcel de Jeu, for sparking my initial interest in academic research.

"*Never confuse education with intelligence, you can have a PhD and still be an idiot.*"
- Richard Feynman

Regarding the insight above, I would like to thank these idiots of PhDs (to-be): Leo, for being my buddy as well and inviting yourself to my MSc graduation. Robert, for being my mathematical sounding board and your delightful humor. Ujjwal, for all your comical answers to '*What would Yannik do?*' Rob, for your enthusiasm, funny stories and positive, yet skeptical, mindset.

"*They say that on your deathbed you never wish you spent more time at the office – but I will.*"
- Michael Scott

Before the COVID-19 pandemic, I never dreaded going to the office. During the pandemic, I enjoyed my sporadic visits and e-lunches. For all this, I would like to thank Alex, Bart, Chintan, Dick, Julien, Marit, Reza, Ronald, Stevan, the PhDs and everyone else at the ABS for the lunches, coffee breaks, plopmoments, drinks, outings and events.

"*I'll be there for you.*"
- The Rembrandts

To all my friends and family from Breda, Leiden, Amsterdam and The Hague, thank you for simply being there for me.

"*I can't carry it for you, but I can carry you!*"
- Samwise Gamgee

Lastly, I would like to thank my favorite person, Tara, for your support and being my companion for all these years. Thank you for being my inspiration to study hard, to remain focused and to take on new challenges. Moreover, thank you for all the fun and happy moments over the years and in all the years to come. Life would have been very dull without you.

About the Author

Yannik Peeters (1992) was born in Breda. He obtained his BSc degree in Mathematics from Leiden University and his MSc degree (*cum laude*) in Stochastics and Financial Mathemics from the University of Amsterdam (UvA). During his studies, he served on various committees within his student association, tutored students in secondary school and worked as a teaching assistant for the UvA.

In September 2017, he started as a PhD student at the Amsterdam Business School of the UvA in the department Operations Management (currently called Business Analytics). Here, he worked on his dissertation under the supervision of dr. Arnoud den Boer and prof. dr. Michel Mandjes. Moreover, he taught courses in statistics and business processes during this time.

Currently, Yannik works at Rabobank within the Know Your Customer (KYC) domain. This domain ensures that persons and institutions involved in fraud, money laundering, financing terrorism or violating sanctions are denied access to financial systems. In his work, he focuses on the performance and risk management of quantitative methods and models used for KYC purposes.