UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Continuous assortment optimization with logit choice probabilities under incomplete information

Peeters, Y.; den Boer, A.V.; Mandjes, M.

DOI
10.1287/opre.2021.2235

Publication date
2022

Document Version
Submitted manuscript

Published in
Operations Research

Link to publication

Citation for published version (APA):
Peeters, Y., den Boer, A. V., & Mandjes, M. (2022). Continuous assortment optimization with logit choice probabilities under incomplete information. *Operations Research*, *70*(3), 1613-1628. https://doi.org/10.1287/opre.2021.2235

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:11 Feb 2023

arXiv:1807.06497v4 [stat.ML] 14 Apr 2021

Continuous Assortment Optimization with Logit Choice Probabilities under Incomplete Information

Yannik Peeters, Arnoud V. den Boer, Michel Mandjes*

March 14, 2021

Abstract. We consider 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 model. We consider the seller's problem under incomplete information, propose a stochastic-approximation type of policy, and show that its regret – its performance loss compared to the optimal policy – 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: we construct a 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. Numerical illustrations show that our policies outperform or are on par with alternatives.

Keywords: assortment optimization; learning; multi-armed bandit; continuous assortment

*The authors are with Amsterdam Business School, University of Amsterdam, the Netherlands. A.V. den Boer and M. Mandjes are in addition affiliated with Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands.

1 Introduction

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 and Lerman 1985, Mahajan and Van Ryzin 2001, and the references therein). In this model, a nonnegative preference value is associated to each product (and also to 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 esti-

mate 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 paper 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 paper.

1.2 Contributions

The contributions in this paper 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 multinomial logit (MNL) model, in the sense that the continuous model arises as a limit of discrete MNL models where the number of products grow 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 to 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 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 a numerical study we compare our algorithms against alternatives from the literature that are designed for discrete assortment optimization, and show that our algorithms outperform or are on par with these alternatives. Additional numerical experiments included in the Appendix show that our continuous assortment model has good predictive properties compared to its discrete counterpart, even if the true data-generating model is discrete.

1.3 Organization of the paper

After providing an overview of relevant literature in Section 2, we introduce our model for continuous assortment optimization in Section 3. In Section 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 5: we propose a 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. Section 6 contains our numerical study. Mathematical proofs, a discussion of the relation between the continuous and discrete logit choice model, a bisection algorithm to compute the optimal continuous assortment, and additional numerical experiments are collected in the Appendix.

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 and 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 and 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 and 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. (2020) 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 and Honhon (2006) and Fisher and 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. (2020) and Keskin and Birge (2019), the literature mentioned above assumes that the model primitives are known to the seller.

The continuous choice model studied in the present paper aligns well with the widely studied discrete multinomial logit (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 and 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 and Wang (2018) to $\Omega(\sqrt{NT})$, under the assumption that $K \leqslant N/4$. Without capacity constraint, Chen et al. (2018) 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 and 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 and 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 consists 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.

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 $\emptyset$. The *total collection of products* $\mathcal{X}$ is the union of the unit interval and the no-purchase option:

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

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:

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

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

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

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

$$S_t = \pi_t(S_1, X_1, \ldots, S_{t-1}, X_{t-1}) \quad \text{for all } t = 1, \ldots, T; \tag{2}$$

here, we write $S_1 = \pi_1(\emptyset)$ 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

$$\Delta_\pi(T, v) := \sum_{t=1}^{T} \mathbb{E}_\pi \left[\max_{S \in \mathcal{S}} r(S, v) - r(S_t, v) \right], \tag{3}$$

where $S_1, \ldots, S_T$ satisfy (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 (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 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 4 for details. The assumption $\overline{v} > w(0)/\int_0^1 (w(x) - w(0))\mathrm{d}x$ is used in Section 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 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* multinomial logit (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 $\sum_{x \in A} v(x)/(1 + \sum_{x \in S} v(x))$, 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. The discrete multinomial logit choice model can be derived from an assumed underlying random utility model in which a customer assigns utility $u(x) = \log(v(x)) + \epsilon(x)$ to each product x and utility $\epsilon(0)$ to the no-purchase option; here $\{\epsilon(x)\}$ and $\epsilon(0)$ 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 $\sum_{x \in A} v(x)/(1 + \sum_{x \in S} v(x))$ (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 the current paper. 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 Appendix C for details). Furthermore, the policy that we propose in Section 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.

4 Uncapacitated continuous 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 Appendix A.

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 expensive products' for some integer k (cf. Talluri and 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):

$$\begin{aligned}
\max\{r(S, v) : S \in \mathcal{S}\} &= \max\{\varrho \in [0, 1] : \exists S \in \mathcal{S} : r(S, v) \geqslant \varrho\} \\
&= \max\left\{\varrho \in [0, 1] : \exists S \in \mathcal{S} : \int_S v(x)\big(w(x) - \varrho\big)\mathrm{d}x \geqslant \varrho\right\} \\
&= \max\left\{\varrho \in [0, 1] : \max_{S \in \mathcal{S}} \int_S v(x)\big(w(x) - \varrho\big)\mathrm{d}x \geqslant \varrho\right\}. \qquad (4)
\end{aligned}$$

The inner maximization problem in (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 $y \in [0, 1]$.

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 $\alpha > 0$ and $\beta \geqslant 0$.

`Stochastic Approximation Policy SAP`(α, β)

`1. Initialization. Let` $\alpha > 0$, $\beta \geqslant 0$ `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 and Monro 1951, Kushner and 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.

4.3 Regret upper bound

We proceed by showing that the worst-case regret of $\mathrm{SAP}(\alpha, \beta)$ grows at most logarithmically in T.

Theorem 1. *Let π correspond to $\mathrm{SAP}(\alpha, \beta)$ with $\alpha \geqslant \overline{v} + 1$ and $\beta \geqslant \max\{0, \alpha - 1\}$. 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 3. The logarithmic growth rate of the regret in Theorem 1 holds for all choices of $\alpha \geqslant \overline{v} + 1$ and $\beta \geqslant \max\{0, \alpha - 1\}$. 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 4. Theorem 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 1(ii) can be relaxed to v being continuously differentiable: this ensures inequality (24) in the proof of Lemma 4.

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. *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 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 $\emptyset$, 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 and 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.

5 Capacitated continuous 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 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 Appendix B.

5.1 Full information optimal solution

As shown in Section 4.1, the assortment optimization problem under full information can be written as

$$\max\{r(S,v) : S \in \mathcal{S}\} = \max\left\{\varrho \in [0,1] : \max_{S\in\mathcal{S}} \mathcal{I}(S,\varrho) \geqslant \varrho\right\}, \tag{5}$$

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 (5), i.e., how to construct an S_ϱ, for each $\varrho \in [0,1]$, such that

$$\mathcal{I}(S_\varrho,\varrho) = \max\{\mathcal{I}(S,\varrho) : S \in \mathcal{S}\}. \tag{6}$$

Next, we utilize this result to obtain an optimal solution for (5). To this end, let

$$h(x,\varrho) := v(x)\big(w(x)-\varrho\big), \quad x \in [0,1], \varrho \in [0,1], \tag{7}$$

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 (6).

Lemma 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 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 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 (6) given in Lemma 1, we now characterize an optimal solution to (5).

Proposition 1. *For each* $\varrho \in [0,1]$ *let* $S_\varrho \in \mathcal{S}$ *satisfy* (6). *Then there is a unique solution* $\varrho^* \in [0,1]$ *to the fixed-point equation*

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

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 the equality (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 5. The optimal assortment can be efficiently computed up to any desired accuracy via a bisection method. In Appendix D we present an implementation of such a bisection algorithm.

Remark 6. In contrast to the setting discussed in Section 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 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.

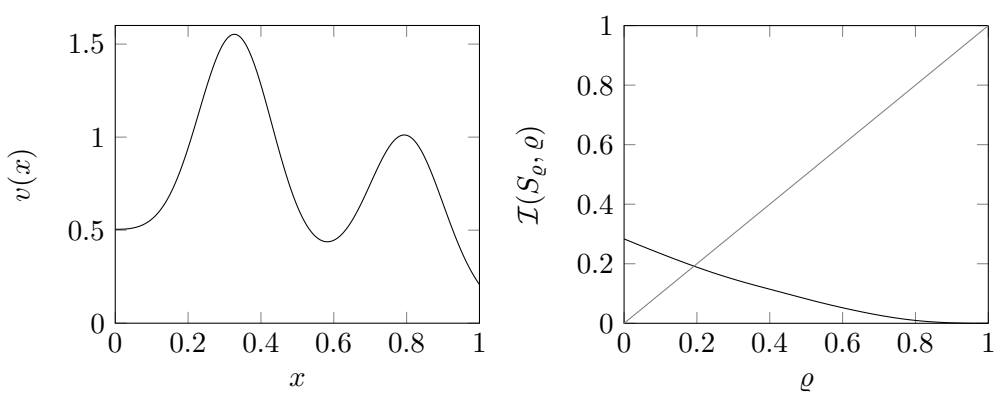

Figure 1: The left-hand panel shows the bi-modal preference function $v(x) = \frac{1}{10} + \frac{1}{5}(2 + x)(1 - x) + \frac{2}{7}\phi(x; 0.33, 0.1) + \frac{1}{5}\phi(x; 0.8, 0.1)$, $x \in [0, 1]$, where $\phi(\,\cdot\,; \mu, \sigma)$ is the normal probability density function with parameters μ and 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 $\mathcal{I}(S_\varrho, \varrho)$ is equal to ϱ.

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 (7) is a piece-wise constant function that attains the values $Nv_i(w_i - \varrho)$, for $i = 1, \ldots, N$. The ordering of the quantities $\{Nv_i(w_i - \varrho) : i = 1, \ldots, N\}$ does not change when ϱ is slightly changed, except possibly if ϱ is of the form

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

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 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.

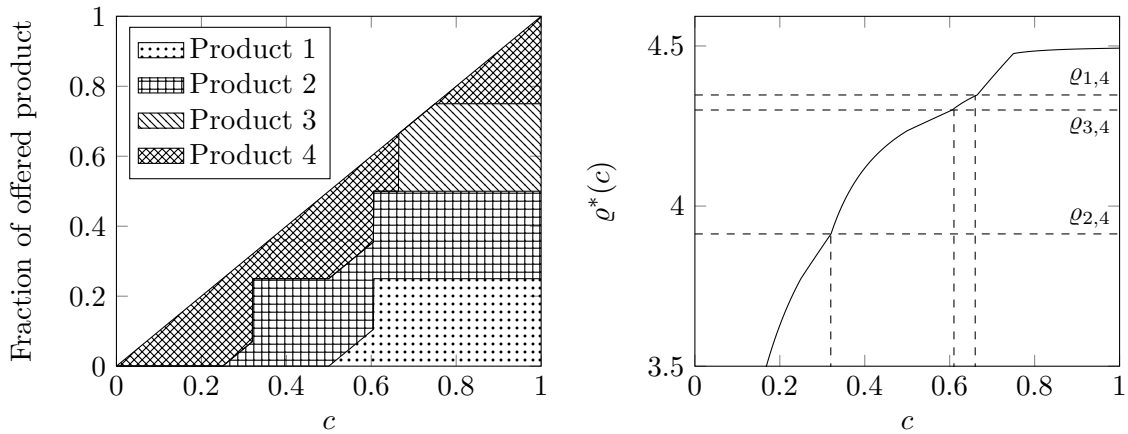

Figure 2: The left-hand panel shows the optimal amount of each products, as function of c. The right-hand panel shows the corresponding optimal expected profit $\varrho^*(c)$.

5.2 A policy for incomplete information

We proceed by presenting a 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). In what follows, we use $[n]$ as a compact notation for the set $\{1, \ldots, n\}$ (where $n \in \mathbb{N}$).

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 multinomial logit (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

$$B_i := \left[\frac{i-1}{N}, \frac{i}{N}\right) \tag{8}$$

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

$$B_N := \left[\frac{N-1}{N}, 1\right], \tag{9}$$

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 $X_t \in S_t \cup \{\emptyset\}$, and translate this X_t to a discrete purchase Y_t by

$$Y_t := \sum_{i=1}^{N} i \mathbf{1}_{B_i}(X_t).$$

Observe that $X_t \in B_{Y_t}$ if $X_t \in S_t$ and $Y_t = 0$ if $X_t = \emptyset$. The policy at time t computes upper confidence parameters $v_{1,t}^{\mathrm{UCB}}, \ldots, v_{N,t}^{\mathrm{UCB}}$ of the parameters $v_1, \ldots, v_N$ using observed discrete purchases $Y_1, \ldots, Y_t$. 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 $v_{1,t}^{\mathrm{UCB}}, \ldots, v_{N,t}^{\mathrm{UCB}}$ for $t = 1, \ldots, T$, the observed discrete purchases $Y_1, \ldots, Y_t$ 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., $X_t = \emptyset$ or, equivalently, $Y_t = 0$. Specifically, let $t_0 := 0$ and recursively define

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

and $t_\ell := T$ if $\{t \in \{t_{\ell-1}+1, \ldots, T\} : Y_t = 0\} = \emptyset$. 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

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

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 $s, t \in \{t_{\ell-1}, \ldots, t_\ell - 1\}$. 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 $v_{1,t}^{\mathrm{UCB}}, \ldots, v_{N,t}^{\mathrm{UCB}}$ become

$$v_{i,t}^{\mathrm{UCB}} := \begin{cases} \bar{v}_{i,\ell} + \sqrt{\bar{v}_{i,\ell} \dfrac{48 \log(\sqrt{N}\ell + 1)}{|\mathcal{T}_i(\ell)|}} + \dfrac{48 \log(\sqrt{N}\ell + 1)}{|\mathcal{T}_i(\ell)|} & \text{if } t = t_\ell \text{ for some } \ell \in [L] \text{ and } i \in D^\ell, \\ v_{i,t-1}^{\mathrm{UCB}} & \text{otherwise.} \end{cases} \tag{10}$$

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 (10) there exists an $\ell \in [L]$ such that $t = t_\ell$ if and only if $Y_t = 0$.

After the verbal description of our DUCB(N) policy, we now present the formal algorithm.

`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` (8) `and` (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`

$$D_t \in \operatorname*{arg\,max}_{D \subseteq [N] : |D| \leqslant K} \frac{\sum_{i \in D} v_{i,t-1}^{\mathrm{UCB}} w_i}{1 + \sum_{i \in D} v_{i,t-1}^{\mathrm{UCB}}}, \tag{11}$$

`and`

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

`Determine` $v_{1,t}^{\mathrm{UCB}}, \ldots, v_{N,t}^{\mathrm{UCB}}$ `as in` (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 (11) is not unique, ties are dealt with by applying an arbitrary fixed ordering of assortments.

5.3 Regret upper bound

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 3. *Let $T \geqslant 2$, $\gamma = \max\{\overline{v}, 1/c+1\}$, $N = \lfloor \gamma T^{1/3} \rfloor$, and let π correspond to* DUCB(N)*. There is a $\overline{C} > 0$, independent of T, such*

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

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

$$\check{v}(x) := N \sum_{i=1}^{N} \mathbf{1}_{B_i}(x) \int_{B_i} v(y)\,\mathrm{d}y, \qquad x \in [0,1], \tag{12}$$

$$\check{w}(x) := N \sum_{i=1}^{N} \mathbf{1}_{B_i}(x) \int_{B_i} w(y)\,\mathrm{d}y, \qquad x \in [0,1]. \tag{13}$$

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 below.

Proposition 2. *Let $\check{v}$ and $\check{w}$ be as in* (12) *and* (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,*

$$r(S^*, v, w) = \max_{S \in \mathcal{S}} r(S, v, w) \quad \textit{and} \quad r(\check{S}, \check{v}, \check{w}) = \max_{S \in \mathcal{S}} r(S, \check{v}, \check{w}). \tag{14}$$

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

$$r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) \leqslant \|v - \check{v}\|_1 + \overline{v}\,\|w - \check{w}\|_1, \tag{15}$$

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 below.

Lemma 2. *Let $\check{v}$ and $\check{w}$ be as in* (12) *and* (13)*, respectively. Let $\mathcal{A}_K$ be the set of all collections of at most $K = \lfloor cN \rfloor$ bins B_i, that is,*

$$\mathcal{A}_K := \left\{ \bigcup_{i \in D} B_i : D \subset [N] \ \textit{ and } \ |D| \leqslant K \right\}. \tag{16}$$

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,

$$r(\check{S}, \check{v}, \check{w}) = \max_{S \in \mathcal{S}} r(S, \check{v}, \check{v}) \quad \textit{and} \quad r(S^d, \check{v}, \check{w}) = \max_{S \in \mathcal{A}_K} r(S, \check{v}, \check{v}). \tag{17}$$

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 (16), we present the result below for a general set in $\mathcal{A}_K$.

Lemma 3. *Let $\check{v}$ and $\check{w}$ be as in* (12) *and* (13)*, respectively. Let $\mathcal{A}_K$ be as in* (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 (14) and (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

$$\begin{aligned} r(S^*, v, w) - r(S_t, v, w) = {} & r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) + && (18)\\ & r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) + && (19)\\ & r(S^d, \check{v}, \check{w}) - r(S_t, \check{v}, \check{w}) + && (20)\\ & r(S_t, \check{v}, \check{w}) - r(S_t, v, w); && (21) \end{aligned}$$

the idea is to apply the triangle inequality. For the right-hand side of (18), (19), and (21), we apply Proposition 2, Lemma 2 and Lemma 3, respectively. Note that the term in (20) corresponds to the instantaneous regret of UCB. The remainder of the proof of Theorem 3 consists of showing that both the L_1-distances $\|v - \check{v}\|_1$ and $\|w - \check{w}\|_1$ are of the order $1/N$ and applying Theorem 1 from Agrawal et al. (2019).

Remark 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-distance 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 are of order $N^{-(d+1)}$, similar as in (34). 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).

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 4.1 and 5.1) translate into a different complexity of the corresponding data-driven optimization problem, characterized by the growth rate of regret.

We consider the following instance. Let $\underline{v} \in (0, 0.79)$, $\overline{v} \geqslant 4$, let $c \in (0, \frac{1}{4}]$, $s = 0.8c$, $\delta = \frac{1}{2}$, and consider the marginal profit function

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

To obtain a lower bound on 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

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

This preference function has the property that $\varrho_0^* := \max_{S \in \mathcal{S}} r(S, v_0)$ is equal to s (see Appendix B, Lemma 8), 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 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 $\sigma = 0.3$:

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

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

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

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

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

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

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

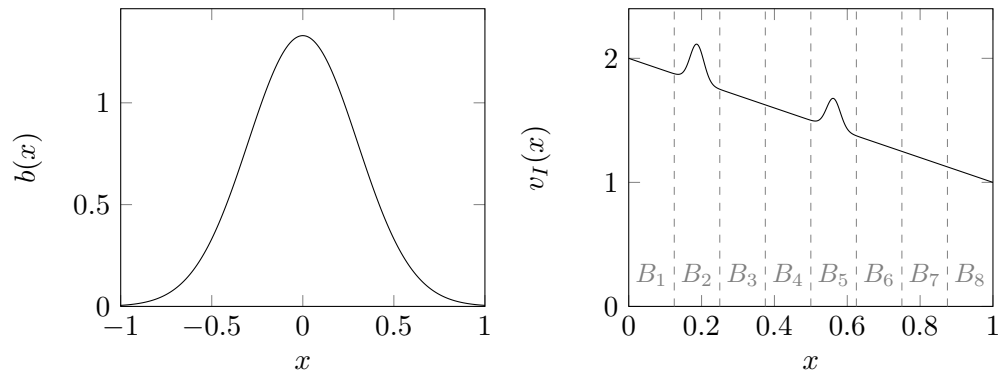

Figure 3: Left: bump function $b(\cdot)$. Right: preference function $v_I(\cdot)$ for $c = 0.25$, $K = 2$ and $I = \{2, 5\}$.

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.

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 to 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

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

and let

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

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

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

The proposition is proven by exploiting the structure of the optimal assortment as outlined in Section 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 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 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*

$$\Big|\mathbb{E}_I[k(x)] - \mathbb{E}_J[k(x)]\Big| \leqslant C_c \left(\frac{T}{K}\right)^{3/2}. \tag{22}$$

This bound is proven by relating the left-hand side of (22) to the Kullback-Leibler divergence of $\mathbb{P}_I$ and $\mathbb{P}_J$, using Pinsker's inequality, and subsequently bounding this expression from above by carefully analysing its dependence on v_I and v_J. The constant C_c is given explicitly in the proof of Proposition 4.

With Propositions 3 and 4 at hand, we finally arrive at our regret lower bound.

Theorem 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 1. This implies that the worst-case regret $\sup_{v\in\mathcal{V}} \Delta_\pi(T,v)$ 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 3. Proceeding in a similar fashion as in the proof of the regret lower bound obtained by Chen and Wang (2018) for discrete assortments, while dealing with all the intricacies of having a continuum product space, we connect the expression in Proposition 3 to the statement (22) of Proposition 4. By carefully selecting K, we arrive at the stated lower bound.

6 Numerical experiments

In this section we compare the numerical performance of the policies proposed in this study to alternative policies that are specifically designed for the discrete assortment problem. We use the notations and concepts introduced in Sections 5.2 and 5.3. 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. (2018), 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 algorithm DUCB to DTS but not to DTR, since the Trisection-based algorithm of Chen et al. (2018) 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 (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 paper. In this section we report numerical results on the regret behavior for these different algorithms; Appendix E contains additional numerical experiments on the predictive performance of our continuous model.

We set the preference function v as the bi-modal function that is plotted in Figure 1. This function is defined as

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

where $\phi(\,\cdot\,;\mu,\sigma)$ 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 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 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 1 and 2.

	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	0.49	0.500	0.507
N	19	25	28	31	34	36	38	39	41	43

Table 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
N	29	37	43	47	51	54	57	59	62	64

Table 2: Simulated average regret of the policies with $c = 0.5$ based on 100 simulations.

Table 1 shows that our algorithm SAP outperforms the alternatives DTR and DTS by a significant margin. The top row of Figure 4 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 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 2 records the regret of DTS, DUCB, and ADUCB; the results are visualized in the middle and bottom row of Figure 4. The figure illustrates that the regrets of both DUCB and ADUCB grow sublinearly. The adjusted policy ADUCB performs on par with DTS, while both ADUCB 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 and $\gamma_2 = 0.67$ for ADUCB. 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 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.

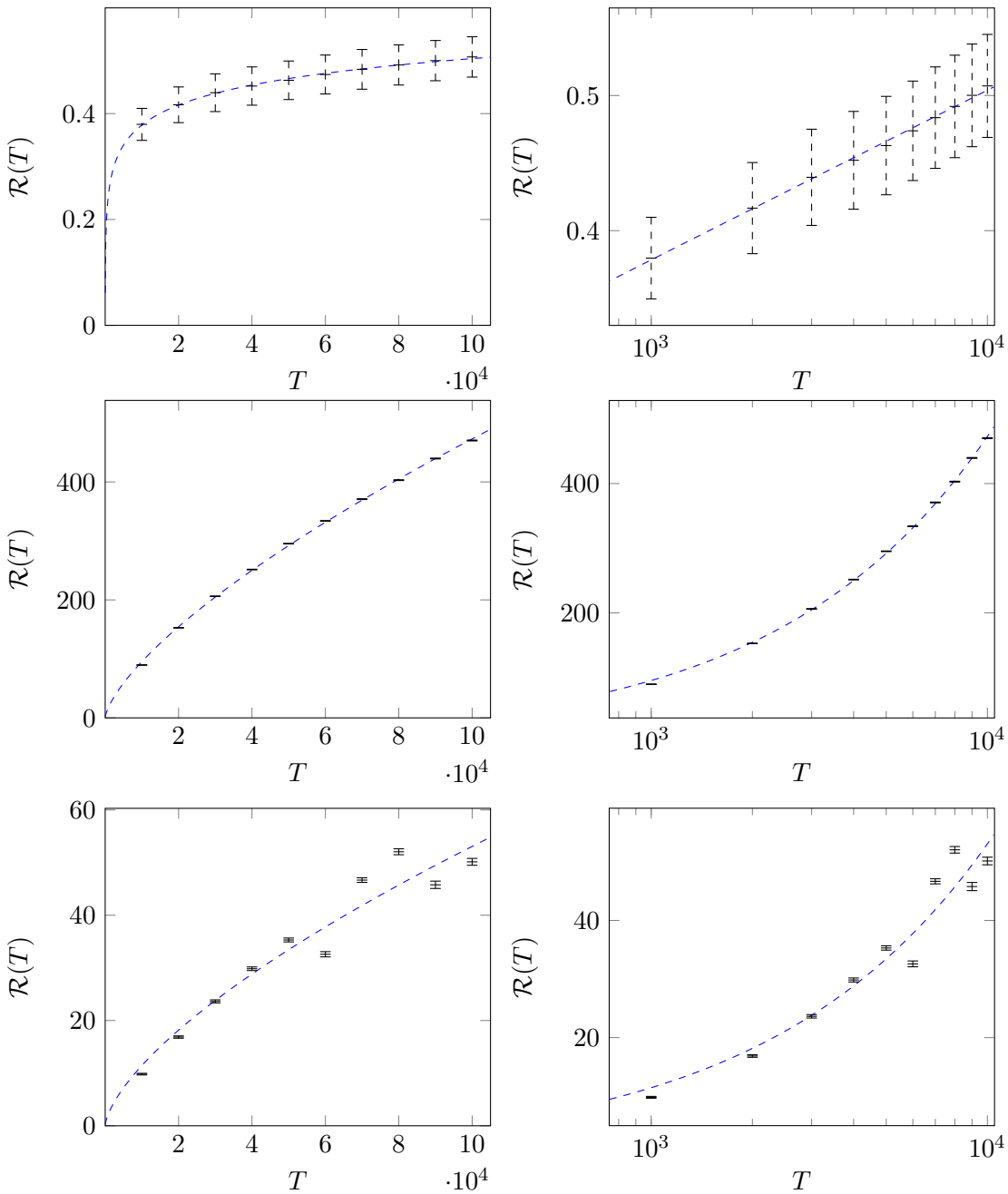

Figure 4: The whiskers show the 95% confidence interval of the mean cumulative regret for SAP (top row), DUCB (middle row) and ADUCB (bottom row) with regular axes (left panels) and a logarithmic axis for T (right panels), based on 100 simulations. The blue solid line shows the fitted curves $\gamma_1 + \gamma_2 \log T$ with $\gamma_1 = 0.00171$ and $\gamma_2 = 0.0545$ (first row), $\gamma_1 T^{\gamma_2}$ with $\gamma_1 = 0.783$ and $\gamma_2 = 0.700$ (second row) and $\gamma_1 T^{\gamma_2}$ with $\gamma_1 = 0.115$ and $\gamma_2 = 0.666$ (third row).

7 Discussion

In this paper we have introduced the concept of continuous assortment optimization with demand learning. We distinguish between the capacitated and uncapacitated case, 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 paper 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 paper 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 setup 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.

References

A. Agarwal, D. P. Foster, D. J. Hsu, S. M. Kakade, and A. Rakhlin. Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1035–1043, 2011.

R. Agrawal. Sample mean based index policies with O(log n) regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995.

S. Agrawal, V. Avadhanula, V. Goyal, and A. Zeevi. Thompson sampling for the MNL-bandit. In *Conference on Learning Theory (COLT)*, pages 76–78, 2017.

S. Agrawal, V. Avadhanula, V. Goyal, and A. Zeevi. MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485, 2019.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *Conference on Learning Theory (COLT)*, pages 454–468, 2007.

M. Ben-Akiva and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge, MA, 1985.

G. Berbeglia, A. Garassino, and G. Vulcano. A comparative empirical study of discrete choice models in retail operations. *SSRN eLibrary*, 2018. URL `https://ssrn.com/abstract=3136816`.

M. Broadie, D. Cicek, and A. Zeevi. General bounds and finite-time improvement for the Kiefer-Wolfowitz stochastic approximation algorithm. *Operations Research*, 59(5):1211–1224, 2011.

S. Bubeck, G. Stoltz, C. Szepesvári, and R. Munos. Online optimization in x-armed bandits. In *Advances in Neural Information Processing Systems (NIPS)*, pages 201–208, 2009.

S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011a.

S. Bubeck, G. Stoltz, and J. Y. Yu. Lipschitz bandits without the Lipschitz constant. In *International Conference on Algorithmic Learning Theory (ALT)*, pages 144–158, 2011b.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78 (5):1404–1422, 2012.

W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning (ICML)*, pages 151–159, 2013.

X. Chen and Y. Wang. A note on a tight lower bound for MNL-bandit assortment selection models. *Operations Research Letters*, 46(5):534–537, 2018.

X. Chen, Y. Wang, and Y. Zhou. An optimal policy for dynamic assortment planning under uncapacitated multinomial logit models. *ArXiv e-print*, 2018. URL `https://arxiv.org/abs/1805.04785`.

R. Combes, M. S. Talebi Mazraeh Shahi, A. Proutiere, and M. Lelarge. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2116–2124, 2015.

E. W. Cope. Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Transactions on Automatic Control*, 54(6):1243–1253, 2009.

A. V. den Boer, B. Chen, and Y. Wang. Pricing and positioning of horizontally differentiated products with incomplete demand information. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn.3682921. URL `https://doi.org/10.2139/ssrn.3682921`.

R. Dewan, B. Jing, and A. Seidmann. Product customization and price competition on the internet. *Management science*, 49(8):1055–1070, 2003.

M. Fisher and R. Vaidyanathan. A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, 60(10):2401–2415, 2014.

A. D. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 385–394, 2005.

F. S. Fogliatto, G. J. Da Silveira, and D. Borenstein. The mass customization decade: An updated review of the literature. *International Journal of Production Economics*, 138(1):14–25, 2012.

V. Gaur and D. Honhon. Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52(10):1528–1543, 2006.

R. Gill and B. Levit. Applications of the van Trees inequality: A Bayesian Cramér-Rao bound. *Bernoulli*, 1(1/2):59–79, 1995.

N. B. Keskin and J. R. Birge. Dynamic selling mechanisms for product differentiation and learning. *Operations Research*, 67(4):1069–1089, 2019.

R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems (NIPS)*, pages 697–704, 2005.

R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. *Fourtieth Annual ACM Symposium on Theory of Computing (STOC)*, pages 681–690, 2008.

H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer-Verlag, New York, NY, 1997.

T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

S. Mahajan and G. Van Ryzin. Inventory competition under dynamic consumer choice. *Operations Research*, 49(5):646–657, 2001.

K. S. Moorthy. Market segmentation, self-selection, and product line design. *Marketing Science*, 3(4): 288–307, 1984.

H.-G. Müller. Smooth optimum kernel estimators near endpoints. *Biometrika*, 78(3):521–530, 1991.

M. Mussa and S. Rosen. Monopoly and product quality. *Journal of Economic Theory*, 18(2):301–317, 1978.

M. Ou, N. Li, S. Zhu, and R. Jin. Multinomial logit bandit with linear utility functions. *Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2602–2608, 2018.

X. A. Pan and D. Honhon. Assortment planning for vertically differentiated products. *Production and Operations Management*, 21(2):253–275, 2012.

B. J. Pine. *Mass customization*. Harvard Business School Press, Boston, MA, 1993.

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, Sept. 1951.

P. Rusmevichientong, Z.-J. M. Shen, and D. B. Shmoys. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, 58(6):1666–1680, 2010.

D. Sauré and A. Zeevi. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, 15(3):387–404, 2013.

O. Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory (COLT)*, pages 3–24, 2013.

D. W. Stroock. *A Concise Introduction to the Theory of Integration*. Birkhäuser, Boston, MA, 1994.

K. Talluri and G. Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

K. E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, United Kingdom, 2009.

S. Zhang, R. J. Karunamuni, and M. C. Jones. An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, 94(448):1231–1241, 1999.

Appendix A: Mathematical proofs of Section 4

A.1. Proofs of the results in Section 4.3

Proof of Theorem 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 4 follows after the proof of Theorem 1.

Lemma 4. *It holds that $h(\varrho^*) = \varrho^*$. Moreover, for $\varrho \in [0,1]$, the following properties hold:*

$$(\varrho - \varrho^*)(h(\varrho) - \varrho) \leqslant -\frac{1}{1+\overline{v}}(\varrho - \varrho^*)^2, \tag{23}$$

$$h(\varrho^*) - h(\varrho) \leqslant C(\varrho - \varrho^*)^2, \tag{24}$$

for a universal constant $C > 0$.

Note that by our choice of $\beta \geqslant \max\{0, \alpha - 1\}$ it follows that $\varrho_t \in [0,1]$ for all $t \in \mathbb{N}$. 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 $t \in \mathbb{N}$ that

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

where the first inequality follows from $R_t - \varrho_t \in [-1,1]$ and the second inequality from Lemma 4, i.e., (23). 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 $t \in \mathbb{N}$,

$$\delta_{t+1} \leqslant \delta_t \left(1 - \frac{2}{1+\overline{v}} \cdot \frac{\alpha}{t+\beta}\right) + \frac{\alpha^2}{(t+\beta)^2}. \tag{25}$$

From the inequality in (25) one can derive the following lemma in a relatively straightforward way. Its (inductive) proof follows after the proof of Theorem 1.

Lemma 5. *There exists a $\kappa > 0$ such that for all $t \in \mathbb{N}$*

$$\delta_t \leqslant \frac{\kappa}{t+\beta}. \tag{26}$$

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 5. Let C denote the constant as in Lemma 4. 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 (24), the second inequality by (26), 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 4.

We prove the three claims separately.

▷ Following the reasoning at (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 $h(\varrho)=\varrho$.

$\triangleright$ For $\varrho=\varrho^*$, (23) 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

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

$\triangleright$ Firstly, note that

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

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

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

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

$$\sup_{y\in(0,1)}\big\{-g''(y)\big\}\leqslant C_0. \tag{27}$$

To prove (27) 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. Firstly, 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 (27) 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 (24) holds for any constant $C \geqslant 0$. Now, let $\varrho \in [w(0), 1]$. Then note that by (27)

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

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 (24) for all $\varrho \in [0,1]$ with

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

□

Proof of Lemma 5.

We show, by induction, that the inequality (25) implies that, for some $\kappa > 0$, for all $t \in \mathbb{N}$ 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 (25) in combination with the induction hypothesis, yields $\delta_{t+1}\leqslant\kappa/(t+1+\beta)$, so that we have proven the lemma. □

A.2. Proofs of the results in Section 4.4

Proof of Theorem 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 $\overline{v} > w(0)/\int_0^1 (w(x)-w(0))\mathrm{d}x$.

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 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 $S \in \mathcal{S}$,

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

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

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

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

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

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

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

In addition,

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

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

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

$\triangleright$ Step 2. For $S \in \mathcal{S}$ and $\theta \in \Theta$, let Z^S_θ be the random variable with support $[0, 2]$ and probability

density function

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

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 $\emptyset$ if $Z_\theta^S \in [0,2]\backslash S$. Hence, for each $t \in \{1, \ldots, T\}$ there is a function $\pi_t : [0,2]^{t-1} \to \mathcal{S}$ such that $S_t = \pi_t(Z_1, \ldots, Z_t)$ a.s., where $Z_t \stackrel{\mathrm{d}}{=} Z_\theta^{S_t}$ for all $t = 1, \ldots, T$, and where we write $\pi_1(\emptyset) := S_1$. In other words: to prove the regret lower bound we may assume that assortments are a function of the observations $Z_1, Z_2, \ldots$ instead of the purchase observations $X_1, X_2, \ldots$.

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

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

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 $\boldsymbol{z}_0 := \emptyset$. We have

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

and

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

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

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

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 and 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 $a \geqslant 0$. □

Appendix B: Mathematical proofs of Section 5

B.1. Proofs of the results in Section 5.1

Proof of Lemma 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 6. *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 6, $m_\varrho(\ell) \to 0$ 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 6 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 $S = \emptyset$, 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). $\square$

Proof of Lemma 6.

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

$$m_\varrho(\ell) - m_\varrho(\ell_n) = \mathrm{vol}\big(\{x \in [0,1] : h(x,\varrho) \in [\ell_n, \ell)\}\big) = \sum_{k=n}^{\infty} \mathrm{vol}\big(\{x \in [0,1] : h(x,\varrho) \in [\ell_n, \ell_{n+1})\}\big).$$

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. $\square$

Proof of Proposition 1.

Firstly, we show that there exists a unique solution to the fixed-point equation

$$g(\varrho) = \varrho, \tag{28}$$

where $g(\varrho) := \mathcal{I}(S_\varrho, \varrho)$ for $\varrho \in [0,1]$. As the right-hand side of (28) 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

$$\mathcal{I}(S_{\varrho_1}, \varrho_1) - \mathcal{I}(S_{\varrho_2}, \varrho_2) \leqslant \mathcal{I}(S_{\varrho_1}, \varrho_1) - \mathcal{I}(S_{\varrho_1}, \varrho_2) = (\varrho_2 - \varrho_1) \int_{S_{\varrho_1}} v(x)\mathrm{d}x \leqslant |\varrho_1 - \varrho_2| \int_{[0,1]} v(x)\mathrm{d}x,$$

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

$$g(1) = \max_{S\in\mathcal{S}} \int_S v(x)(w(x) - 1)\mathrm{d}x = 0.$$

Secondly, 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 (4) by invoking Lemma 1, we obtain

$$\max\left\{\varrho \in [0,1] : \max_{S\in\mathcal{S}} \mathcal{I}(S_\varrho, \varrho) \geqslant \varrho\right\} = \max\left\{\varrho \in [0,1] : g(\varrho) \geqslant \varrho\right\} = \varrho^* = r(S_{\varrho^*}).$$

□

B.2. Proofs of the results in Section 5.3

Proof of Proposition 2.

In addition to optimal assortments S^* and $\check{S}$ as in (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 (15) as follows:

$$r(S^*, v, w) - r(\check{S}, \check{v}, \check{w}) = r(S^*, v, w) - r(S^p, \check{v}, w) + \tag{29}$$

$$r(S^p, \check{v}, w) - r(\check{S}, \check{v}, \check{w}). \tag{30}$$

We start by bounding the right-hand side of (29) 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 (6)). Next, denote the L_1-distance between v and $\hat{v}$ as $\delta := \|v - \tilde{v}\|_1$. 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 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 (4) entails $\varrho^p \geqslant \varrho^* - \delta$. Thus, (29) is bounded from above as

$$r(S^*, v, w) - r(S^p, \check{v}, w) \leqslant \|v - \check{v}\|_1.$$

Bounding (30) 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.

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

$$\check{\mathcal{M}} := \operatorname*{arg\,max}_{S \in \mathcal{S}} \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \quad \text{and} \quad \mathcal{M}^d := \operatorname*{arg\,max}_{S \in \mathcal{A}_K} \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x, \qquad \varrho \in [0, 1].$$

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

$$\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \geqslant \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \geqslant 0. \tag{31}$$

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

$$\varrho^d \geqslant \int_S \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x. \tag{32}$$

Then, for any $S_1 \in \check{\mathcal{M}}$ and $S_2 \in \mathcal{M}^d$, it follows that

$$\begin{aligned}
r(\check{S}, \check{v}, \check{w}) - r(S^d, \check{v}, \check{w}) &= \frac{\int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} - \varrho^d \\
&= \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{\check{S}} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \varrho^d \right) \\
&\leqslant^{(*)} \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \varrho^d \right) \\
&\leqslant^{(**)} \frac{1}{1 + \int_{\check{S}} \check{v}(x)\check{w}(x)\mathrm{d}x} \left(\int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x \right) \\
&\leqslant^{(***)} \int_{S_1} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x - \int_{S_2} \check{v}(x)(\check{w}(x) - \varrho^d)\mathrm{d}x. \qquad (33)
\end{aligned}$$

Here at $(*)$ we use that $S_1 \in \check{\mathcal{M}}$, at $(**)$ we use (32) and $(***)$ holds because of (31).

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

$$h_i := \check{v}(y_i)\big(\check{w}(y_i) - \varrho^d\big), \quad i \in [N].$$

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 1, we first consider the case that $\mathrm{vol}(W_{\varrho^d}) \leqslant c$. Then we know by Lemma 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 1 and note that, as $\check{v}$ and $\check{w}$ are constant on each bin,

$$m_{\varrho^d}(\ell) = \frac{i}{N}, \quad \ell \in (h_{\sigma(i+1)}, h_{\sigma(i)}] \cap [0, \infty), \;\; i = 1, \ldots, N-1.$$

As a result, $c = K/N$ implies $\ell_{\varrho^d} = h_{\sigma(K)}$ and $R = \emptyset$, 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 1 that $S_2 := S_1 \cup R \in \check{\mathcal{M}}$. This concludes the claim for $\mathrm{vol}(W_{\varrho^d}) > c$.

From (33), the shown claim and the fact that $\check{w}(x) - \varrho^d \leqslant 1$, it follows that

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

□

Proof of Lemma 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 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,

$$\|v-\check{v}\|_1 = \sum_{i=1}^N \int_{B_i} |v(x)-b_i|\mathrm{d}x = \sum_{i=1}^N \int_{B_i} |v(x)-v(c_i)|\mathrm{d}x \leqslant L\sum_{i=1}^N \int_{B_i} |x-c_i|\mathrm{d}x \leqslant L\sum_{i=1}^N \frac{1}{2N^2} \leqslant \frac{L}{2N}, \tag{34}$$

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 (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 (18)–(21), in combination with the above and Proposition 2, Lemma 2 and Lemma 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

$$\overline{C} := \frac{C_1}{(\gamma - 1)(\log 2)^{1/2}} + C_5.$$

□

B.3. Proofs of the results in Section 5.4

Before stating the proofs of the results in Section 5.4, we recollect the notations and concepts introduced in that section. Let $c \in (0, \frac{1}{4}]$, $s = 0.8c$, $\delta = \frac{1}{2}$ and $\sigma = 0.3$. Let $K \geqslant 2$ be an integer, chosen at the end of the proof of Theorem 4. Furthermore, for all $x \in [0, 1]$, $i \in \{1, \ldots, N_K\}$, and $I \subseteq \{1, \ldots, N_K\}$, let

$$N_K = \lfloor K/c \rfloor, \qquad [N_K] = \{1, \ldots, N_K\},$$

$$\mathcal{D}_K = \{I \subseteq [N_K] : |I| = K\}, \qquad B_i = \left[c\frac{i-1}{K}, c\frac{i}{K}\right),$$

$$w(x) = (1 - s)\frac{1 - \delta}{1 - \delta x} + s, \qquad v_0(x) = \frac{s}{c(w(x) - s)},$$

$$b(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \qquad \phi_i(x) = \frac{2Kx}{c} - 2i + 1,$$

$$\tau_i(x) = \frac{c}{K} b\big(\phi_i(x)\big), \qquad \beta = \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}} e^{-(2n-1)^2/2\sigma^2},$$

$$\epsilon_I(x) = \sum_{i\in I}\tau_i(x) - \beta, \qquad v_I(x) = v_0(x)\big(1+\epsilon_I(x)\big).$$

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.

$$H := \frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}} e^{-2n^2/\sigma^2},$$

$$L := \frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}} e^{-(2n-1)^2/2\sigma^2} \qquad \text{and}$$

$$P := \mathbb{P}(-1/\sigma \leqslant Z \leqslant 1/\sigma),$$

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 7 contains a number of inequalities related to the quantities defined above, and Lemma 8 shows that the optimal expected profit under v_0 is precisely equal to s. The proof of these lemmas is given below.

Lemma 7. *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. $\displaystyle\int_0^1 (\tau_i(x))^2 \mathrm{d}x \leqslant \frac{c^3}{4\sigma\sqrt{\pi}K^3}$,

(vi) *for any* $i \in I$, $x \in B_i$ *and* $\beta' \geqslant \beta$, *it holds that* $|\epsilon_I(x;\beta')| \leqslant \tau_i(x) + \beta'$.

Lemma 8. *The optimal expected revenue under the preference function* $v_0(\cdot)$ *equals* s*:*

$$\max_{S\in\mathcal{S}} r(S, v_0) = s.$$

Proof of Proposition 3.

Let

$$C_1 := \frac{s(1-s)(1-\delta)}{c(1-s)(1-\delta)+cs} \qquad \text{and} \qquad C_2 := \frac{s^2(c+2L)}{\big((1-s)(1-\delta)+s\big)(1-s)(1-\delta)}.$$

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 (6) for $\varrho = \varrho^* = \max_{S\in\mathcal{S}} r(S, v)$. Therefore,

$$\int_{S^*} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x \geqslant \int_{I^\dagger} v(x)\big(w(x) - \varrho^*\big)\mathrm{d}x. \tag{35}$$

Observe in addition that

$$\int_S v_0(x)\mathrm{d}x \leqslant \frac{s}{(1-s)(1-\delta)}. \tag{36}$$

It now follows that, for all $S \in \mathcal{S}$,

$$\begin{aligned}
r(S^*, v) - r(S, v) &= \varrho^* - \frac{\int_S v(x)w(x)\mathrm{d}x}{1+\int_S v(x)\mathrm{d}x} \\
&= \frac{1}{1+\int_S v(x)\mathrm{d}x}\left(\varrho^* - \int_S v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x\right) \\
&=^{(*)} \frac{1}{1+\int_S v(x)\mathrm{d}x}\left(\int_{S^*} v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x\right) \\
&\geqslant^{(**)} \frac{(1-s)(1-\delta)}{(1-s)(1-\delta)+s}\left(\int_{S^*} v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x\right) \\
&\geqslant^{(***)} \frac{(1-s)(1-\delta)}{(1-s)(1-\delta)+s}\left(\int_{I^\dagger} v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x\right).
\end{aligned} \tag{37}$$

Here $(*)$ follows from $\varrho^* = \mathcal{I}(S^*, \varrho^*)$ by Proposition 1, $(**)$ follows by (36), and $(***)$ follows by (35). The terms within the large parentheses in (37) can be bounded from below as

$$\begin{aligned}\int_{I^\dagger} v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x - \int_S v(x)\big(w(x)-\varrho^*\big)\mathrm{d}x \\ &= \int_{I^\dagger} v(x)\big(w(x)-s\big)\mathrm{d}x - \int_S v(x)\big(w(x)-s\big)\mathrm{d}x \\ &\quad + (s-\varrho^*)\left(\int_{I^\dagger} v(x)\mathrm{d}x - \int_S v(x)\mathrm{d}x\right) \\ &\geqslant^{(*)} \frac{s}{c}\int_{I^\dagger\setminus S}\big(1+\epsilon_I(x)\big)\mathrm{d}x - \frac{s}{c}\int_{S\setminus I^\dagger}\big(1+\epsilon_I(x)\big)\mathrm{d}x \\ &\quad - |s-\varrho^*|\frac{2s}{(1-s)(1-\delta)},\end{aligned}$$

where at $(*)$ we use that by design $v(x)\big(w(x)-s\big) = \frac{s}{c}(1+\epsilon_I(x))$, together with inequality (36). The absolute difference between ϱ^* and s can be bounded from above by the L_1-difference between v and v_0, as follows. For $S \in \mathcal{S}$ and $\varrho \in [0,1]$, let

$$\mathcal{I}_0(S,\varrho) = \int_S v_0\big(w(x)-\varrho\big) \qquad \text{and} \qquad \mathcal{I}(S,\varrho) = \int_S v\big(w(x)-\varrho\big),$$

and let $S_0 := [0,c]$. As a consequence of Proposition 1 and Lemma 8, we obtain that $s = \mathcal{I}_0(S_0, s)$. 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 (4) this entails $s \geqslant \varrho^* - \|v-v_0\|_1$. Likewise, we derive $\varrho^* \geqslant s - \delta$ 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:

$$\begin{aligned}\int_0^1 \big|v(x)-v_0(x)\big|\mathrm{d}x &= \int_0^1 v_0(x)\big|\epsilon_I(x)\big|\mathrm{d}x \\ &\leqslant \frac{s}{c(1-s)(1-\delta)}\int_0^1 \big|\epsilon_I(x)\big|\mathrm{d}x\end{aligned}$$

$$\leqslant \frac{s}{c(1-s)(1-\delta)}\left(\sum_{i\in I}\int_0^1 \tau_i(x)\mathrm{d}x+\beta\right)$$
$$\leqslant^{(*)} \frac{s}{c(1-s)(1-\delta)}\left(\frac{c^2}{2}+Lc\right)\frac{1}{K}.$$

Here $(*)$ is justified by Lemma 7.(v).1. In addition, since $\epsilon_I(x)\leqslant 0$ for $x\notin I^\dagger$ by Lemma 7.(iii) and $\mathrm{vol}(S\backslash I^\dagger)\leqslant \mathrm{vol}(I^\dagger\backslash S)$ by Lemma 7.(iv), we conclude that

$$\begin{aligned}\int_{I^\dagger\backslash S}(1+\epsilon_I(x))\,\mathrm{d}x-&\int_{S\backslash I^\dagger}(1+\epsilon_I(x))\,\mathrm{d}x\\ &\geqslant \int_{I^\dagger\backslash S}(1+\epsilon_I(x))\,\mathrm{d}x-\mathrm{vol}(S\backslash I^\dagger)\\ &\geqslant \int_{I^\dagger\backslash S}(1+\epsilon_I(x))\,\mathrm{d}x-\mathrm{vol}(I^\dagger\backslash S)=\int_{I^\dagger\backslash S}\epsilon_I(x)\mathrm{d}x.\end{aligned}$$

Hence,

$$\begin{aligned}r(S^*,v)-r(S,v)\geqslant&\frac{s(1-s)(1-\delta)}{c(1-s)(1-\delta)+cs}\int_{I^\dagger\backslash S}\epsilon_I(x)\mathrm{d}x\\ &-\frac{s^2(c+2L)}{\big((1-s)(1-\delta)+s\big)(1-s)(1-\delta)}\frac{1}{K}.\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}_I\left[\sum_{t=1}^T\int_{I^\dagger\backslash S_t}\epsilon_I(x)\mathrm{d}x\right]&=\int_{I^\dagger}\mathbb{E}_I\left[\sum_{t=1}^T(1-1_{S_t}(x))\epsilon_I(x)\mathrm{d}x\right]\\ &=\int_{I^\dagger}(T-\mathbb{E}_I[k(x)])\epsilon_I(x)\mathrm{d}x.\end{aligned}$$

□

Proof of Proposition 4.

Let $x\in[0,1]$, $I\in\mathcal{D}_K$, $i\in I$, and $J=I\backslash\{i\}$. It suffices to show that there is a $C_c>0$ such that

$$\Big|\mathbb{E}_I[k(x)]-\mathbb{E}_J[k(x)]\Big|\leqslant T\sqrt{2\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J)},\tag{38}$$

and

$$\mathrm{KL}(\mathbb{P}_I\|\mathbb{P}_J)\leqslant \tfrac{1}{2}C_c^2\frac{T}{K^3}.\tag{39}$$

We first prove (38), 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,

$$\sup_{A \in \mathcal{F}} \Big| \mathbb{P}(A) - \mathbb{Q}(A) \Big| \leqslant \sqrt{\tfrac{1}{2} \mathrm{KL}(\mathbb{P} \| \mathbb{Q})}. \tag{40}$$

Consider the probability measures p and q on $\{0, \ldots, T\}$, defined by

$$p(n) := \mathbb{P}_I(k(x) = n) \qquad \text{and} \qquad q(n) := \mathbb{P}_J(k(x) = n), \quad (n \in \{0, \ldots, T\}).$$

From the equality

$$\sup_{n=0,\ldots,T} |p(n) - q(n)| = \frac{1}{2} \sum_{n=0}^{T} |p(n) - q(n)|. \tag{41}$$

we obtain

$$\begin{aligned} \Big| \mathbb{E}_I[k(x)] - \mathbb{E}_J[k(x)] \Big| &= \left| \sum_{n=0}^{T} n(p(n) - q(n)) \right| \\ &\leqslant \sum_{n=0}^{T} n \, |p(n) - q(n)| \leqslant T \sum_{n=0}^{T} |p(n) - q(n)| \\ &=^{(*)} 2T \sup_{n=0,\ldots,T} |p(n) - q(n)| \leqslant^{(**)} T \sqrt{2 \, \mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J)}, \end{aligned}$$

where $(*)$ follows by (41), and $(**)$ follows by (40). This proves (38).

We now prove (39). 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 7.(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

$$\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) = \mathbb{E}_I \sum_{t=1}^{T} \left(p_t \log \frac{p_t}{q_t} + \int_{S_t} \log \left(\frac{p_t v(x)}{q_t u(x)} \right) p_t v(x) \mathrm{d}x \right)$$

$$= \mathbb{E}_I \sum_{t=1}^{T} \left(p_t \log\left(1 + \frac{p_t - q_t}{q_t}\right) + \int_{S_t} \log\left(1 + \frac{p_t v(x) - q_t u(x)}{q_t u(x)}\right) p_t v(x) \mathrm{d}x \right).$$

Since $\log(1+x) \leqslant x$ for all $x > -1$, we find the following upper bound:

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) &= \mathbb{E}_I \sum_{t=1}^{T} \left(p_t \log\left(1 + \frac{p_t - q_t}{q_t}\right) + \int_{S_t} \log\left(1 + \frac{p_t v(x) - q_t u(x)}{q_t u(x)}\right) p_t v(x) \mathrm{d}x \right) \\
&\leqslant \mathbb{E}_I \sum_{t=1}^{T} \left(p_t \frac{p_t - q_t}{q_t} + \int_{S_t} \frac{p_t v(x) - q_t u(x)}{q_t u(x)} p_t v(x) \mathrm{d}x \right) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\quad + \mathbb{E}_I \sum_{t=1}^{T} \left(p_t - q_t + \int_{S_t} \big(p_t v(x) - q_t u(x)\big) \mathrm{d}x \right) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\quad + \mathbb{E}_I \sum_{t=1}^{T} \big(p_t - q_t + (1 - p_t) - (1 - q_t)\big) \\
&= \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right).
\end{aligned}$$

Note that $q_t \geqslant p_0$ and $u(x) \geqslant 1/C_1$ for all $x \in [0,1]$, where

$$C_1 := \frac{c(1-s)}{s(1-\beta)} > 0.$$

Hence, we can bound the KL divergence further as

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) &\leqslant \mathbb{E}_I \sum_{t=1}^{T} \left(\frac{(p_t - q_t)^2}{q_t} + \int_{S_t} \frac{\big(p_t v(x) - q_t u(x)\big)^2}{q_t u(x)} \mathrm{d}x \right) \\
&\leqslant \frac{1}{p_0} \mathbb{E}_I \sum_{t=1}^{T} \left(\underbrace{(p_t - q_t)^2}_{(a)} + C_1 \underbrace{\int_{S_t} \big(p_t v(x) - q_t u(x)\big)^2 \mathrm{d}x}_{(b)} \right). \qquad (42)
\end{aligned}$$

We bound both (a) and (b) in (42) from above. Let $t \in \{1, \ldots, T\}$. For (a), observe that

$$\begin{aligned}
(p_t - q_t)^2 &=^{(*)} \frac{\left(\int_{S_t} (v(x) - u(x)) \mathrm{d}x\right)^2}{\left(1 + \int_{S_t} v(x) \mathrm{d}x\right)^2 \left(1 + \int_{S_t} u(x) \mathrm{d}x\right)^2} \\
&\leqslant \left(\int_{S_t} (v(x) - u(x)) \mathrm{d}x \right)^2
\end{aligned}$$

$$\leqslant \left(\frac{s}{c(1-s)(1-\delta)} \int_{S_t} \tau_i(x) \mathrm{d}x \right)^2 \leqslant^{(**)} C_2 \frac{c^2}{4K^4}, \tag{43}$$

where

$$C_2 := \frac{s^2}{c^2(1-s)^2(1-\delta)^2},$$

and where $(*)$ holds since the cross terms cancel out and $(**)$ follows by Lemma 7.(v).1.

We now bound (b) in (42) from above. Observe that

$$\begin{aligned}
\int_{S_t} \big(p_t v(x) - q_t u(x)\big)^2 \mathrm{d}x &= \int_{S_t} \big(p_t v(x) - q_t v(x) + q_t v(x) - q_t u(x)\big)^2 \mathrm{d}x \\
&= (p_t - q_t)^2 \int_{S_t} v(x)^2 \mathrm{d}x && (44) \\
&\quad + 2q_t(p_t - q_t) \int_{S_t} v(x)\tau_i(x) \mathrm{d}x && (45) \\
&\quad + q_t^2 \int_{S_t} (\tau_i(x))^2 \, \mathrm{d}x. && (46)
\end{aligned}$$

The integral in (44) can be bounded by applying Lemma 7.(v).2. Combining that with the bound for $(p_t - q_t)^2$ from (43), gives

$$(p_t - q_t)^2 \int_{S_t} v(x)^2 \mathrm{d}x \leqslant C_2^2 (1+H) \frac{c^3}{4K^4}.$$

For the term (45), Lemma 7.(i) shows that $\tau_i(x) \leqslant Hc/K$. Together with (43) and Lemma 7.(ii).1 we find

$$\begin{aligned}
2q_t(p_t - q_t) \int_{S_t} v(x)\tau_i(x) \mathrm{d}x &\leqslant 2|p_t - q_t| \int_{S_t} v(x)\tau_i(x) \mathrm{d}x \\
&\leqslant 2|p_t - q_t| \left(\max_{x \in [0,1]} \tau_i(x) \right) \int_S v(x) \mathrm{d}x \\
&\leqslant 2\sqrt{C_2} \frac{c}{2K^2} \left(H \frac{c}{K} \cdot c\sqrt{C_2}(1+H) \right) \\
&= C_2 H (1+H) \frac{c^3}{K^3}
\end{aligned}$$

Finally, we bound the term (46). As a consequence of Lemma 7.(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 (44), (45), (46) in (42), we obtain

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_J) &\leqslant \frac{1}{p_0}\mathbb{E}_I \sum_{t=1}^{T} \left((p_t - q_t)^2 + C_1 \int_{S_t} \big(p_t v(y) - q_t u(y)\big)^2 \mathrm{d}y \right) \\
&\leqslant \frac{1}{p_0}\mathbb{E}_I \sum_{t=1}^{T} \left(C_2 \frac{c^2}{4K^4} + C_1 \left(C_2^2 (1+H) \frac{c^3}{4K^4} + C_2 H(1+H) \frac{c^3}{K^3} + \frac{c^3}{4\sigma\sqrt{\pi}K^3} \right) \right) \\
&\leqslant \frac{1}{4p_0} \left(c^2 C_2 + C_1 \left(c^3 C_2^2 (1+H) + 4c^3 C_2 H(1+H) + \frac{c^3}{\sigma\sqrt{\pi}} \right) \right) \frac{T}{K^3}.
\end{aligned}$$

This implies (39). □

Proof of Theorem 4.

We first show that the preference functions v_0 and $\{v_I : I \in \mathcal{D}_K, K \geqslant 2\}$ satisfy Assumption 1. To see this observe that the choice $c \in (0, \frac{1}{4}]$, $s = 0.8c$ and $\delta = \frac{1}{2}$ implies

$$v_0(x) \in \left[\frac{0.8}{1-s}, \frac{1.6}{1-s} \right] \subseteq [0.8, 2]\,.$$

Moreover, for all $K \geqslant 2$ and $I \in \mathcal{D}_K$ we have $\beta \leqslant L/8 \leqslant 0.0013$, and therefore

$$v_I(x) \geqslant v_0(x)(1-\beta) \geqslant 0.79,$$

and Lemma 7.(i) implies

$$v_I(x) \leqslant v_0(x) \left(1 + \frac{H}{8} \right) \leqslant 2.01 \leqslant 2.56 \leqslant \frac{w(0)}{\int_0^1 (w(x) - w(0))\mathrm{d}x} \leqslant 4,$$

for all choices of $c \in (0, \frac{1}{4}]$. This shows that Assumption 1(i) is satisfied with $\overline{v} = 4$ and $\underline{v} = 0.79$.

We now show that $v_I'(\cdot)$ is uniformly bounded and hence Assumption 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 7.(i) it suffices to show that $\sum_{i \in I} \tau_i'(x)$ is uniformly bounded. Note that

$$\tau_i'(x) = -\frac{2}{\sigma^2} \phi_i(x) b\big(\phi_i(x)\big).$$

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 $B_{N_K+1} :=$

$[0,1]\setminus\bigcup_{i\in[N_K]} B_i$, and $\phi_i(B_{i_x}) = \big[2(i_x-i)-1, 2(i_x-i)+1\big)$. Since $|yb(y)|$ is decreasing for $y \geqslant 1$ and increasing for $y \leqslant -1$, we obtain that, for all $i < i_x$,

$$0 < \phi_i(x)b\big(\phi_i(x)\big) \leqslant \big(2(i_x-i)-1\big)b\big(2(i_x-i)-1\big),$$

and for all $i > i_x$,

$$0 < -\phi_i(x)b\big(\phi_i(x)\big) \leqslant \big(2(i_x-i)+1\big)b\big(2(i_x-i)+1\big).$$

From this we conclude that

$$\begin{aligned}
\left|\sum_{i\in I}\tau_i'(x)\right| &= \left|\sum_{i\in I}\frac{2}{\sigma^2}\phi_i(x)b\big(\phi_i(x)\big)\right| \\
&\leqslant \frac{2}{\sigma^2}\left(\big|\phi_{i_x}(x)b\big(\phi_{i_x}(x)\big)\big| + \sum_{i=1}^{i_x-1}\phi_i(x)b\big(\phi_i(x)\big) - \sum_{i=i_x+1}^{N_K}\phi_i(x)b\big(\phi_i(x)\big)\right) \\
&\leqslant \frac{2}{\sigma^2}\left(\big|\phi_{i_x}(x)b\big(\phi_{i_x}(x)\big)\big| + \sum_{i=1}^{i_x-1}\big(2(i_x-i)-1\big)b\big(2(i_x-i)-1\big)\right. \\
&\qquad \left. - \sum_{i=i_x+1}^{N_K}\big(2(i_x-i)+1\big)b\big(2(i_x-i)+1\big)\right) \\
&\leqslant \frac{2}{\sigma^2}\left(\frac{\sigma}{\sqrt{e}} + 2\sum_{n=1}^{\infty}(2n-1)b\big(2n-1\big)\right) < \infty.
\end{aligned}$$

As a result, v_0 and $\{v_I : I \in \mathcal{D}_K, K \geqslant 2\}$ satisfy Assumption 1. This implies

$$\begin{aligned}
\Delta_\pi(T) &= \sup_{v\in\mathcal{V}}\Delta_\pi(T,v) \\
&\geqslant \frac{1}{|\mathcal{D}_K|}\sum_{I\in\mathcal{D}_K}\Delta_\pi(T,v_I) \\
&\geqslant \frac{1}{|\mathcal{D}_K|}\sum_{I\in\mathcal{D}_K}\left(C_1\int_{I^\dagger}(T-\mathbb{E}_I[k(x)])\epsilon_I(x)\mathrm{d}x - C_2\frac{T}{K}\right),
\end{aligned} \tag{47}$$

where C_1 and C_2 are as in Proposition 3.

The integral $\int_{I^\dagger}\epsilon_I(x)\mathrm{d}x$ can be bounded from below as

$$\begin{aligned}
\int_{I^\dagger}\epsilon_I(x)\mathrm{d}x &= \sum_{i\in I}\int_{I^\dagger}\tau_i(x)\mathrm{d}x - \beta c = \sum_{i\in I}\int_{B_i}\tau_i(x)\mathrm{d}x - \beta c \\
&\geqslant^{(*)} P\frac{c^2}{2K} - L\frac{c^2}{K} = \frac{c^2(P-2L)}{2K},
\end{aligned}$$

where at $(*)$ we used Lemma 7.(v).1. We use this lower bound to analyze (47). To this end, let $C_3 := c^2 C_1(P-2L)/2$. Then

$$\Delta_\pi(T) \geqslant (C_3 - C_2)\frac{T}{K} - \frac{C_1}{|\mathcal{D}_K|} \underbrace{\sum_{I\in\mathcal{D}_K} \int_{I^\dagger} \mathbb{E}_I[k(x)]\epsilon_I(x)\mathrm{d}x}_{(a)}. \tag{48}$$

We now bound the term (a) in (48) from above, using Proposition 4. Let C_c denote the constant from Proposition 4, and let $I \in \mathcal{D}_K$ and $J = I\backslash\{i\}$ for some $i \in I$. Then, for $x \in B_i$,

$$\mathbb{E}_I[k(x)]\epsilon_I(x) \leqslant \left(\mathbb{E}_J[k(x)] + C_c\left(\frac{T}{K}\right)^{3/2}\right)|\epsilon_I(x)|. \tag{49}$$

To apply (49) in order to bound (a) in (48), 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,

$$\begin{aligned}
\sum_{I\in\mathcal{D}_K} \int_{I^\dagger} \mathbb{E}_I[k(x)]\epsilon_I(x)\mathrm{d}x &= \int_{x\in U} \sum_{I\in E_K^x} \mathbb{E}_I[k(x)]\epsilon_I(x)\mathrm{d}x \\
&= \int_{x\in U} \sum_{J\in F_{K-1}^x} \mathbb{E}_{J\cup\{i_x\}}[k(x)]\epsilon_{J\cup\{i_x\}}(x)\mathrm{d}x \\
&\leqslant^{(*)} \int_{x\in U} \sum_{J\in F_{K-1}^x} \mathbb{E}_J[k(x)]\big|\epsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x \qquad (50) \\
&\quad + C_c\left(\frac{T}{K}\right)^{3/2} \int_{x\in U} \sum_{J\in F_{K-1}^x} \big|\epsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x, \qquad (51)
\end{aligned}$$

where at $(*)$ we apply (49). We now bound (50) and (51) from above. For (50), $|\epsilon_I(x)|$ is bounded uniformly in x by Lemma 7.(i):

$$\begin{aligned}
&\int_{x\in U} \sum_{J\in F_{K-1}^x} \mathbb{E}_J[k(x)]\big|\epsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x \\
&\leqslant (H+L)\frac{c}{K}\int_{x\in U} \sum_{J\in F_{K-1}^x} \mathbb{E}_J[k(x)]\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\int_{x\in U\setminus J^\dagger}\mathbb{E}_J[k(x)]\mathrm{d}x\\
&\leqslant (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\int_0^1\mathbb{E}_J[k(x)]\mathrm{d}x\\
&\leqslant (H+L)\,\frac{c}{K}\sum_{J\in\mathcal{D}_{K-1}}\sum_{t=1}^{T}\mathbb{E}_J[\mathrm{vol}(S_t)]\\
&\leqslant (H+L)\,\frac{c^2}{K}|\mathcal{D}_{K-1}|T.
\end{aligned}
$$

We now consider (51). Observe that $|\epsilon_I(x)|$ is bounded locally on B_i:

$$
\begin{aligned}
&\int_{x\in U}\sum_{J\in F^x_{K-1}}\big|\epsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x = \sum_{J\in\mathcal{D}_{K-1}}\int_{x\in U\setminus J^\dagger}\big|\epsilon_{J\cup\{i_x\}}(x)\big|\mathrm{d}x\\
&= \sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\int_{B_i}|\epsilon_{J\cup\{i\}}(x)|\mathrm{d}x \leqslant^{(*)} \sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\int_{B_i}\big(\tau_i(x)+\beta\big)\mathrm{d}x\\
&\leqslant^{(**)} \sum_{J\in\mathcal{D}_{K-1}}\sum_{i\notin J}\frac{c^2(1+2L)}{2K^2} = \frac{c^2(1+2L)}{2K^2}|\mathcal{D}_{K-1}|(N_K-K+1),
\end{aligned}
$$

where we apply Lemma 7.(vi) at $(*)$ and Lemma 7.(v).1 at $(**)$. After inserting these upper bounds for (50) and (51) into (48), we conclude

$$
\begin{aligned}
\Delta_\pi(T) &\geqslant (C_3-C_2)\frac{T}{K} - \frac{C_1}{|\mathcal{D}_K|}\left((H+L)\,\frac{c^2}{K}|\mathcal{D}_{K-1}|T + \frac{c^2C_c(1+2L)}{2}|\mathcal{D}_{K-1}|(N-K+1)\frac{T^{3/2}}{K^{7/2}}\right)\\
&= \left(C_3-C_2-\frac{c^2C_1\,(H+L)\,|\mathcal{D}_{K-1}|}{|\mathcal{D}_K|}\right)\frac{T}{K} - \frac{c^2C_1C_c(1+2L)|\mathcal{D}_{K-1}|}{2|\mathcal{D}_K|}(N-K+1)\frac{T^{3/2}}{K^{7/2}}.
\end{aligned}
$$

Next, note that

$$
\frac{|\mathcal{D}_{K-1}|}{|\mathcal{D}_K|} = \frac{K}{N_K-K+1},
$$

and therefore

$$
\Delta_\pi(T) \geqslant \underbrace{\left(C_3-C_2-\frac{(H+L)\,c^2C_1K}{N_K-K+1}\right)}_{(b)}\frac{T}{K} - \frac{c^2C_1C_c(1+2L)}{2}\frac{T^{3/2}}{K^{5/2}}. \tag{52}
$$

We abbreviate the constant $C_4 := c^2C_1C_c(1+2L)/2$. 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 (52) 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 $c \in (0, \frac{1}{4}]$ we obtain $C_6 \geqslant (P-2L)/2 - (H-L)/3 = 0.043 > 0$. In addition, our choice of $s = 0.8c$ implies

$$\frac{s}{1-s} < \frac{\sqrt{cC_6}(1-\delta)}{\sqrt{c+L}},$$

and therefore $C_5 = c^2 C_1 C_6 - C_2 > 0$. Now, choose

$$\gamma = \left(\frac{5C_4}{C_5}\right)^{2/3} \qquad \text{and} \qquad K = \max\{2, \left\lceil \gamma T^{1/3} \right\rceil\}.$$

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

$$\underline{C} = \min\left\{\left(\tfrac{1}{2} - \tfrac{\sqrt{2}}{40}\right)C_5, \left(\tfrac{1}{2}\left(\tfrac{1}{5}\right)^{2/3} - \left(\tfrac{1}{5}\right)^{5/3}\right)\frac{C_5^{5/3}}{C_4^{2/3}}\right\} > 0.$$

□

Proof of Lemma 7.

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

$$\sum_{i\in I} \tau_i(x) = \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}} \sum_{i\in I} \exp\big(-\tfrac{1}{2\sigma^2}(y+2i_0-2i)^2\big) \leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \exp\big(-\tfrac{1}{2\sigma^2}(y-2n)^2\big)$$
$$\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \exp\left(-\frac{2n^2}{\sigma^2}\right) = H\frac{c}{K}.$$

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 $B_{N_K+1} := [0,1]\backslash \bigcup_{i\in[N_K]} B_i$. Note that τ_i is either increasing or decreasing on B_{i_x} for $i \neq i_x$. Then

$$\tau_i(x) \leqslant \max\left\{\tau_i\left(c\frac{i_x-1}{K}\right), \tau_i\left(c\frac{i_x}{K}\right)\right\} = \frac{c}{K}\max\Big\{b\big(2(i_x-i)+1\big), b\big(2(i_x-i)-1\big)\Big\},$$

for $i \neq i_x$. From this, we derive for any $x \notin I^\dagger$,

$$\sum_{i\in I}\tau_i(x) \leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{i\in I}\max\Big\{\exp\big(-\tfrac{1}{2\sigma^2}(2(i_x-i)+1)^2\big), \exp\big(-\tfrac{1}{2\sigma^2}(2(i_x-i)-1)^2\big)\Big\}$$
$$\leqslant \frac{c}{K}\frac{1}{\sigma\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\exp\big(-\frac{1}{2\sigma^2}(2n-1)^2\big),$$

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

$$\int_0^1 \tau_i(x)\mathrm{d}x = \frac{c}{K}\int_0^1 b\left(\frac{2Kx}{c} - 2i + 1\right)\mathrm{d}x$$
$$\leqslant \frac{c}{K}\int_{\mathbb{R}} b\left(\frac{2Kx}{c} - 2i + 1\right)\mathrm{d}x = \frac{c^2}{2K^2}\int_{\mathbb{R}} b(y)\mathrm{d}y = \frac{c^2}{2K^2}.$$

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 $x \in B_i$,

$$-\beta' \leqslant \epsilon_{I\setminus\{i\}}(x;\beta') \leqslant 0,$$

since $x \notin (I\setminus\{i\})^\dagger$. Hence,

$$|\epsilon_I(x;\beta')| = \left|\tau_i(x) + \epsilon_{I\setminus\{i\}}(x;\beta')\right| \leqslant \tau_i(x) + \left|\epsilon_{I\setminus\{i\}}(x;\beta')\right| \leqslant \tau_i(x) + \beta'.$$

□

Proof of Lemma 8.

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 (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 1. □

Appendix C: Relation to discrete multinomial logit choice probabilities

The choice probabilities in our continuous assortment optimization model are closely connected to the discrete multinomial logit (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 $v : [0,1] \to (0, \infty)$. 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 (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

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

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

Appendix D: Bisection algorithm for Section 5

According to Proposition 1, the optimal assortment can be computed up to any desired accuracy $\epsilon > 0$. The algorithm COA(n) below shows how this is done, where $n := -\log \epsilon$. Recall that

$$\mathcal{I}(S, \varrho) := \int_S v(x)\big(w(x) - \varrho\big)\mathrm{d}x. \tag{53}$$

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 8. 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 IB(n, ϱ) is $\mathcal{O}(-\log \epsilon)$, and hence the run time of COA(n) is $\mathcal{O}((\log \epsilon)^2)$.

`Capacitated Optimal Assortment COA`(n)

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

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

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

(ii) `If` $\mathrm{vol}(W_{\texttt{piv}}) \leqslant c$, `then put` $S_{\texttt{piv}} := W_{\texttt{piv}}$ `and` $I_{\texttt{piv}} := \mathcal{I}(S_{\texttt{piv}}, \texttt{piv})$ `as in` (53) `and go to 5.`

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

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

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

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 (53). 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 $S^* := S_{\mathtt{piv}}$. 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 $x_{\mathtt{piv}}$ 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 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 6.

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

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

2. Level set. Put

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

Go to 3.

3. Pivot.

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

(ii) If $\mathrm{vol}(L^{\mathtt{piv}}) \leqslant c$, 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 4.

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

```
5. Terminate.
```

Appendix E: Additional numerical experiments

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 multinomial logit (MNL) model. Section E.1 describes the experimental set-up, and in Section E.2 we report our results. Section E.3 contains additional details on the derivation of the maximum-likelihood estimator, and in Section E.4 we specify the kernel density estimator used in these numerical experiments.

E.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 develop a kernel density estimator (KDE). Throughout this section we use the same notations and concepts as in Sections 5.2 and 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. (2018), 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 $N \in \{10, 30, 50\}$,

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

and where $\phi(\,\cdot\,; \mu, \sigma)$ 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

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

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 $Y_t = \sum_{i=1}^{N} i \mathbf{1}_{B_i}(X_t)$.

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 $\boldsymbol{v} = (v_1, \ldots, v_N)$ of the discrete model corresponds to an instance of the continuous model, by letting the discrete purchase Y_t coincide with the continuous purchase $X_t \in B_{i_t}$ (and $X_t = \emptyset$ if $Y_t = 0$) and by setting the preference function $v(x)$ equal to

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

Conversely, each instance of the continuous model with preference function v that is constants on bins $B_1, \ldots, B_N$ corresponds to an instance of the discrete model by setting $v_i = \int_{B_i} v(x) \mathrm{d}x$,

for all $i = 1, \ldots, N$. 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 $(v_1^{(3)}, \ldots, v_N^{(3)})$ 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 $v^{(j)}(x)$ (defined in more detail in Appendix E.4) and let $\hat{v}^{(j),\mathrm{MLE}}(x)$ denote the stepwise constant function

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

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

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

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 Appendix E.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. (2018), 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 $\check{w}$, in line with Equation (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

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

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 $c = 0.5$.

E.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 5 through 10.

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.

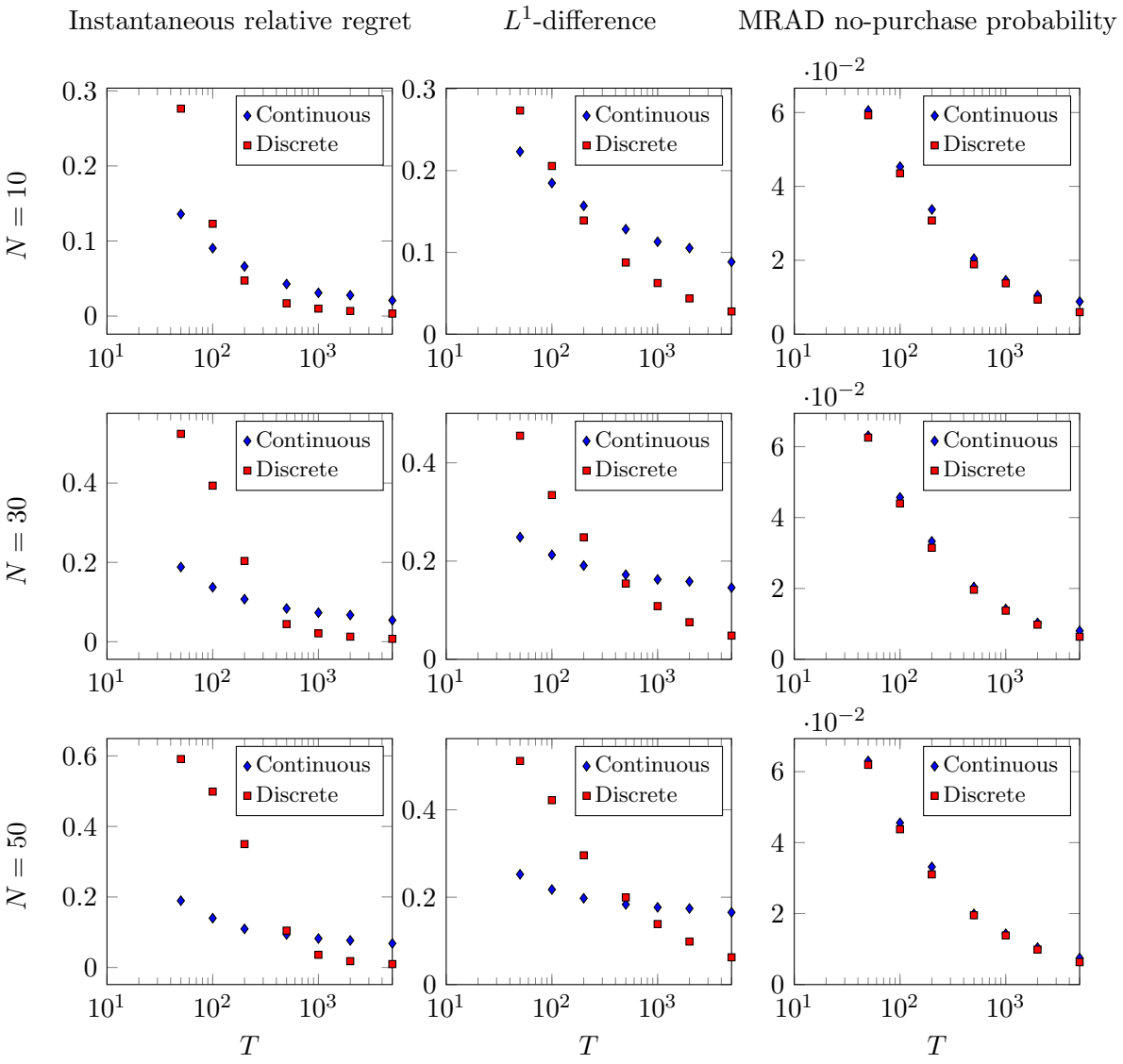

Figure 5: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 1 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

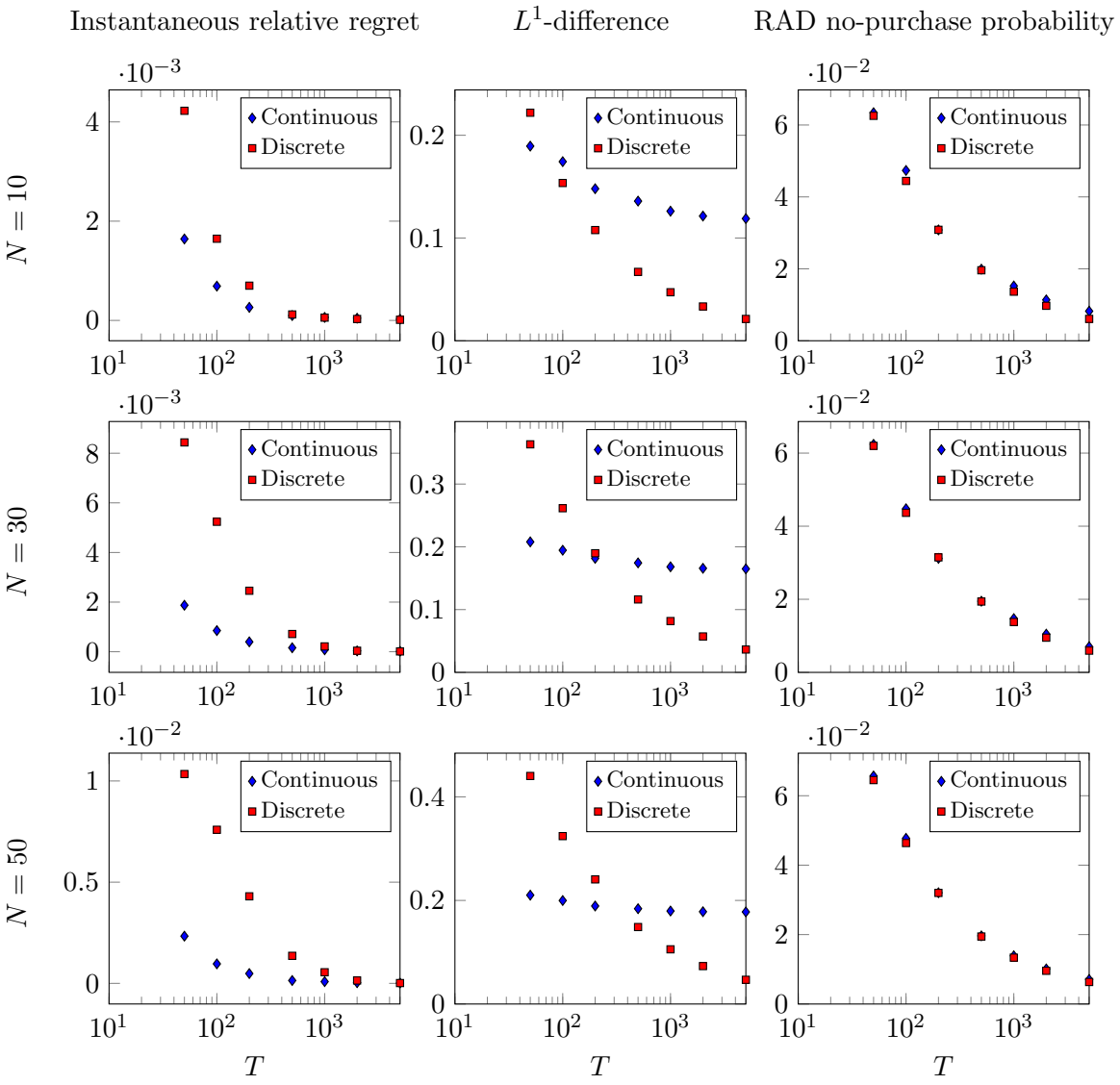

Figure 6: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 1 with $c = 1$ and $K = N$ based on 1 000 simulations.

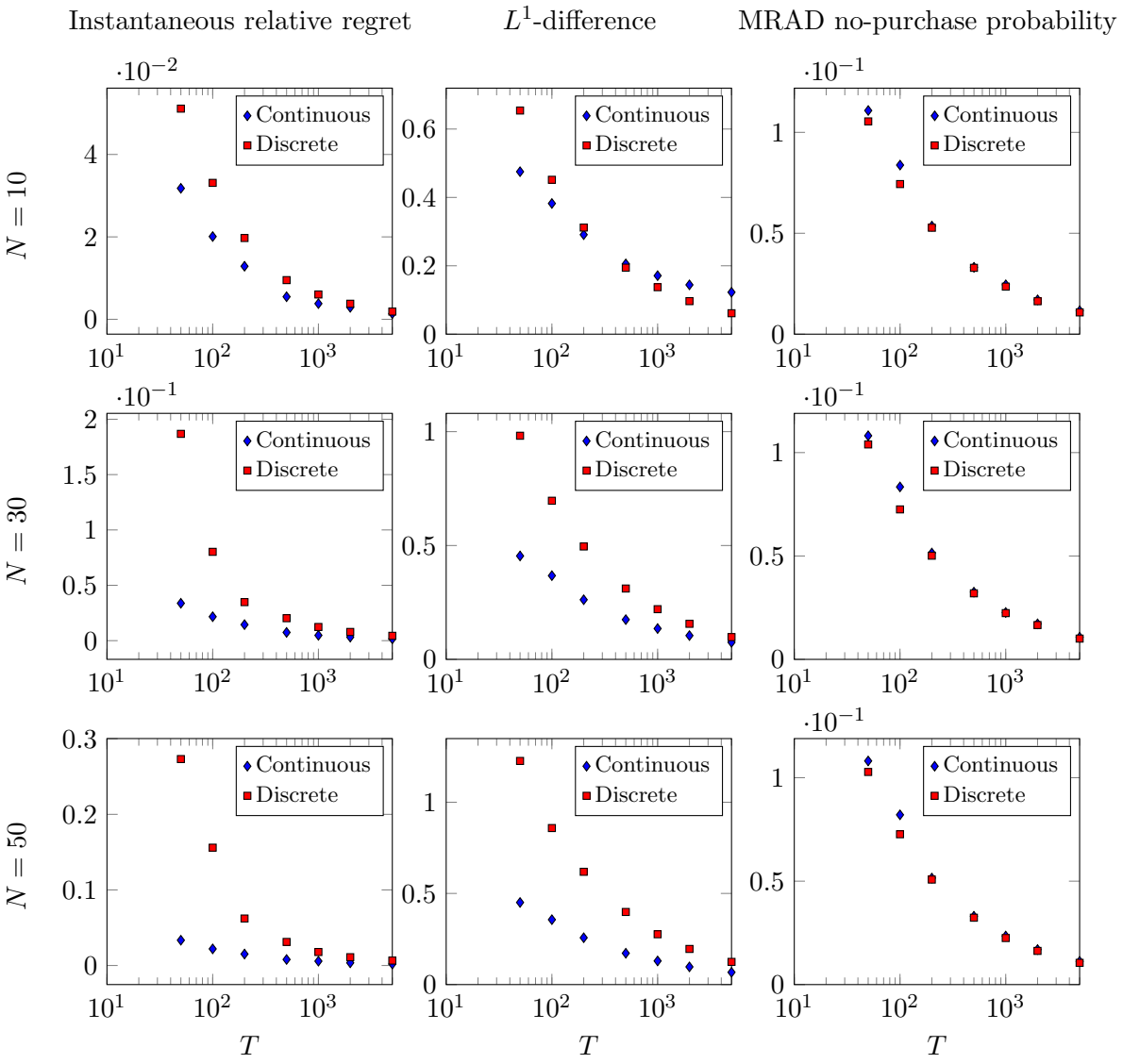

Figure 7: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 2 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

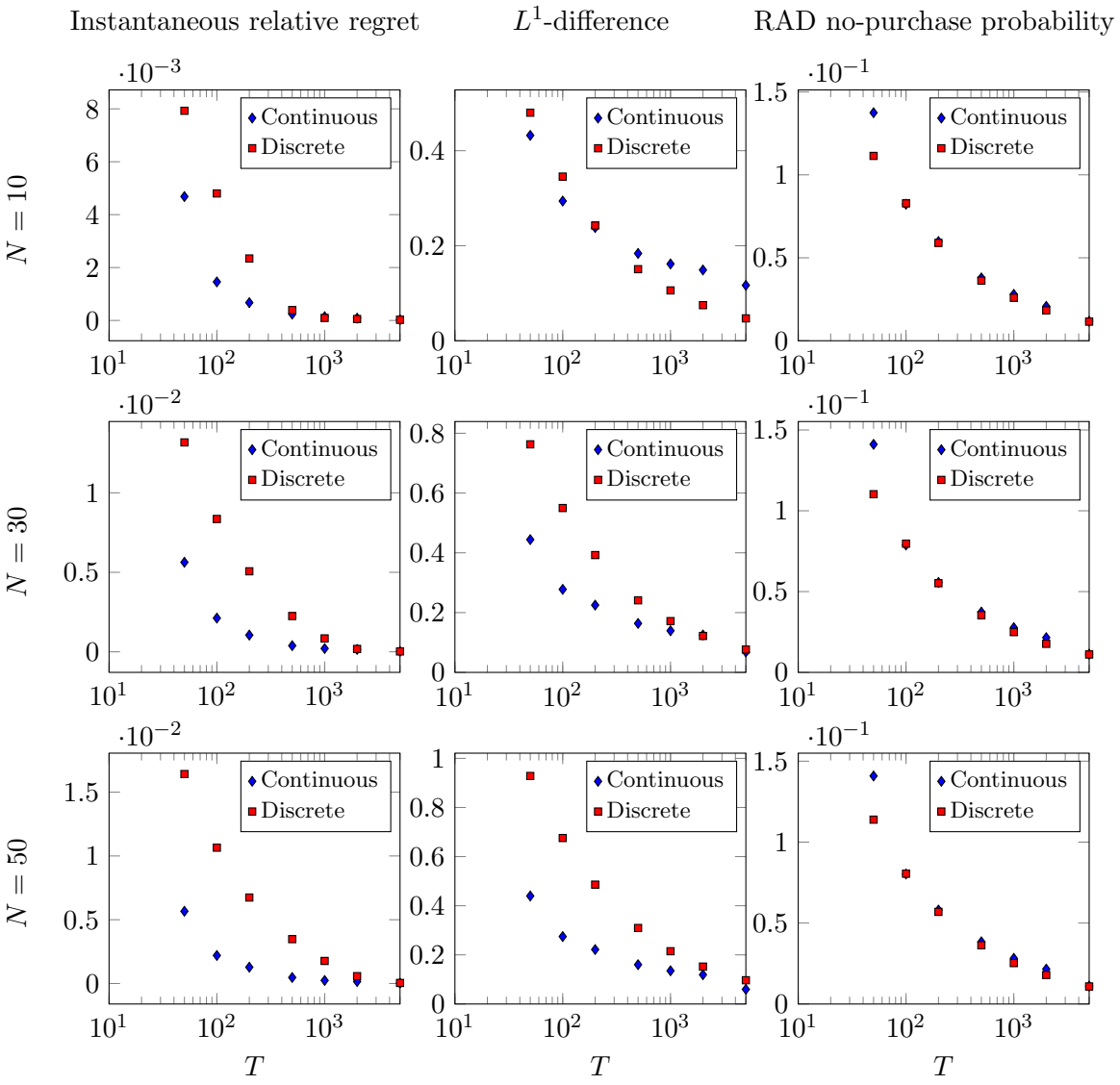

Figure 8: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 2 with $c = 1$ and $K = N$ based on 1 000 simulations.

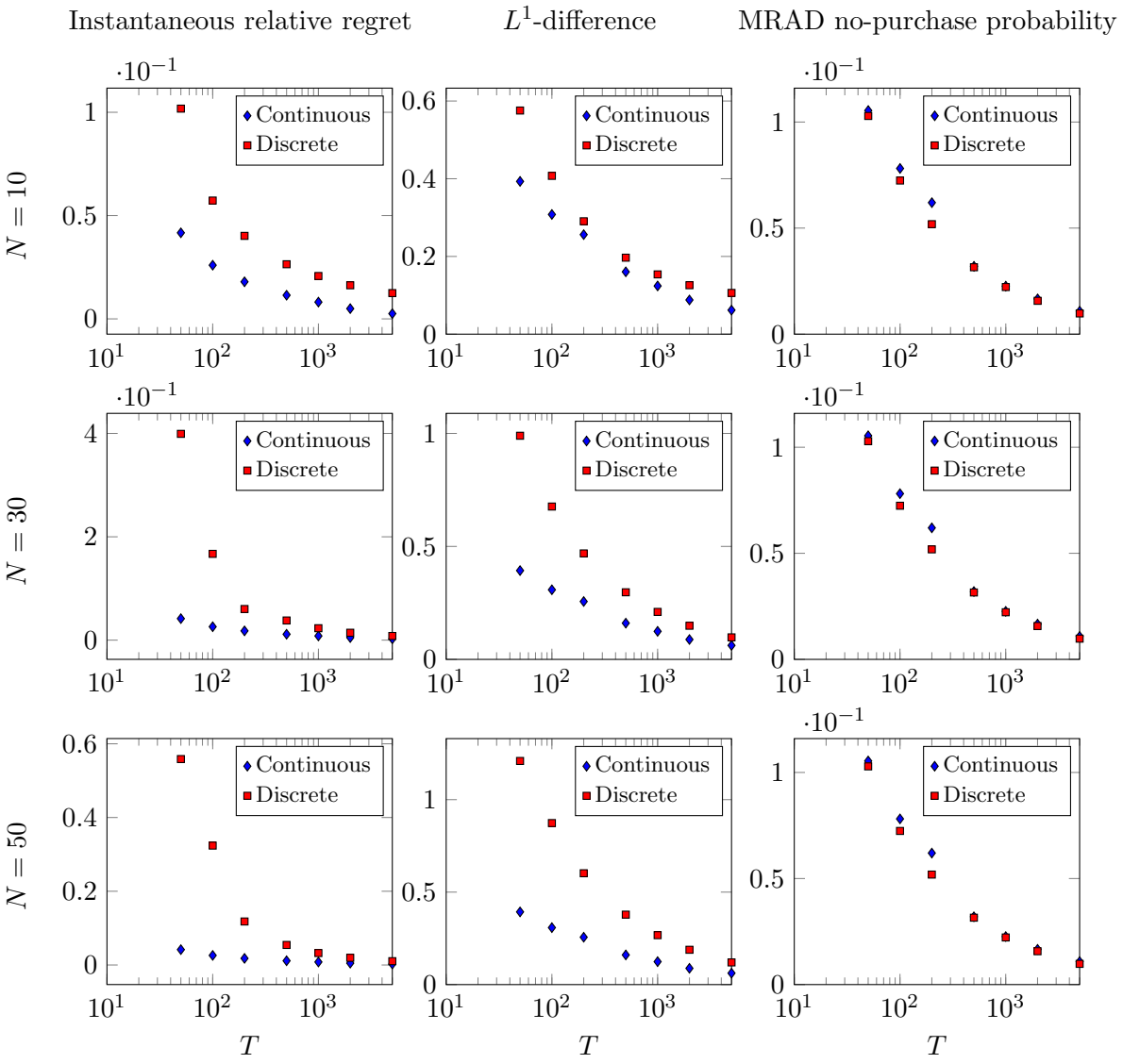

Figure 9: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 3 with $c = 0.5$ and $K = N/2$ based on 1 000 simulations.

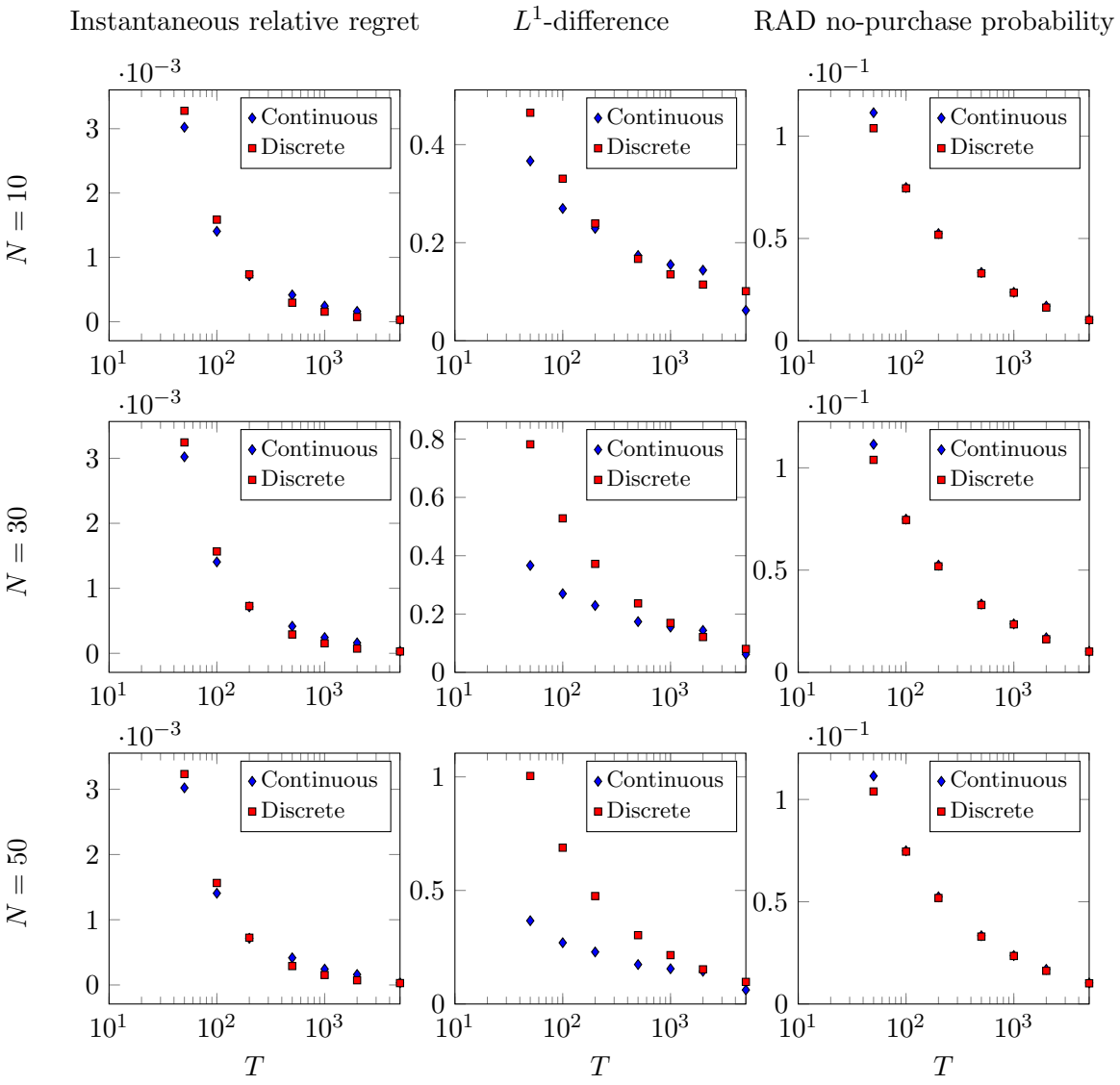

Figure 10: The performance metrics comparing the predictive performance for of the continuous and the discrete logit choice model for scenario 3 with $c = 1$ and $K = N$ based on 1 000 simulations.

E.3. 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 $\hat{v}_1, \ldots, \hat{v}_N$. Following Appendix E.1 and E.2, 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 $\{1, \ldots, N/2\}$ and $\{N/2+1, \ldots, N\}$ (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 i_t denote the discrete purchase observed at time t when offering $D_t \subseteq [N]$. Then the log likelihood is

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

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

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

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

$$\sum_{t=1}^{T} \mathbf{1}\{i_t = j\} = \sum_{t=1}^{T} \frac{\hat{v}_j}{1 + \sum_{i=1}^{N} \hat{v}_i}. \tag{54}$$

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

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

or, equivalently,

$$\sum_{t=1}^{T} \mathbf{1}\{i_t = 0\} = \sum_{t=1}^{T} \frac{1}{1 + \sum_{i=1}^{N} \hat{v}_i}. \tag{55}$$

Combining (54) and (55), we obtain

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

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

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

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

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

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

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

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

$$\sum_{t \in \mathcal{T}^k} \mathbf{1}\{i_t = j\} = \sum_{t \in \mathcal{T}^k} \frac{\hat{v}_j}{1 + \sum_{i = D^k} \hat{v}_i}. \tag{56}$$

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

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

or, equivalently,

$$\sum_{t \in \mathcal{T}^k} \mathbf{1}\{i_t = 0\} = \sum_{t \in \mathcal{T}^k} \frac{1}{1 + \sum_{i \in D^k} \hat{v}_i}. \tag{57}$$

Combining (56) and (57), we obtain

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

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

E.4. Kernel density estimator

In this section we define the kernel density estimator used to estimate the preference function v in the continuous assortment model.

Because traditional kernel density estimation does not perform well near endpoints of the support, we construct the KDE 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 kernel density estimation, 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.

We define our estimator $\hat{v}$ of v based on continuous purchases $X_1, \ldots, X_T \in [0,1] \cup \{\emptyset\}$. If our estimator is applied to scenario 1 or 2, in which case the observed purchases $Y_1, \ldots, Y_T \in [N] \cup \{0\}$ are discrete, we draw X_t uniformly at random from B_{Y_t} if $Y_t \neq 0$ and set $X_t := \emptyset$ if $Y_t = 0$, for all $t \in [T]$. We define the estimator $\hat{v}$ for the situation that there are $L \in \mathbb{N}$ so-called test assortments $S^1, \ldots, S_L$ each of which is offered during exactly $M \in \mathbb{N}$ time periods, and each of which has volume c. More precisely, the offered assortment at time $t \in [T]$ is $S_t = S^k$ if $t \in \{(k-1)M+1, \ldots, kM\}$. For all $x \in [0,1]$, let $e(x)$ denote the number of times that product x is contained in the test assortments $S^1, \ldots, S^L$:

$$e(x) := \sum_{k=1}^{L} \mathbf{1}_{S^k}(x).$$

We assume that the test assortments $S^1, \ldots, S^L$ cover the entire set of products $[0,1]$, that is, $e(x) > 0$ for all $x \in [0,1]$. For each test assortment S^k we construct a corresponding estimate $\hat{v}_k(x)$ of $v(x)\mathbf{1}_{S^k}(x)$, and then combine these into our estimate $\hat{v}$, as follows:

$$\hat{v}(x) := \frac{1}{e(x)} \sum_{k=1}^{L} \hat{v}_k(x), \qquad x \in [0,1]. \tag{58}$$

To define $\hat{v}_k$, 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 $L_2([-1,1])$. Let a_k and b_k be such that $S^k = [a_k, b_k]$, for all $k \in [L]$, let $h \in (0, c/2]$ be a bandwidth parameter and for all $k \in [L]$ and $x \in \mathbb{R}$ define the shifted support I_x^k as

$$I_x^k = \left[-\min\left\{1, \frac{x-a_k}{h}\right\}, \min\left\{1, \frac{b_k - x}{h}\right\} \right].$$

In addition, we define two shift coefficients γ_x^k and ζ_x^k as

$$(\,\gamma_x^k\,,\,\zeta_x^k\,) = \begin{cases} \left(\dfrac{2h}{h+x-a_k}\,,\quad -\dfrac{h-(x-a_k)}{h+x-a_k}\right) & \text{for } x \in [a_k, a_k+h), \\ (1,\quad 0) & \text{for } x \in [a_k+h, b_k-h], \\ \left(\dfrac{2h}{h+b_k-x}\,,\quad \dfrac{h-(b_k-x)}{h+b_k-x}\right) & \text{for } x \in (b_k-h, b_k], \end{cases}$$

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

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

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

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

$$E_k := \{X_t : X_t = \emptyset \ \text{ and } \ (k-1)M + 1 \leqslant t \leqslant kM\},$$

and let

$$A_k := \{X_t : X_t \neq \emptyset \ \text{ and } \ (k-1)M + 1 \leqslant t \leqslant kM\}$$

denote the actual purchases observed when S^k is offered. Then we estimate $v(x)\mathbf{1}_{S^k}(x)$ by

$$\hat{v}_k(x) := \frac{1}{(|E_k|+1)h} \sum_{X \in A_k} K_x^k\left(\frac{X-x}{h}\right), \qquad x \in S^k,$$

and set $\hat{v}_k(x) := 0$ for $x \notin S^k$. These estimates are combined into one estimate $\hat{v}(x)$ of $v(x)$, as given by (58). Analysis of the convergence rates reveals that an appropriate choice for the bandwidth parameter h and order parameters ℓ is

$$h^* := \min\left\{\frac{c}{2}, \frac{1}{e}\right\} \qquad \text{and} \qquad \ell^* := \left[\frac{1}{2}\log\big(-2M\log h^*\big) - \frac{1}{2}\right],$$

respectively, where $[x]$ denotes the rounded value of $x \in \mathbb{R}$.