The Approximability of Assortment Optimization under Ranking Preferences

Ali Aouad*   Vivek Farias †   Retsef Levi‡   Danny Segev§

Abstract

The main contribution of this paper is to provide best-possible approximability bounds for assortment planning under a general choice model, where customer choices are modeled through an arbitrary distribution over ranked lists of their preferred products, subsuming most random utility choice models of interest. From a technical perspective, we show how to relate this optimization problem to the computational task of detecting large independent sets in graphs, allowing us to argue that general ranking preferences are extremely hard to approximate with respect to various problem parameters. These findings are complemented by a number of approximation algorithms that attain essentially best-possible factors, proving that our hardness results are tight up to lower-order terms. Surprisingly, our results imply that a simple and widely studied policy, known as revenue-ordered assortments, achieves the best possible performance guarantee with respect to the price parameters.

Keywords: Assortment optimization, choice models, hardness of approximation, independent set, approximation algorithms.

---

*London Business School, London NW1 4SA, United Kingdom. Email: aouad.ma@gmail.com.

†Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, Email: vivekf@mit.edu.

‡Sloan School of Management, Massachusetts Institute of Technology,77 Massachusetts Avenue, Cambridge, Massachusetts 02139, Email: retsef@mit.edu.

§Department of Statistics, University of Haifa, Haifa 31905, Israel, Email: segevd@stat.haifa.ac.il.

1 Introduction

Assortment planning is paramount to revenue management in highly differentiated markets, such as offline and online retail. The typical computational problem in this context is that of identifying a selection of products that maximizes revenue (assuming no stock-out events) based on previously-estimated random and heterogeneous customer preferences over the underlying set of products. The extensive literature in economics, marketing, and operation management proposes numerous approaches to modeling customer choice preferences, which are then used for predicting the variations in market shares in response to how the product mix changes.

This paper is focused on studying the computational complexity of a very general problem formulation, where customer choices are modeled through an arbitrary distribution over ranked preference lists. The incorporation of this choice model into revenue management settings was first proposed by Mahajan and Van Ryzin (2001) and later on by Rusmevichientong et al. (2006), with different objectives in mind. Specifically, Mahajan and Van Ryzin (2001) considered a sequential model, where the objective is to compute an assortment that maximizes the expected revenue generated by a sequence of arriving customers. They established basic structural properties, devised a sample-path gradient algorithm (converging to a local optimum), and conducted an extensive numerical study. Rusmevichientong et al. (2006) aimed at optimizing prices with respect to a sample of consumer data. They investigated sample-size complexity, proved that the corresponding pricing problem is NP-complete, devised an efficient heuristic algorithm, and tested this approach within a case study. Subsequently, the question of model estimation from data was examined through various methodologies, including the robust formulation of Farias et al. (2013), the column generation algorithm by van Ryzin and Vulcano (2014), and the expectation-maximization method of van Ryzin and Vulcano (2017). This non-parametric modeling approach, whose specifics are given in Section 1.3, subsumes most models of practical interest as special cases. In particular, ranked preference lists are equivalent to a general random utility model, in which a representative agent maximizes his random utility function over a set of alternatives to derive his preferences.

In the context of assortment planning, these choice models are subject to a fundamental tradeoff between model expressiveness and computational tractability. Indeed, assortment optimization was shown to be tractable under specific choice models proposed in the revenue management literature, where various structural and probabilistic assumptions are made. Probably the most well-known settings that still admit polynomial-time solution methods are the widespread multinomial-logit (MNL) model and variants of the nested-logit (NL) model. In the specific context of ranking-based models, the work of Honhon et al. (2012) identifies classes of simple combinatorial structures enabling polynomial-time algorithms. Since an exhaustive survey of these results is beyond the scope of this paper, we refer the reader to the work of Mahajan and Van Ryzin (2001), Talluri and Van Ryzin (2004), Blanchet et al. (2016), Davis et al. (2014), and Li et al. (2015). The references therein provide an excellent overview of tractable approaches in assortment optimization.

Despite an increasing stream of positive results for specific classes of instances, assortment planning initiates computationally-hard problems in more general settings. This was formally corroborated by several intractability results, such as that of Davis et al. (2014) and Gallego

and Topaloglu (2014), who demonstrated that natural extensions of the NL model are NP-hard. Under mixtures of logits, this problem is known to be strongly NP-hard even for two customer classes, as shown by Bront et al. (2009) and Rusmevichientong et al. (2014). For ranking-based models, while Honhon et al. (2015) developed practical pruning heuristics, their algorithms are is still exponential in general settings. As a result, beyond attraction-based models with a single customer class, the family of tractable choice models remains quite limited.

It is worth noting that the above-mentioned results merely state that the problems in question cannot be solved to optimality in polynomial-time (unless P = NP), and in fact, very little is known about hardness of approximation in this context. To our knowledge, the only result in this spirit was given by Goyal et al. (2016), showing that under ranking preferences, the capacitated assortment planning problem is NP-hard to approximate within factor better than $1 - 1/e$.

1.1 Our results

The main contribution of this paper is to provide best-possible inapproximability bounds for assortment planning under ranking preferences and to reveal hidden connections to other fundamental branches of discrete optimization. From a technical perspective, we show how to relate this model to the computational task of detecting large independent sets in graphs, allowing us to argue that general ranking preferences are extremely hard to approximate with respect to various problem parameters. These findings are complemented by a number of approximation algorithms that attain essentially best-possible performance guarantees with respect to various parameters, such as the ratio between extremal prices and the maximum length of any preference list. Our results provide a tight characterization (up to lower-order terms) of the approximability of assortment planning under a general model specification, as we briefly summarize next.

Hardness of approximation. By proposing a reduction from the maximum independent set problem, we prove that assortment planning under ranking preferences is NP-hard to approximate within factor $O(n^{1-\epsilon})$ for any fixed $\epsilon > 0$, where n stands for the number of products. This is the first strong inapproximability bound for the ranking preferences model, which is surprisingly established even in the uncapacitated setting. As previously mentioned, the hardness result of Goyal et al. (2016) only proves a constant lower bound and makes use of an additional capacity constraint. In fact, our $O(n^{1-\epsilon})$ bound holds even when all preference lists are derived from a common permutation over the set of products, meaning that all customers rank their alternatives consistently according to a unique order. Moreover, our reduction also gives an inapproximability bound of $O(\log^{1-\epsilon}(P_{\max}/P_{\min}))$, where $P_{\min}$ and $P_{\max}$ designate the minimal and maximal prices, respectively. Finally, through a reduction from the Min-Buying pricing problem, we establish APX-hardness even when there are only two distinct prices, with uniform probability of customer arrivals. The specifics of these results are given in Section 2.

Approximation algorithms. On the positive side, we devise approximation algorithms showing that the above-mentioned inapproximability bounds are best possible. By examin-

ing revenue-ordered assortments, we propose an efficient algorithm that attains performance guarantees of $O(\lceil \log(P_{\max}/P_{\min}) \rceil)$ and $O(\lceil \log(1/\tilde{\lambda}) \rceil)$, where $\tilde{\lambda}$ denotes the combined arrival probability of all customers who have the highest price item on their list. In particular, when all customer arrival probabilities are polynomially bounded away from 0, this bound translates to a logarithmic approximation (for example, under a uniform distribution). Finally, we devise a tight approximation algorithm in terms of the maximum length of any preference list. We prove that an $e\Delta$-approximation can be obtained via randomly generated assortments under a well-chosen distribution, where Δ denotes the maximal size of any preference list. Consequently, an immediate implication is that, when all preference lists are comprised of $O(1)$ products, we can approximate the optimal revenue within a constant factor. By derandomization, the resulting algorithm asymptotically matches the $O(\Delta^{1-\epsilon})$ inapproximability bound hiding within our reduction from the independent set problem. Additional details on these algorithms are provided in Section 3.

1.2 Subsequent work

The techniques developed in this paper have spurred new complexity results for related assortment planning problems. In particular, after communicating our reduction from the maximum independent set problem to Antoine Désir, Vineet Goyal, and Jiawei Zhang, they observed that ideas in this spirit provide tight inapproximability bounds for the mixture-of-MNL model (Désir et al. 2014). The basic ideas behind our reduction have also been utilized by Feldman and Topaloglu (2017) to prove strong inapproximability results for assortment optimization under the MNL model with arbitrary consideration sets.

In addition, shortly after our work appeared online (Aouad et al. 2015), a working paper of Berbeglia and Joret (2015) focused on the performance analysis of revenue-ordered assortments. In comparison to the $O(\lceil \log(P_{\max}/P_{\min}) \rceil)$ approximation we provide in Section 3.1, they were able to improve on the constant hiding within the $O(\cdot)$-notation. However, unlike our tight inapproximability bound (see Section 2.1), they prove only constant-factor hardness, similar to the previously known result by Goyal et al. (2016). Their algorithmic results hold in a broader setting that generalizes the class of random utility choice models. Indeed, the only technical assumption required is the regularity axiom, stating that the probability of choosing a specific product does not increase when the assortment is enlarged. It is worth noting that the latter observation also holds for the analysis we develop in Section 3.1.

1.3 The ranking preferences model

We are given a collection of n items (or products), where the per-unit selling price of item i is denoted by P_i. In addition, we model a population consisting of k customer types, one of which arrives at random, such that customer j is assumed to arrive with probability λ_j. Each customer type is defined by a preference list over the underlying set of products, according to which purchasing decisions are made. For any customer j, the preference list L_j is a subset of the products along with a linear order on these products. In other words, L_j can be viewed as a vector of products, ordered from the most preferred to the least preferred item that a customer

type is willing to purchase. For ease of notation, L_j will refer in the sequel to both the ranked list and the corresponding subset of products, $L_j \subseteq [n]$.

We define an assortment as a selection of products that is made available to customers. When faced with the assortment $S \subseteq [n]$, a customer type purchases the most preferred item in his list that is made available by S. If none of these products is available, he leaves without purchasing any item. Under this decision mechanism, we use $R_j(S)$ to denote the revenue obtained should customer type j arrive, for the assortment S. Conditional on the arrival of customer type j, the resulting revenue is equal to the price of the product purchased according to L_j, or to 0 when none of these products has been made available. The objective is to compute an assortment of products whose expected revenue is maximized, i.e., to identify a subset $S \subseteq [n]$ that maximizes

$$\mathcal{R}(S) = \sum_{j=1}^{k} \lambda_j \cdot R_j(S) \ .$$

2 Hardness Results

2.1 Relation to maximum independent set

Our main inapproximability result proceeds from unraveling a well-hidden connection between assortment planning and the maximum independent set problem (henceforth, Max-IS). To this end, we begin by recalling how the latter problem is defined, and state known hardness of approximation results due to Håstad (1996).

An instance of Max-IS is defined by an undirected graph $G = (V, E)$, where V is a set of n vertices, and E is the set of edges. A subset of vertices $U \subseteq V$ is said to be independent if no pair of vertices in U is connected by an edge. The objective is to compute an independent set of maximal cardinality. The most useful inapproximability result for our purposes is that of Håstad (1996), who proved that for any fixed $\epsilon > 0$, Max-IS cannot be approximated in polynomial time within factor $O(n^{1-\epsilon})$ unless P = NP.

Theorem 2.1. *Assortment planning under ranking preferences is NP-hard to approximate within $O(n^{1-\epsilon})$, for any fixed $\epsilon > 0$.*

Proof. In what follows, we describe an approximation-preserving reduction Φ that maps any instance $\mathcal{I}$ of Max-IS, defined on an n-vertex graph, to an assortment planning instance $\Phi(\mathcal{I})$, consisting of n products and n customers.

We begin by introducing some notation. Given a Max-IS instance $\mathcal{I}$ defined on an undirected graph $G = (V, E)$, let $V = \{v_1, \ldots, v_n\}$, each vertex being designated by an arbitrary label v_i. For each vertex $v_i \in V$, we use $N^-(i)$ to designate the indices of v_i's neighbors that are smaller than i, namely,

$$N^-(i) = \{j \in [n] : (v_i, v_j) \in E \text{ and } j < i\} \ .$$

The assortment planning instance $\Phi(\mathcal{I})$ is defined as follows:

- For each vertex $v_i \in V$, we introduce a product indexed by i, with price $P_i = n^{2i}/\alpha$, where $\alpha = 1/\sum_{i=1}^{n} n^{-2i}$.

4

- Also, for each vertex $v_i \in V$, there is a corresponding customer type whose preference list is L_i. This list consists of the products $N^-(i) \cup \{i\}$, and the preference order is set such that i is the least preferable product. Any order between the remaining products $N^-(i)$ works for our purposes, but to have a concrete definition, we assume that L_i orders these products by increasing indices (or equivalently, by increasing price).

- The probability (or arrival rate) of customer type i is $\lambda_i = \alpha/n^{2i}$. Note that, by definition of α, these probabilities indeed sum to 1.

Based on the above-mentioned hardness results of Håstad (1996), in order to establish our inapproximability bound, it is sufficient to prove that Φ satisfies two properties:

1. For any independent set $U \subseteq V$ in $\mathcal{I}$ there exists a corresponding assortment S_U in $\Phi(\mathcal{I})$ with $\mathcal{R}(S_U) \geq |U|$.

2. Reciprocally, given any assortment S in $\Phi(\mathcal{I})$, we can efficiently construct a corresponding independent set $U_S \subseteq V$ in $\mathcal{I}$ of size at least $\lfloor \mathcal{R}(S) \rfloor$.

Claim 2.2. *For any independent set $U \subseteq V$, the assortment defined by $S_U = \{i : v_i \in U\}$ guarantees that $\mathcal{R}(S_U) \geq |U|$ for the assortment planning instance $\Phi(\mathcal{I})$.*

Proof. We begin by observing that for any vertex $v_i \in U$, the only item made available by S_U within the preference list L_i is product i. To see this, note that L_i consists of the products $N^-(i) \cup \{i\}$, and since U is an independent set, none of v_i's neighbors belongs in U, meaning in particular that $N^-(i) \cap S_U = \emptyset$. Therefore, conditional on the arrival of the customer corresponding to list L_i, the revenue obtained by the assortment S_U is exactly P_i. Thus, we can lower bound the expected revenue due to S_U by

$$\mathcal{R}(S_U) = \sum_{i=1}^n \lambda_i \cdot R_i(S_U) \geq \sum_{i \in S_U} \lambda_i \cdot P_i = \sum_{i \in S_U} \frac{\alpha}{n^{2i}} \cdot \frac{n^{2i}}{\alpha} = |U| \ .$$

∎

Claim 2.3. *For any assortment $S \subseteq [n]$, we can compute in polynomial time an independent set $U_S \subseteq V$ whose cardinality is at least $\lfloor \mathcal{R}(S) \rfloor$.*

Proof. When faced with assortment S, the collection of customers can be partitioned into two groups: Those who purchase their most expensive product, and those who do not. We let $U_S \subseteq [n]$ denote the former subset. By definition, for all $i \in U_S$, customer i purchases product i, which is the most expensive one in L_i. The contribution of this purchase to the expected revenue is therefore $\lambda_i P_i = 1$. On the other hand, the contribution of each customer $i \in [n] \setminus U_S$ to the expected revenue is at most

$$\lambda_i \cdot \max_{j \in N^-(i)} P_j \leq \lambda_i \cdot P_{i-1} = \frac{\alpha}{n^{2i}} \cdot \frac{n^{2(i-1)}}{\alpha} = \frac{1}{n^2} \ .$$

Consequently, the total contribution of the latter customers (of which there are at most n) to the expected revenue is upper bounded by $1/n$. This means that precisely $\lfloor \mathcal{R}(S) \rfloor$ customers generate an expected revenue of 1, and therefore, $|U_S| = \lfloor \mathcal{R}(S) \rfloor$.

We now claim that the vertex set $\{v_i : i \in U_S\}$ forms an independent set in G. Indeed, if $i < j$ are both in U_S and $(v_i, v_j) \in E$, then $i \in N^-(j)$ and v_i is preferred over v_j by the preference list L_j. As a consequence, the contribution of customer j to the expected revenue is strictly less than 1, contradicting the fact that $j \in U_S$. ∎

∎

Additional observations. It is worth noting that the maximum and minimum prices in our reduction, denoted $P_{\max}$ and $P_{\min}$ respectively, satisfy

$$\log\left(\frac{P_{\max}}{P_{\min}}\right) = \log\left(\frac{n^{2n}/\alpha}{n^2/\alpha}\right) = O(n\log n) \ .$$

Therefore, as an immediate corollary, we also obtain an inapproximability bound in terms of $P_{\max}$ and $P_{\min}$.

Corollary 2.4. *Assortment planning under ranking preferences is NP-hard to approximate within $O(\log^{1-\epsilon}(P_{\max}/P_{\min}))$ for any fixed $\epsilon > 0$.*

Finally, as pointed out during the construction of L_i, our reduction does not require a specific order within each preference list, as long as the most expensive product is the least desirable one. As a result, the inapproximability bounds we have just established hold even when all preference lists are derived from a common permutation over the set of products. That is, customer types rank their alternatives consistently with respect to a single permutation.

2.2 Relation to the Min-Buying problem

In the previous reduction, we used distinct selling prices for products, as well as distinct arrival probabilities for customer types. In fact, we constructed assortment planning instances wherein both of these parameters have very large variability. Thus, motivated by practical choice specifications, an interesting question is whether the problem is rendered tractable under a small number of distinct prices, possibly with uniform arrival probabilities.

We resolve this question by proving that, for some constant $\alpha > 0$, assortment planning is NP-hard to approximate within factor better than $1 + \alpha$ even when there are only two distinct selling prices, and preference lists occur according to a uniform distribution. It is worth mentioning that, when all products have identical prices, the problem becomes trivial. Specifically, by selecting all products in the assortment, we ensure that each preference list picks its maximal price item.

Our proof relies on a hardness result obtained by Aggarwal et al. (2004) in the context of multi-product pricing under the Min-Buying choice mode. We begin by formally introducing the latter problem.

An instance of the (uniform) Min-Buying pricing problem can be described as follows. Given a collection of n items, we assume there are k customer types, each of which arrives at random with probability $1/k$. For all $j \in [k]$, customer type j is characterized by a subset of products $S_j \subseteq [n]$ she is willing to purchase and by a budget B_j. She buys the least expensive item in

S_j that meets her budget constraint. The objective is to determine a pricing vector $p \in \mathbb{R}^n_+$ to maximize the expected revenue under a random customer arrival, i.e.,

$$\max_{p \in \mathbb{R}^n_+} \frac{1}{k} \sum_{j=1}^{k} \min \{p_i : i \in S_j \text{ and } p_i \leq B_j\} \ .$$

Aggarwal et al. (2004) proved that the Min-Buying problem is APX-hard even for instances with only two distinct budget values. Thus, the two-budget case of Min-Buying is NP-hard to approximate within $1 + \alpha$, for some constant $\alpha > 0$.

Theorem 2.5. *Assortment planning under ranking preferences is NP-hard to approximate within $1 + \alpha$, for some constant $\alpha > 0$, even with two distinct selling prices and with uniform customer arrival probabilities.*

Proof. In what follows, we construct an efficiently-computable mapping Φ of each instance $\mathcal{I}$ of the Min-Buying problem to an instance $\Phi(\mathcal{I})$ of the assortment planning problem satisfying the next two claims:

1. $\text{OPT}(\Phi(\mathcal{I})) \geq \text{OPT}(\mathcal{I})$.

2. Given any assortment for $\Phi(\mathcal{I})$, we can compute in polynomial time a pricing vector for $\mathcal{I}$ whose expected revenue is at least as good.

These properties jointly imply that our reduction translates the APX-hardness result of Aggarwal et al. (2004) to the assortment planning problem, thus proving the desired claim.

We begin by noting that, without any loss in the expected revenue, any pricing vector of the Min-Buying problem can be transformed into another vector such that the price of each product is identical to the budget of at least one customer type. In other terms, we can restrict the feasible pricing vectors to reside within $\mathcal{B}^\setminus$, where $\mathcal{B} = \{\mathcal{B}_\infty, \ldots, \mathcal{B}_\|\}$.

Given an instance $\mathcal{I}$ of the Min-Buying problem, we define a corresponding assortment planning instance $\Phi(\mathcal{I})$ as follows:

- The collection of products in $\Phi(\mathcal{I})$ is $[n] \times \mathcal{B}$, meaning that each combination of product $i \in [n]$ and price $B \in \mathcal{B}$ is represented by a distinct 'copy' product in $\Phi(\mathcal{I})$.

- There are k customer types with uniform arrival probabilities.

- For every customer j, the preference list L_j is derived from S_j by considering all copies of products in S_j that meet the budget constraint B_j, namely,

$$L_j = \{(i, B) \in [n] \times \mathcal{B} : \rangle \in \mathcal{S}_| \text{ and } \mathcal{B} \leq \mathcal{B}_|\} \ .$$

Here, the preference order in L_j is based on decreasing prices. That is, a less expensive product is always preferred over a more expensive one; when there are ties (equal prices), the relative ranking of products is set arbitrarily.

Proof of Claim 1. Let $p \in \mathcal{B}^n$ be a pricing vector in $\mathcal{I}$. We build an assortment that generates as much revenue in $\Phi(\mathcal{I})$ as the price vector p in $\mathcal{I}$. The idea is to determine an assortment

7

where each customer buys the same combination of price and product as in $\Phi(\mathcal{I})$. Specifically, for each product $i \in [n]$, we select in the assortment product (i, p_i), i.e., the copy of i with price p_i, which is possible since $p_i \in \mathcal{B}$.

We now claim that this assortment generates as much revenue as the pricing vector p in the Min-Buying instance. Indeed, in this assortment, each customer type $j \in [k]$ chooses the least expensive product that intersects his list L_j, noting that ties between products do not have any impact since such products generate identical revenues. By construction of L_j, the purchase price of customer j is thus equal to that of the least expensive product in S_j under pricing p, assuming that the budget constraint is satisfied. We therefore get $\text{OPT}(\Phi(\mathcal{I})) \geq \text{OPT}(\mathcal{I})$. ∎

Proof of Claim 2. Reciprocally, let S be an assortment of the instance $\Phi(\mathcal{I})$. We prove that S can be translated in polynomial time into a pricing vector whose revenue in the Min-Buying instance is at least $\mathcal{R}(S)$. First, let us remark that although several of copies of the same product $i \in [n]$, with different prices, have been selected in S, all customers would only buy the least expensive copy. Indeed, if product i belongs to L_j then any cheaper copy belongs to L_j as well, and customer type j only picks the cheapest. Therefore, we can eliminate from S all redundant copies that are not picked by any customer, and keep only one copy per product. By considering the remaining items, the assortment defines a partial assignment of prices to products: If the copy (i, B), of item i with price B, has been selected – we assign B as the price in $\mathcal{I}$, i.e., set $p_i = B$.

On the other hand, for any product of which no copy has been selected, we set its price to $\max(\mathcal{B})$. We observe that any customer type j in $\mathcal{I}$, under pricing p, would purchase a product whose price is larger than that of the product she purchases in $\Phi(\mathcal{I})$, when faced with the assortment S. Indeed, if she purchases a product of price B in $\mathcal{I}$, then, either there exists $(i, B) \in L_j \cap S$ and customer j purchases a product of price lower than B in $\Phi(\mathcal{I})$, or $B = \max(\mathcal{B})$ and this customer generates a lower revenue in $\Phi(\mathcal{I})$. This yields the desired result. ∎

∎

3 Approximation Algorithms

3.1 Approximation in terms of price ratio

In this section, we show that a natural algorithm, often used by practitioners and proposed in related literature for various models, attains the best-possible approximation ratio up to lower order terms under our general choice model. A revenue-ordered assortment consists in selecting all products whose price is greater or equal to a given threshold (Talluri and Van Ryzin 2004, Rusmevichientong et al. 2014). In what follows, we use S_p to designate the revenue-ordered assortment corresponding to a minimum price of p, i.e., $S_p = \{i \in [n] : P_i \geq p\}$. As the next theorem shows, by limiting attention to such assortments and selecting the one with largest expected revenue, we are able to match the inapproximability bound established in Corollary 2.4.

Theorem 3.1. *The optimal revenue-ordered assortment approximates the optimal expected revenue within factor $O(\lceil \ln(P_{\max}/P_{\min}) \rceil)$.*

8

Proof. Without loss of generality, we may assume that empty preference lists have been discarded, and that the remaining arrival probabilities sum up to 1. Indeed, this can be achieved by renormalizing the distribution, which results in multiplying the expected revenue of any assortment by the same constant.

Let OPT designate the expected revenue obtained by the optimal assortment. For each customer $j \in [k]$, we define a corresponding budget B_j as the highest price on his list, i.e., $B_j = \max_{i \in L_j} P_i$. Without loss of generality, we can assume that customer indices are arranged so that $B_1 \geq \cdots \geq B_k$. Finally, we define $j^* \in [k]$ to be the customer j for which $B_j \cdot \sum_{r=1}^{j} \lambda_r$ is maximized, picking j^* arbitrarily, when the maximum value is attained by two or more customers.

We proceed by considering the assortment $S_{B_{j^*}}$, formed by all products whose price is greater or equal to B_{j^*}. Since $B_1 \geq \cdots \geq B_k$, any preference list in $[j^*]$ contains at least one product with a per-selling price of at least B_{j^*}. As a result, any such preference list generates a revenue greater or equal to B_{j^*} when faced with the assortment $S_{B_{j^*}}$, and therefore,

$$\mathcal{R}(S_{B_{j^*}}) = \sum_{j=1}^{k} \lambda_j \cdot R_j(S_{B_{j^*}}) \geq B_{j^*} \cdot \sum_{r=1}^{j^*} \lambda_r \ . \tag{1}$$

In order to relate this quantity to OPT, we define

$$u^* = \min \left\{ u \in [k] : \sum_{j=1}^{u} \lambda_j \geq \frac{1}{2} \cdot \frac{P_{\min}}{P_{\max}} \right\} \ ,$$

noting that u^* is well defined, since $\sum_{j=1}^{k} \lambda_j = 1$. By remarking that B_j corresponds to the maximal revenue that can be extracted from each customer type j, we can upper bound the optimal expected revenue by

$$\text{OPT} \leq \sum_{j=1}^{k} \lambda_j \cdot B_j \leq \sum_{j=1}^{u^*-1} \lambda_j \cdot B_j + \lambda_{u^*} \cdot B_{u^*} + \sum_{j=u^*+1}^{k} \lambda_j \cdot B_j \ . \tag{2}$$

By definition of u^*, the first sum on the right is upper bounded by $B_1 \cdot P_{\min}/(2P_{\max}) \leq P_{\min}/2$. For the middle term, Equation (1) implies in particular that $\lambda_{u^*} \cdot B_{u^*} \leq \mathcal{R}(S_{B_{j^*}})$. Finally, we can upper-bound the last sum as follows:

$$\sum_{j=u^*+1}^{k} \lambda_j \cdot B_j = \sum_{j=u^*+1}^{k} \frac{\lambda_j}{\sum_{r=1}^{j} \lambda_r} \cdot \left(B_j \cdot \sum_{r=1}^{j} \lambda_r \right)$$

$$\leq \sum_{j=u^*+1}^{k} \frac{\lambda_j}{\sum_{r=1}^{j} \lambda_r} \cdot \left(B_{j^*} \cdot \sum_{r=1}^{j^*} \lambda_r \right)$$

$$\leq \sum_{j=u^*+1}^{k} \frac{\lambda_j}{\sum_{r=1}^{j} \lambda_r} \cdot \mathcal{R}(S_{B_{j^*}})$$

$$= \sum_{j=u^*+1}^{k} \left(\int_{\sum_{r=1}^{j-1} \lambda_r}^{\sum_{r=1}^{j} \lambda_r} \frac{1}{\sum_{r=1}^{j} \lambda_r} dx \right) \cdot \mathcal{R}(S_{B_{j^*}}) \ ,$$

where the first inequality follows from the definition of j^*, and the second inequality is derived from Equation (1). By the monotonicity of $x \mapsto \frac{1}{x}$, we obtain:

$$
\begin{aligned}
\sum_{j=u^*+1}^{k} \lambda_j \cdot B_j &\leq \sum_{j=u^*+1}^{k} \left(\int_{\sum_{r=1}^{j-1} \lambda_r}^{\sum_{r=1}^{j} \lambda_r} \frac{1}{x} dx \right) \cdot \mathcal{R}(S_{B_{j^*}}) \\
&= \left(\int_{\sum_{r=1}^{u^*} \lambda_r}^{1} \frac{1}{x} dx \right) \cdot \mathcal{R}(S_{B_{j^*}}) \\
&\leq \left(\int_{\frac{1}{2} \cdot \frac{P_{\min}}{P_{\max}}}^{1} \frac{1}{x} dx \right) \cdot \mathcal{R}(S_{B_{j^*}}) \\
&= \ln \left(2 \cdot \frac{P_{\max}}{P_{\min}} \right) \cdot \mathcal{R}(S_{B_{j^*}}) \ ,
\end{aligned}
$$

where the second inequality follow from the definition of u^*.

As a result, we can now infer from inequality (2) that the assortment $S_{B_{j^*}}$ indeed approximates the optimal expected revenue within factor $O(\lceil \ln(P_{\max}/P_{\min}) \rceil)$, since

$$
\begin{aligned}
\text{OPT} &\leq \frac{P_{\min}}{2} + \left(1 + \ln \left(2 \cdot \frac{P_{\max}}{P_{\min}} \right) \right) \cdot \mathcal{R}(S_{B_{j^*}}) \\
&\leq \left(\frac{3}{2} + \ln \left(2 \cdot \frac{P_{\max}}{P_{\min}} \right) \right) \cdot \mathcal{R}(S_{B_{j^*}}) \\
&\leq \frac{5}{2} \cdot \left\lceil \ln \left(\frac{P_{\max}}{P_{\min}} \right) \right\rceil \cdot \mathcal{R}(S_{B_{j^*}}) \ .
\end{aligned}
$$

Here, the second inequality is obtained by observing that $P_{\min} \leq \mathcal{R}(S_{B_{j^*}})$, since by the choice of j^* and by our initial assumption that all empty lists have been eliminated, we have

$$
\mathcal{R}(S_{B_{j^*}}) \geq B_{j^*} \cdot \sum_{j=1}^{j^*} \lambda_j \geq B_k \cdot \sum_{j=1}^{k} \lambda_j \geq P_{\min} \ .
$$

∎

As a corollary, we prove that revenue-ordered assortment also achieve an approximation ratio of $O(\lceil \log(1/\tilde{\lambda}) \rceil)$, where $\tilde{\lambda}$ denotes the combined arrival probability of all customers who have the highest price item on their list. In particular, when all arrival probabilities are polynomially bounded away from 0, i.e. $\Omega(1/\text{poly}(k))$, this bound translates to an $O(\log k)$ approximation (for example, under a uniform distribution).

Corollary 3.2. *The assortment planning problem under ranking preferences can be approximated within factor* $O(\lceil \log(1/\tilde{\lambda}) \rceil)$.

Proof. Proof. We prove that, when all products with price smaller than $(\tilde{\lambda}/2) \cdot P_{\max}$ are eliminated, there is still an assortment that generates an expected revenue of at least OPT/2. This transformation guarantees that all remaining prices are within factor $2/\tilde{\lambda}$ of each other, in which case the upper bound given in Theorem 3.1 becomes $O(\lceil \log(1/\tilde{\lambda}) \rceil)$.

Let $\bar{S}$ designate the subset of products that have been eliminated, i.e., $\bar{S} = \{i \in [n] : P_i \leq (\tilde{\lambda}/2) \cdot P_{\max}\}$. When we eliminate products from an assortment, the probability that a

customer purchases each of the remaining products (and consequently, the expected revenue from the remainder selection) can only increase. For this reason, it is sufficient to consider the contribution of $\bar{S}$ to the expected revenue of the optimal assortment, which can be upper bounded by

$$\sum_{j=1}^{k} \lambda_j \cdot R_j(\bar{S}) \leq \sum_{j=1}^{k} \lambda_j \cdot \frac{\tilde{\lambda}}{2} \cdot P_{\max} = \frac{\tilde{\lambda}}{2} \cdot P_{\max} \leq \frac{\text{OPT}}{2} \ ,$$

where the last inequality holds since $\text{OPT} \geq \tilde{\lambda} \cdot P_{\max}$. Indeed, this is the expected revenue of the assortment formed by stocking only the highest price product. ∎

3.2 Approximation in terms of list length

A close inspection of our reduction from Max-IS (see Theorem 2.1) reveals that the maximal size of any preference list was equivalent to the maximal degree Δ in the original graph. As a consequence, this inapproximability result gives an $O(\Delta^{1-\epsilon})$ hardness for assortment planning with preference lists of size at most Δ. Since there are numerous algorithms for approximating Max-IS in terms of Δ (Karger et al. 1998, Alon and Kahale 1998, Halperin 2002), it is natural to investigate whether improved approximation guarantees can be obtained in terms of the maximum length of any list. In fact, the underlying assumption that each preference list is comprised of relatively few products finds behavioral and empirical support, and subsumes practical choice modeling specifications (Hauser et al. 2009).

In this setting, we analyze the expected revenue of random assortments arising from an appropriate generative distribution. By derandomization, we obtain a polynomial-time algorithm that is asymptotically tight, as asserted by the following theorem.

Theorem 3.3. *The assortment planning problem under ranking preferences can be approximated within factor $e\Delta$, where Δ is the maximal size of a preference list.*

Proof. For any customer type j, let $M(j)$ be the item with maximal price within the preference list L_j. The optimal expected revenue is naturally bounded by

$$\text{OPT} \leq \sum_{j=1}^{k} \lambda_j \cdot P_{M(j)} \ .$$

We construct a random assortment S_X through the following procedure: First, we independently draw values for $X_1, \ldots, X_n$, which are n i.i.d. Bernoulli variables with probability of success $1/\Delta$. Then, we pick each product to the assortment if and only if its corresponding variable is successful, meaning that $S_X = \{i \in [n] : X_i = 1\}$.

The important observation is that, for any preference list L_j, the probability that customer type j would purchase product $M(j)$ when faced with the assortment S_X is at least

$$\frac{1}{\Delta} \cdot \left(1 - \frac{1}{\Delta}\right)^{|L_j|-1} \geq \frac{1}{\Delta} \cdot \left(1 - \frac{1}{\Delta}\right)^{\Delta-1} \geq \frac{1}{e\Delta} \ ,$$

where the last inequality holds since the function $[x \mapsto (1 - 1/x)^{x-1}]$ is monotone-decreasing over $(1, \infty)$, and converges to $1/e$. Indeed, this is precisely the probability that $M(j)$ belongs to

S_X, and that all other products in L_j are unavailable. We conclude that the expected revenue of S_X is

$$\mathbb{E}_X \left[\sum_{j=1}^{k} \lambda_j \cdot R_j(S_X) \right] = \sum_{j=1}^{k} \lambda_j \cdot \mathbb{E}_X [R_j(S_X)] \geq \frac{1}{e\Delta} \cdot \sum_{j=1}^{k} \lambda_j \cdot P_{M(j)} \geq \frac{1}{e\Delta} \cdot \text{OPT} \ .$$

This algorithm can be derandomized through the method of conditional expectations (see, for example, Chapter 16.1 in Alon and Spencer (2004)). Indeed, conditional on any partial assortment, i.e., a sequence of fixed binary values for the variables $X_1, \ldots, X_\ell$, the expected revenue can be computed exactly in polynomial time. Specifically, the independence between the Bernoulli variables allows to compute the probability that each customer type picks a given product in his list. By applying the method of conditional expectations iteratively over $\ell = 1, \ldots, n$, we retrieve a deterministic assortment that approximates OPT within factor $e\Delta$. ∎

4 Concluding Remarks

Cardinality constraints. From a technical point of view, the approximation algorithms we propose in Section 3 make use of the freedom in picking assortments of any possible cardinality. An interesting direction for future research is to investigate whether our algorithms can be extended to the capacitated setting, where at most C distinct products can be stocked. Results in this spirit have previously been attained for several tractable models (see, for instance, Rusmevichientong et al. (2010), Davis et al. (2013)), although the computational difficulties here appear to be of significantly different nature.

Specification of the choice model. A particularly desirable property of revenue-ordered assortments is that an explicit description of the preference list distribution is not required, as long as one has access to an efficient oracle for computing the expected revenue of any given assortment. Therefore, the approximation guarantees we provide in Section 3.1 extend to a broader class of random utility choice models, where the distribution over preference lists potentially has a large support, such as Mixture of Multinomial Logits (Bront et al. 2009, Méndez-Díaz et al. 2010, Rusmevichientong et al. 2014, Désir et al. 2014, Feldman and Topaloglu 2015).

Uniform distribution. An interesting open question is that of determining the best approximation possible for uniform preference list distributions, i.e., when each customer type is picked with equal probability. Such models are of practical importance, since in many applications, the distribution probabilities are conditioned by the number of samples used to estimate the model parameters. For this special case, one could try to narrow the gap between our APX-hardness results, given in Theorem 2.5, and the $O(\log k)$ approximation that follows from Corollary 3.2.

References

Aggarwal, G., Feder, T., Motwani, R. and Zhu, A. (2004), Algorithms for multi-product pricing, *in* 'Proceedings of the 31st International Colloquium on Automata, Languages and Programming', pp. 72–83.

Alon, N. and Kahale, N. (1998), 'Approximating the independence number via the theta-function', *Mathematical Programming* **80**, 253–264.

Alon, N. and Spencer, J. H. (2004), *The probabilistic method*, John Wiley & Sons.

Aouad, A., Farias, V., Levi, R. and Segev, D. (2015), 'The approximability of assortment optimization under ranking preferences', Working paper. Available at SSRN 2612947 (June 3rd, 2015).

Berbeglia, G. and Joret, G. (2015), 'Assortment optimisation under a general discrete choice model: A tight analysis of revenue-ordered assortments', *Working paper. Available at SSRN 2620165 (June 19th, 2015)* .

Blanchet, J. H., Gallego, G. and Goyal, V. (2016), 'A markov chain approximation to choice modeling', *Operations Research* **64**(4), 886–905.

Bront, J. J. M., Méndez-Díaz, I. and Vulcano, G. (2009), 'A column generation algorithm for choice-based network revenue management', *Operations Research* **57**(3), 769–784.

Davis, J. M., Gallego, G. and Topaloglu, H. (2013), 'Assortment planning under the multinomial logit model with totally unimodular constraint structures', Working Paper.

Davis, J. M., Gallego, G. and Topaloglu, H. (2014), 'Assortment optimization under variants of the nested logit model', *Operations Research* **62**(2), 250–273.

Désir, A., Goyal, V. and Zhang, J. (2014), 'Near-optimal algorithms for capacity constrained assortment optimization'. Working paper. Available at SSRN 2543309.

Farias, V. F., Jagabathula, S. and Shah, D. (2013), 'A nonparametric approach to modeling choice with limited data', *Management Science* **59**(2), 305–322.

Feldman, J. B. and Topaloglu, H. (2015), 'Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits', *Production and Operations Management* **24**(10), 1598–1620.

Feldman, J. and Topaloglu, H. (2017), 'Capacitated assortment optimization under the multinomial logit model with nested consideration sets'. Operations Research (forthcoming).

Gallego, G. and Topaloglu, H. (2014), 'Constrained assortment optimization for the nested logit model', *Management Science* **60**(10), 2583–2601.

Goyal, V., Levi, R. and Segev, D. (2016), 'Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand', *Operations Research* **64**(1), 219–235.

Halperin, E. (2002), 'Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs', *SIAM Journal on Computing* **31**(5), 1608–1623.

Håstad, J. (1996), Clique is hard to approximate within $n^{1-\epsilon}$, *in* 'Proceedings of the 37th Annual Symposium on Foundations of Computer Science', pp. 627–636.

Hauser, J. R., Ding, M. and Gaskin, S. P. (2009), 'Non compensatory (and compensatory) models of consideration-set decisions', Sawtooth Software Conference Proceedings.

Honhon, D., Jonnalagedda, S. and Pan, X. A. (2012), 'Optimal algorithms for assortment selection under ranking-based consumer choice models', *Manufacturing & Service Operations Management* **14**(2), 279–289.

Honhon, D., Pan, X. and Jonnalagedda, S. (2015), Heuristics for the assortment planning problem under ranking-based consumer choice models, Technical report, Working paper, Naveen Jindal School of Management, UT Dallas, TX.

Karger, D. R., Motwani, R. and Sudan, M. (1998), 'Approximate graph coloring by semidefinite programming', *Journal of The ACM* **45**(2), 246–265.

Li, G., Rusmevichientong, P. and Topaloglu, H. (2015), 'The d-level nested logit model: Assortment and price optimization problems', *Operations Research* **63**(2), 325–342.

Mahajan, S. and Van Ryzin, G. (2001), 'Stocking retail assortments under dynamic consumer substitution', *Operations Research* **49**(3), 334–351.

Méndez-Díaz, I., Miranda-Bront, J. J., Vulcano, G. and Zabala, P. (2010), 'A branch-and-cut algorithm for the latent class logit assortment problem', *Electronic Notes in Discrete Mathematics* **36**, 383–390.

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

Rusmevichientong, P., Shmoys, D., Tong, C. and Topaloglu, H. (2014), 'Assortment optimization under the multinomial logit model with random choice parameters', *Production and Operations Management* **23**(11), 2023–2039.

Rusmevichientong, P., Van Roy, B. and Glynn, P. W. (2006), 'A nonparametric approach to multiproduct pricing', *Operations Research* **54**(1), 82–98.

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

van Ryzin, G. and Vulcano, G. (2014), 'A market discovery algorithm to estimate a general class of nonparametric choice models', *Management Science* **61**(2), 281–300.

van Ryzin, G. and Vulcano, G. (2017), 'An expectation-maximization method to estimate a rank-based choice model of demand', *Operations Research* **65**(2), 396–407.