Technical Note: Assortment Optimization with Small Consideration Sets

Jacob Feldman
Olin Business School, Washington University,
Saint Louis, MO 63108, USA
jbfeldman@wustl.edu

Alice Paul
Data Science Initiative, Brown University,
Providence, RI 02912, USA
alice_paul@brown.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech,
New York, NY 11011, USA
topaloglu@orie.cornell.edu

July 31, 2018

Abstract

We study a customer choice model that captures purchasing behavior when there is a limit on the number of times that a customer will substitute among the offered products. Under this model, we assume each customer is characterized by a ranked preference list of products and, upon arrival, will purchase the highest ranking offered product. Since we restrict ourselves to settings in which customers consider a limited number of products, we assume that these rankings contain at most k products. We call this model the k-product nonparametric choice model. We focus on the assortment optimization problem under this choice model. In this problem, the retailer wants to find the revenue maximizing set of products to offer when the buying process of each customer is governed by the k-product nonparametric choice model. First, we show that the assortment problem is strongly NP-hard even for $k = 2$. Motivated by this result, we develop a linear programming-based randomized rounding algorithm that gives the best known approximation guarantee. We tighten the approximation guarantee further when each preference list contains at most two products and consider the case where there is a limit on the number of products that can be offered to the customers.

Keywords: Assortment Optimization, Customer Choice Models, Linear Programming, Approximation Algorithms.

1 Introduction

The ability to accurately model customer demand is critical for any retailer since this model will guide everything from inventory decisions to pricing to promotion strategies. Traditionally, the demand for a product is assumed to be independent of the availability of other products. This assumption fails to capture a typical behavior of customers commonly referred to as substitution: the phenomenon when a customer chooses an alternative product when her most preferred product is not available for purchase due to stock-outs or deliberate inventory controls. When customers substitute, the demand for a particular product should be modeled as a function of the availability of all other substitutable alternatives. In the revenue management literature, models capturing substitution patterns are known as *customer choice models.* A customer choice model maps any assortment of products to the probability that each product in the assortment is purchased. There are a variety of customer choice models, each of which captures different aspects of typical customer substitution behavior. Choosing an appropriate choice model is often a difficult task for the retailer as there is a trade-off between model accuracy and tractability; the choice models that capture the most general forms of customer behavior are often those whose underlying parameters are difficult to estimate and whose corresponding assortment problems are difficult to solve.

One customer choice model to gain attention recently is the nonparametric ranking-based choice model dating back to Mahajan and van Ryzin (2001*a*) and Mahajan and van Ryzin (2001*b*). In the full nonparametric ranking-based choice model, each customer class is distinguished by an arrival probability and a unique ranking on a subset of products referred to as a preference list. When presented with an assortment of products, a customer will purchase the highest ranking offered product in her preference list, and if there is no offered product in her preference list, then she leaves without making a purchase. As stated, this model places no restrictions on the set of potential preference lists, and hence, the number of customer classes grows exponentially in the number of products.

In contrast, we limit all preference lists to be of length at most k. We refer to this choice model as the k-product nonparametric choice model. By restricting the length of preference lists, we capture settings in which customers have small consideration sets, where the consideration set of a customer is simply the set of products they would ever consider purchasing. In other words, customers are only willing to substitute a limited number of times.

There are numerous papers in the marketing and revenue management literature that provide evidence for the existence of limited substitution in settings in which there is either a high-purchasing bias or a low cost of leaving the system. Lapersonne et al. (1995) study customers considering an automobile purchase and find that a large percentage of customers will only consider the brand of their previous car. This sort of customer behavior is also described in Hauser et al. (2009), in which the authors emphasize that for frequently purchased products, customers often only consider a handful of brands. Further, Jeuland (1979) studies a purchasing bias which is referred to as short term brand loyalty or inertia in choice, which captures customers continuing to buy the products or brands they have previously purchased. Crompton and Ankomah (1993) study how travelers pick vacation destinations. They claim that in this setting "consideration sets have size at most four". Finally, Meissner et al. (2013) and Talluri (2014) argue that customers have small consideration sets when purchasing airline tickets due to the fact that there are generally a limited number of itineraries that satisfy a given travelers ideal schedule.

Through the k-product nonparametric choice model, we directly capture the trade-off between higher modeling accuracy and tractability; as k increases we recapture the full nonparametric model but the associated estimation and optimization problems become harder. When k is reasonably small, we show that we can develop algorithms for the assortment problem that give good theoretical guarantees and perform well in practice. Further, we show that estimating the arrival probabilities for relatively small values of k is efficient.

Contributions. We first prove that the assortment optimization problem under the k-product nonparametric choice model is strongly NP-hard even when $k = 2$ and the set of preference lists is derived from a single ordering of the products, meaning that products can be indexed such that a product with a lower index is never ranked below a product with a higher index in any preference list. We show this result via a reduction from the vertex cover problem on cubic graphs.

Motivated by this hardness result, we focus on developing approximation algorithms for the assortment optimization problem under the k-product nonparametric choice model. This optimization problem can be formulated as an integer program with binary decision variables representing whether or not each product is offered. The algorithm that we present considers rounding the values of the optimal decision variables in the linear programming (LP) relaxation of this integer program. Specifically, we offer each product i independently with a probability that is increasing in its corresponding optimal decision variable value and decreasing in k. A worst-case analysis of our approximation algorithm gives a performance guarantee of $2(1 - \frac{1}{k})^{k-1}(\frac{1}{k})$, which improves upon the previously best known guarantee given by Aouad, Farias, Levi and Segev (2015) by a factor of two. When $k = 2$, this yields a 1/2-approximation for the assortment optimization problem. Further, for $k = 3$ or $k = 4$, the LP-based rounding algorithm has performance guarantees of 0.296 and 0.210, respectively. Using standard techniques, our LP-based rounding algorithm can easily be derandomized to produce a deterministic assortment.

For the case when $k = 2$, we are able to improve upon the performance of our LP-based rounding algorithm. In particular, we are able to reduce the assortment optimization problem under the 2-product nonparametric choice model to a maximum directed cut problem. This reduction yields a 0.874-approximation algorithm that rounds the optimal solution to a semidefinite programming relaxation of the problem. Further, when $k = 2$, we show how to modify our LP-based rounding algorithm to the capacitated version of the problem, in which the retailer can only offer a limited number of products.

In order to test the efficacy and practicality of our LP-based rounding algorithm, we run a series of computational experiments on a large collection of instances of the assortment problem under the k-product nonparametric choice model. We compute the optimality gap of the expected revenue of the assortment produced by the LP-based rounding performs by comparing this revenue to a tractable upper bound on the optimal expected revenue. We find that the average percent optimality gap never exceeds 1%, far exceeding the worst-case theoretical performance guarantees. Further, we compare the performance of our LP-based rounding algorithm against the randomized approach of Aouad, Farias, Levi and Segev (2015). We show that our LP-based rounding algorithm provides noticeable improvements over this other approach both in terms of average and worst-case performance over all test cases.

We conclude with another set of computational experiments that study the marginal benefit of fitting increasingly complex nonparametric choice models. We generate synthetic sales data from a nonparametric choice model that captures customers making purchasing decision from a set of vertically differentiated products, meaning the products can be ordered based on some quality metric. In generating this sales data, we vary the number of customer types, the degree to which customers obey the quality ranking, and the average length of the preference lists in the underlying customer population. We fit k-product nonparametric choice models for $k \in \{1, 2, 3, 4\}$ to study the trade-off between the improved accuracy that results from increasing k and the computational effort required to fit these richer nonparametric choice models. We find that the k-product nonparametric choice model for small values of k captures the customer choice process even when, in reality, customers potentially substitute among a larger number of products.

Related Literature. To the best of our knowledge, the first paper to consider assortment optimization under a nonparametric model was Honhon et al. (2012), which also considers a restricted set of potential preferences lists. They provide an optimal algorithm for the assortment problem when the products are nodes in a tree, and each potential preference list can be represented by a path in the tree starting or ending at the root node. This result is extended in Paul et al. (2016), in which the authors work in a more general tree setting in which preference lists are still paths in a tree, but these paths can begin or end anywhere in the tree. Both of these papers solve the assortment optimization problem via dynamic programming methods that take advantage of the special structure of the potential set of preference lists.

Two other papers that study the assortment problem under the nonparametric choice model are

Aouad, Farias, Levi and Segev (2015) and Aouad, Farias and Levi (2015). The former proves various hardness results related to the assortment problem. Specifically, the authors prove that it is NP-hard to find a $O(1/k^{1-\varepsilon})$-approximation algorithm for the assortment optimization problem under the k-product nonparametric choice model. The latter considers the assortment optimization problem under the nonparametric choice model when customer preference lists are associated with structured set systems defined over a single overarching ordering of the products. One such structured set system is a laminar family. In an extended setting, Jagabathula and Rusmevichientong (2016) presents an approximation scheme for the joint assortment and pricing problem in which the retailer chooses both the set of products to offer and the prices for these products with the intention of maximizing expected revenue. In this problem setting, each customer class is distinguished by a price threshold and a preference list. Once prices have been set, customers purchase the highest ranking offered product that is priced below their price threshold.

There have also been a few papers that specifically study the assortment optimization problem under the k-product nonparametric model. In particular, Bertsimas and Mišić (2017) presents an integer program (IP) for this problem, but the authors do not provide an approach to find an approximate solution with a performance guarantee in polynomial time. In addition to the various hardness results, Aouad, Farias, Levi and Segev (2015) shows that randomly offering each product with probability $1/k$ produces a $1/(e \cdot k)$-approximation algorithm. To our knowledge, this is the best previously known guarantee for this problem. The 2-product nonparametric model also closely resembles the substitution model outlined in Kök and Fisher (2007). In this paper, the single substitution event that occurs when a customer's first choice product is unavailable unfolds in two steps. First, the customer decides whether to consider a second product or to leave the store. If she chooses to consider a second product, then she substitutes into one of the other products with a probability that is dependent on the customer's first choice product. The authors give heuristics for the associated assortment optimization problems.

Assortment optimization has also been thoroughly studied under other choice models. Talluri and van Ryzin (2004) solve the assortment optimization problem when customers choose according to the multinomial logit (MNL) choice model. Extending this result, Rusmevichientong et al. (2010), Wang (2012), Davis et al. (2013), and Wang (2013) study various versions of the constrained assortment problems under the MNL model. Méndez-Díaz et al. (2014), Désir and Goyal (2014), and Rusmevichientong et al. (2014) focus on assortment problems when customer choices are governed by a mixture of multinomial logit models. Li and Rusmevichientong (2014), Davis et al. (2014), and Li et al. (2015) develop efficient solution methods for the unconstrained assortment problem when customers choose under the nested logit model. Gallego and Topaloglu (2014) and Feldman and Topaloglu (2015*a*) consider the constrained versions of the assortment problem when customers choose according to the nested logit model. More recently, Blanchet et al. (2016) introduces the Markov chain choice model with additional work on this model by Feldman and Topaloglu (2015*b*) and Désir et al. (2015).

Organization. The rest of the paper is organized as follows. In Section 2, we introduce the k-product nonparametric choice model and we show that the assortment problem is strongly NP-hard. In Section 3, we present the LP-based rounding algorithm for the general k-product nonparametric model. In Section 4, we extend these results for the setting in which $k = 2$. We numerically test the performance of our LP-based rounding algorithm in Section 5 to show that the practical performance is in fact much better than the theoretical guarantees. We then provide our estimation results in Section 6. Lastly, in Section 7, we conclude and discuss how future work can extend our results.

2 Problem Formulation and Complexity

To formulate the assortment optimization problem, we let the set of potential products be $N = \{1, 2, \ldots, n\}$ and let $r_j \geq 0$ be the revenue of product j for $j \in N$. We refer to a product as offered if it is available for purchase and not offered if not. We encode the offered assortment through the vector $x = (x_1, \ldots, x_n) \in \{0, 1\}^n$, where $x_j = 1$ indicates that product j is offered and $x_j = 0$ indicates that it is not offered.

Under the k-product nonparametric choice model, we let $\mathcal{G}$ denote the set of potential customer classes, where an arriving customer is of class $g \in \mathcal{G}$ with probability λ_g. Further, each customer of class g is associated with a ranking σ_g on a subset $S_g \subseteq N$ of products representing their consideration set. The k-product nonparametric choice model represents the setting in which all customers are willing to consider at most k products. Therefore, we assume that $|S_g| = k \leq n$ for all $g \in \mathcal{G}$. The equality assumption can be accomplished by adding in dummy products with zero revenue for customers who consider less than k products, i.e. $|S_g| < k$. For $i \in S_g$, we let $\sigma_g(i)$ give the index of product i in customer class g's preference list, and for $1 \leq j \leq k$, we let g_j be customer class g's jth most preferred product. We use the convention that lower indexed products have a higher ranking, i.e. a product that is first in a customer's preference list has the highest ranking.

An arriving customer will purchase her highest ranked offered item in her consideration set, if any. In particular, if the retailer offers an assortment given by $x \in \{0,1\}^n$, then a customer of type g will purchase product

$$\pi_g(x) := \begin{cases} \arg\min_{i \in S_g, x_i = 1} \sigma_g(i) & \text{if } S_g \cap \{i : x_i = 1\} \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases}$$

where $\pi_g(x) = 0$ indicates that the customer leaves the system without making a purchase. For assortment $x \in \{0,1\}^n$, the probability product j is purchased is given by

$$\text{Pr}_j(x) = \sum_{i=1}^{k} \sum_{g \in \mathcal{G}: g_i = j} \lambda_g (1 - x_{g_1})(1 - x_{g_2}) \ldots (1 - x_{g_{i-1}}) x_j. \tag{1}$$

Given this notation, we denote the expected revenue of a given assortment $x \in \{0,1\}^n$ as $\text{Rev}(x) = \sum_{j \in N} r_j \text{Pr}_j(x)$. Our goal is to find the assortment of products that maximizes the expected revenue from an arriving customer, yielding the problem

$$\text{OPT} = \max_{x \in \{0,1\}^n} \text{Rev}(x). \tag{2}$$

We first provide a hardness result for the assortment optimization problem under the k-product nonparametric choice model. As mentioned previously, Aouad, Farias, Levi and Segev (2015) shows that it is NP-hard to approximate the k-product nonparametric choice model within a factor that is strictly better than $O(1/k)$. We augment this result by showing that even in the simplest case in which $k = 2$ the assortment problem is strongly NP-hard. Further, we prove the result in a setting in which the preference lists all follow some universal ordering, meaning that the products can be indexed such that for any $g \in \mathcal{G}$ such that $i, j \in S_g$ we have that $\sigma_g(i) < \sigma_g(j)$ if $i < j$. Our proof, shown in Appendix A.1, involves a reduction from the minimum vertex cover problem on cubic graphs, which is a well known APX-hard problem (see Alimonti and Kahn (2000)).

Theorem 2.1. *Unless P=NP, there does not exist a fully polynomial-time approximation scheme (FPTAS) for the assortment problem under the 2-product nonparametric choice model even when all of the preference lists are derived from a single ordering.*

Motivated by this hardness result, we turn our focus to finding an efficient approximation algorithm for this problem. In the next section, we develop a LP-based rounding algorithm for the k-product nonparametric choice model.

3 LP-Based Rounding Algorithm

In this section, we present the LP-based rounding algorithm for problem (2). We can formulate this optimization problem as an IP. For each product $i \in N$, let $x_i \in \{0,1\}$ be a variable representing whether or not we offer product i. Further, for each customer class $g \in \mathcal{G}$ and $j \in \{1,2,\ldots,k\}$, let $y_{g,g_j} \in \{0,1\}$ be a variable representing whether or not a customer of class g purchases product g_j, the jth preferred product for customers in class g. Then, the assortment optimization problem can be formulated as the IP

$$\begin{aligned}
\text{OPT} = \max \sum_{g} \sum_{j=1}^{k} & \lambda_g r_{g_j} y_{g,g_j} && \text{(IP1)} \\
\text{s.t.}\ \ y_{g,g_j} &\le x_{g_j} && \forall g \in \mathcal{G}, j \in \{1,2,\ldots,k\} \\
y_{g,g_j} &\le 1 - x_{g_i} && \forall g \in \mathcal{G}, j \in \{2,\ldots,k\}, i \in \{1,\ldots,j-1\} \\
y_{g,g_j} &\ge 0 && \forall g \in \mathcal{G}, j \in \{1,2,\ldots,k\} \\
x_i &\in \{0,1\} && \forall i \in N.
\end{aligned}$$

The objective function calculates the total expected revenue. The first constraint states that a customer can only purchase an offered product, and the second constraint states that if a customer's ith preferred product is available then she cannot purchase a less preferred product. Together, these constraints imply that each customer class can buy at most one product and hence there is no need to explicitly enforce $\sum_{j=1}^{k} y_{g,g_j} \le 1 \ \forall\ g \in \mathcal{G}$. We obtain the LP relaxation of this IP by replacing the constraints $x_i \in \{0,1\}$ with $x_i \in [0,1]$. Let (x^*, y^*) be an optimal solution to the LP relaxation of the above IP and let z^* be the optimal objective value. Note that any optimal basic feasible solution satisfies $y^*_{g,g_j} = \min(x^*_{g_j}, 1 - x^*_{g_1}, \ldots, 1 - x^*_{g_{j-1}})$. In the following lemma, we show that the optimal basic feasible solutions for this LP turn out to be half-integral. The proof appears in Appendix A.2. We will use this structure later in developing the approximation guarantee for the LP-based rounding algorithm that we present next.

Lemma 3.1. *Let (x^*, y^*) be any optimal basic solution to the LP relaxation of (IP1). Then this solution is half-integral in x^*, i.e. $x^*_i \in \{0, \frac{1}{2}, 1\}$ for all $i = 1, \ldots, n$, which implies that y^* is also half-integral.*

We use the LP solution to inform our decisions about which products to offer by constructing a random assortment $\bar{x}$ from the optimal decision variables. For any product i such that $x^*_i \in \{0,1\}$, we set $\bar{x}_i = x_i^*$. On the other hand, if $0 < x_i^* < 1$, we offer product i independently with probability $\alpha + \beta x_i^*$, where $0 \leq \alpha + \beta/2 \leq 1/2$. That is, we set $\bar{x}_i = 1$ with probability $\alpha + \beta x_i^*$, and set $\bar{x}_i = 0$ with probability $1 - (\alpha + \beta x_i^*)$. Since (x^*, y^*) is half-integral, this is equivalent to offering product i with probability $\alpha + \beta/2$. Through the analysis of the algorithm's performance guarantee, we will be able to show how to set α and β optimally. Our goal is to compare the expected revenue of the assortment derived from our randomized algorithm with z^*, the optimal objective value of the LP relaxation for the assortment problem. We make this comparison term by term by comparing the expected revenue of $\bar{x}$ for a single customer class g and product g_j to the LP value y^*_{g,g_j} as illustrated by the following lemma.

Lemma 3.2. *For a customer class $g \in \mathcal{G}$ and $j \in \{1, 2, \ldots, k\}$,*

$$\mathbb{E}[\lambda_g r_{g_j}(1 - \bar{x}_{g_1})(1 - \bar{x}_{g_2}) \ldots (1 - \bar{x}_{g_{j-1}})\bar{x}_{g_j}] \geq \gamma_j(\alpha, \beta)\lambda_g r_{g_j} y^*_{g,g_j},$$

where

$$\gamma_j(\alpha, \beta) = 2\left(1 - (\alpha + \beta/2)\right)^{j-1}(\alpha + \beta/2).$$

Proof. From Lemma 3.1, we know that $y^*_{g,g_j} \in \{0, \frac{1}{2}, 1\}$. When $y^*_{g,g_j} = 0$, the theorem is trivially true. In the case in which $y^*_{g,g_j} = 1$, we must have that $x^*_{g_i} = 0$ for $i = 1, \ldots, j-1$ and $x^*_{g_j} = 1$ due to the first two constraints of the LP. In this scenario, we get that $\mathbb{E}[\lambda_g r_{g_j}(1 - \bar{x}_{g_1})(1 - \bar{x}_{g_2}) \ldots (1 - \bar{x}_{g_{j-1}})\bar{x}_{g_j}] = \lambda_g r_{g_j} y^*_{g,g_j} \geq \gamma_j(\alpha, \beta) \cdot \lambda_g r_{g_j} y^*_{g,g_j}$, where the inequality follows because $\gamma_j(\alpha, \beta) \leq 1$ since $\alpha + \beta/2 \leq 1/2$. In the remaining case in which $y^*_{g,g_j} = \frac{1}{2}$, the LP constraints imply that $x^*_{g_i} < 1$ for all $i = 1, \ldots, j-1$ and $x^*_{g_j} \geq \frac{1}{2}$. Therefore, the LP-based rounding algorithm offers products $g_1, \ldots, g_{j-1}$ with probability either equal to 0 or $\alpha + \beta x^*_{g_i}$ and product g_j with probability equal to 1 or $\alpha + \beta x^*_{g_j}$. As a result, we get that

$$\begin{aligned}
&\mathbb{E}[\lambda_g r_{g_j}(1 - \bar{x}_{g_1})(1 - \bar{x}_{g_2}) \ldots (1 - \bar{x}_{g_{j-1}})\bar{x}_{g_j}] \\
&\geq \lambda_g r_{g_j}\left(1 - \alpha - \beta x^*_{g_1}\right)\left(1 - \alpha - \beta x^*_{g_2}\right) \ldots \left(1 - \alpha - \beta x^*_{g_{j-1}}\right)\left(\alpha + \beta x^*_{g_j}\right) \\
&= \lambda_g r_{g_j}\left(1 - \alpha - \beta + \beta(1 - x^*_{g_1})\right)\left(1 - \alpha - \beta + \beta(1 - x^*_{g_2})\right) \ldots \left(1 - \alpha - \beta + \beta(1 - x^*_{g_{j-1}})\right)\left(\alpha + \beta x^*_{g_j}\right) \\
&\geq \lambda_g r_{g_j}\left(1 - \alpha - \beta + \beta y^*_{g,g_j}\right)^{j-1}\left(\alpha + \beta y^*_{g,g_j}\right) \\
&= 2\lambda_g r_{g_j}\left(1 - (\alpha + \beta/2)\right)^{j-1}(\alpha + \beta/2)\, y^*_{g,g_j} \\
&= \gamma_j(\alpha, \beta)\lambda_g r_{g_j} y^*_{g,g_j},
\end{aligned}$$

where the second inequality comes from the first and second constraints of the LP and the second equality uses the fact that $y^*_{g,g_j} = \frac{1}{2}$. □

We can use Lemma 3.2 to derive the optimal performance guarantee of our LP-based rounding algorithm. Let

$$(\alpha^*, \beta^*) = \underset{\alpha,\beta:\alpha+\beta/2\leq 1/2}{\operatorname{argmax}} 2\left(1-(\alpha+\beta/2)\right)^{k-1}(\alpha+\beta/2)$$

and $\gamma_k = \gamma_k(\alpha^*, \beta^*)$. Let $\bar{z}$ be the expected revenue from the random assortment $\bar{x}$ with $\alpha = \alpha^*$ and $\beta = \beta^*$. The following proposition relates OPT and $\bar{z}$ through γ_k.

Proposition 3.3. $\bar{z} \geq \gamma_k \mathrm{OPT}$

Proof. We have that $\bar{z} = \sum_{g\in\mathcal{G}} \lambda_g \sum_{j=1}^{k} \mathbb{E}[r_{g_j}(1-\bar{x}_{g_1})(1-\bar{x}_{g_2})\dots(1-\bar{x}_{g_{j-1}})\bar{x}_{g_j}] \geq \sum_{g\in\mathcal{G}} \lambda_g \sum_{j=1}^{k} \gamma_j(\alpha^*,\beta^*) r_{g_j} y^*_{g,g_j} \geq \gamma_k \sum_{g\in\mathcal{G}} \lambda_g \sum_{j=1}^{k} r_{g_j} y^*_{g,g_j} = \gamma_k z^* \geq \gamma_k \mathrm{OPT}$, as desired. The first inequality follows since Lemma 3.2 holds for any values of α and β. The second inequality holds because $\gamma_j(\alpha^*,\beta^*) \geq \gamma_k$ by definition of γ_k. □

Thus, by returning the LP rounded solution we get a γ_k-approximation algorithm in expectation. Further, we can derandomize this LP-based rounding algorithm in $O(k|\mathcal{G}|)$ time using the standard method of conditional expectations while only improving the approximation ratio (Erdõs and Selfridge (1973), Spencer (1987)).

All that is left is to find γ_k to compute the guarantee of our LP-based rounding algorithm. Note that we can rewrite γ_k as

$$\max_{0\leq q\leq 1/2} 2\left(1-q\right)^{k-1}(q),$$

which is trivially optimized at $q^* = 1/k$. Hence any value of α^*, β^* satisfying $\alpha^* + \beta^*/2 = 1/k$ will suffice. One such combination that works is $\alpha^* = 1/(2k)$ and $\beta^* = 1/k$. The randomized rounding can thus be summarized as follows. When $x_i^* = 0$ we do not offer product i, when $x_i^* = 1$ we do offer offer product i, and when $x_i^* = 1/2$ we offer product i with probability $1/k$. Theorem 3.4 gives the approximation ratios we achieve for arbitrary k under this strategy.

Theorem 3.4. *We can efficiently find an assortment with revenue at least*

$$2\left(1-\frac{1}{k}\right)^{k-1}\left(\frac{1}{k}\right)\cdot \mathrm{OPT}.$$

Thus, there exists a 1/2-approximation algorithm for the assortment optimization under the 2-product nonparametric choice model, a 0.296-approximation algorithm under the 3-product nonparametric choice model, and a 0.210-approximation algorithm under the 4-product nonparametric choice model.

Aouad, Farias, Levi and Segev (2015) give an algorithm that provides a $1/(e \cdot k)$ approximation ratio, which is the best known previous guarantee for problem (2). Noting that $2\left(1-\frac{1}{k}\right)^{k-1}\left(\frac{1}{k}\right) \geq 2/(e \cdot k)$, we have that the performance guarantee presented in Theorem 3.4 improves upon the previous best guarantee by a factor of two. In Section 5, we show that this theoretical improvement also leads to noticeable improvements in practical performance. Through a series of computational experiments, we demonstrate that the LP-based rounding algorithm outperforms the rounding algorithm of Aouad, Farias, Levi and Segev (2015) on a series of randomly generated instances of the k-product nonparametric assortment problem. It turns out that we can give a tighter LP relaxation for our assortment problem, and building on this LP relaxation, we can give an LP-based rounding algorithm that has the same approximation guarantee. This tighter LP relaxation provides a tighter upper bound on the optimal expected revenue, which helps us to get a better estimate of the optimality gaps in our numerical experiments.

3.1 Tightening the LP Relaxation

A natural way to measure the optimality gap of the assortment recommended by our randomized approach is to use the upper bound provided by the LP relaxation. However, in Appendix A.3, we give an instance of the k-product nonparametric choice model for which the LP relaxation of (IP1) has an optimality gap of $O(k)$. Consequently, in this section, we study a tighter formulation of the problem. We show that we can use the LP relaxation of this modified formulation to construct an LP-based approximation algorithm with the same performance guarantee. Consider the following modified IP for the assortment optimization problem under the k-product nonparametric choice model

$$
\begin{aligned}
\text{maximize} \quad & \sum_{g} \sum_{j=1}^{k} \lambda_g r_{g_j} y_{g,g_j} && && \text{(IP2)} \\
\text{subject to} \quad & \sum_{j=1}^{k} y_{g,g_j} \leq 1 && \forall g \in \mathcal{G} \\
& y_{g,g_j} \leq x_{g_j} && \forall g \in \mathcal{G}, j \in \{1,2,\ldots,k\} \\
& \sum_{i=j+1}^{k} y_{g,g_i} \leq 1 - x_{g_j} && \forall g \in \mathcal{G}, j \in \{1,2,\ldots,k-1\} \\
& x_i \in \{0,1\} && \forall i \in N \\
& y_{g,g_j} \in \{0,1\} && \forall g \in \mathcal{G}, j \in \{1,2,\ldots,k\}.
\end{aligned}
$$

The objective function calculates the total expected revenue. The first constraint states that a customer of class g can make at most one purchase, the second constraint states that a customer can only purchase an offered product, and the third constraint states that if a customer's jth preferred product is available then they cannot purchase a less preferred product. This formulation was previously introduced by Bertsimas and Mišić (2017).

We round the LP relaxation of this IP exactly as before. In particular, let (x^*, y^*) be an optimal LP solution with optimal objective value z^*. From this fractional solution, we will construct an assortment $\bar{x}$. For any product i such that $x_i^* \in \{0, 1\}$, we set $\bar{x}_i = x_i^*$. On the other hand, if $0 < x_i^* < 1$, then we offer product i independently with probability $\frac{1}{2k} + \frac{1}{k}x_i^*$. To analyze this revised algorithm, we will compare the expected revenue of $\bar{x}$ for a single customer class g and product g_j to the LP value y^*_{g,g_j}. This result is given in the next theorem. We defer the proof to Appendix A.4. Then, applying the same proof technique in Proposition 3.3 yields the desired result.

Theorem 3.5. *For a customer class $g \in \mathcal{G}$ and $j \in \{1, 2, \ldots, k\}$,*

$$\mathbb{E}[\lambda_g r_{g_j}(1 - \bar{x}_{g_1})(1 - \bar{x}_{g_2}) \ldots (1 - \bar{x}_{g_{j-1}})\bar{x}_{g_j}] \geq 2\left(1 - \frac{1}{k}\right)^{k-1}\left(\frac{1}{k}\right)\lambda_g r_{g_j} y^*_{g,g_j}.$$

When we test our LP-based rounding algorithm in Section 5, we use the tighter formulation given by (IP2). Further, in Appendix A.5, we show that this approximation guarantee is tight for $k = 2$. In the next section, we extend and improve our results for this case.

4 Extensions to Two Product Preference Lists

When $k = 2$, we are able to exploit the simpler substitution patterns to extend the results in the previous section. First, we show that we can improve the approximation guarantee from 1/2 to 0.874 by exploiting existing algorithms for maximum directed cut problem. Second, we show that we can use a more nuanced version of the LP-based rounding algorithm to provide an approximation algorithm for the cardinality constrained version of the assortment problem, in which there is a limit on the total number of products that can be offered.

4.1 Improved Approximation Guarantee

When each preference list includes at most two products, we can reduce the assortment optimization problem to a well-known discrete optimization problem known as the maximum directed cut problem. In the maximum directed cut problem, we are given a directed graph $G = (V, E)$ along with edge weights $w_e \geq 0$ for all $e \in E$. The goal is to find a cut $C \subseteq V$ that maximizes the weight of edges directed from C to $V - C$. We can reduce our assortment optimization problem to maximum directed cut problem for which Lewin et al. (2002) use a semidefinite programming relaxation to give a 0.874-approximation algorithm for this problem. It immediately follows that there exists a 0.874 approximation algorithm for the assortment optimization problem. We state this result in the next theorem and defer its proof, which shows the reduction to the maximum directed cut problem, to Appendix A.6.

Theorem 4.1. *There exists a 0.874-approximation ratio for the assortment optimization problem under the 2-product nonparametric choice model.*

The algorithm in Theorem 4.1 rounds the solution to a semidefinite program, and it can be derandomized using the techniques proposed in Mahajan and Ramesh (1999). This approach, however, does not extend to larger values of k.

4.2 Cardinality Constrained Assortment Problem

A natural extension of the work in Section 3 is to consider the cardinality constrained version of the assortment problem. In this problem, the firm chooses an assortment $S \subseteq N$ to maximize the expected revenue with the added constraint that $|S| \leq c$ for some fixed integer c. This constraint might occur in online settings in which the retailer wants to choose which products to display when there is limited website real estate or in a setting in which there is a fixed physical space to display products. In this section, we present a 1/4-approximation algorithm for the capacitated assortment problem under the 2-product nonparametric choice model. Our approach uses the LP relaxation of an IP formulation of the problem, but we need to be careful to ensure that the derandomized assortment satisfies the cardinality constraint.

An IP for this problem under the 2-product nonparametric choice model is given by

$$
\begin{aligned}
\max \quad & \sum_{g} \sum_{j=1}^{k} \lambda_g r_{g_j} y_{g,g_j} \\
\text{s.t.} \quad & y_{g,g_j} \leq x_{g_j} && \forall g \in \mathcal{G}, j \in \{1,2\} \\
& y_{g,g_2} \leq 1 - x_{g_1} && \forall g \in \mathcal{G} \\
& \sum_{i=1}^{n} x_i \leq c \\
& y_{g,g_j} \geq 0 && \forall g \in \mathcal{G}, j \in \{1,2\} \\
& x_i \in \{0,1\} && \forall i \in N.
\end{aligned}
$$

This is the same IP as in Section 3 with the added constraint that we can offer at most c products.

As before, we will use the LP relaxation of this IP to inform our decisions on which products to offer. However, we have to be careful when rounding the LP solution that we do not go over the capacity c. Let (x^*, y^*) be an optimal LP solution with optimal objective value z^*. We first construct a corresponding random assortment $\bar{x}$ that offers each product i independently with probability $\frac{1}{2}x_i^*$.

Lemma 4.2.

$$\mathbb{E}[\mathrm{Rev}(\bar{x})] \geq \frac{1}{4} \cdot \mathrm{OPT}.$$

Proof. We analyze the expected revenue of $\bar{x}$ by looking at the expected revenue for each customer class g. Consider a single customer class $g \in \mathcal{G}$. We first consider the expected revenue generated from the first choice product g_1. We offer product g_1 with probability $\frac{1}{2}x_{g_1}^*$. The associated expected revenue is $\mathbb{E}[\lambda_g r_{g_1} \bar{x}_{g_1}] = \lambda_g r_{g_1} \left[\frac{1}{2}x_{g_1}^*\right] \geq \frac{1}{2}\lambda_g r_{g_1} y_{g,g_1}^*$, where the inequality follows from the first constraint of the LP.

Now consider the expected revenue from the second choice product g_2 for customer class g. We offer product g_1 with probability $\frac{1}{2}x_{g_1}^*$ and offer product g_2 with probability $\frac{1}{2}x_{g_2}^*$. This yields expected revenue $\mathbb{E}[\lambda_g r_{g_2}(1-\bar{x}_{g_1})\bar{x}_{g_2}] = \lambda_g r_{g_2} \left(1 - \frac{x_{g_1}^*}{2}\right)\left(\frac{x_{g_2}^*}{2}\right) = \lambda_g r_{g_2} \left(\frac{1}{2} + \frac{1-x_{g_1}^*}{2}\right)\left(\frac{x_{g_2}^*}{2}\right) \geq \lambda_g r_{g_2} \left(\frac{1}{2} + \frac{y_{g,g_2}^*}{2}\right)\left(\frac{y_{g,g_2}^*}{2}\right) \geq \frac{1}{4} \cdot \lambda_g r_{g_2} y_{g,g_2}^*$, where the second to last inequality comes from the first and second constraints of the LP. Thus, the expected revenue from the LP-based rounding algorithm is $\mathbb{E}[\sum_g \lambda_g [r_{g_1}\bar{x}_{g_1} + r_{g_2}(1-\bar{x}_{g_1})\bar{x}_{g_2}] \geq \sum_g \frac{1}{2}\lambda_g r_{g_1} y_{g,g_1}^* + \frac{1}{4}\lambda_g r_{g_2} y_{g,g_2}^* \geq \frac{1}{4} \cdot z^* \geq \frac{1}{4} \cdot \mathrm{OPT}$. This shows that the expected revenue of $\bar{x}$ is at least 1/4 of the optimal revenue. $\square$

The lemma above shows that the random assortment $\bar{x}$ obtains at least 1/4 of the optimal revenue in expectation. Further, the expected size of $\bar{x}$ is $\frac{1}{2}\sum_{i=1}^n x_i^* \leq c/2 \leq c - 1$. Here, we assume that $c > 1$ since otherwise we can solve this problem trivially by enumeration. However, we cannot just hope to derandomize the algorithm as before using the method of conditional expectations without possibly violating the cardinality constraint. Instead, we will look at rounding pairs of products to transform the solution into a deterministic assortment while only improving the expected revenue. The proof of the following lemma, which describes this rounding technique, is given in Appendix A.7.

Lemma 4.3. *In polynomial time, we can find an assortment $x \in \{0,1\}^n$ such that* $\mathrm{Rev}(x) \geq \mathbb{E}[\mathrm{Rev}(\bar{x})]$ *and* $\sum_{i=1}^{n} x_i \leq c$.

Combining Lemma 4.2 and 4.3 yields the following theorem.

Theorem 4.4. *There exists a 1/4-approximation algorithm for the capacitated assortment optimization problem under the 2-product nonparametric choice model.*

5 Testing the Performance of our LP-Based Rounding Algorithm

In this section, we provide computational experiments that demonstrate the effectiveness of the randomized rounding procedure from Section 3 used to solve the assortment optimization problem under the k-product nonparametric choice model. We generate a series of random test cases for $k = 3, 4$ and compare the performance of the assortments produced by our LP-based rounding algorithm with those produced by the randomized algorithm of Aouad, Farias, Levi and Segev (2015). Not only does our algorithm perform better than this benchmark, but its performance is near optimal in the majority of test cases.

We generate each test instance using the following procedure. First, we set the maximum length of any preference list to be $k \in \{3, 4\}$, the number of products to be $n \in \{50, 100\}$, and the number of customer classes to be $m \in \{1000, 5000, 10000\}$. Next, we sample the revenues of each product independently from a uniform distribution over $[1, 100]$. We choose the collection of m preference lists by uniformly sampling from the list of all possible preference lists for the given value of k. In other words, for fixed $k \in \{3, 4\}$ and $n \in \{50, 100\}$, we generate all possible preference lists of length at most k. From this set, we uniformly sample without replacement m preference lists. Lastly, we generate the arrival probabilities for each customer class by first generating $(\beta_1, \dots, \beta_m)$ independently from a uniform distribution over $[0, 1]$ and then setting $\lambda_g = \beta_g / \sum_{k=1}^{m} \beta_k$.

The variation of $(k, n, m) \in \{3, 4\} \times \{50, 100\} \times \{1000, 5000, 10000\}$ gives 12 parameter combinations. For each set of parameters, we generate 100 independent problem instances using the approach outlined above. For each problem instance, we employ two different approaches. The first is the randomized LP rounding algorithm presented in Section 3.1. We refer to this algorithm as ALG. The second approach is the randomized algorithm of Aouad, Farias, Levi and Segev (2015), which randomly offers each product with probability $1/k$. We refer to this algorithm as RAND. For both algorithms, we use the method of conditional expectations to derandomize which only improves the algorithms' performance in expectation. To check the quality of the proposed assortments given by the two approaches, we find the gap between the expected revenue of these assortments and the upper bound provided by the linear programming formulation in Section 3.1.

5.1 Results

Table 1a focuses on ALG, whereas Table 1b focuses on RAND. Columns 1-3 give the parameter combinations that define each test case. Recall that for each test case, we generate 100 problem instances. Let ALG^p and RAND^p be the expected revenues of the assortments produced by ALG and RAND on problem instance p. Further, let LP^p be the objective of the LP relaxation of the assortment problem for problem instance p, which is trivially an upper bound on the optimal expected revenue. Then, the percent optimality gaps for the two algorithms are given by $100 \times (\text{LP}^p - \text{ALG}^p)/\text{LP}^p$ and $100 \times (\text{LP}^p - \text{RAND}^p)/\text{LP}^p$, respectively. Column 4 in Tables 1a and 1b reports the average optimality gap over all 100 problem instances, Column 5 reports the gap at the 75th percentile, and Column 6 reports the largest gap seen. The running times for any single instance never exceeded a few minutes. We also tested our algorithm on instances of the assortment problem with $n = 50$, k=5 and $m = 100,000$. Even on these very large instances of the assortment problem, the running time of our algorithm never exceeded ten minutes and continued to produce near optimal assortments. By contrast, the running time to solve the problem exactly on these large instances using Gurobi's integer programming solver could exceeded 5 hours. All experiments used Python 2.7 on an Intel Core i5 with 3.2 GHz CPU and 32GB of RAM and Gurobi 6.5.1 as the linear/integer programming solver.

The results in Table 1 indicate that the randomized LP-based rounding algorithm is noticeably better than the randomized algorithm of Aouad, Farias, Levi and Segev (2015). Over all parameter combinations, the average performance of ALG is never below 0.52% of optimal. Further, over all 1200 problem instances, the percent optimality gap of ALG never exceeds 3.7%. On the other hand, the average percent optimality gap of RAND is never below 2.90% and this algorithm suffers from optimality gaps of over 10% for certain problem instances.

As k grows, substitution behavior becomes more complicated, and as a result, we would expect the performance of both algorithms to deteriorate. This trend is precisely what we observe; for ALG the average percent optimality gap grows from 0.055% to 0.36% as k goes from 3 to 4 and the percent gap for RAND also grows from to 3.10% to 4.84%. Surprisingly, we observe that both algorithms perform at their worst for smaller problem instances. In fact, the worst optimality gaps for both algorithms occur when $n = 50$. It might be that in these smaller instances there is a smaller margin for error and hence there are fewer assortments that have expected revenue close to optimal.

			% Optimality Gap		
k	n	m	Avg.	75th	Max.
3	50	1000	0.11	0.0	0.86
3	50	5000	0.01	0.0	1.25
3	50	10000	0.05	0.0	1.01
3	100	1000	0.06	0.0	0.59
3	100	5000	0.05	0.0	0.98
3	100	10000	0.05	0.0	0.51
4	50	1000	0.37	0.0	2.62
4	50	5000	0.41	0.0	2.60
4	50	10000	0.52	0.0	3.66
4	100	1000	0.22	0.0	2.42
4	100	5000	0.32	0.0	1.94
4	100	10000	0.33	0.0	3.29

(a) Performance of ALG on 1200 randomly generated instances.

			% Optimality Gap		
k	n	m	Avg.	75th	Max.
3	50	1000	3.29	2.07	8.34
3	50	5000	2.93	1.97	6.45
3	50	10000	3.08	2.04	6.85
3	100	1000	3.51	2.88	6.00
3	100	5000	2.91	2.35	5.73
3	100	10000	2.90	2.20	5.45
4	50	1000	4.97	3.51	9.93
4	50	5000	5.06	3.75	10.37
4	50	10000	4.88	3.87	8.94
4	100	1000	4.95	4.04	8.79
4	100	5000	4.74	4.17	8.25
4	100	10000	4.45	3.45	8.45

(b) Performance of RAND on 1200 randomly generated instances.

Table 1: Comparing the performance of the two randomized algorithms.

6 Estimating the k-Product Nonparametric Model

In this section, we fit k-product nonparametric choice models for $k \in \{1, 2, 3, 4\}$ to historical sales data. Throughout this section, we refer to the model that generates the sales data as the ground choice model. Our approach for generating the ground choice model is reflective of a setting in which customers make purchasing decisions from a set of vertically differentiated products, meaning that there is some overarching order on the quality of products. In particular, we assume that the ground choice model is a nonparametric choice model in which each customer's preference list takes the form of a slight perturbation of the consensus quality ranking.

6.1 Experimental Set-up

In the ground choice model, we let the set of products be given by the set $N = \{1, 2, \ldots, n\}$ and the set of customer classes be given by $\mathcal{G} = \{1, \ldots, m\}$. Throughout the experiments, we fix $n = 20$. The preference lists and the arrival probability of each customer class $g \in \mathcal{G}$ are respectively given by σ_g and λ_g. Given the set of preference lists, we set the arrival probabilities for each customer class by first generating $(\beta_1, \ldots, \beta_m)$ independently from a uniform distribution over $[0, 1]$ and then setting $\lambda_g = \beta_g / \sum_{k=1}^{m} \beta_k$. We note that this process does not ensure that the arrival probabilities will be generated uniformly from the probability simplex $\{\lambda \in \mathbb{R}^m : \sum_{g=1}^{m} \lambda_g = 1\}$. Nonetheless, we believe that our approach for generating the arrival probabilities is equally valid for accessing the quality of the fits of the k-product nonparametric choice model.

Next, we briefly motivate our process for generating the preference lists in the ground choice model. We assume that the products are indexed in decreasing order of quality (i.e. lower indexed products have higher qualities). Each customer's preference list σ_g consists of products from an associated quality interval $Q_g = [i_g, i_g + 1, \ldots, j_g]$ for $i_g, j_g \in N$ such that $i_g \leq j_g$. Under the assumption that higher quality products are generally associated with higher prices, we assume that each customer $g \in \mathcal{G}$ does not consider purchasing higher quality products $\{1, \ldots, i_g - 1\}$ because these products are priced above her willingness to pay. On the other hand, she does not consider purchasing products $\{j_g + 1, \ldots, n\}$ because these products are not of high enough quality. So each customer has a willingness to pay, which determines i_g, and a minimum quality threshold, which determines j_g. We then randomly delete products from each customer's quality interval to capture the notion that customers might have particular brands or styles that they will never consider purchasing. Finally, we also randomly swap the order of products in each customer's quality interval to capture the notion that quality is rarely perceived in the same way by every customer.

We conduct two sets of experiments, which are differentiated by how we sample the quality intervals. In our first set of experiments, for each customer type $g \in \mathcal{G}$, we generate i_g uniformly from the set $\{1, \ldots, n\}$ and then generate j_g uniformly from the set $\{i_g, \ldots, n\}$. The second, shorter set of experiments serves as a test of the more extreme case in which customers substitute far more than is described by the fitted k-product nonparametric model. To accomplish this, we sample quality intervals in the same way, but we only keep quality intervals that contain at least eight products - double the largest value of k that we fit. After the quality interval has been sampled, we set $\sigma_g = Q_g$ and then drop each product from σ_g with probability p_d. Next, we consider $\mathcal{F}$ flip events, which each occur with probability 0.5. If a flip event is executed, we uniformly sample a product $k \in \sigma_g$ and flip its position in σ_g with the product ranked immediately ahead of it.

We note that even though the ground choice models are nonparametric choice models, they are generated so that they differ substantially from the k-product nonparametric choice models that we fit. Tables 2a and 2b show the percentage of preference lists in each ground choice model that contain more than eight products in the first set of experiments. We note again that our threshold length of eight is double the largest value of k that is used in fitting our k-product nonparametric choice models. Given that all of the percentages exceed 23%, we can be fairly certain that the substitution patterns dictated by the ground choice models are distinct from those described by the choice models that we fit.

p_d	$\mathcal{F}$	$\%\lvert\sigma_g\rvert \geq 8$
0	1	36.4
0	2	37.7
0	4	35.4
0.25	1	25.1
0.25	2	23.2
0.25	4	24.2

(a) $m = 100$.

p_d	$\mathcal{F}$	$\%\lvert\sigma_g\rvert \geq 8$
0	1	47.5
0	2	46.1
0	4	45.3
0.25	1	30.0
0.25	2	29.2
0.25	4	28.6

(b) $m = 500$

Table 2: Average percentage of preference lists satisfying $|\sigma_g| \geq 8$ in the ground choice model in the first set of experiments.

Once the ground choice model has been generated, we then generate the historical sales data under the assumption that the purchasing behavior of all arriving customers is governed by the ground choice model. We assume that we have access to the past purchasing history of τ customers given by $\text{PH}_\tau = \{(S_t, z_t) : t = 1, \ldots, \tau\}$, where S_t is the subset of products offered to customer in time period t and z_t is the product purchased in time period t. We form the subset S_t in each time period by randomly including each product with probability 0.5. In this way, we uniformly sample an assortment. The class g_t that customer t belongs to is then sampled from the distribution $(\lambda_1, \ldots, \lambda_m)$. Lastly, we set $z_t = 0$ if $S_t \cap \sigma_{g_t} = \emptyset$ so that the customer leaves without making a purchase and set $z_t = \text{argmin}_{i \in S_t} \sigma_{g_t}(i)$ otherwise.

In the first set of experiment, we vary $(m, p_d, \mathcal{F}, \tau) \in (100, 500) \times (0, 0.25) \times (1, 2, 4) \times (5000, 10000)$. For the second set of experiments, which tests settings in which customers have longer preference lists, we fix $p_d = 0.25$, $\mathcal{F} = 4$, and $\tau = 10000$ but vary $m \in (100, 500)$. For each combination of parameters, we generate 10 ground choice models. Then, for each of these choice models we generate PH_{5000} and PH_{10000} using the sales data generation process described above such that $\text{PH}_{5000} \subseteq \text{PH}_{10000}$. In this way, PH_{10000} can be viewed as additional data that should, and in our experiments, always does, improve the accuracy of the fitted models. For each data set generated from the ground choice model GC, we fit nonparametric choice models NP_i for $i \in \{1, 2, 3, 4\}$, where NP_i is a nonparametric choice model that includes all preference lists with length less than or equal to i, and we also fit an MNL model ML, which we use as a benchmark.

For each of the five models, we use maximum likelihood estimation (MLE) to fit the corresponding model. It is well documented (see Train (2009) for the MNL model and van Ryzin and Vulcano (2015) for the nonparametric model) that the log-likelihood functions under both the nonparametric choice model and the MNL model are concave and thus MLE is tractable. For the MLE problem under the k-product nonparametric choice model, we use a piecewise linear approximation to the log-likelihood. In this case, the MLE problem can be written as an LP whose size and accuracy is determined by the precision of the piecewise linear approximation. We use equally spaced grid points of length 0.02 to generate our piecewise linear approximation, which results in a loss in precision of at most 0.26%. The MLE procedure was implemented in Java 1.8 on an Intel Core i5 with 3.2 GHz CPU and 32 GB of RAM and the LPs were all solved in Gurobi 6.5.1. We elect to use this piecewise linear approximation due to the efficiency with which Gurobi can solve large scale LPs. An alternative approach would be to directly maximize the log-likelihood via a nonlinear constrained solver such as MATLAB's *fmincon* or Python's *CVXPY*. In Appendix A.8, we report the average runtimes for fitting the k-product nonparametric choice models for $\tau = 10000$.

Last, we describe how we measure the accuracy of the fitted models. Suppose that for the data set generated from ground choice model GC, we have fitted choice models $\{\text{NP}_1, \ldots, \text{NP}_4, \text{ML}\}$. For choice model CM $\in \{\text{NP}_1, \ldots, \text{NP}_4, \text{ML}\}$, let $\text{Pr}_i^{\text{CM}}(S)$ be the probability that product i is purchased under choice model CM when assortment S is offered. We measure the efficacy of these fitted models using the Mean Absolute Error (MAE) of the predicted purchase probabilities of each choice model. In particular, for fitted choice model CM, $\text{MAE(CM)} = \frac{1}{2^n} \sum_{S \subseteq N} \sum_{i \in S} |\text{Pr}_i^{\text{CM}}(S) - \text{Pr}_i^{\text{GC}}(S)|/|S|$ is the average error in predicted purchase probability over every product and every assortment.

6.2 Results

The results for the first set of experiments are given in Tables 3a and 3b for $m = 100$ and $m = 500$, respectively. Column 1 and 2 in Tables 3a and 3b give the values of p_d and $\mathcal{F}$ that were used to generate the ground choice models, Column 3 denotes the fitted model and Columns 4 and 5 report the MAE of these fitted models for each value of $\tau \in \{5000, 10000\}$, averaged over the 10 different purchase histories.

There are a few key things to note about the results. First, both NP_3 and NP_4 have lower average MAE than ML for all combinations of p_d, $\mathcal{F}$, and m. In fact, for NP_3 and NP_4, the improvement over ML can be as high as 70% and 100%, respectively. There is a slight decline in the performance of the nonparametric models' performance for $m = 500$, which reflects the more complex substitution behavior. Further, as expected, NP_4 provides the best fit over all parameter combinations, but the improvement in MAE decreases as k increases, implying that there is diminishing returns for fitting more complex models. It is also interesting to note that all

p_d	$\mathcal{F}$	Model	τ 5000	10000
0	1	NP_1	0.088	0.088
0	1	NP_2	0.043	0.043
0	1	NP_3	0.023	0.022
0	1	NP_4	0.018	0.015
0	1	ML	0.039	0.039
0	2	NP_1	0.088	0.088
0	2	NP_2	0.043	0.042
0	2	NP_3	0.023	0.021
0	2	NP_4	0.018	0.015
0	2	ML	0.037	0.037
0	4	NP_1	0.087	0.087
0	4	NP_2	0.042	0.041
0	4	NP_3	0.023	0.021
0	4	NP_4	0.018	0.015
0	4	ML	0.034	0.034
0.25	1	NP_1	0.080	0.080
0.25	1	NP_2	0.035	0.035
0.25	1	NP_3	0.019	0.017
0.25	1	NP_4	0.016	0.013
0.25	1	ML	0.030	0.030
0.25	2	NP_1	0.079	0.079
0.25	2	NP_2	0.035	0.034
0.25	2	NP_3	0.019	0.017
0.25	2	NP_4	0.017	0.014
0.25	2	ML	0.029	0.028
0.25	4	NP_1	0.079	0.079
0.25	4	NP_2	0.035	0.034
0.25	4	NP_3	0.019	0.017
0.25	4	NP_4	0.017	0.014
0.25	4	ML	0.027	0.027

(a) Average MAE of the fitted models for $m = 100$.

p_d	$\mathcal{F}$	Model	τ 5000	10000
0	1	NP_1	0.093	0.093
0	1	NP_2	0.049	0.048
0	1	NP_3	0.027	0.026
0	1	NP_4	0.019	0.017
0	1	ML	0.039	0.039
0	2	NP_1	0.092	0.092
0	2	NP_2	0.048	0.047
0	2	NP_3	0.027	0.025
0	2	NP_4	0.019	0.017
0	2	ML	0.037	0.037
0	4	NP_1	0.092	0.092
0	4	NP_2	0.047	0.047
0	4	NP_3	0.026	0.025
0	4	NP_4	0.019	0.017
0	4	ML	0.034	0.034
0.25	1	NP_1	0.085	0.085
0.25	1	NP_2	0.040	0.040
0.25	1	NP_3	0.021	0.020
0.25	1	NP_4	0.017	0.014
0.25	1	ML	0.029	0.029
0.25	2	NP_1	0.085	0.085
0.25	2	NP_2	0.040	0.039
0.25	2	NP_3	0.021	0.019
0.25	2	NP_4	0.017	0.014
0.25	2	ML	0.027	0.027
0.25	4	NP_1	0.084	0.084
0.25	4	NP_2	0.040	0.039
0.25	4	NP_3	0.021	0.019
0.25	4	NP_4	0.017	0.014
0.25	4	ML	0.025	0.025

(b) Average MAE of the fitted models for $m = 500$.

Table 3: Estimation results for the first set of experiments.

the fitted models perform worst when $p_d = 0$ and $\mathcal{F} = 1$. For these test cases, each customer's preference list most closely resembles a quality interval. However, the decrease in performance is most pronounced for ML. These results suggest that retailers should be wary in attempting to use the MNL choice model to capture customer purchasing behavior in a setting in which the products are vertically differentiated. Last, we note that the additional data in PH_{10000} improves the performance for NP_2, NP_3, and NP_4 but not for NP_1 or ML. This is likely due to the fact that the former models have significantly more parameters to estimate.

p_d	$\mathcal{F}$	Model	τ 10000
0.25	4	NP_1	0.110
0.25	4	NP_2	0.060
0.25	4	NP_3	0.035
0.25	4	NP_4	0.022
0.25	4	ML	0.032

(a) Average MAE of the fitted models for $m = 100$.

p_d	$\mathcal{F}$	Model	τ 10000
0.25	4	NP_1	0.110
0.25	4	NP_2	0.060
0.25	4	NP_3	0.035
0.25	4	NP_4	0.021
0.25	4	ML	0.030

(b) Average MAE of the fitted models for $m = 500$.

Table 4: Estimation results for the second set of experiments.

The corresponding results for the second set of experiments are given in Tables 4a and 4b for $m = 100$ and $m = 500$, respectively. We only provide results for when $p_d = 0.25, \mathcal{F} = 4$ and $\tau = 10000$, since this was the test case where the MNL fits performed best in the first set of experiments. In this set of experiments, the ground choice models exhibit more complex patterns of substitution behavior, and hence, it is no surprise that the MAE of each of our fitted models increases slightly in comparison to the first set of experiments. Nonetheless, the performance of NP_4 is again significantly better than ML, producing improvements of up to 50% with regard to the MAE.

7 Conclusions and Future Work

In this paper, we study the k-product nonparametric choice model, which captures settings in which there is limited substitution between products. We prove that the assortment optimization problem under this model is NP-hard even for $k = 2$. Motivated by this result, we develop a general approximation scheme whose approximation ratio decreases as k increases. Through a series of computational experiments, we show that the performance of this algorithm far exceeds the theoretical guarantee. We also show that this model has the potential to capture general customer choice behavior for small values of k. In order to establish the exact extent of this generality, future work is needed to test the accuracy of fitted k-product nonparametric choice models when the historical sales data is generated from a more diverse array of ground choice models. Further, the accuracy of the fitted models could be assessed on additional metrics beyond MAE, which measures the extent to which the estimated purchase probabilities align with the true purchase probabilities but does not directly give a sense of the profitability of subsequent assortment or inventory decisions that will eventually be made using the estimated choice models.

From an algorithmic perspective, there are also several interesting directions for future research. One extension of our result for the assortment problem under the k-product nonparametric choice model would be to improve the approximation guarantees for smaller fixed values of k. A separate, yet equally important, direction of future work could consider the cardinality constrained assortment problem for $k > 2$.

ACKNOWLEDGEMENTS. We would like to thank the associate editor and two anonymous referees whose comments improved our paper substantially.

AUTHOR BIOS. **Jacob B. Feldman** is an assistant professor at Olin Business School at Washington University in St. Louis. His research interests include customer choice modeling and approximation algorithms. **Alice Paul** is postdoctoral researcher at the Data Science Initiative at Brown University. Her research focuses on developing approximation algorithms for NP-Hard problems. **Huseyin Topaloglu** is a professor at the School of Operations Research and Information Engineering at Cornell Tech. His research interests include revenue management and pricing.

References

Alimonti, P. and Kahn, V. (2000), 'Some apx-completeness results for cubic graphs', *Theoretical Computer* **237**(1), 123–134.

Aouad, A., Farias, V. F. and Levi, R. (2015), Assortment optimization under consider-then-choose choice models. Available at SSRN: `https://ssrn.com/abstract=2618823`.

Aouad, A., Farias, V. F., Levi, R. and Segev, D. (2015), The approximability of assortment optimization under ranking preferences. Available at SSRN: `https://ssrn.com/abstract=2612947`.

Bertsimas, D. and Mišić, V. (2017), Exact first-choice product line optimization. Working Paper, MIT Sloan School. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3020502` (Accessed January 8, 2018).

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

Crompton, J. L. and Ankomah, P. K. (1993), 'Choice set propositions in destination decisions', *Annals of Tourism Research* **20**(3), 461–476.

Davis, J. M., Gallego, G. and Topaloglu, H. (2013), Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working Paper, Department of IEOR, Columbia University. Available at `http://www.columbia.edu/gmg2/logit_const.pdf` (accessed on May 25, 2017).

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. and Goyal, V. (2014), Near-optimal for capacity constrained assortment optimization. Available at SSRN: `https://ssrn.com/abstract=2543309`.

Désir, A., Goyal, V., Segev, D. and Ye, C. (2015), Capacity constrained assortment optimization under the Markov chain based choice model.

Erdõs, P. and Selfridge, J. L. (1973), 'On a combinatorial game', *Journal of Combinatorial Theory B* **14**, 293–301.

Feldman, J. and Topaloglu, H. (2015*a*), 'Capacity constraints across nests in assortment optimization under the nested logit model', *Operations Research* **63**(4), 812–822.

Feldman, J. and Topaloglu, H. (2015*b*), Revenue management under the Markov chain choice model. Unpublished Manuscript.

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

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

Hochbaum, D. S., Megiddo, N., Naor, J. and Tamir, A. (1993), 'Tight bounds and 2-approximation algorithms for integer programs with two variables per inequality', *Mathematical Programming* **62**(1), 69–83.

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

Jagabathula, S. and Rusmevichientong, P. (2016), 'A nonparametric joint assortment and price choice model', *To appear, Management Science* .

Jeuland, A. P. (1979), 'Brand choice inertia as one aspect of the notion of brand loyalty', *Management Science* **25**(7), 671–682.

Kök, A. G. and Fisher, M. L. (2007), 'Demand estimation and assortment optimization under substitution: Methodology and application', *Operations Research* **55**(6), 1001–1021.

Lapersonne, E., Laurent, G. and Le Goff, J.-J. (1995), 'Consideration sets of size one: An empirical investigation of automobile purchases', *International Journal of Research in Marketing* **49**(1), 55–66.

Lewin, M., Livnat, D. and Zwick, U. (2002), Improved rounding techniques for the MAX 2-SAT and MAX DI-CUT problems, *in* W. J. Cook and A. S. Schulz, eds, 'International Conference on Integer Programming and Combinatorial Optimization: IPCO 2002', Lecture Notes in Computer Science, vol 2337, Springer, Berlin, Heidelberg, pp. 67–82.

Li, G. and Rusmevichientong, P. (2014), 'A greedy algorithm for the two-level nested logit model', *Operations Research Letters* **42**(5), 319–324.

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 Ramesh, H. (1999), 'Derandomizing approximation algorithms based on semidefinite programming', *SIAM Journal of Computing* **28**(5), 1641–1663.

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

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

Meissner, J., Strauss, A. and Talluri, K. (2013), 'Enhanced concave program relaxation for choice network revenue management. production and operations management', *Production and Operations Management* **22**, 71–87.

Méndez-Díaz, I., Miranda-Bront, J. J., Vulcano, G. and Zabala, P. (2014), 'A branch-and-cut algorithm for the latent-class logit assortment problem', *Discrete Applied Mathematics* **164**, 246–263.

Nemhauser, G. L. and Trotter, L. E. (1975), 'Vertex packings: structural properties and algorithms', *Mathematical Programming* **8**(1), 232–248.

Paul, A., Feldman, J. and Davis, J. (2016), Assortment optimization and pricing under a nonparametric tree choice model. Unpublished Manuscript.

Rusmevichientong, P., Shen, Z.-J. 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. B., 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.

Spencer, J. H. (1987), *Ten lectures on the probabilistic method*, Vol. 52, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Talluri, K. (2014), 'New formulations for choice network revenue management', *INFORMS Journal of Computing* **26**, 401–413.

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

Train, K. (2009), *Discrete Choice Models with Simulation*, Cambridge University Press.

van Ryzin, G. and Vulcano, G. (2015), 'A market discovery algorithm to estimate a general class of nonparametric choice models', *Management Science* **61**(2), 281–300.

Wang, R. (2012), 'Capacitated assortment and price optimization under the multinomial logit model', *Operations Research Letters* **40**(6), 492–497.

Wang, R. (2013), 'Assortment management under the generalized attraction model with a capacity constraint', *Journal of Revenue and Pricing Management* **12**(3), 254–270.

A Appendix

A.1 Proof of Theorem 2.1

We define the minimum vertex cover problem on cubic graphs as follows.

Minimum vertex cover on cubic graphs. In this problem, we are given an undirected graph $G = (V, E)$ in which each vertex has degree three. The goal of the minimum vertex cover problem is to choose the minimum cardinality set of vertices $S \subseteq V$ such that each edge in G has at least one endpoint in S.

Proof. Consider an instance of vertex cover on a cubic graph $G = (V, E)$ with $V = \{v_1, v_2, \ldots, v_n\}$ and constant k. We construct an instance of the assortment optimization problem by creating products $\{1, 2, \ldots, n+1\}$ where $r_i = 1$ if $i \leq n$ and $r_i = 2$ if $i = n+1$. For every edge $(v_i, v_j) \in E$, we create a customer class g that first considers product $g_1 = \min(i, j)$ and then considers product $g_2 = \max(i, j)$. We refer to these customers as "edge" customers. Similarly, for each vertex $v_i \in V$, we create a customer class g that first considers product $g_1 = i$ and then considers product $g_2 = n + 1$. We refer to these customers as "vertex" customers. Edge customers will arrive with unnormalized probability 1 and vertex customers will arrive with unnormalized probability $\frac{1}{3}$. It is without loss of generality that we consider unnormalized arrival probabilities.

We will first show the claim that there exists a vertex cover of size $\leq k$ in G if and only if the optimal assortment of products in the corresponding assortment optimization problem generates an expected revenue of at least $|E| + \frac{1}{3}(k + 2(n - k))$.

Suppose there exists a vertex cover $U \subseteq V$ such that $|U| \leq k$. Then, let $S = \{i : v_i \in U\} \cup \{n+1\}$. For every edge customer (v_i, v_j) either v_i or v_j is in U so either i or j is in S, and thus we accumulate an expected revenue of 1 from each edge customer. Next, every vertex customer corresponding to $v_i \in U$ generates an expected revenue of $\frac{1}{3}$ since these customers will purchase product i. On the other hand, every vertex customer corresponding to $v_i \notin U$ generates an expected revenue $\frac{2}{3}$ since these customers purchase product $n + 1$. Thus, $\text{Rev}(S) = |E| + \frac{1}{3}(|U| + 2(n - |U|)) \geq |E| + \frac{1}{3}(k + 2(n - k))$.

Let S^* be the optimal solution to the assortment problem and assume that $\text{Rev}(S^*) \geq |E| + \frac{1}{3}(k + 2(n - k))$. Without loss of generality, S^* contains product $n + 1$ since adding this product can only increase the revenue. Suppose that there exists an edge customer (v_i, v_j) such that neither i nor j are in S^*. Then, by adding i to S^* we gain an expected revenue of at least 1 from this customer. However, now the vertex customer associated with vertex v_i also purchases product i yielding expected revenue $\frac{1}{3}$ instead of product $n+1$ with expected revenue $\frac{2}{3}$. Overall the net gain in revenue is at least $\frac{1}{3}$ which contradicts the optimality of S^*. This shows that all edge customer classes must make a purchase when S^* is offered and thus S^* will be a vertex cover. Furthermore, we know that

$$\begin{aligned}\text{Rev}(S^*) &= |E| + \frac{1}{3}(|S^*| + 2(n - |S^*|)) \\ &\geq |E| + \frac{1}{3}(k + 2(n-k)),\end{aligned}$$

where the inequality follows from our assumption on the value of $\text{Rev}(S^*)$. Thus, we must have that $|S^*| \leq k$, and we have found a vertex cover of size $\leq k$. This establishes the claim.

In particular, this shows that if k is the size of a minimum vertex cover then the optimal revenue in the assortment problem that we create above is $\text{OPT} = |E| + \frac{1}{3}(k + 2(n-k))$. Let S_2 be an assortment that includes product $n+1$ and achieves the second largest revenue amongst such assortments. In other words, we have that $n+1 \in S_2$ and $\text{Rev}(S_2) = \text{OPT}_2 < \text{OPT}$. Suppose S_2 is a vertex cover in the original vertex cover problem. Then, $\text{OPT}_2 = |E| + \frac{1}{3}(|S_2| + 2(n - |S_2|)) < |E| + \frac{1}{3}(k+2(n-k))$ since $\text{Rev}(S_2) < \text{OPT}$. This implies that $\text{OPT}_2 \leq |E| + \frac{1}{3}(k+1+2(n-k-1))$. Now suppose that S_2 is not a vertex cover. If more than two edge customers do not make a purchase, then we can improve the revenue while maintaining that S_2 is not a valid vertex cover. Hence S_2 could not have been the assortment including product $n+1$ that achieves the second highest revenue. Let $(v_i, v_j) \in E$ be the edge associated with the edge customer that does not make a purchase. By the argument made above, adding the product corresponding to v_i or v_j to S_2 improves the revenue and is not optimal since S_2 will still not be a vertex cover. Therefore, S_2 must have exactly one edge customer not making a purchase, which implies that $|S_2| \geq k-1$ since the minimum vertex cover has size k. Thus, $\text{OPT}_2 \leq |E| - 1 + \frac{1}{3}(k-1+2(n-k+1)) < |E| + \frac{1}{3}(k+1+2(n-k-1))$. Combining these two cases we get that $\text{OPT}_2 \leq |E| + \frac{1}{3}(k+1+2(n-k-1))$. Thus,

$$\begin{aligned}\frac{\text{OPT}_2}{\text{OPT}} &\leq \frac{|E| + \frac{1}{3}(k+1+2(n-k-1))}{|E| + \frac{1}{3}(k+2(n-k))} \\ &= 1 - \frac{1/3}{3n/2 + \frac{1}{3}k + \frac{2}{3}(n-k)} \\ &\leq 1 - \frac{1/3}{3n/2 + \frac{1}{3}n + \frac{2}{3}n} \\ &= 1 - \frac{1}{7.5n}\end{aligned}$$

where the equality comes from the fact that G is a cubic graph and $|E| = 3n/2$. Thus, if one can produce an assortment that is within a factor of $1 - \frac{1}{7.5n}$ of OPT, then this assortment must actually be optimal.

Suppose there exists an FPTAS for the assortment problem under the 2-product nonparametric choice model. We can assume without loss of generality that the FPTAS always returns an assortment that includes product $n+1$ since adding this product to any assortment improves the revenue. For any input to the vertex cover problem on a cubic graph, we can transform the problem into an assortment optimization problem as above and set $\varepsilon < \frac{1}{7.5n}$. Note that $1/\varepsilon = \mathrm{O}(n)$. Then, the FPTAS would be guaranteed to return the optimal solution, which in turn will reveal the optimal solution to the vertex cover problem in polynomial time. This is a contradiction since vertex cover on cubic graphs is NP-hard. □

A.2 Proof of Lemma 3.1

Proof. Given the form of this integer program, which contains at most two variables in each inequality, Hochbaum et al. (1993) show that there exists an optimal half-integral solution to the linear programming relaxation. We additionally show that all basic feasible solutions are half-integral. The proof will follow a similar argument to that by Nemhauser and Trotter (1975), who show that basic feasible solutions for the vertex cover problem LP are half-integral.

Assume by way of contradiction that (x^*, y^*) is not half-integral. We will show that (x^*, y^*) is not a basic solution. First, for any $\varepsilon > 0$, let

$$\bar{x}_i = \begin{cases} x_i^* + \varepsilon & \text{if } 0 < x_i^* < \frac{1}{2}, \\ x_i^* - \varepsilon & \text{if } \frac{1}{2} < x_i^* < 1, \\ x_i^* & \text{otherwise.} \end{cases} \qquad x_i' = \begin{cases} x_i^* - \varepsilon & \text{if } 0 < x_i^* < \frac{1}{2}, \\ x_i^* + \varepsilon & \text{if } \frac{1}{2} < x_i^* < 1, \\ x_i^* & \text{otherwise.} \end{cases}$$

Furthermore, for all $g \in \mathcal{G}$ and $j \in \{1, 2, \ldots, k\}$,

$$\bar{y}_{g,g_j} = \begin{cases} y_{g,g_j}^* + \varepsilon & \text{if } 0 < y_{g,g_j}^* < \frac{1}{2}, \\ y_{g,g_j}^* - \varepsilon & \text{if } \frac{1}{2} < y_{g,g_j}^* < 1, \\ y_{g,g_j}^* & \text{otherwise.} \end{cases} \qquad y_{g,g_j}' = \begin{cases} y_{g,g_j}^* - \varepsilon & \text{if } 0 < y_{g,g_j}^* < \frac{1}{2}, \\ y_{g,g_j}^* + \varepsilon & \text{if } \frac{1}{2} < y_{g,g_j}^* < 1, \\ y_{g,g_j}^* & \text{otherwise.} \end{cases}$$

Clearly, $x^* = \frac{1}{2}\bar{x} + \frac{1}{2}x'$ and $y^* = \frac{1}{2}\bar{y} + \frac{1}{2}y'$. Therefore, we just need to show both $(\bar{x}, \bar{y})$ and (x', y') are feasible solutions to the LP for small enough ε to show that (x^*, y^*) cannot be a basic feasible solution.

We start with $(\bar{x}, \bar{y})$. Note that by setting ε small enough we maintain that $0 \leq \bar{x} \leq 1$ and $0 \leq \bar{y} \leq 1$. It remains to show the feasibility of the two LP inequalities. Consider the first inequality for $g \in \mathcal{G}$ and $1 \le j \le k$. Note that if $y^*_{g,g_j} < x^*_{g_j}$, then for small enough ε this constraint will continue to hold for $(\bar{x}, \bar{y})$. On the other hand, if $y^*_{g,g_j} = x^*_{g_j}$ then either neither LP value is adjusted or y^*_{g,g_j} is adjusted in the same direction as $x^*_{g_j}$. Again, this implies that the constraint continues to hold for $(\bar{x}, \bar{y})$. Similarly, consider the second LP inequality for $g \in \mathcal{G}$ and $1 \le i < j \le k$. If $y^*_{g,g_j} < 1 - x^*_{g_i}$, then for small enough ε this constraint will continue to hold for $(\bar{x}, \bar{y})$. On the other hand, if $y^*_{g,g_j} = 1 - x^*_{g_i}$, then either neither LP value is adjusted or y^*_{g,g_j} is adjusted in the opposite direction as $x^*_{g_i}$ and the constraint will continue to hold. Thus, for small enough ε, $(\bar{x}, \bar{y})$ is feasible. A similar argument can be made for (x', y'). □

A.3 Optimality Gap of IP1

Consider an instance of the k-product nonparametric choice model with a single customer class that arrives with probability 1 and whose preference list is $[1, 2, \ldots, k]$. We also assume that the revenue of each product is 1. In this case, we trivially get that OPT = 1. Further, the solution $y_{1,i} = x_i = 1/2,\ i = 1, \ldots, k$ is feasible to the LP-relaxation of (IP1) and this solution has an objective value of $k/2$ giving us the desired result.

A.4 Proof of Theorem 3.5

Proof. For customer class g, and consider the revenue the random assortment garners from product g_j. If $x^*_{g_i} = 1$ for any $i < j$, then the algorithm always offers product g_i and does not gain any revenue from g_j. Similarly, by the third constraint in the LP, we have that $y^*_{g,g_j} = 0$ and the LP does not gain any revenue from this product either. Similarly, if $x^*_{g_j} = 0$, then the algorithm does not gain any revenue from this product and, by the second constraint in the LP, we have that $y^*_{g,g_j} = 0$ and so neither does the LP.

In the remaining cases, we offer product g_i with probability 0 if $x^*_{g_i} = 0$ and probability $\frac{1}{2k} + \frac{1}{k} x^*_{g_i}$ if $0 < x^*_{g_i} < 1$, for all $i < j$. Additionally, we offer product g_j with probability 1 if $x^*_{g_j} = 1$ and probability $\frac{1}{2k} + \frac{1}{k}x^*_{g_j}$ if $0 < x^*_{g_j} < 1$. This yields expected revenue

$$\begin{aligned}
&\mathbb{E}[\lambda_g r_{g_j}(1-\bar{x}_{g_1})(1-\bar{x}_{g_2})\dots(1-\bar{x}_{g_{j-1}})\bar{x}_{g_j}] \\
&\geq \lambda_g r_{g_j}\left(1-\frac{1}{2k}-\frac{x^*_{g_1}}{k}\right)\left(1-\frac{1}{2k}-\frac{x^*_{g_2}}{k}\right)\dots\left(1-\frac{1}{2k}-\frac{x^*_{g_{j-1}}}{k}\right)\left(\frac{1}{2k}+\frac{x^*_{g_j}}{k}\right) \\
&= \lambda_g r_{g_j}\left(1-\frac{3}{2k}+\frac{1-x^*_{g_1}}{k}\right)\left(1-\frac{3}{2k}+\frac{1-x^*_{g_2}}{k}\right)\dots\left(1-\frac{3}{2k}+\frac{1-x^*_{g_{j-1}}}{k}\right)\left(\frac{1}{2k}+\frac{x^*_{g_j}}{k}\right) \\
&\geq \lambda_g r_{g_j}\left(1-\frac{3}{2k}+\frac{y^*_{g,g_j}}{k}\right)^{j-1}\left(\frac{1}{2k}+\frac{y^*_{g,g_j}}{k}\right) \\
&\geq \lambda_g r_{g_j}\left(1-\frac{3}{2k}+\frac{y^*_{g,g_j}}{k}\right)^{k-1}\left(\frac{1}{2k}+\frac{y^*_{g,g_j}}{k}\right),
\end{aligned}$$

where the second to last inequality is implied by the second and third constraints of the LP.

From customer class g and product g_j, the LP gets $\lambda_g r_{g_j} y^*_{g,g_j}$. Thus, the worst-case approximation ratio with respect to the LP value is given by the following expression:

$$\frac{\lambda_g r_{g_j}\left(1-\frac{3}{2k}+\frac{y^*_{g,g_j}}{k}\right)^{k-1}\left(\frac{1}{2k}+\frac{y^*_{g,g_j}}{k}\right)}{\lambda_g r_{g_j} y^*_{g,g_j}} \geq \min_{0\leq y\leq 1}\left(1-\frac{3}{2k}+\frac{y}{k}\right)^{k-1}\left(\frac{1}{2ky}+\frac{1}{k}\right).$$

The derivative of the objective function of the minimization problem above with respect to y is given by

$$\frac{k-1}{k}\left(1-\frac{3}{2k}+\frac{y}{k}\right)^{k-2}\left(\frac{1}{2ky}+\frac{1}{k}\right)-\frac{1}{2ky^2}\left(1-\frac{3}{2k}+\frac{y}{k}\right)^{k-1}.$$

Setting this expression equal to zero and simplifying,

$$\frac{k-1}{k}\left(\frac{1}{2ky}+\frac{1}{k}\right)-\frac{1}{2ky^2}\left(1-\frac{3}{2k}+\frac{y}{k}\right)=0.$$

Multiplying by $2k^2y^2$,

$$2(k-1)y^2+(k-2)y-k+\frac{3}{2}=0.$$

This quadratic has roots $\frac{1}{2}$ and $-\frac{k-1/2}{k-1}$. Further, the second derivative of the objective in the minimization problem is

$$\frac{(k-2)(k-1)}{k^2}\left(\frac{y}{k}-\frac{3}{2k}+1\right)^{k-3}+\frac{1}{ky^3}\geq 0.$$

Since the latter root is negative and the function is convex, we can easily see that the function is minimized at $y=\frac{1}{2}$ which gives approximation ratio

$$2\left(1-\frac{1}{k}\right)^{k-1}\frac{1}{k}.$$

□

Thus, by tightening the LP constraints we maintain that the performance guarantee of our LP-rounding algorithm is $2(1-\frac{1}{k})^{k-1}\frac{1}{k}$.

A.5 Tightness of the Approximation Guarantee

Theorem 3.5 shows that rounding the LP-relaxation of IP2 yields a $[2(1-\frac{1}{k})^{k-1}\frac{1}{k}]$-approximation algorithm. In contrast, in this section we show that for $k = 2$ and all $\epsilon > 0$, it is possible to construct an instance such that the approximation ratio is $1/2 + \epsilon$. This implies that the approximation guarantee is tight. Moreover, adding a cardinality constraint with $c = 2$ in this example immediately shows that it is not possible to improve the performance guarantee of our LP-based rounding for the cardinality constrained assortment problem with $k = 2$ beyond 1/2, although our analysis establishes a performance guarantee of only 1/4 for the cardinality constrained case.

Let $0 < p < 1$ and let $N = \{1, 2, 3, 4\}$ with revenues $[0, 0, 1-p, 1]$, respectively. Customer classes consist of the following preference lists $[[1, 2], [1, 3], [2, 3], [3, 4]]$ with the following associated probabilities $[0, p, 0, 1-p]$. In this case, one optimal solution to the assortment problem is $S^* = \{1, 4\}$ (although there are other optimal solutions) with expected revenue $1 - p$. On the other hand, the solution that places value 1/2 on all products is an optimal basic feasible solution to the LP-relaxation of IP2 with objective value also equal to

$$p(1-p)/2 + (1-p)[(1-p)/2 + 1/2] = 1-p.$$

Further, the expected revenue of the rounded assortment is

$$\frac{1}{4}p(1-p) + \frac{1}{2}(1-p)^2 + \frac{1}{4}(1-p) = (1-p)(\frac{3}{4} - \frac{1}{4}p).$$

Thus, the approximation ratio is $\frac{3}{4} - \frac{1}{4}p$, which goes to $\frac{1}{2}$ as $p \mapsto 1$.

A.6 Reduction to Maximum Directed Cut

In this section, we show that the 2-product nonparametric choice model can be reduced to an instance of the maximum directed cut problem. Consider an instance of the assortment optimization problem under the 2-product nonparametric choice model. We construct a corresponding directed graph G. First, for each product i we construct a vertex i. We also add a special vertex 0. Next, for each customer class $g \in \mathcal{G}$ with first choice product g_1 and second choice product g_2, we construct a directed edge $(g_1, 0)$ of weight $\lambda_g r_{g_1}$ and a directed edge (g_2, g_1) of weight $\lambda_g r_{g_2}$. We show there is a bijection between cuts in our constructed instance of the maximum directed cut problem and assortments in the assortment optimization problem under the 2-product nonparametric choice model. Further, the objective function value of each cut is equal to the revenue of the corresponding assortment.

Let $C \subseteq V$ be an arbitrary directed cut. Without loss of generality, we may assume that $0 \notin C$ since vertex 0 has no outward edges and it can only improve the solution to remove 0 from C. For any customer class $g \in \mathcal{G}$, the edge $(g_1, 0)$ of weight $\lambda_g r_{g_1}$ is in the cut if and only if $g_1 \in C$, and the edge (g_2, g_1) of weight $\lambda_g r_{g_2}$ is in the cut if and only $g_1 \notin C$ and $g_2 \in C$. Thus, the weight of the cut C is equal to the revenue of the assortment that includes each of the products corresponding to the vertices in C.

Similarly, we can show that the revenue for any assortment $S \subseteq N$ is equal to the weight of the cut C that includes all of the vertices of the products in S. Consider the revenue gained from arbitrary customer class $g \in \mathcal{G}$ under assortment S in relation to the weight gained by the cut C from the edges corresponding to g. If product $g_1 \in S$, then the revenue gain is $\lambda_g r_{g_1}$ and the cut similarly gains a weight of the same magnitude. If $g_1 \notin S$ and $g_2 \in S$, then the contribution of customer class g to the objective function in each problem is $\lambda_g r_{g_2}$. Finally, if $g_1, g_2 \notin S$ then this customer class does not contribute to either objective. Thus, finding the maximum directed cut is equivalent to finding an optimal assortment.

A.7 Proof of Lemma 4.3

Proof. Slightly abusing notation let $x \in [0,1]^n$ be the probability that product i is offered. Then, the expected revenue of the corresponding random assortment is given by $F(x) := \sum_g \lambda_g r_{g_1} x_{g_1} + \lambda_g r_{g_2}(1 - x_{g_1}) x_{g_2}$ and the expected size of the assortment is $\sum_{i=1}^n x_i$. To start, set x_i to be the probability i is offered in $\bar{x}$. Then, the expected size of the assortment is at most $c - 1$. While x is not an integral assortment, choose two products i and j such that x_i and x_j are both fractional. We account for the scenario in which there is only one fractional variable remaining at the end of the proof. Consider the following modified solution for a carefully chosen ε: $x_\varepsilon = x + \varepsilon(e_i - e_j)$, where $e_l \in \mathbb{R}^n$ is a vector of all zeros except for a one in the lth position. If $\varepsilon > 0$, this new solution adds ε to x_i and subtracts ε from x_j and hence the expected size of the assortment is unchanged. If $\varepsilon < 0$, then this solution subtracts ε from x_i and adds ε to x_j. Again, the expected size of the assortment is unchanged.

We now analyze the change $F(x_\varepsilon) - F(x)$ resulting from moving to this new solution. For any customer class $g \in \mathcal{G}$ and product $g_1 \in \{i, j\}$, the change in $\lambda_g r_{g_1} x_{g_1}$ is $+\lambda_g r_{g_1} \varepsilon$ if $g_1 = i$ and $-\lambda_g r_{g_1} \varepsilon$ if $g_1 = j$. In either case, the change is linear in ε. Now consider the change in $\lambda_g r_{g_2}(1 - x_{g_1}) x_{g_2}$, the expected revenue from the second choice product. If $g_2 \in \{i, j\}$ but $g_1 \notin \{i, j\}$, the change is either $+\lambda_g r_{g_2}(1 - x_{g_1})\varepsilon$ or $-\lambda_g r_{g_2}(1 - x_{g_1})\varepsilon$, and if $g_1 \in \{i, j\}$ but $g_2 \notin \{i, j\}$, the change is either $+\lambda_g r_{g_2} x_{g_2}\varepsilon$ or $-\lambda_g r_{g_2} x_{g_2}\varepsilon$. In either case, the change is linear in ε. Lastly, if $g_1 = i$ and $g_2 = j$, the change is $\lambda_g r_{g_j}[-(1-x_i)\varepsilon - x_j\varepsilon + \varepsilon^2]$, and if $g_1 = j$ and $g_2 = i$, the change is $\lambda_g r_{g_j}[(1-x_i)\varepsilon + x_j\varepsilon + \varepsilon^2]$.

Thus, $F(x_\varepsilon) - F(x)$ is a quadratic function in ε in which the quadratic coefficient is positive and the constant term is zero. We can therefore chose ε to ensure that $F(x_\varepsilon) - F(x) > 0$ by simply considering the sign of the linear coefficient. If the linear coefficient of ε is positive, we set $\varepsilon = \min\{1 - x_i, x_j\}$. By the equation above, this only increases the expected revenue and either x_i or x_j becomes integer valued. Otherwise, if the linear coefficient of ε is negative, we set $\varepsilon = -\min\{1 - x_j, x_i\}$. Again, this can only increase the expected revenue and at least one of the two variable becomes integer. Further, we have not increased the overall size of the assortment.

Thus, after at most n iterations there is at most one fractional value x_i. At this point, setting $x_\varepsilon = x + \varepsilon e_i$ implies that $F(x_\varepsilon) - F(x)$ is a linear function in ε with a constant term equal to zero. Hence, either setting $x_i = 0$ or $x_i = 1$ improves the expected revenue. In the end, we have transformed x into an integer solution of size not exceeding c while only increasing the expected revenue. □

A.8 Average Runtime for MLE Procedure

The average runtimes (in seconds) of the MLE procedure used to fit the k-product nonparametric choice models in the first set of experiments are given in Tables 5a and 5b for $m = 100$ and $m = 500$ respectively. Columns 1-3 give the parameter combinations that define each test case, and Column 4 reports the average computation time needed to perform the MLE procedure for $\tau = 10000$, averaged over the 10 distinct purchase histories. We report this data so that we can better assess the marginal gain from fitting increasingly rich choice models; more complex nonparametric choice models capture a broader set of purchasing behavior but require more time to estimate. We observe that the runtime increases exponentially as k increases. Nonetheless, the average runtime for NP_4 never exceeds one hour for any set of parameter combinations. It is interesting to note that as m increases from 100 to 500 there is very little change in the runtime of the MLE procedure. This shows that a greater diversity in preferences in the ground choice model does not necessarily complicate the estimation procedure.

The corresponding results for the second set of experiments are given in Tables 6a and 6b for $m = 100$ and $m = 500$, respectively. Recall that in this set of experiments, the ground choice

p_d	$\mathcal{F}$	Fitted Model	τ 10000
0	1	NP_1	10.80
0	1	NP_2	111.83
0	1	NP_3	683.16
0	1	NP_4	3082.14
0	2	NP_1	8.69
0	2	NP_2	151.99
0	2	NP_3	400.82
0	2	NP_4	3048.83
0	4	NP_1	12.09
0	4	NP_2	91.97
0	4	NP_3	373.37
0	4	NP_4	3048.83
0.25	1	NP_1	12.02
0.25	1	NP_2	78.77
0.25	1	NP_3	375.98
0.25	1	NP_4	2924.39
0.25	2	NP_1	11.94
0.25	2	NP_2	54.16
0.25	2	NP_3	391.12
0.25	2	NP_4	2968.26
0.25	4	NP_1	12.37
0.25	4	NP_2	57.22
0.25	4	NP_3	376.82
0.25	4	NP_4	2908.62

(a) Average runtimes (sec.) for $m = 100$.

p_d	$\mathcal{F}$	Fitted Model	τ 10000
0	1	NP_1	9.55
0	1	NP_2	75.42
0	1	NP_3	437.10
0	1	NP_4	3173.09
0	2	NP_1	8.82
0	2	NP_2	122.12
0	2	NP_3	378.84
0	2	NP_4	3135.45
0	4	NP_1	9.33
0	4	NP_2	77.57
0	4	NP_3	372.95
0	4	NP_4	2960.57
0.25	1	NP_1	10.11
0.25	1	NP_2	56.13
0.25	1	NP_3	382.50
0.25	1	NP_4	2912.36
0.25	2	NP_1	10.36
0.25	2	NP_2	54.65
0.25	2	NP_3	385.27
0.25	2	NP_4	2945.25
0.25	4	NP_1	11.59
0.25	4	NP_2	49.57
0.25	4	NP_3	369.83
0.25	4	NP_4	3002.94

(b) Average runtimes (sec.) for $m = 500$.

Table 5: Runtimes of the MLE procedure for the first set of experiments.

p_d	$\mathcal{F}$	Fitted Model	τ 10000
0.25	1	NP_1	8.92
0.25	1	NP_2	36.27
0.25	1	NP_3	276.70
0.25	1	NP_4	4287.07
0.25	2	NP_1	8.77
0.25	2	NP_2	35.59
0.25	2	NP_3	262.54
0.25	2	NP_4	3494.14
0.25	4	NP_1	8.34
0.25	4	NP_2	33.56
0.25	4	NP_3	287.90
0.25	4	NP_4	3550.54

(a) Average runtimes (sec.) for $m = 100$.

p_d	$\mathcal{F}$	Fitted Model	τ 10000
0.25	1	NP_1	8.55
0.25	1	NP_2	34.12
0.25	1	NP_3	302.83
0.25	1	NP_4	3639.75
0.25	2	NP_1	9.04
0.25	2	NP_2	35.48
0.25	2	NP_3	265.80
0.25	2	NP_4	3481.51
0.25	4	NP_1	8.93
0.25	4	NP_2	34.72
0.25	4	NP_3	290.09
0.25	4	NP_4	3660.24

(b) Average runtimes (sec.) for $m = 500$.

Table 6: Runtimes of the MLE procedure for the second set of experiments.

model consists of customers whose preference lists contain a larger number of products, and hence, substitution patterns are more complex. However, the runtimes of our MLE procedure remain quite reasonable, never exceeding 90 minutes.