Assortment Optimization Under the Mallows model

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 Introduction

A common problem that arises in many applications in practice is to determine the subset (or assortment) of items to offer to a user. We focus on the concrete setting in which a retailer must decide the subset of products to offer to each arriving customer (either online or in a brick-and-mortar store). The retailer chooses the subset from a large universe \( U \) of products with the objective of maximizing either the expected revenue/profit earned from each customer. Determining the best offer set requires: (a) a demand model and (b) a set optimizer. The demand model specifies the expected revenue from each offer set, and the set optimizer finds (an approximation of) the revenue maximizing subset.

The demand for each offered product very often depends on entire offer set due to product substitution behavior, characterized by a customer substituting to an available product (say, a dark blue shirt) if her most preferred product (say, a black one) is not offered. Incorporating such substitution effects can significantly improve the accuracy of demand predictions. To capture substitution, demand is modeled through a choice model that specifies the demand for product \( a \) from offer set \( S \) as the probability \( P(a|S) \) of a random customer choosing product \( a \) from offer set \( S \). A general and popularly studied class of choice models is the rank-based class, which model the mechanism through which consumers make purchase decisions through a distribution over preference lists or permutations. Particularly, in each choice occasion, the customer samples a preference list, which specifies a preference ordering over a subset of products. Then, given an offer set, she purchases the first available product on her list, which could very well be the no-purchase option.

The general rank-based model can capture complex substitution patterns but results in computationally intractable estimation and decision problems because the most general specification can have an exponentially-large number of preference lists in the support. Therefore, existing work has focused on particular parametric models over rankings and exploiting their structure to design tractable algorithms both for estimation and assortment optimization. The most commonly studied models in this context are the Plackett-Luce (PL) model and its variants, the nested logit (NL) model and the mixture of PL models. The reason for their popularity is that the assumptions made in these models (such as the Gumbel assumption) are geared towards obtaining closed-form expressions for the choice probabilities \( P(a|S) \). On the other hand, other popular models in the machine learning literature such as the Mallows model have largely been ignored because computing choice probabilities under these models is generally considered to be computationally challenging, requiring marginalization of a distribution with an exponentially-large support size.

In this paper, we focus on solving the assortment optimization problem under the Mallows model, when given the parameters of the model. The Mallows distribution was introduced in the mid-
1950’s [11] and is the most popular member of the so-called distance-based ranking models, which are characterized by a model ranking $\omega$ and a concentration parameter $\theta$. The probability that a ranking $\sigma$ is sampled falls exponentially as $e^{-\theta \cdot d(\sigma, \omega)}$. Different distance functions result in different models. The Mallows model uses the Kendall-Tau distance, which measures the number of pairwise disagreements between the two rankings. Intuitively, the Mallows model assumes that consumer preferences are concentrated around a central permutation, with the likelihood of large deviations being low.

Most existing work in machine learning has focused on estimating the parameters of the Mallows model when the observations are complete rankings [7], partitioned preferences [8] (which include top-k/bottom-k items), or a general partial-order specified in the form of a collection of pairwise preferences [9]. We focus on settings in which observations are of the form of a collection of choices (item $i$ chosen from $S$), which are often collected as part of purchase transactions. While the techniques based on complete rankings and partitioned preferences don’t apply to this context, the techniques proposed in [9] can be applied to infer the model parameters. Therefore, we primarily focus on the assortment problem, when given the model parameters. Nevertheless, we briefly discuss how our results allow us to improve the technique in [9] when the data consist of only choice observations. [VERIFY THIS].

Our results. The paper focuses on addressing the two key computational challenges: (a) computing the choice probabilities and hence, the expected revenue/profit, for a given offer set $S$ and (b) finding the optimal offer set $S^*$. Our main contribution is to propose two alternate procedures to to efficiently compute the choice probabilities $\mathbb{P}(a|S)$ under the Mallows model. As elaborated below, this is a non-trivial computational task because, even though specifying the underlying distribution requires a small number of parameters, evaluating $\mathbb{P}(a|S)$ according to its definition requires summing an exponential number of terms. In Section 3 we exploit the combinatorial structure of the Mallows distribution to derive a closed-form expression for the choice probabilities that takes the form of a discrete convolution. Using the fast Fourier transform, this yields a $O(n^2 \log n)$ time (see Theorem 3.2) method for computing the choice probabilities, where $n$ is the number of products in the universe $\mathcal{U}$.

In Section 4 we exploit the repeated insertion method (RIM) [4] for sampling rankings according to the Mallows distribution to obtain a dynamic program (DP) for computing the choice probabilities in $O(n^3)$ time (see Theorem 4.1). The key advantage of the DP specification is that the choice probabilities are expressed as the unique solution to a system of linear equations. Based on this specification, we formulate the assortment optimization problem as a compact mixed linear integer program (MIP) with $O(n)$ binary variables and $O(n^3)$ continuous variables and constraints. Using a simulation study, we show that the MIP scales well to large problem sizes and provides accurate decisions. Finally, we discuss how our techniques extend to a mixture of Mallows models.

Literature review. Numerous parametric choice models have extensively been studied in diverse areas including marketing, transportation, economics, and operations management (see [14, 2, 19], for example). The Multinomial Logit model is by far the most popular model in practical applications. It was introduced independently by Luce [10] and Plackett [16], where it was referred to as the Plackett-Luce model, later to be known as the multinomial logit model, following the work of McFadden [13]. The popularity of this model arises from the tractability of its parameter estimation and the corresponding assortment optimization problem [17].

Non-parametric choice models have also been considered in the literature. The focus has been on estimating sparse distribution over permutations [1] or on studying the assortment optimization under various non-parametric choice models [11, 6, 7]. Nevertheless, there is only a handful of models whose parameter estimation and assortment optimization problem can both be solved efficiently.

2 Model and problem statement

Notation. We consider a universe $\mathcal{U}$ of $n$ products. Specifically, in order to distinguish products from their corresponding ranks, we let $\mathcal{U} = \{a_1, \ldots, a_n\}$ denote the universe of alternatives, under an arbitrary indexing. Preferences over this universe are captured by an anti-reflexive, anti-symmetric, and transitive relation $\succ$, which induces a total ordering (or ranking) over all products; specifically, $a \succ b$ means that $a$ is preferred to $b$. We also represent preference through rankings or permutations. In particular, for any pair of integers $i \leq j$, let $[i, j]$ denote the set $\{i, i+1, \ldots, j\}$. When $i = 1,
we simply use \([j]\) to denote \([1, 2, \ldots, j]\). A complete ranking (or simply a ranking) is a bijection \(\sigma: \mathcal{W} \to [n]\) that maps each product \(a \in \mathcal{W}\) to its rank \(\sigma(a) \in [n]\). Lower ranks indicate higher preference so that \(\sigma(a) < \sigma(b)\) if and only if \(a \succ_\sigma b\), where \(\succ_\sigma\) denotes the preference relation induced by the ranking \(\sigma\). For simplicity of notation, we also let \(\sigma_i\) denote the product ranked at position \(i\). Thus, \(\sigma_1 \sigma_2 \cdots \sigma_n\) is the list of the products written by increasing order of their ranks.

**Mallows model.** The Mallows model is a member of the distance-based ranking family models \([15]\). This model is described by a location parameter \(\omega\), which denotes the central permutation, and a scale parameter \(\theta \in \mathbb{R}_+\), such that the probability of each permutation \(\sigma\) is given by

\[
\lambda(\sigma) = \frac{e^{-\theta \cdot d(\sigma, \omega)}}{\psi(\theta)},
\]

where \(\psi(\theta) = \sum_\sigma \exp(-\theta \cdot d(\sigma, \omega))\) is the normalization constant, and \(d(\cdot, \cdot)\) is the Kendall-Tau metric of distance between permutations defined as

\[
d(\sigma, \omega) = \sum_{i < j} 1[(\sigma(a_i) - \sigma(a_j)) \cdot (\omega(a_i) - \omega(a_j)) < 0].
\]

In other words, the \(d(\sigma, \omega)\) counts the number of pairs of products that are oppositely ranked by the permutations \(\sigma\) and \(\omega\). It is easy to verify that \(d(\cdot, \cdot)\) is a distance function, which is right-invariant under the composition of the symmetric group, i.e., \(d(\pi_1, \pi_2) = d(\pi_1 \pi, \pi_2 \pi)\) for every \(\pi, \pi_1, \pi_2\), where the composition \(\pi \pi\) is defined as \(\pi \pi(a) = \pi(\pi(a))\). This symmetry can be exploited to show that the normalization constant \(\psi(\theta)\) has a closed-form expression \([12]\) (Srikanth, since this is a book, you can add a specific section or page number reference) given by

\[
\psi(\theta) = \prod_{i=1}^{n+1} \frac{1 - e^{-\theta}}{1 - e^{-\theta}}.
\]

Note that \(\psi(\theta)\) depends only on the scale parameter \(\theta\) and does not depend on the location parameter \(\omega\).

Intuitively, the Mallows model defines a set of consumers whose preferences are “similar”, in the sense of being centered around a common permutation, where the probability for deviations thereof are decreasing exponentially. The similarity of consumer preferences is captured by the Kendall-Tau distance metric.

**Problem statement.** We focus on efficiently deriving the probability that a given product \(a\) will be chosen from an offer set \(S \subseteq \mathcal{W}\) under the Mallows model. Here, when offered a subset \(S\) of products, the customer initially samples a preference list according to the Mallows model and then chooses the most preferred product from \(S\) according to this list. Therefore, the probability of choosing product \(a\) from the offer set \(S\) is given by

\[
\mathbb{P}(a|S) = \sum_\sigma \lambda(\sigma) \cdot 1[\sigma, a, S],
\]

where \(1[\sigma, a, S]\) indicates whether \(\sigma(a) < \sigma(a')\) for all \(a' \in S, a' \neq a\). It is worth noting that the above sum runs over \(n!\) preference lists, meaning that it is not clear whether \(\mathbb{P}(a|S)\) can be computed efficiently.

Without loss of generality, we assume from this point on that the products are indexed such that the central permutation \(\omega\) ranks product \(a_i\) at position \(i\), for all \(i \in [n]\).

## 3 Choice probabilities: closed-form expression

In this paper, we prove that a closed-form expression can indeed be obtained for the choice probabilities. The next theorem shows that, when the offer set is contiguous, the choice probabilities enjoy a rather simple form. Using these expressions as building blocks, we further derive a closed-form expression for general offer sets.
Theorem 3.1 (Contiguous offer set) Suppose $S = a_{[i,j]} = \{a_i, \ldots, a_j\}$ for some $1 \leq i \leq j \leq n$. Then, the probability of choosing product $a_k \in S$ under the Mallows model with location parameter $\omega$ and scale parameter $\theta$ is given by

$$\mathbb{P}(a_k|S) = \frac{e^{-\theta(k-i)}}{1 + e^{-\theta} + \cdots + e^{-\theta(j-i)}}.$$ 

The choice probability under a general offer set has a more involved structure for which additional notation will be needed. For a pair of integers $1 \leq m \leq q \leq n$, define

$$\psi(q, \theta) = \prod_{s=1}^{q-1} e^{-\theta s} \quad \text{and} \quad \psi(q, m, \theta) = \psi(m, \theta) \cdot \psi(q - m, \theta).$$

In addition, for a collection of $M$ discrete functions $h_m : \mathbb{Z} \to \mathbb{R}$, $m = 1, \ldots, M$ such that $h_m(r) = 0$ for any $r < 0$, their discrete convolution is defined as

$$(h_1 * \cdots * h_m)(r) = \sum_{r_1, \ldots, r_m = r} h_1(r_1) \cdots h_M(r_M).$$

Theorem 3.2 (General offer set) Suppose $S = a_{[i_1,j_1]} \cup \cdots \cup a_{[i_m,j_m]}$ where $i_m \leq j_m$ for $1 \leq m \leq M$ and $j_m < i_{m+1}$ for $1 \leq m \leq M - 1$. Let $G_m = a_{[j_m,i_{m+1}]}$ for $1 \leq m \leq M - 1$, $\alpha = G_1 \cup \cdots \cup G_M$, and $C = a_{[i_1,j_M]}$. Then, the probability of choosing $a_k \in a_{[i_k,j_k]}$ can be written as

$$\mathbb{P}(a_k|S) = e^{-\theta(k-i_k)} \cdot \prod_{m=1}^{M-1} \frac{\psi(|G_m|, \theta)}{\psi(|C|, \theta)} \cdot \left(f_0 * \tilde{f}_1 * \cdots * \tilde{f}_\ell * f_{\ell+1} * \cdots * f_M\right)(|G|),$$

where:

- $f_m(r) = e^{-\theta r} \cdot \frac{1}{\psi(|G_m|, r, \theta)}$, if $0 \leq r \leq |G_m|$, for $1 \leq m \leq M$.
- $\tilde{f}_m(r) = e^{\theta r} f_m(r)$, for $1 \leq m \leq M$.
- $f_0(r) = \psi(|C|, |G| - r, \theta) \cdot \frac{e^{\theta(|G|-r)^2/2}}{1 + e^{-\theta} + \cdots + e^{-\theta(|G|-1)}}$, for $0 \leq r \leq |G|$. 
- $f_m(r) = 0$, for $0 \leq m \leq M$ and any $r$ outside the ranges described above.

Proof (Sketch). At a high level, deriving the expression for a general offer set involves breaking down the probabilistic event of choosing $a_k \in S$ into simpler events for which we can use the expression given in Theorem 3.1 and then combining these expressions using the symmetries of the Mallows distribution.

For a given vector $R = (r_0, \ldots, r_M) \in \mathbb{R}^{M+1}$ such that $r_0 + \cdots + r_M = |G|$, let $h(R)$ be the set of permutations which satisfy the following two conditions:

- among all the products of $S$, $a_k$ is the most preferred.
- for all $m \in [M]$, there are exactly $r_m$ products from $G_m$ which are preferred to $a_k$. We denote this subset of products by $G_m$ for all $m \in [M]$.

This implies that there are $r_0$ products from $G$ which are less preferred than $a_k$. With this notation,

$$\mathbb{P}(a_k|S) = \sum_{R: r_0 + \cdots + r_M = |G|} \sum_{\sigma \in h(R)} \lambda(\sigma).$$

Recall that for all $\sigma$, we have

$$\lambda(\sigma) = \frac{e^{-\theta \sum_{i,j} \xi(i,j)}}{\psi(\theta)},$$
where \( \xi(\sigma, i, j) = 1[\sigma(a_i) - \sigma(a_j)] \cdot (\omega(a_i) - \omega(a_j)) < 0 \). For all \( \sigma \), we can break down the sum in the exponential as follows:

\[
\sum_{i,j} \xi(\sigma, i, j) = C_1(\sigma) + C_2(\sigma) + C_3(\sigma),
\]

where,

- \( C_1(\sigma) \) contains pairs of products \((i, j)\) such that \( a_i \in \tilde{G}_m \) for some \( m \in [M] \) and \( a_j \in S \).
- \( C_2(\sigma) \) contains pairs of products \((i, j)\) such that \( a_i \in \tilde{G}_m \) for some \( m \in [M] \) and \( a_j \in G_{m'} \setminus \tilde{G}_m \) for some \( m \neq m' \).
- \( C_3(\sigma) \) contains the remaining pairs of products.

For a fixed \( R \), we show that \( C_1(\sigma) \) and \( C_2(\sigma) \) are constant for all \( \sigma \in h(R) \).

**Part 1.** \( C_1(\sigma) \) counts the number of disagreements (i.e., number of pairs of products that are oppositely ranked in \( \sigma \) and \( \omega \) ) between some product in \( S \) and some product in \( \tilde{G}_m \) for any \( m \in [M] \). For all \( m \in [M] \), a product in \( a_i \in \tilde{G}_m \) induces a disagreement with all product \( a_j \in S \) such that \( j < i \).

Therefore, the sum of all these disagreements is equal to,

\[
C_1(\sigma) = \sum_{m=1}^{M} \sum_{a_i \in S} \sum_{a_j \in \tilde{G}_m} \xi(\sigma, i, j) = \sum_{m=1}^{M} m \cdot (j_m - i_1 + 1).
\]

**Part 2.** \( C_2(\sigma) \) counts the number of disagreements between some product in any \( \tilde{G}_m \), and some product in any \( G_{m'} \setminus \tilde{G}_m \) for \( m' \neq m \). The sum of all these disagreements is equal to,

\[
C_2(\sigma) = \sum_{m \neq m'} \sum_{a_i \in \tilde{G}_m} \sum_{a_j \in G_{m'} \setminus \tilde{G}_m} \xi(\sigma, i, j)
\]

Consequently, for all \( \sigma \in h(R) \), we can write \( d(\sigma, \omega) = C_1(R) + C_2(R) + C_3(\sigma) \) and therefore,

\[
\mathbb{P}(a_k|S) = \sum_{R: r_0, \ldots, r_M = |G|} e^{-\theta C_1(R) + C_2(R)} \cdot \sum_{\sigma \in h(R)} e^{-\theta C_3(\sigma)}. 
\]

Computing the inner sum requires a similar but more involved partitioning of the permutations as well as using Theorem 3.1. We defer the details to a full version of the paper. In particular, we can show that for a fixed \( R, \sum_{\sigma \in h(R)} e^{-\theta C_3(\sigma)} \) is equal to

\[
\psi(|G| - m_0, \theta) \cdot \psi(|S| + m_0, \theta) \cdot \frac{e^{-\theta (k-1 - \sum_{m=1}^{M} r_m)}}{1 + \ldots + e^{-\theta (|S| + m_0 - 1)}} \cdot \prod_{m=1}^{M} \frac{\psi(|G_m|, \theta)}{\psi(r_m, \theta) \cdot \psi(|G| - r_m, \theta)}.
\]

Putting all the pieces together yields the desired result.

Due to representing \( \mathbb{P}(a_k|S) \) as a discrete convolution, we can efficiently compute this probability using fast Fourier transform in \( O(n^2 \log n) \) time [3], which is a dramatic improvement over the exponential sum (1) that defines the choice probabilities. Although Theorem 3.2 allows us to compute the choice probabilities in polynomial time, it is not directly useful in solving the assortment optimization problem under the Mallows model. To this end, we present an alternative (and slightly less efficient) method for computing the choice probabilities by means of dynamic programming.
4 Choice probabilities: a dynamic programming approach

In what follows, we present an alternative algorithm for computing the choice probabilities. Our approach is based on an efficient procedure to sample a random permutation according to a Mallows distribution with location parameter $\omega$ and scale parameter $\theta$. The random permutation is constructed sequentially, as explained in Algorithm 1.

Algorithm 1 Repeated insertion procedure
1: Let $\sigma = \{a_1\}$.
2: For $i = 2, \ldots, n$, insert $a_i$ into $\sigma$ at position $s = 1, \ldots, i$ with probability

$$p_{i, s} = \frac{e^{-\theta (i-s)}}{1 + e^{-\theta} + \cdots + e^{-\theta (i-1)}}.$$ 

3: Return $\sigma$.

Lemma 4.1 (Theorem 3 in [9]) The repeated insertion procedure generates a random sample from a Mallows distribution with location parameter $\omega$ and scale parameter $\theta$.

Based on the correctness of this procedure, we describe a dynamic program to compute the choice probabilities of a general offer set $S$. The key idea is to decompose these probabilities to include the position at which a product is chosen. In particular, for $i \leq k$ and $s \in [k]$, let $\pi(i, s, k)$ be the probability that product $a_i$ is chosen (i.e., first among products in $S$) at position $s$ after the $k$-th step of Algorithm 1. In other words, $\pi(i, s, k)$ corresponds to a choice probability when restricting $S$ to the first $k$ products, $a_1, \ldots, a_k$. With this notation, we have for all $i \in [n],$

$$\mathbb{P}(a_i | S) = \sum_{s=1}^{n} \pi(i, s, n).$$

We compute $\pi(i, s, k)$ iteratively for $k = 1, \ldots, n$. In particular, in order to compute $\pi(i, s, k + 1)$, we use the correctness of the sampling procedure. Specifically, starting from a permutation $\sigma$ that includes the products $a_1, \ldots, a_k$, the product $a_{k+1}$ is inserted at position $j$ with probability $p_{k+1, j}$, and we have two cases to consider.

Case 1: $a_{k+1} \notin S$. In this case, $\pi(k + 1, s, k + 1) = 0$ for all $s = 1, \ldots, k + 1$. Consider a product $a_i$ for $i \leq k$. In order for $a_i$ to be chosen at position $s$ after $a_{k+1}$ is inserted, one of the following events has to occur:

- $a_i$ was already chosen at position $s$ before $a_{k+1}$ is inserted, and $a_{k+1}$ is inserted at a position $\ell > s$.
- $a_i$ was chosen at position $s - 1$, and $a_{k+1}$ is inserted at a position $\ell \leq s - 1$.

Consequently, we have for all $i \leq k$

$$\pi(i, s, k + 1) = \sum_{\ell=s+1}^{k+1} p_{k+1, \ell} \cdot \pi(i, s, k) + \sum_{\ell=1}^{s-1} p_{k+1, \ell} \cdot \pi(i, s-1, k)$$

$$= (1 - \gamma_{k+1, s}) \cdot \pi(i, s, k) + \gamma_{k+1, s-1} \cdot \pi(i, s-1, k)$$

where $\gamma_{k, s} = \sum_{\ell=1}^{s} p_{k, \ell}$ for all $k, s$.

Case 2: $a_{k+1} \in S$. Consider a product $a_i$ with $i \leq k$. This product is chosen at position $s$ only if it was already chosen at position $s$ and $a_{k+1}$ is inserted at a position $\ell > s$. Therefore, for all $i \leq k$,

$$\pi(i, s, k + 1) = (1 - \gamma_{k+1, s}) \cdot \pi(i, s, k).$$

For product $a_{k+1}$, it is chosen at position $s$ only if all products $a_i$ for $i \leq k$ are at positions $\ell \geq s$ and $a_{k+1}$ is inserted at position $s$, implying that

$$\pi(k + 1, s, k + 1) = p_{k+1, s} \cdot \sum_{i=k}^{n} \sum_{\ell=s}^{n} \pi(i, \ell, k).$$

Algorithm 2 summarizes this procedure.
Algorithm 2 Computing choice probabilities

1: Let $S$ be a general offer set. Without loss of generality, we assume that $a_1 \in S$.
2: Let $\pi(1, 1, 1) = 1$.
3: For $k = 1, \ldots, n - 1$,
   (a) For all $i \leq k$ and $s = 1, \ldots, k + 1$,
   $$\pi(i, s, k + 1) = (1 - \gamma_{k+1, s}) \cdot \pi(i, s, k) + \mathbb{I}[a_{k+1} \notin S] \cdot \gamma_{k+1, s} \cdot \pi(i, s, k + 1).$$
   (b) For $s = 1, \ldots, k + 1$,
   $$\pi(k + 1, s, k + 1) = \mathbb{I}[a_{k+1} \in S] \cdot p_{k+1, s} \cdot \sum_{i \leq k} \sum_{\ell = s}^{n} \pi(i, \ell, k).$$
4: For all $i \in [n]$, return $\mathbb{P}(a_i | S) = \sum_{s=1}^{n} \pi(i, s, n)$.

Theorem 4.2 For any offer set $S$, Algorithm 2 returns the choice probabilities under a Mallows distribution with location parameter $\omega$ and scale parameter $\theta$.

This dynamic programming approach provides an $O(n^3)$ time algorithm for computing $\mathbb{P}(a | S)$ for all products $a \in S$ simultaneously. Moreover, as explained in the next section, these ideas lead to an algorithm to solve the assortment optimization problem.

5 Assortment optimization: integer programming formulation

In the assortment optimization problem, each product $a$ has an exogenously fixed price $r_a$. Moreover, there is an additional product $a_q$ that represents the outside option (no-purchase), with price $r_q = 0$. The goal is to determine the subset of products that maximizes the expected revenue, that is,

$$\max_{S \subseteq S} \sum_{a \in S} \mathbb{P}(a | S \cup \{a_q\}) \cdot r_a.$$

Building on Algorithm 2 and introducing a binary variable for each product, we get the following reformulation of the assortment optimization problem.

Theorem 5.1 Conditional on $a_1 \in S$, the following integer program computes an optimal solution to problem (2):\[
\begin{align*}
\max & \quad \sum_{i,s} r_i \cdot \pi(i, s, n) \\
\text{s.t.} & \quad \pi(1, 1, 1) = 1, \\
& \quad \pi(1, s, 1) = 0, \quad \forall s = 2, \ldots, n, \\
& \quad \pi(i, s, k + 1) = (1 - w_{k+1, s}) \cdot \pi(i, s, k) + y_{i,s,k+1}, \quad \forall i, s, \forall k \geq 2, \\
& \quad \pi(k + 1, s, k + 1) = z_{s,k+1}, \quad \forall s, \forall k \geq 2, \\
& \quad y_{i,s,k} \leq w_{k,s-1} \cdot \pi(i, s - 1, k - 1), \quad \forall i, s, \forall k \geq 2, \\
& \quad y_{i,s,k} \leq w_{k,s-1} \cdot (1 - x_k), \quad \forall i, s, \forall k \geq 2, \\
& \quad z_{s,k} \leq p_{k,s} \cdot \sum_{\ell = s}^{n} \pi(i, \ell, k - 1), \quad \forall s, \forall k \geq 2, \\
& \quad z_{s,k} \leq p_{k,s} \cdot x_k, \quad \forall s, \forall k \geq 2, \\
& \quad x_1 = 1, \\
& \quad x_q = 1, \\
& \quad x_k \in \{0, 1\}.
\end{align*}
\]

Note that this reformulation involves only $O(n^3)$ variables and constraints, with only $n$ integer variables. Moreover, we assume for simplicity that the first product of $S$ (here $a_1$) is known. Since
this product is generally not known a-priori, in order to obtain an optimal solution to problem (2),
we need to guess the first offered product and solve the above integer program for each of the \( O(n) \)
guesses.

6 Numerical experiments

References


[2] M. Ben-Akiva and S. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand,


1978.


[17] K. Talluri and G. Van Ryzin. Revenue management under a general discrete choice model of consumer

[18] G. van Ryzin and G. Vulcano. A market discovery algorithm to estimate a general class of nonparametric