Approximation Algorithms for Dynamic Assortment Optimization Models

Ali Aouad* Retsef Levi† Danny Segev‡

Abstract

We consider the single-period joint assortment and inventory planning problem with stochastic demand and dynamic substitution across products, motivated by applications in highly differentiated markets, such as online retailing and airlines. This class of problems is known to be notoriously hard to deal with from a computational standpoint. In fact, prior to the present paper, only a handful of modeling approaches were shown to admit provably-good algorithms, at the cost of strong restrictions on customers' choice outcomes. Our main contribution is to provide the first efficient algorithms with provable performance guarantees for a broad class of dynamic assortment optimization models. Under general rank-based choice models, our approximation algorithm is best-possible with respect to the price parameters, up to lower-order terms. In particular, we obtain a constant-factor approximation under horizontal differentiation, where product prices are uniform. In more structured settings, where the customers' ranking behavior is motivated by price and quality cues, we derive improved guarantees through tailor-made algorithms. In extensive computational experiments, our approach dominates existing heuristics in terms of revenue performance, as well as in terms of speed, given the myopic nature of our methods. From a technical perspective, we introduce a number of novel algorithmic ideas of independent interest, and unravel hidden relations to submodular maximization.

Keywords: Assortment Planning, Inventory Management, Choice Models, Dynamic Optimization, Approximation Algorithms, Submodularity.

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

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

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

1 Introduction

The challenge of designing an offer set to meet uncertain demand, with heterogeneous preferences across products, is a key driver of strategic and operational success in many industries [9]. In layman's terms, an offer set corresponds to choosing an assortment, that specifies the subset of products offered to customers, as well as deciding on their inventory levels, i.e., the initial number of units stocked of each product. Typically, inventory and assortment decisions are very likely to inform each other. Indeed, as demand evolves during the selling season, some products may be fully consumed (or stock-out) due to inventory limitations. Consequently, customers arriving at different times could very well be facing different assortments, thereby affecting their purchasing behavior. Despite this fundamental interplay between inventory and assortment, the corresponding decision-making problems have been studied for the most part in separate frameworks, arguably due to their computational hardness.

In this paper, we consider the single-period joint assortment planning and inventory management problem, whereby a retailer wishes to maximize expected revenue in the face of stochastic demand, consisting of a random number of customers with dynamic substitution across products. A complete mathematical description of this model, along with additional technical terms mentioned below (including the IFR property), are given in Section 1.2. Somewhat informally, the retailer selects an assortment of products, and determines their initial inventory levels, given a capacity constraint on the total number of units to be stocked. These decisions are made at the beginning of the selling period. Next, the consumption process consists in a random sequence of arriving customers, each of which purchases at most one unit. The problem formulation largely depends on the distribution of the number of arriving customers, also named the demand, and on the probabilistic structure of their purchasing preferences. This preference distribution, also called a choice model, describes the probability that a unit of any given product is purchased, amongst the assortment products available upon the customer's arrival.

In recent years, the availability of highly detailed purchase data has motivated a generic approach to modeling the customer choice preferences through a distribution over ranked preference lists [6, 21]. Generally speaking, each list describes a ranking over the product alternatives considered by a customer, among which she chooses her most preferred product out of those available upon arrival, or leaves without making any purchase if none is currently stocked. Unfortunately, as choice models become more detailed, and subsume more general families of preference list distributions, the corresponding optimization problems rapidly become intractable. In this paper, we develop general-purpose algorithms under distributions over preference lists. Our approach enjoys best-possible theoretical guarantees and demonstrates superior practical performance against existing heuristics. Moreover, we show that additional structure on customer choices, motivated by real-life purchasing behaviors, can be leveraged to obtain improved guarantees.

Directly-related work. With a single customer arrival, the fundamental problem considered in this paper reduces to the well-known 'static' assortment optimization problem. Here, the objective is to compute an assortment of products that maximizes the expected revenue due to one representative customer. However, even in this seemingly simple context, the problem was proven to be APX-hard by Goyal et al. [11], and in fact, was recently shown by Aouad et al. [1] to be inapproximable within any factor sub-linear in the number of products. On the positive side, there has been an ever-growing line of work in recent revenue management literature on the static variant, investigating tractable approaches for special cases of the aforementioned choice model [2, 4, 14, 16, 23, 27]. In contrast, under multiple customer arrivals, the problem becomes considerably harder due to the additional 'dynamic' aspect. Indeed, the initial assortment is altered along the sequence of arrivals due to stock-out events, and the dynamic substitution behavior of customers depends on each sample path realization. In fact, even the evaluation of the expected revenue for a given offer set is by itself an open and challenging question.

For the dynamic models in question, most of the work we are aware of makes use of heuristics based on continuous relaxations and probabilistic assumptions [13, 15, 17, 19, 26, 28]. In particular, these approaches either give rise to exponential-time algorithms, do not admit provable non-trivial approximation guarantees, or apply to inventory models of very different nature. Interestingly, even in rather restricted settings, such as that of horizontally differentiated products (where prices are uniform), not much is known at present time. In this context, Gaur and Honhon [10] studied a newsvendor-like inventory model with dynamic substitution, and devised a heuristic for the locational choice model, assuming a parametric demand distribution. While they propose lower and upper bounds on the optimal revenue, the analysis thereof does not translate into an efficient algorithm with theoretical guarantees. Chen and Bassok [3] considered a similar setting where prices are uniform, assuming a less-realistic allocation rule, where products are assigned to customers at the retail's discretion after seeing the entire sequence of arriving customers, instead of sequentially.

To our knowledge, prior to this paper, the work of Goyal et al. [11] and Segev [24] are the only papers that study dynamic optimization models with discrete demand realizations and a fully stochastic sequence of arrivals through the lens of approximation algorithms. Specifically, Goyal et al. devised a polynomial-time approximation scheme (PTAS) assuming that the demand follows an increasing failure rate (IFR) distribution, when the choice model consists in a distribution over nested preference lists, referred to in the sequel as the nested choice model. This model assumes that customers always prefer cheaper products, and have a budget constraint that dictates the most expensive product they are willing to purchase. Their algorithm is based on efficient enumeration methods, by observing that there are near-optimal assortments comprised of a constant number of products.

While this algorithm approximates the optimal revenue within factor $1 - \epsilon$, the overall approach suffers from two major drawbacks. First, since it resorts to enumerating all inventory levels over a predetermined assortment with $\mathrm{poly}(1/\epsilon)$ products, the resulting running time is exponential in $1/\epsilon$, and becomes impractical even for medium-scale instances. Second, in some practical settings the demand is 'heavy tailed', while retailers wish to hedge against extreme demand realizations, especially for newly launched products with limited data and forecast accuracy. However, the algorithm and its analysis do not carry over when the IFR property is relaxed. The latter drawback has been bypassed by Segev [24], who proposed a quasi-PTAS for general demand distributions (again, under the nested choice model), based on an dynamic programming approach with an approximate state space representation. However, this result is accuracy. However, the algorithm and its analysis do not carry over when the IFR property is relaxed. The latter drawback has been bypassed by Segev Segev (2015), who proposed a quasi-PTAS for general demand distributions (again, under the nested choice model), based on an dynamic programming approach with an approximate state space representation. However, this result is more theoretical in nature, and still leaves open the question of efficiently approximating nested preference lists under general demand distributions.

1.1 Our results

Our main contribution is to provide the first polynomial-time algorithms with provable approximation guarantees for a broad class of demand and choice model specifications in dynamic assortment optimization. From a technical perspective, we introduce a number of novel algorithmic ideas of independent interest, that could very well be applicable in a wide range of settings, and unravel hidden relations to submodular maximization. In addition, our algorithms employ a mixture of greedy procedures and low-dimensional dynamic programs, that are suitable for solving instances of practical nature and scale. In computational experiments, these algorithms are shown to be faster than existing heuristics by an order of magnitude and to result in substantially better expected revenues. Our main results can be briefly summarized as follows.

General rank-based choice model. We first devise an approximation algorithm in the context of horizontal differentiation, where product prices are identical, assuming an IFR demand distribution. As previously mentioned, a similar modeling approach has been taken by Gaur and Honhon Gaur and Honhon (2006) and Chen and Bassok Chen and Bassok (2008). In this setting, we obtain a constant-factor approximation, without any structural assumptions on the preference list distribution. Our algorithm, formally described in Section 2, is based on a two-step "selective-greedy" approach. In the selection step, we restrict attention to a subset of products, identified by approximately solving the underlying static problem (without inventory limitations). Next, the inventory levels are determined by greedily optimizing a multi-item newsvendor problem, whose optimal solution provides a lower bound on the optimal expected revenue. To analyze this approach, we explicitly construct a feasible solution to the latter problem that generates an $(e-1)/(4e-2)$ fraction of the optimal expected revenue.

This result extends to the general class of random utility choice models (preference list distributions with potentially exponential support), as long as there exists an efficient oracle to evaluate the purchase probabilities of products in any given assortment. As explained later on, this setting subsumes most choice models studied in the revenue management literature. Moreover, when the static version of the problem admits an approximation ratio of $\alpha \geq 1-1/e$, we argue that the above-mentioned guarantee can be improved to $\alpha/4$.

In the presence of price differentiation, the selection step is complemented by price thresholding. That is, we eliminate all products cheaper than an appropriate price threshold, and apply our horizontal differentiation algorithm on the remaining products, assuming identical prices. This approach yields an approximation guarantee of $O(\log(p_{\max}/p_{\min}))$, where $p_{\max}$ and $p_{\min}$ stand for the maximum and minimum price of any product, respectively. The latter ratio is best possible in this setting, due to the work of Aouad et al. Aouad et al. (2015), who established an $\Omega(\log(p_{\max}/p_{\min}))$ inapproximability bound even for the static problem formulation.

Intervals choice model. We proceed by proving that the above-mentioned logarithmic ratio, obtained under price differentiation, can be beaten in more structured settings. We first investigate a generalization of nested preference lists, known as the intervals choice model. This model subsumes any distribution over lists that are comprised of an interval of products, ranked by increasing price order. Such preference lists find behavioral justification in capturing common screening rules, used by customers to generate their choice set (see the survey by Hauser Hauser et al. (2009) in the marketing literature). Indeed, the intervals structure naturally arises when the customers' choice set is formed by the conjunction of a budget constraint and a quality requirement, assuming that price and perceived quality are inversely related, which is commonly observed in practical settings Zeithalm (1988).

Assuming an IFR distribution of customer arrivals, we develop in Section 3 an algorithmic approach with an approximation guarantee of $O(\log\log(p_{\max}/p_{\min}))$ under the intervals choice model. First, we show that the problem can be approximated within a factor logarithmic in the number of products. This is achieved through a recursive decomposition of the preference lists into a logarithmic number of classes, thereby creating independent and highly-structured instances, that can be approximated within constant factors. The log-logarithmic ratio is attained via a refined decomposition, where products are initially grouped into nearly-uniform price buckets, which allow us to employ our algorithm for horizontally differentiated products as a subroutine.

Nested choice model with general demand distributions. For arbitrary (non-IFR) demand distributions, we provide the first polynomial-time approximation algorithm under the nested choice model, attaining a performance guarantee of $1 - 1/e$. Our algorithm, whose specifics are given in Section 4, reveals a hidden submodular structure within this setting, and relies on a distinct selective-greedy approach (very dissimilar to the one described in Section 2). We initially eliminate sub-optimal products by leveraging solutions to multiple instances of the static version. Next, the resulting problem formed by the residual products is cast as a capacity-constrained maximization of a certain submodular function. In contrast to existing enumeration-based approaches Goyal et al. (2016), Segev (2015), our algorithm is very efficient in practice.

Computational experiments. In Section 5, we conduct extensive computational experiments on randomly-generated instances, showing that our general-purpose algorithm largely outperforms existing heuristics, in terms of both revenue and efficiency. Specifically, our approach is compared to the following heuristics:

1. A local-search heuristic based on greedily exchanging units between pairs of products.

2. A gradient-descent approach based on a continuous extension of the revenue function.

3. A discrete-greedy algorithm, where in each step a single unit is added to the product with the largest marginal expected revenue.

4. The approximation scheme of Goyal et al. (2016) for the nested choice model.

On average, the expected revenue is (on average) increased by a factor ranging between 8% and 30% relative to the best heuristic, while reducing the running time in most configurations. In structured settings, with interval and nested preference lists, our comparative experiments again validate the practicality of the algorithms we propose against existing heuristics.

Counter-examples. For each of the settings considered in Sections 2-4, we construct carefully-made instances, showing that the objective function of the respective formulations is neither concave, nor submodular. These counter-examples, described in Appendix A.3, suggest that generic optimization methods do not directly apply to the problems considered in this paper.

1.2 Problem formulation

Products and inventories. We are given a collection of n products, with per-unit selling prices $p_1 \leq \cdots \leq p_n$. In addition, there is a capacity bound of C on the total number of units to be stocked. In the (single-period) dynamic assortment problem, the retailer has to jointly decide on an assortment, i.e., a subset products to be stocked, as well as on the initial inventory levels of these products, which are not replenished later on. In other words, a feasible solution specifies the initial inventory levels of all products, represented by an integer-valued vector $U = (u_1, \dots, u_n)$ that meets the capacity constraint, $\sum_{i=1}^{n} u_i \leq C$.

The consumption process. Independently of stocking decisions, a random number of customers M arrive sequentially, where the distribution of M is assumed to be known to the decision-maker. Each customer j picks a random preference list L_j that describes a sub-collection of products in decreasing order of preference. This list is drawn from a common known distribution over a collection of preference lists, independently of the other lists and the number of customers M. Unless mentioned otherwise, this distribution is encoded explicitly as a collection of preference lists $\mathcal{L}$, that are specified as an input along with their respective probabilities. Upon arrival, each customer purchases a single unit of the most preferred product on her list available at that time. In other words, the customer first attempts to purchase her most preferred product, and if that product has stocked out or was not initially selected in the assortment, the customer substitutes to the second most preferred product, so forth and so on. If none of the products in her preference list is available, this customer leaves without purchasing any product. Therefore, at each step, the inventory vector is decremented by at most one unit, corresponding to the customer's purchasing decision.

Objective. When the sequence of customer arrivals ends, we use $\mathcal{R}(U)$ to denote the revenue resulting from an initial inventory vector U. Based on the preceding discussion, this revenue is clearly random, due to the stochasticity in the number of customers and in their choice of preference lists. The objective is to compute a feasible inventory vector, so that the expected revenue is maximized, i.e.,

$$\max_{(u_1,\dots,u_n)\in\mathbb{Z}_+^n}\left\{\mathbb{E}\left[\mathcal{R}(u_1,\dots,u_n)\right]:\sum_{i=1}^n u_i\le C\right\}\ .$$

Structural properties. We have previously discussed several structural properties that give rise to different settings studied in this paper. Below, detailed definitions of these modeling assumptions are provided.

- *Nested choice model:* This model describes distributions over a collection of preference lists $\mathcal{L}$ that consists of intervals of the form $(1,\dots,\ell)$, where $\ell\in[n]$. Namely, there are $n+1$ possible preferences lists, $(),(1),(1,2),\dots,(1,2,\dots,n)$, where the respective probabilities of these lists are arbitrary. Here, $()$ denotes the empty preference list, for customers who are not interested in purchasing any product.

- *Intervals choice model:* This model describes a more general class of preference list distributions, where $\mathcal{L}$ consists of intervals of the form $(\ell,\dots,k)$, with $1\le\ell\le k\le n$. Again, the respective probabilities of these lists can be arbitrary.

- *Increasing Failure Rate:* Here, the distribution of the number of customers M is assumed to have an increasing failure rate (IFR), meaning that $\Pr\left[M=k\right]/\Pr\left[M\ge k\right]$ is non-decreasing over the integer domain. This definition is equivalent to requiring that the sequence of random variables $[M-k|M\ge k]_{k\in\mathbb{Z}}$ is stochastically non-increasing in k. For definitions of stochastic orders and stochastic monotonicity, we refer the reader to Shaked and Shanthikumar Shaked and Shanthikumar (1994). It is worth mentioning that the IFR property is satisfied by many distributions considered in operations management applications, including Normal, Exponential, Geometric, Poisson, and Beta (for certain parameters).

Additional notation. In certain settings, we will treat the preference lists in $\mathcal{L}$ as subsets of products. Also, for each list $L_\ell\in\mathcal{L}$, we use θ_ℓ to denote the probability that it is picked by an arriving customer. Finally, we allow $\mathcal{L}$ to interchangeably designate the set of preference lists, as well as the corresponding collection of indices, $\ell=1,\dots,|\mathcal{L}|$.

2 General choice model

In what follows, we consider the most general setting, where the underlying choice model is expressed as a distribution over an arbitrary collection of preference lists. As discussed subsequently, this setting coincides with the class of random-utility choice models, and therefore subsumes most models of practical interest proposed in the literature (parametric and non-parametric).

Our approximation algorithm is introduced in an incremental way. In Sections 2.1-2.3, we begin by investigating the setting of horizontally differentiated products. Here, products are assumed to be associated with uniform prices, meaning that without loss of generality $p_1=\cdots=p_n=1$, and the retailer wishes to maximize his expected sales quantity. When the number of customers M follows an IFR distribution, we show how to efficiently compute an inventory vector that approximates the optimal expected revenue within a constant factor. For ease of presentation, we do not attempt to optimize the latter constant. It is worth noting that the uniform-price problem is NP-hard to approximate within factor larger than $1 - 1/e$, since it subsumes the maximum coverage problem Feige (1998) as a special case, even with a single customer.

Theorem 2.1. *When M is IFR-distributed and the product prices are uniform, the dynamic assortment planning problem under general preference list distributions can be approximated within factor $(e-1)/(4e-2)$ in polynomial time.*

In Section 2.4, this constant-factor approximation is leveraged as a subroutine to solve the general problem with price differentiation. Here, without loss of generality, we assume that products are indexed such that $p_1 \leq \cdots \leq p_n$. As stated in the next theorem, our performance guarantee scales logarithmically with the ratio of extremal prices. The latter ratio is best possible in this setting (up to constant terms), due a matching inapproximability bound, established by Aouad et al. Aouad et al. (2015) even for the static problem formulation.

Theorem 2.2. *When M is IFR-distributed, the dynamic assortment planning problem under general preference list distributions can be approximated within factor $O(\log(p_n/p_1))$ in polynomial time.*

2.1 Algorithm under horizontal differentiation

At a high level, our algorithm is based on two-step approach, termed selective-greedy:

- *Selection step.* First, we ignore the inventory limitations and select an assortment of products that approximates the static (single-customer) problem. To this end, we observe that the static problem is equivalent to computing a weighted maximum coverage of a set system defined by the collection of preference lists. Thus, a constant-factor approximation for the static case is obtained by a greedy allocation rule.

- *Greedy step.* Next, to allocate the inventory capacity over the assortment products in the dynamic setting, we consider a lower bound on the expected revenue, that can be viewed a multi-item newsvendor formulation. The latter (simplified) objective is then optimized greedily.

In what follows, we provide a more detailed description of the algorithm.

Step 1 (selection): approximating the static solution. Recall that $\mathcal{L}$ designates the original collection of preference lists, where each list L_ℓ is picked by a single customer with probability θ_ℓ. We begin by considering the static variant of the problem, that seeks to maximize the expected revenue extracted from a single arriving customer. Since all prices are identical, out of all subsets of products with cardinality at most C, we wish to pick one that maximizes the total probability of all preference lists in $\mathcal{L}$ that are being hit. In other words, we would like to identify a subset of products that satisfies the cardinality constraint and achieves a maximum coverage of the preference lists, where each product covers (or hits) the subset of lists that contain it. This is precisely an instance of the maximum coverage problem, that can be approximated within factor $1 - 1/e$ by a classic greedy procedure Nemhauser et al. (1978). Specifically, a single product is greedily added at each step to maximize the coverage quantity, defined as the combined probability of all lists intersecting the assortment. As a result, we use $\mathcal{Q} \subseteq [n]$ to denote the corresponding assortment picked for the static problem. In addition, we assume without loss of generality that the assortment $\mathcal{Q}$ is minimal with respect to inclusion, that is, removing any product would decrease the combined probability of preference lists that contain any of the products in $\mathcal{Q}$.

Newsvendor-like lower bound. We begin by defining $\mathcal{L}_{\mathcal{Q}} \subseteq \mathcal{L}$ as the subset of lists that intersect with the assortment $\mathcal{Q}$. We construct an assignment $\mathcal{A} : \mathcal{L}_{\mathcal{Q}} \to \mathcal{Q}$ that maps each list in $\mathcal{L}_{\mathcal{Q}}$ to its most preferred product in $\mathcal{Q}$, which exists by definition of $\mathcal{L}_{\mathcal{Q}}$. Hence, $\mathcal{A}^{-1}(i)$ is the subset of lists in which product $i \in \mathcal{Q}$ is the most preferred when faced with the assortment $\mathcal{Q}$. Note that, since $\mathcal{Q}$ is minimal with respect to inclusion (see step 1), each product in this assortment is necessarily assigned with at least one preference list in $\mathcal{L}_{\mathcal{Q}}$, meaning that $\mathcal{A}^{-1}(i) \neq \emptyset$ for every $i \in \mathcal{Q}$. We highlight a basic property attained by the assignment $\mathcal{A}$. Consider some product $i \in \mathcal{Q}$, and suppose that we are looking on a customer who has just arrived. The key observation is that, if product i has at least one unit in stock at the moment, the current customer will purchase i with probability at least $\psi_i = \sum_{\ell \in \mathcal{A}^{-1}(i)} \theta_\ell$, regardless of the inventory levels of all other products. The reason is that, for any list $\ell \in \mathcal{A}^{-1}(i)$, which occurs with probability θ_ℓ, product i is preferred over any other product in the assortment $\mathcal{Q}$, and we are not stocking any of the products in $[n] \setminus \mathcal{Q}$. Therefore, the number of units purchased from i if this product had an infinite (unlimited) inventory level is stochastically larger than $Y_i \sim B(M, \psi_i)$. However, assuming that u_i units of product i are initially stocked, we would actually be considering the truncated random variable $\bar{Y}_i(u_i) = \min\{Y_i, u_i\}$. Therefore, letting $(u_1, \dots, u_n)$ be an inventory vector stocking only products in $\mathcal{Q}$, an immediate lower bound on its expected revenue is given by:

$$\mathbb{E}\left[\mathcal{R}(u_1, \dots, u_n)\right] \geq \sum_{i \in \mathcal{Q}} \mathbb{E}\left[\bar{Y}_i(u_i)\right] . \tag{1}$$

Since the above lower bound is separable by products, it can be viewed as the revenue function of a multi-item newsvendor problem (without stock-out substitution), over the products in $\mathcal{Q}$.

Step 2 (greedy): solving a multi-item newsvendor problem. The inventory levels of products in $\mathcal{Q}$ are now set in order to optimize the lower bound described by inequality (1). As noted above, this formulation is a special case of the multi-item newsvendor problem subject to a single cardinality constraint, with salvage value and cost 0. It is well-known that an exact solution can be derived in polynomial time (see, e.g., (Muckstadt and Sapra 2010), Chapter 5). For instance, one may employ a greedy procedure that, at each step, augments the current inventory vector by a single unit that incurs the largest marginal increase of the expected revenue. It is worth noting that our lower bound on the revenue contribution of any given unit can easily be computed in polynomial time. In contrast, evaluating the exact expected revenue it generates is an open question by itself, as explained in Section 1, and our algorithm is surprisingly able to bypass this difficulty.

Running time. The selection step requires n greedy increments, each with at most n evaluations of the static expected revenue, whereas the greedy step involves C increments of the inventory vector, each leading to n evaluations of the multi-item newsvendor objective function. The static expected revenue of a given assortment can be computed in time $O(n \cdot |\mathcal{L}|)$ while the newsvendor objective can be evaluated through dynamic programming in time $O(nC\bar{M})$, where $\bar{M}$ is the maximal number of arrivals (see Goyal et al. Goyal et al. (2016)). Hence, the overall running time is $O(n^3 \cdot |\mathcal{L}| + n^2C^2\bar{M})$. In fact, by observing that each incremental action only affects a constant number of the revenue terms (and their corresponding purchase probabilities), a refined implementation of the evaluation oracles leads to a running time of $O(n^2 \cdot |\mathcal{L}| + nC\bar{M})$. Practically speaking, for common parametric random variables (e.g., gaussian) where $\bar{M}$ is infinite, the newsvendor objective can be evaluated using numerical integration techniques, with high accuracy.

2.2 Analysis under horizontal differentiation

To analyze our algorithm, we explicitly construct a 'good' candidate solution, making use of products in $\mathcal{Q}$. The subsequent analysis reveals that, under the choice of such inventory levels, the newsvendor-like lower bound approximates the optimal revenue within a constant factor. This candidate solution is constructed in two steps.

Step 1: Rounding up the capacity. We begin by modifying the original capacity C. Specifically, we round the capacity value up to the nearest multiple of $2 \cdot |\mathcal{Q}|$, denoted by $\bar{C}$. Note that since $|\mathcal{Q}| \leq C$, we must have $\bar{C} \leq 2C$. Our approach to design a feasible solution first creates a solution under the relaxed capacity $\bar{C}$. We overload notation by reusing C as the current capacity, instead of $\bar{C}$, in the next algorithmic steps. However, once the final inventory vector is computed, it remains to restore the original capacity by selecting the 'best' C units. Namely, we select the C units with largest contribution to the lower bound described in Section 2.1. This leaves us with at least half of the lower bound attained by the relaxed solution.

Step 2: Setting inventory levels. Based on the assignment $\mathcal{A}$ defined earlier, we proceed by explaining how to spread the capacity of C over the underlying set of products $\mathcal{Q}$. Intuitively, we would like the number of units stocked from each product $i \in \mathcal{Q}$ to be proportional to $\psi_i/\Theta_{\mathcal{Q}}$, where $\Theta_{\mathcal{Q}} = \sum_{i \in \mathcal{Q}} \psi_i$. Namely,

$$\tilde{u}_i = \frac{\psi_i}{\Theta_{\mathcal{Q}}} \cdot C \ .$$

However, this quantity may not be integral, and is therefore rounded down to the nearest multiple of $C/(2 \cdot |\mathcal{Q}|)$, which is necessarily integral by step 1. For this purpose, we can uniquely write

$$\tilde{u}_i = \mu_i \cdot \frac{C}{2 \cdot |\mathcal{Q}|} + \alpha_i \ ,$$

for some integer $\mu_i \in [0, 2 \cdot |\mathcal{Q}|]$ and some real $\alpha_i \in [0, C/(2 \cdot |\mathcal{Q}|))$. With these definitions in place, for each product $i \in \mathcal{Q}$, the number of units to be stocked is

$$u_i = \mu_i \cdot \frac{C}{2 \cdot |\mathcal{Q}|} \ ,$$

while other products are not stocked at all. This way, each product indeed has an integer number of units stocked, and furthermore, we do not exceed the overall capacity, since

$$\sum_{i \in \mathcal{Q}} u_i \leq \sum_{i \in \mathcal{Q}} \tilde{u}_i = \frac{C}{\Theta_{\mathcal{Q}}} \cdot \sum_{i \in \mathcal{Q}} \psi_i = C \ .$$

Deriving the approximation ratio. For the remainder of this section, let $(u_1, \ldots, u_n)$ be the inventory vector that has just been constructed. Since this vector stocks only products in $\mathcal{Q}$, it is a feasible solution to the multi-item newsvendor instance solved (exactly) by our algorithm. Therefore, to prove Theorem 2.1, it remains to show that the expected revenue generated by $(u_1, \ldots, u_n)$ can be lower bounded in terms of the optimal expected revenue.

We begin by stating two technical lemmas, that prove useful for analyzing the consumption process of the inventory vector $(u_1, \ldots, u_n)$. Although written in slightly different terms, the first lemma has been proven by Goyal et al. (Goyal et al. 2016, Lem. 4). The second lemma is easy to establish, as shown in Appendix A.2.

Lemma 2.3. *Let M be a non-negative integer-valued IFR random variable. For any $\alpha \in [0, 1]$, the random variable $X \sim B(M, \alpha)$ also follows an IFR distribution.*

Lemma 2.4. *Let X be a non-negative IFR random variable, and let $\bar{X} = \min\{X, C\}$, for some constant C. Suppose that $\mathbb{E}[\bar{X}] \leq \delta C$ for some $\delta \in [0, 1]$. Then, $\mathbb{E}[\bar{X}] \geq (1 - \delta) \cdot \mathbb{E}[X]$.*

Upper bounds on the optimal revenue. The important observation is that, for any inventory vector with a total capacity of at most C, an arriving customer will purchase a unit with probability at most $(e/(e-1)) \cdot \Theta_{\mathcal{Q}}$. This follows by noting that, as explained in step 1, the assortment $\mathcal{Q}$ approximates the optimal maximal coverage solution within factor $1 - 1/e$. Therefore, the expected revenue of the optimal inventory vector $(u_1^*, \ldots, u_n^*)$ can be bounded by

$$\mathbb{E}\left[\mathcal{R}(u_1^*, \ldots, u_n^*)\right] \leq \min\left\{C, \frac{e}{e-1} \cdot \mathbb{E}\left[M\right] \cdot \Theta_{\mathcal{Q}}\right\} \ . \tag{2}$$

Frequent and rare products. The key idea of our analysis is to distinguish between two types of products. For a parameter $\delta \in [0, 1]$ whose value will be optimized later, we say that product $i \in \mathcal{Q}$ is frequent when, in expectation, at least a δ-fraction of the units stocked are purchased in the consumption process, i.e., $\mathbb{E}[\bar{Y}_i(u_i)] \geq \delta u_i$. Otherwise, this product is said to be rare. We denote the sets of frequent and rare products by $\mathcal{F}$ and $\mathcal{R}$, respectively. Note that, by the relation between u_i and $\tilde{u}_i$,

$$\sum_{i\in\mathcal{F}} u_i + \sum_{i\in\mathcal{R}} u_i = \sum_{i\in\mathcal{Q}} u_i = \sum_{i\in\mathcal{Q}} \tilde{u}_i - \sum_{i\in\mathcal{Q}} \alpha_i \geq C - |\mathcal{Q}| \cdot \frac{C}{2\cdot|\mathcal{Q}|} = \frac{C}{2} \ . \tag{3}$$

We separately examine the contribution of each product type to the lower bound stated in inequality (1). For the contribution of frequent products, by definition we clearly have

$$\sum_{i\in\mathcal{F}} \mathbb{E}\left[\bar{Y}_i(u_i)\right] \geq \delta \cdot \sum_{i\in\mathcal{F}} u_i \ . \tag{4}$$

We now lower bound the contribution of rare products. Based on Lemma 2.3, since the number of customers M is assumed to be IFR distributed, we know that $Y_i \sim B(M,\psi_i)$ follows an IFR distribution as well. As a result, by Lemma 2.4, we infer that the expectations of Y_i and $\bar{Y}_i(u_i)$ are closely-related for every rare product i, meaning that $\mathbb{E}[\bar{Y}_i(u_i)] \geq (1-\delta)\cdot\mathbb{E}[Y_i]$, and therefore,

$$\sum_{i\in\mathcal{R}} \mathbb{E}\left[\bar{Y}_i(u_i)\right] \geq (1-\delta) \cdot \sum_{i\in\mathcal{R}} \mathbb{E}\left[Y_i\right] \ . \tag{5}$$

Also, by definition of u_i and $\tilde{u}_i$, we observe that

$$\mathbb{E}\left[Y_i\right] = \psi_i \cdot \mathbb{E}\left[M\right] = \frac{\Theta_{\mathcal{Q}}}{C} \cdot \tilde{u}_i \cdot \mathbb{E}\left[M\right] \geq \frac{\Theta_{\mathcal{Q}}}{C} \cdot u_i \cdot \mathbb{E}\left[M\right] \ .$$

Therefore, combining this with inequality (5), we obtain

$$\sum_{i\in\mathcal{R}} \mathbb{E}\left[\bar{Y}_i(u_i)\right] \geq (1-\delta) \cdot \mathbb{E}\left[M\right] \cdot \frac{\Theta_{\mathcal{Q}}}{C} \cdot \sum_{i\in\mathcal{R}} u_i \ . \tag{6}$$

Conclusion. By substituting (4) and (6) into the lower bound stated in inequality (1) and setting $\delta = (e-1)/(2e-1)$, we infer that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}(u_1,\ldots,u_n)\right] &\geq \delta \cdot \sum_{i\in\mathcal{F}} u_i + (1-\delta)\cdot\mathbb{E}\left[M\right] \cdot \frac{\Theta_{\mathcal{Q}}}{C} \cdot \sum_{i\in\mathcal{R}} u_i \\
&\geq \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] \cdot \left(\frac{\delta}{C} \cdot \sum_{i\in\mathcal{F}} u_i + \frac{1-\delta}{C} \cdot \left(1-\frac{1}{e}\right) \cdot \sum_{i\in\mathcal{R}} u_i \right) \\
&= \frac{e-1}{2e-1} \cdot \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] \cdot \frac{1}{C} \cdot \left(\sum_{i\in\mathcal{R}} u_i + \sum_{i\in\mathcal{F}} u_i \right) \\
&\geq \frac{e-1}{4e-2} \cdot \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] \ ,
\end{aligned}$$

where the second inequality is derived from the upper bound in (2), and the last inequality follows from (3). Finally, as explained in step 1, we restore the original capacity by selecting at least half of the units stocked, based on their individual contributions to the lower bounds in (4) and (6). This alteration yields an approximation guarantee of $(e-1)/(8e-4)$.

For a more careful analysis, we need to distinguish in the lower bound above between the rounded capacity $\bar{C}$ and the initial capacity C, yielding

$$\mathbb{E}\left[\mathcal{R}(u_1,\ldots,u_n)\right] \geq \frac{e-1}{2e-1} \cdot \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] \cdot \frac{1}{\bar{C}} \cdot \left(\sum_{i \in \mathcal{R}} u_i + \sum_{i \in \mathcal{F}} u_i \right) .$$

We now observe that, when restoring the original capacity C, our lower bound scales-down by a factor of C/β, where β is the total number of units in the relaxed solution, i.e., $\beta = \sum_{i \in \mathcal{R}} u_i + \sum_{i \in \mathcal{F}} u_i$. For this reason, we obtain an expected revenue of at least

$$\frac{e-1}{2e-1} \cdot \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] \cdot \frac{\beta}{\bar{C}} \cdot \frac{C}{\beta} \geq \frac{e-1}{4e-2} \cdot \mathbb{E}\left[\mathcal{R}(u_1^*,\ldots,u_n^*)\right] ,$$

concluding the proof of Theorem 2.1.

2.3 Refined performance guarantee

A close investigation of our algorithm shows that the factor of $1 - 1/e$ is incurred due to employing a general-purpose maximum coverage algorithm to solve the static problem in step 1. However, for numerous special cases of preference lists (such as nested, intervals, laminar, just to name a few), this variant can be solved either exactly or within a greater degree of accuracy. The following claim explicitly relates between our approximation guarantee for the dynamic model and the best achievable one for the static variant.

Corollary 2.5. *Suppose that, for a certain class of preference lists, the static variant can be efficiently approximated within factor α. Then, the corresponding dynamic formulation, with identical prices and IFR demand distribution, admits an $\alpha/(2\alpha+2)$-approximation in polynomial time.*

It is important to point out that our results extend to the case where the distribution over preference lists is not explicitly specified as part of the input, potentially having an exponentially-large support. In fact, to efficiently implement our algorithm, we only require a polynomial-time procedure for computing the probability that each product in a given assortment is purchased under a single customer arrival. Indeed, this property is sufficient to greedily approximate the static problem (step 1 in Section 2.1) and to compute each of the ψ_i-probabilities defining the newsvendor objective function. In particular, such procedures can easily be devised for most choice models proposed in the revenue management literature, including mixtures of logits Talluri and Van Ryzin (2004), Rusmevichientong and Topaloglu (2012), nested logit Li et al. (2015), Davis et al. (2014), as well as the Markov chain model Blanchet et al. (2016), Feldman and Topaloglu (2017), Désir et al. (2015).

2.4 Price differentiation

Algorithm. In what follows, we explain how the algorithm of Section 2.1 can be adapted in the presence of price differentiation to obtain an $O(\log(p_{\max}/p_{\min}))$-approximation. The basic idea is based on the classify-and-select paradigm, where products are initially partitioned into classes with nearly-uniform prices, and then, we employ for each class our constant-factor approximation as a subroutine, treating these products as if they are associated with uniform prices. Specifically, assuming that products are indexed such that $p_1 \leq \cdots \leq p_n$, our algorithm picks the most profitable inventory vector out of $U^1, \ldots, U^K$, where $K = \lceil \log(p_n/p_1) \rceil$. Each vector U^k is generated as follows:

1. Let a_k be the minimum index of a product whose price is at least $p_1 \cdot 2^{k-1}$, that is, $a_k = \min\{i \in [n] : p_i \geq p_1 \cdot 2^{k-1}\}$. Given this parameter, we define the collection of products $\mathcal{A}_k = [a_k, a_{k+1} - 1]$.
2. The inventory vector $U^k = (u_1^k, \ldots, u_n^k)$ is constructed by applying the horizontal differentiation procedure (see Section 2.1) to the subproblem formed by products in $\mathcal{A}_k$, with identical prices of $\tilde{p}_{a_k} = \cdots = \tilde{p}_{a_{k+1}-1} = 1$.

It is worth noting that, in order to pick the "most profitable" inventory vector, we make the final comparisons in terms of the multi-item newsvendor lower bound, that can be computed in polynomial time.

Analysis. In order to establish the performance guarantee attained by our algorithm, we begin by highlighting a fundamental property of the consumption process under an arbitrary collection of preference lists. To avoid deviating from the overall discussion, we prove the next claim in Appendix A.1.

Lemma 2.6. *The expected revenue function* $\mathbb{E}[\mathcal{R}(\cdot)]$ *is subadditive.*

Now, for every $k \in [K]$, let $U^{*(k)}$ be the projection of the optimal inventory vector U^* on the products $\mathcal{A}_k$. That is, $U^{*(k)}$ is obtained from U^* by setting to 0 the inventory levels of all products in $[n] \setminus \mathcal{A}_k$. We proceed by showing that U^k, the inventory vector constructed earlier by our algorithm, generates a constant fraction of the expected revenue of $U^{*(k)}$.

Lemma 2.7. $\mathbb{E}[\mathcal{R}(U^k)] \geq \frac{e-1}{8e-4} \cdot \mathbb{E}[\mathcal{R}(U^{*(k)})]$.

Proof. Recall that the inventory vector U^k constitutes an $(e-1)/(4e-2)$-approximation for the optimal sales quantity (i.e., expected number of units purchased) when the underlying set of products is $\mathcal{A}_k$. Therefore, since $U^{*(k)}$ stocks only products in $\mathcal{A}_k$, the expected sales quantity with respect to U^k is at least $(e-1)/(4e-2)$ times the analogous quantity with respect to $U^{*(k)}$. The claim follows by observing that, by definition of $\mathcal{A}_k$, the ratio between the extremal prices within this class is at most 2. ■

Based on the preceding discussion, we are now ready to show that the most profitable inventory vector out of $U^1, \ldots, U^K$ guarantees an expected revenue within factor $O(\log(p_n/p_1))$ of optimal. Indeed,

$$\begin{aligned}
\max_k \left\{ \mathbb{E}\left[\mathcal{R}\left(U^k \right) \right] \right\} &\geq \frac{e-1}{8e-4} \cdot \max_k \left\{ \mathbb{E}\left[\mathcal{R}\left(U^{*(k)} \right) \right] \right\} \\
&\geq \frac{e-1}{8e-4} \cdot \frac{1}{K} \cdot \sum_{k=1}^{K} \mathbb{E}\left[\mathcal{R}\left(U^{*(k)} \right) \right] \\
&\geq \frac{e-1}{8e-4} \cdot \frac{1}{K} \cdot \mathbb{E}\left[\mathcal{R}\left(U^* \right) \right]
\end{aligned}$$

$$= \Omega\left(\frac{1}{\log(p_n/p_1)}\right) \cdot \mathbb{E}\left[\mathcal{R}\left(U^*\right)\right] ,$$

where the first inequality follows from Lemma 2.7 and the third inequality is implied by the subadditivity of the expected revenue function (see Lemma 2.6).

3 Approximation Algorithms for the Intervals Choice Model

In what follows, we consider the dynamic assortment planning problem under interval preference lists. When the number of customers M satisfies the IFR property, we show how to efficiently compute an inventory vector that approximates the optimal expected revenue within factor $O(\log\log(p_n/p_1))$, where p_1 and p_n are the minimal and maximal prices, respectively.

Since our approach employs recursive decompositions of the preference lists, it is instructive to start off by presenting some of the high-level ideas, followed by a simpler $O(\log n)$ approximation. We then explain how to make use of our algorithm for uniform prices, given in Section 2, to establish the main result of this section.

3.1 General outline

The main algorithmic idea, exploited in different forms in Sections 3.2 and 3.3, consists in partitioning the collection of preference lists $\mathcal{L}$ into a small number of classes, L. By separating customer purchases according to their different classes, the expected revenue function decomposes into L terms. Hence, to obtain an $O(L)$-approximation, we propose the following approach:

1. Consider separately each of the L subproblems, where the consumption process is limited to customers picking preference lists from a single class of the partition.
2. Approximately solve each of these L subproblems. The crux would be to design a partition such that the corresponding subproblems have a simplified structure, admitting constant-factor approximations.
3. Pick the best solution among these L inventory vectors.

However, this approach is generally insufficient to claim the desired approximation ratio. Indeed, the decomposition into separate subproblems does not take into account the dependency of the revenue functions across the different preference list classes, in the joint sequence of arrivals. In other words, the expected revenue restricted to a single class in the full consumption process (i.e., when all preference lists in $\mathcal{L}$ could be picked) is different from the one generated by the distribution induced on that class. Consequently, our decomposition approach may very well under-estimate the potential expected revenue.

Motivated by this observation, for any class of preference lists $\mathcal{V} \subseteq \mathcal{L}$ and inventory vector U, we distinguish between two types of revenues, captured by the following random variables:

- *Original model:* $\mathcal{R}^+_{\mathcal{V}}(U)$ designates the revenue generated by the arrival of M customers who draw a preference list in $\mathcal{V}$, assuming that the consumption process is formed by the original model, where all preference lists in $\mathcal{L}$ occur according to the initial distribution.

- *$\mathcal{V}$-restricted model:* $\mathcal{R}^-_{\mathcal{V}}(U)$ denotes the revenue generated by the arrival of M customers who draw a preference list in $\mathcal{V}$, assuming that the consumption process is formed by the $\mathcal{V}$-restricted model. Here, only preference lists in $\mathcal{V}$ can occur and their probabilities remain unchanged, whereas all lists in $\mathcal{L} \setminus \mathcal{V}$ are replaced by an empty list.

Now assume that the classes of our partition are denoted by $\mathcal{V}_1, \dots, \mathcal{V}_L$. Generally speaking, the expected revenues in the restricted and original models are unrelated; elementary examples demonstrate that neither one dominates the other. What we need to argue to utilize this approach is that these revenues are within constant factors of each other, due to the specific properties of our decomposition. Formally, for every $\ell \in [L]$, we construct a feasible inventory vector U_ℓ satisfying

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] = \Omega(1) \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] \ , \tag{7}$$

where U^* is the optimal inventory vector for the original model.

As a result, the best inventory vector out of $U_1, \dots, U_L$ guarantees an $O(L)$-approximation for the original model. Indeed, by considering any realization of the consumption process, it is easy to verify that the revenue generated in the original model is stochastically larger than that of the $\mathcal{V}_\ell$-restricted model. By combining this observation and equation (7),

$$\begin{aligned}
\max_{\ell \in [L]} \mathbb{E}\left[\mathcal{R}(U_\ell)\right] &\geq \max_{\ell \in [L]} \mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \\
&\geq \frac{1}{L} \cdot \sum_{\ell=1}^{L} \mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \\
&= \Omega\left(\frac{1}{L}\right) \cdot \sum_{\ell=1}^{L} \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] \\
&= \Omega\left(\frac{1}{L}\right) \cdot \mathbb{E}\left[\mathcal{R}\left(U^*\right)\right] \ .
\end{aligned}$$

3.2 $O(\log n)$-approximation

We begin by describing our partition of the preference lists into $L = O(\log n)$ classes $\mathcal{V}_1, \dots, \mathcal{V}_L$. For the resulting partition, we devise a polynomial-time algorithm for computing an inventory vector U_ℓ satisfying

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] \ .$$

3.2.1 The recursive decomposition

In order to formalize our decomposition approach, we first introduce a sequence of increasingly refined partitions of the products in $[n]$, denoted by $\mathcal{S}_1, \dots, \mathcal{S}_L$. In turn, this sequence induced the desired partition of preference lists into $\mathcal{V}_1, \dots, \mathcal{V}_L$.

Partitions of products. We define the middle product of a segment $[a, b] \subseteq [n]$ as the product $\lceil (a + b)/2 \rceil$. The sequence $\mathcal{S}_1, \dots, \mathcal{S}_L$ is obtained by the following recursive procedure:

- First, we define $\mathcal{S}_1$ as the trivial partition of $[n]$, comprised of a single segment consisting of all products, i.e., $\mathcal{S}_1 = \{[n]\}$.

- The next partition, $\mathcal{S}_2$, is obtained by breaking the segment $[n]$ at its middle product, that is, $\mathcal{S}_2 = \{[1, \lceil (n+1)/2 \rceil], [\lceil (n+1)/2 \rceil + 1, n]\}$.
- This process continues recursively, that is, we define $\mathcal{S}_\ell$ as the partition of products obtained by breaking each segment of $\mathcal{S}_{\ell-1}$ at its middle product into two parts.

Partition of preference lists. Given any subset of lists $\mathcal{V} \subseteq \mathcal{L}$ and a partition $\mathcal{S}$ of the products $[n]$ into pairwise-disjoint segments, we define $\mathrm{mid}(\mathcal{V}, \mathcal{S})$ as the subset of lists in $\mathcal{V}$ that contain the middle product of at least one segment in $\mathcal{S}$. With this definition at hand, we construct the partition of the preference lists into $\mathcal{V}_1, \ldots, \mathcal{V}_L$ as follows:

- First, we have $\mathcal{V}_1 = \mathrm{mid}(\mathcal{L}, \mathcal{S}_1)$.
- Then, $\mathcal{V}_2 = \mathrm{mid}(\mathcal{L} \setminus \mathcal{V}_1, \mathcal{S}_2)$.
- This process continues recursively, as illustrated in Figure 1. That is, we define $\mathcal{V}_\ell$ as the subset of residual preference lists that contain the middle product of a segment in $\mathcal{S}_\ell$, i.e., $\mathcal{V}_\ell = \mathrm{mid}(\mathcal{L} \setminus (\cup_{j=1}^{\ell-1} \mathcal{V}_j), \mathcal{S}_\ell)$.

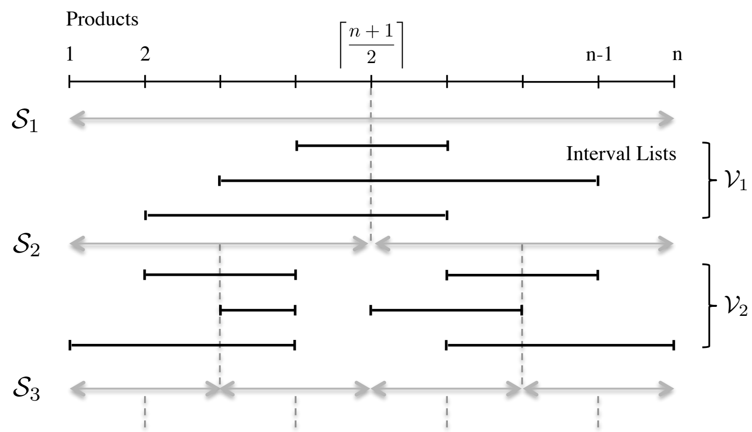

Figure 1: The recursive decomposition of $\mathcal{L}$ into $\mathcal{V}_1, \ldots, \mathcal{V}_L$.

Structural properties. Since the maximum length of any segment shrinks by a constant factor at each level of the decomposition, it immediately follows that the resulting number of classes is $L = O(\log n)$. In addition, each partition of products $\mathcal{S}_\ell$ can be viewed as a collection of pairwise-disjoint segments, satisfying the next two properties:

- *Property 1:* each interval list in $\mathcal{V}_\ell$ is fully contained in precisely one of the segments in $\mathcal{S}_\ell$.
- *Property 2:* for each segment $S \in \mathcal{S}_\ell$, there exists a product in S that intersects all intervals in $\mathcal{V}_\ell$ contained in this segment.

These are precisely the sufficient properties that will enable us to compute a feasible inventory vector U_ℓ satisfying

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] .$$

3.2.2 Proving the existence of U_ℓ

Single segment analysis. In order to construct U_ℓ, it is sufficient to show that, for every segment of products S in the partition $\mathcal{S}_\ell$, there exists an inventory vector U_ℓ^S such that:

- The vector U_ℓ^S only makes use of products in S.
- Letting $\mathcal{V}_\ell^S$ be the set of interval preference lists in $\mathcal{V}_\ell$ that are fully contained in S, we have
$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell^S}\left(U_\ell^S\right)\right] \geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell^S}\left(U^*\right)\right] \ . \tag{8}$$
- The combined number of units stocked in $\{U_\ell^S : S \in \mathcal{S}_\ell\}$ is at most C.

Indeed, given property 1, since the segments in $\mathcal{S}_\ell$ are pairwise disjoint, the expected revenue of the combined vector $\sum_{S\in\mathcal{S}_\ell} U_\ell^S$ decomposes into the sum of expected revenues generated by each vector U_ℓ^S. This decomposition applies in both the original model and the $\mathcal{V}_\ell$-restricted model. In other words, assuming that each U_ℓ^S satisfies inequality (8), we obtain the desired inequality:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}\left(\sum_{S\in\mathcal{S}_\ell} U_\ell^S\right)\right] &= \sum_{S\in\mathcal{S}_\ell} \mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell^S}\left(U_\ell^S\right)\right] \\
&\geq \frac{1}{8} \cdot \sum_{S\in\mathcal{S}_\ell} \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell^S}\left(U^*\right)\right] \\
&= \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}\left(U^*\right)\right] \ .
\end{aligned}$$

Simplified problem. The preceding discussion implies that we can focus on a single segment S from this point on. We prove the existence of an inventory vector U_ℓ^S that satisfies the above properties by analyzing the revenue generated under the optimal vector U^* by the lists in $\mathcal{V}_\ell^S$. In fact, in the course of proving the existence of U_ℓ^S, we implicitly describe an efficient algorithmic procedure to construct such a vector. To simplify the presentation, the corresponding algorithm is made explicit in Section 3.2.3. Also, we use simplified notation throughout this section, where $\mathcal{V}_\ell^S$ and U_ℓ^S are replaced by $\tilde{\mathcal{V}}$ and $\tilde{U}$, respectively, i.e., we do not explicitly mention the dependency of these variables on S and ℓ.

Revenue decomposition. Based on property 2, we define J to be the highest index product that intersects all intervals in $\tilde{\mathcal{V}}$. We now break the segment S into a left part $S_{\text{left}} = S \cap [1, J]$ and a right part $S_{\text{right}} = S \cap [J+1, n]$, noting that the latter part could be empty. These definitions, in turn, are used to further divide the expected revenue $\mathbb{E}[\mathcal{R}^+_{\tilde{\mathcal{V}}}(U^*)]$ based on whether units are purchased from the left or right part of S, that is,

$$\mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}}}(U^*)\right] = \mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{left}}}(U^*)\right] + \mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{right}}}(U^*)\right] \ .$$

The proof proceeds by considering two cases, depending on whether most of the expected revenue is coming from the left or right parts. To better understand this case analysis, we advise the reader to consult Figure 2.

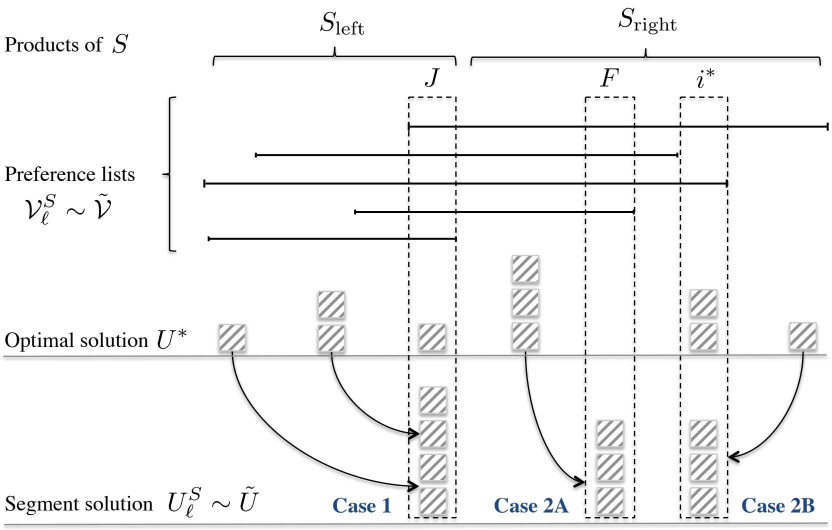

Figure 2: The inventory vectors examined by the algorithm to construct U_ℓ^S.

Case 1: $\boldsymbol{E[R^+_{\tilde{V},S_{left}}(U^*)]\geq E[R^+_{\tilde{V}}(U^*)]/2}$. This case can be handled rather easily. To construct the inventory vector $\tilde{U}$, we consider the restriction of the optimal vector U^* to the left segment $S_{\text{left}} = S \cap [1, J]$, and relocate all stocked units to product J. It is not difficult to verify that, in every realization of the consumption process, the number of units of the left segment S_{left} consumed in the $\tilde{\mathcal{V}}$-restricted model is greater or equal to the number of units consumed in the original model by preference lists in $\tilde{\mathcal{V}}$. In addition, J is the most expensive product in S_{left}, meaning that $\mathcal{R}^-_{\tilde{\mathcal{V}},S_{\text{left}}}(\tilde{U}) \geq_{\text{st}} \mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{left}}}(U^*)$, and therefore

$$\mathbb{E}\left[\mathcal{R}^-_{\tilde{\mathcal{V}}}\left(\tilde{U}\right)\right] = \mathbb{E}\left[\mathcal{R}^-_{\tilde{\mathcal{V}},S_{\text{left}}}\left(\tilde{U}\right)\right] \geq \mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{left}}}\left(U^*\right)\right] \geq \frac{1}{2}\cdot\mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}}}\left(U^*\right)\right] \ ,$$

where the first equality holds since $\tilde{U}$ only contains the product J, and the last inequality is due to the case hypothesis.

Case 2: $\boldsymbol{E[R^+_{\tilde{V},S_{right}}(U^*)]\geq E[R^+_{\tilde{V}}(U^*)]/2}$. This case is more involved. Let us focus on some product i in the segment $S_{\text{right}} = S \cap [J+1, n]$, and let ψ_i be the probability that an arriving customer picks one of the intervals in $\tilde{\mathcal{V}}$ that contains i (necessarily on its right part). Note that since all non-empty right parts have $J+1$ as a left endpoint, it follows that $\psi_{J+1} \geq \psi_{J+2} \geq \cdots$. We begin by defining a pair of random variables, whose exact meaning will be revealed later on. These are $Y_i \sim B(M, \psi_i)$ and $\bar{Y}_i = \min\{Y_i, C^*_S\}$, where C^*_S is the total capacity used by the optimal inventory vector U^* over the segment S_{right}.

With these random variables at hand, we say that product i is frequent when $\mathbb{E}[\bar{Y}_i] \geq C^*_S/2$. Otherwise, this product is rare. Since $\psi_{J+1} \geq \psi_{J+2} \geq \cdots$, it follows that there is a product F such that the set of frequent products $\mathcal{F}$ is precisely those in $[J+1, F]$, whereas the rare ones $\mathcal{R}$ are those in $S_{\text{right}} \setminus [J+1, F]$. As a result, we can break the revenue $\mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{right}}}(U^*)$ into purchases of frequent and rare products, obtaining that

$$\mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},\mathcal{F}}\left(U^*\right)\right] + \mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},\mathcal{R}}\left(U^*\right)\right] = \mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}},S_{\text{right}}}\left(U^*\right)\right] \geq \frac{1}{2}\cdot\mathbb{E}\left[\mathcal{R}^+_{\tilde{\mathcal{V}}}\left(U^*\right)\right] \ ,$$

where the last inequality follows from the case hypothesis. It remains to consider two cases, depending on which set of products (frequent or rare) is contributing more in the above inequality.

Case 2A: $\boldsymbol{E[R^{+}_{\tilde{V},F}(U^*)]\geq E[R^{+}_{\tilde{V}}(U^*)]/4}$. Note that $C^*_S \cdot p_F$ is a trivial upper bound on the random variable $\mathcal{R}^{+}_{\tilde{\mathcal{V}},\mathcal{F}}(U^*)$, and consequently on its expectation. This follows by observing that, in each realization, at most C^*_S units are purchased among frequent products, and each purchase generates a revenue of at most p_F, given that F is the right endpoint of $\mathcal{F}$. To construct the inventory vector $\tilde{U}$, we simply stock C^*_S units of F, the most expensive frequent product. We observe that in the $\tilde{\mathcal{V}}$-restricted model, the distribution of the number of units consumed is identical to that of $\bar{Y}_F$. Indeed, similar to the consumption process considered in Section 2, under the $\tilde{\mathcal{V}}$-restricted model, a unit of product F is consumed with probability ψ_F as long as this product has not stocked-out, corresponding to a sequence of M independent Bernoulli trials whose sum is capped by the capacity C^*_S. This means that the expected revenue under the restricted model would be

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}^{-}_{\tilde{\mathcal{V}}}\left(\tilde{U}\right)\right] &= p_F \cdot \mathbb{E}\left[\bar{Y}_F\right] \\
&\geq p_F \cdot \frac{C^*_S}{2} \\
&\geq \frac{1}{2} \cdot \mathbb{E}\left[\mathcal{R}^{+}_{\tilde{\mathcal{V}},\mathcal{F}}\left(U^*\right)\right] \\
&\geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^{+}_{\tilde{\mathcal{V}}}\left(U^*\right)\right] \ ,
\end{aligned}$$

where the first inequality follows from the definition of frequent products while the last inequality is due to the case hypothesis.

Case 2B: $\boldsymbol{E[R^{+}_{\tilde{V},R}(U^*)]\geq E[R^{+}_{\tilde{V}}(U^*)]/4}$. Here, we derive an upper bound on the expected revenue of $\mathcal{R}^{+}_{\tilde{\mathcal{V}},\mathcal{R}}(U^*)$ by considering an unrealistic model, where prior to the arrival of any customer, the current inventory vector can be re-optimized. Specifically, suppose that we may substitute to any vector (still, using only rare products), without any capacity restrictions. In this model, since we are only interested in maximizing the expected revenue due to purchases made by (the right part of) intervals in $\tilde{\mathcal{V}}$, the optimal strategy is to stock a single unit of i^*, which is the product that maximizes $\psi_i p_i$ over all rare products. Indeed, assuming i is the minimal-index product stocked, the expected revenue generated by a single arrival of the lists in $\tilde{\mathcal{V}}$ is exactly $\psi_i p_i$. Therefore,

$$\mathbb{E}\left[\mathcal{R}^{+}_{\tilde{\mathcal{V}},\mathcal{R}}\left(U^*\right)\right] \leq \mathbb{E}\left[M\right] \cdot \psi_{i^*} p_{i^*} \ .$$

Now, to construct the inventory vector $\tilde{U}$, we simply stock C^*_S units of product i^*. Once again, in the $\tilde{\mathcal{V}}$-restricted model, the distribution of the number of units consumed will be identical to that of $\bar{Y}_{i^*} \sim B(M, \psi_{i^*})$. Indeed, similar to the consumption process considered in Section 2, under the $\tilde{\mathcal{V}}$-restricted model, a unit of product i^* is consumed with probability ψ_{i^*} as long as this product has not stocked-out. Therefore, the resulting expected revenue is

$$\mathbb{E}\left[\mathcal{R}^{-}_{\tilde{\mathcal{V}}}\left(\tilde{U}\right)\right] = p_{i^*} \cdot \mathbb{E}\left[\bar{Y}_{i^*}\right]$$

$$
\begin{aligned}
&\geq p_{i^*} \cdot \frac{\mathbb{E}\left[Y_{i^*}\right]}{2} \\
&= \frac{1}{2} \cdot \mathbb{E}\left[M\right] \cdot \psi_{i^*} p_{i^*} \\
&\geq \frac{1}{2} \cdot \mathbb{E}\left[\mathcal{R}^{+}_{\tilde{\mathcal{V}},\mathcal{R}}\left(U^*\right)\right] \\
&\geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^{+}_{\tilde{\mathcal{V}}}\left(U^*\right)\right] \ ,
\end{aligned}
$$

where the first inequality follows from Lemmas 2.3 and 2.4, recalling that i^* is a rare product, and the last inequality is due to the case hypothesis.

3.2.3 Dynamic program

A careful review of the arguments used to prove the existence of U_ℓ reveals that we actually describe an efficient way to construct this vector, assuming that the number of units C_S of the optimal solution within each interval $S \in \mathcal{S}_\ell$ is known a-priori. We prove that this assumption is not needed, explaining why a similar approximation ratio can be attained by means of dynamic programming. In the following, J, F, and i^* play precisely the same roles as in the previous section. It is not difficult to verify that each of these products can easily be identified in polynomial time.

The general idea is to formulate a dynamic program that tries out all feasible capacities for each single segment $S \in \mathcal{S}_\ell$, and chooses the best vector among those constructed in cases 1, 2A, and 2B. Specifically, for any positive capacity c, we define $U^{(1)}_{S,c}$ as the inventory vector described in case 1 that stocks c units of product J. Similarly, $U^{(2A)}_{S,c}$ is the inventory vector described in case 2A that stocks c units of product F, and $U^{(2B)}_{S,c}$ is the vector of case 2B that stocks c units of product i^*. One important observation is that we can efficiently compute the expected revenue associated with each of these vectors under the $\mathcal{V}^S_\ell$-restricted model, as it is equivalent to computing the expected value of a truncated binomial random variable. Finally, we let $\{S_1, \ldots, S_r\}$ designate the segments of products in the partition $\mathcal{S}_\ell$, numbered in increasing order of product indices.

For any $j \in [r]$ and $\bar{c} \in [C]$, we define the objective function $G(j, \bar{c})$ as the maximal expected revenue in the restricted model, generated by a vector with at most $\bar{c}$ units, obtained by concatenating the candidate solutions $U^{(1)}_{S,c}$ or $U^{(2A)}_{S,c}$ or $U^{(2B)}_{S,c}$, over the first j segments of the partition, $S_1, \ldots, S_j$. It is not difficult to verify that the function G satisfies the following recursion formula:

$$
G(j, \bar{c}) = \max_{c \leq \bar{c}} \left\{ G\left(j-1, \bar{c}-c\right) + \max\left\{ \mathbb{E}\left[\mathcal{R}^{-}_{\mathcal{V}^S_\ell}\left(U^{(1)}_{S_j,c}\right)\right], \mathbb{E}\left[\mathcal{R}^{-}_{\mathcal{V}^S_\ell}\left(U^{(2A)}_{S_j,c}\right)\right], \mathbb{E}\left[\mathcal{R}^{-}_{\mathcal{V}^S_\ell}\left(U^{(2B)}_{S_j,c}\right)\right] \right\} \right\} .
$$

By solving this recursion forward, we infer the quantity $G(r, C)$. Given the optimality conditions satisfied by the above dynamic program, it follows that $G(r, C) \geq \mathbb{E}[\mathcal{R}^{-}_{\mathcal{V}^S_\ell}(U_\ell)]$, where U_ℓ is the vector constructed in Section 3.2.2. Indeed, the allocation of capacity across the different segments, as described in the previous section, can be replicated by the dynamic program, and for each segment, the dynamic program selects a vector that maximizes the expected revenue.

3.2.4 Running time analysis

The recursive decomposition is computed in time $O(|\mathcal{L}|\cdot \log n)$. Indeed, we identify the location of each interval list with respect to at most $\lceil \log n \rceil$ middle products along the recursion. Next, the dynamic program in Section 3.2.3 is solved in time $O(|\mathcal{S}_\ell|\cdot C^2 \bar{M})$ for the collection of preference lists $\mathcal{V}_\ell$. Indeed, the state space has size $O(|\mathcal{S}_\ell|\cdot C)$. Each recursive step is computed in $O(\bar{M}C)$ time by comparing the expected revenues of three inventory vectors, each stocking C units. The latter require to evaluate the expectations of truncated binomial variables, with at most $\bar{M}$ trials. Summing over all $\ell \in [L]$, the overall running time is $O(|\mathcal{L}|\cdot \log n + nC^2\bar{M})$.

3.3 $O(\log\log(p_n/p_1))$-approximation

In this section, we explain how the main technical ideas of Section 3.2 can be utilized in order to attain an approximation guarantee of $O(\log\log(p_n/p_1))$, where p_n and p_1 stand for the maximum and minimum price of any product. Here, we employ a different decomposition of $\mathcal{L}$, that allows us to make use of the constant-factor approximation for uniform prices (see Section 2) as a subroutine.

3.3.1 The recursive decomposition

We initially break the interval $[1,n]$ into $K = O(\log(p_n/p_1))$ buckets $B_1,\ldots,B_K$ according to prices, geometrically by powers of 2. That is, the first bucket B_1 consists of products with prices in $[p_1, 2p_1)$, the second bucket B_2 corresponds to prices in $[2p_1, 2^2 p_1)$, so forth and so on.

The recursive partition here resembles the one in Section 3.2.1, at the exception that segments are now defined with respect to the indexing $1,\ldots,K$, and the middle product depends on the collection of buckets $B_1,\ldots,B_K$. Specifically, the middle product associated with a segment $[a,b] \subseteq [K]$ is defined as the right-most product of the middle bucket $B_{\lceil (a+b)/2 \rceil}$. Given any subset of lists $\mathcal{V} \subseteq \mathcal{L}$ and a partition $\mathcal{K}$ of $[K]$ into pairwise-disjoint segments, we define $\mathrm{mid}(\mathcal{V},\mathcal{K})$ as the set of interval lists in $\mathcal{V}$ that contain the middle product of at least one segment in $\mathcal{K}$. With this definition at hand, we define the classes of lists $\mathcal{V}_1,\ldots,\mathcal{V}_L,\mathcal{V}_{\mathrm{in}}$ as follows:

- The special class $\mathcal{V}_{\mathrm{in}}$ is comprised of all intervals in $\mathcal{L}$ that are fully contained in one of the buckets $B_1,\ldots,B_K$.
- The remaining classes are determined as follows:
  - First, we have $\mathcal{V}_1 = \mathrm{mid}(\mathcal{L} \setminus \mathcal{V}_{\mathrm{in}}, \mathcal{K}_1)$, where $\mathcal{K}_1 = \{[K]\}$.
  - Then, $\mathcal{V}_2 = \mathrm{mid}(\mathcal{L} \setminus (\mathcal{V}_1 \cup \mathcal{V}_{\mathrm{in}}), \mathcal{K}_2)$, where $\mathcal{K}_2$ is obtained by breaking the segment $[K]$ at its middle product.
  - This process continues recursively, as illustrated in Figure 3. That is, we define $\mathcal{K}_\ell$ as the partition of $[K]$ obtained by breaking each segment of $\mathcal{K}_{\ell-1}$ at its middle product into two parts. Then, $\mathcal{V}_\ell$ is the residual subset of lists that contain the middle product of at least one segment in $\mathcal{K}_\ell$, i.e., $\mathcal{V}_\ell = \mathrm{mid}(\mathcal{L} \setminus ((\cup_{j=1}^{\ell-1}\mathcal{V}_j) \cup \mathcal{V}_{\mathrm{in}}), \mathcal{K}_\ell)$.

The decomposition above terminates as soon as we reach a level L, where $\mathcal{K}_L$ consists of only singletons of $[1,K]$. Once again, since the maximum length of any segment shrinks by a constant factor at each level, it follows that the depth of this decomposition is $L = O(\log K) = O(\log\log(p_n/p_1))$.

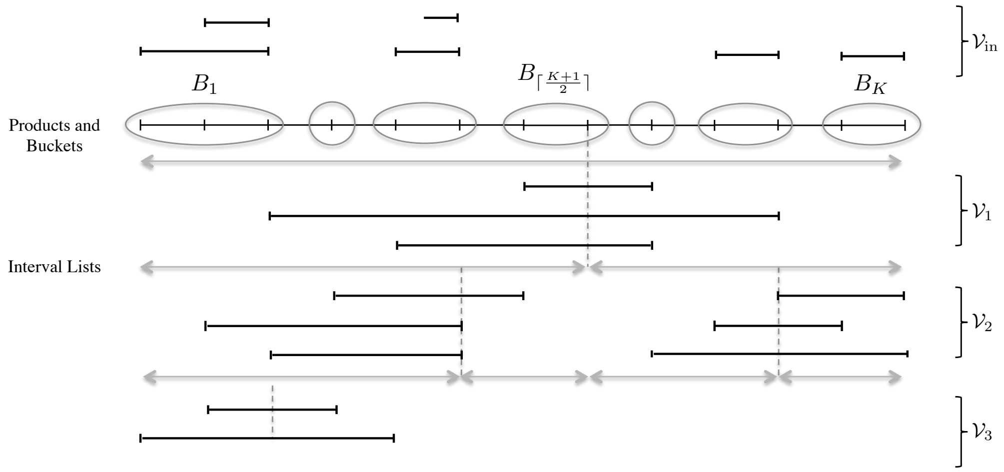

Figure 3: The decomposition of $\mathcal{L}$ into $\mathcal{V}_1, \dots, \mathcal{V}_L, \mathcal{V}_{\text{in}}$.

3.3.2 Proving the existence of U_ℓ and U_{in}

We now argue that there is an efficient way of meeting the fundamental inequality (7) that relates between the restricted and original models for each class of the partition. Formally, for every $\ell \in [L]$, we devise a polynomial-time procedure to compute a feasible inventory vector U_ℓ satisfying

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] \ ,$$

where U^* is the optimal inventory vector for the original model. We also construct U_{in} such that

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_{\text{in}}}(U_{\text{in}})\right] = \Omega(1) \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_{\text{in}}}(U^*)\right] \ .$$

Following the discussion in Section 3.1, we obtain an $O(\log\log(p_n/p_1))$ approximation for the original model by picking the best vector out of $U_1, \dots, U_L, U_{\text{in}}$.

Handling $V_1, \dots, V_L$. The important observation is that the intervals in each class $\mathcal{V}_\ell$ satisfy the sufficient properties mentioned at the end of Section 3.2.1. For this reason, the exact same algorithm, now applied to a different collection of segments, enables us to compute a feasible inventory vector U_ℓ such that

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_\ell}(U_\ell)\right] \geq \frac{1}{8} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_\ell}(U^*)\right] \ .$$

Handling V_{in}. Let us focus attention on a single bucket B, corresponding to a power-of-2 price range, say $[\Delta, 2\Delta)$, where the set of intervals contained in this bucket are denoted by $\mathcal{V}^B_{\text{in}}$. The important observation is that, in every realization, the number of units consumed in the $\mathcal{V}^B_{\text{in}}$-restricted model is greater or equal to the number of units consumed in the original model by these intervals. Indeed, this claim can be proven inductively over the arrival rank of customers and by arguing that, at any point in time during the arrival sequence, the number of units left within each interval in $\mathcal{V}_{\text{in}}^B$ in the former model is greater or equal to the corresponding number of units in the latter model. Therefore, since all products in bucket B have prices in $[\Delta, 2\Delta)$, it follows that $\mathcal{R}^-_{\mathcal{V}_{\text{in}}^B}(U^*) \geq_{\text{st}} \mathcal{R}^+_{\mathcal{V}_{\text{in}}^B}(U^*)/2$, and consequently,

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_{\text{in}}^B}(U^*)\right] \geq \frac{1}{2} \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_{\text{in}}^B}(U^*)\right] . \tag{9}$$

Now, based on our constant-factor approximation for uniform prices (see Section 2), assuming that the capacity used by the optimal vector U^* over the bucket B is known in advance, we can efficiently compute an inventory vector U_{in}^B satisfying

$$\mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_{\text{in}}^B}\left(U_{\text{in}}^B\right)\right] = \Omega(1) \cdot \mathbb{E}\left[\mathcal{R}^-_{\mathcal{V}_{\text{in}}^B}(U^*)\right] = \Omega(1) \cdot \mathbb{E}\left[\mathcal{R}^+_{\mathcal{V}_{\text{in}}^B}(U^*)\right] ,$$

where the last equation follows from (9). By gluing the inventory vectors U_{in}^B over all buckets $B_1, \ldots, B_K$, we obtain an expected revenue of $\Omega(1) \cdot \mathbb{E}[\mathcal{R}^+_{\mathcal{V}_{\text{in}}}(U^*)]$. Finally, since the capacities used by U^* are not known a-priori, the assumption above can be bypassed by means of dynamic programming, similar to that of Section 3.2.3.

4 Nested Choice Model and General Demand Distribution

In this section, we provide a constant-factor approximation for the nested choice model under a general (non-IFR) demand distribution. This result is obtained through a sequence of structural transformations, allowing us to formulate the resulting instance as a (monotone) submodular maximization problem subject to a cardinality constraint. By leveraging the existing machinery in this context, we derive the following result.

Theorem 4.1. *Under the nested choice model, the dynamic assortment planning problem can be approximated within factor* $1 - 1/e$ *in polynomial time.*

4.1 Technical overview

For ease of exposition, we focus here on presenting the overall idea, and defer most of the technicalities to Sections 4.2 and 4.3.

Selection step: elimination of suboptimal products. The first step consists in simplifying the problem by identifying a well-structured collection of products, while preserving the optimal expected revenue. We begin by defining the quantity r_i, for each product $i \in [n]$, that denotes the expected revenue generated by a single customer arrival assuming that i is the most preferred product available. Namely, $r_i = p_i \cdot \sum_{\ell \in \mathcal{L}_i} \theta_\ell$, where $\mathcal{L}_i \subseteq \mathcal{L}$ is the subset of lists containing product i. Next, we define the i-maximal product as the highest-index product that maximizes the quantity r_j over $j \in [i, n]$. We show that, without any loss in optimality, we can restrict our attention to assortments included in the collection of i-maximal products, over all $i \in [n]$. This subset of products is designated by $\mathcal{V}$, while $\mathcal{V}(i)$ denotes the i-maximal product.

Lemma 4.2. *There exists an optimal inventory vector that stocks only products of* $\mathcal{V}$.

The proof of this claim is given in Section 4.2. The main observation is that there is no point in stocking any of the products strictly between two successive products in $\mathcal{V}$. Specifically, we prove that the expected revenue can only increase by shifting any unit of product $i \in [n] \setminus \mathcal{V}$ to the i-maximal product $\mathcal{V}(i)$. As a result, while preserving the optimal revenue, all products in $[n] \setminus \mathcal{V}$ are eliminated. Thus, we assume from this point on that r_i is non-increasing over $i \in [n]$.

Greedy step: set decision formulation. We now argue that the problem can equivalently be recast as the maximization of a set function subject to a capacity constraint. This modified problem is referred to as the 'set decision' formulation hereafter. Specifically, each product is represented by C distinct copies of identical price, that are consecutive in the preference order. (Recall that C represents the maximal number of units due to the capacity constraint.) Thus, there are exactly $N = n \cdot C$ distinct products. Each preference list is now represented by the interval of $[N]$ containing all copies of its initial products. Finally, a decision is made relative to each product, whether to stock it or not in the assortment. In other words, the inventory level decisions are replaced by a set decision over products. In this new formulation, our objective is to maximize the expected revenue over all subsets of products that satisfy the cardinality constraint.

Establishing submodularity. The expected revenue generated by a subset $S \subseteq [N]$ is denoted by $f(S)$. We now state our main technical result, which is established in Section 4.3.

Lemma 4.3. *The set function $f : 2^N \to \mathbb{R}^+$ is submodular and monotone.*

Interestingly, submodularity does not hold for arbitrary instances of the nested choice model, i.e., ones that were not processed by our elimination procedure, as demonstrated in Lemma A.2. In fact, the revenue function is also not concave, as we argue in Lemma A.3. To avoid deviating from the overall discussion, the proofs of these claims are given in Appendix A.3.

Submodular maximization problems have extensively been studied in combinatorial optimization, and in particular, when the input function is also monotone, this problem can be approximated within factor $1 - 1/e$ under a cardinality constraint Nemhauser et al. (1978). Moreover, the algorithm thereof is based on a greedy procedure that admits very efficient implementations when the function has an evaluation oracle. In our particular case, Goyal et al. Goyal et al. (2016) showed that the revenue function can be evaluated by dynamic programming in time $O(\bar{M} \cdot N \cdot k)$, where $\bar{M}$ is the maximal number of arrivals. By leveraging this algorithm, Theorem 4.1 immediately follows.

4.2 Proof of Lemma 4.2

For any inventory vector U and integer m, we define the random variable $\alpha_m(U)$ to denote the most preferred product available (i.e., with positive inventory level) for the m-th arriving customer, when initially stocking the vector U. If there are fewer than m arrivals, or if no units are left, $\alpha_m(U) = \infty$, denoting a dummy product with price $p_\infty = 0$. With this definition, the expected revenue function can be rewritten by conditioning on the most preferred product available upon each arrival, yielding

$$\mathbb{E}\left[\mathcal{R}(U)\right] = \mathbb{E}\left[\sum_{m=1}^{\infty} r_{\alpha_m(U)}\right] . \tag{10}$$

The desired result is proven by mapping any inventory vector $U = (u_1, \ldots, u_n)$ to a vector $\tilde{U} = (\tilde{u}_1, \ldots, \tilde{u}_n)$ that only stocks products in $\mathcal{V}$, has the same number of units, and generates at least as much expected revenue as U. The vector $\tilde{U}$ is constructed as follows: each unit of product i in U is represented in $\tilde{U}$ by a distinct unit of the maximal product $\mathcal{V}(i)$. That is,

$$\tilde{u}_i = \begin{cases} \sum_{k:\mathcal{V}(k)=i} u_k & \text{if } i = \mathcal{V}(i), \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, the number of units in $\tilde{U}$ is identical to that of U. This construction is illustrated in Figure 5.

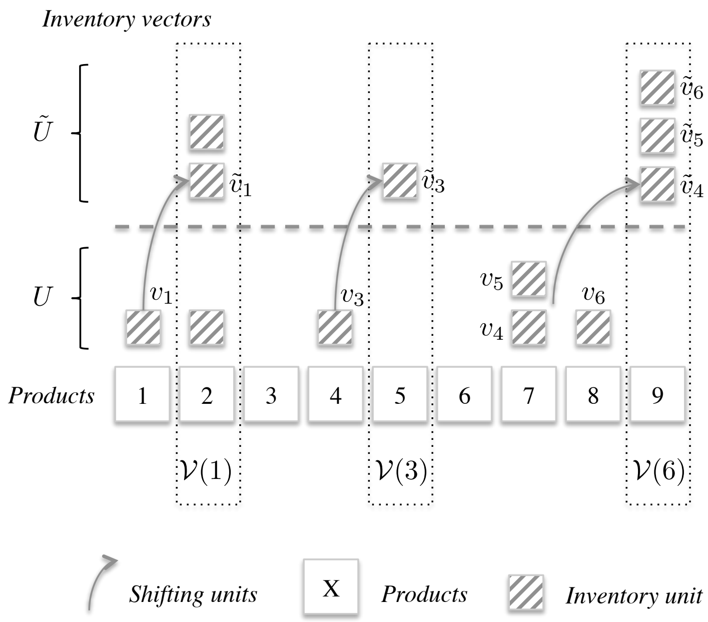

Figure 4: Construction of $\tilde{U}$ by shifting each unit in U towards its corresponding maximal product.

Claim 4.4. *For $m \geq 1$, every realization of the consumption process satisfies $\alpha_m(\tilde{U}) \leq \mathcal{V}(\alpha_m(U))$.*

Proof. To prove this inequality, we interpret each inventory vector as a non-decreasing sequence of products, that enumerates (with repetition) the units in the preference order. In other words, if X units are stocked of a given product, then this product is repeated X times consecutively in the sequence. Let $(v_j)_{j\leq C}$ and $(\tilde{v}_j)_{j\leq C}$ denote the sequences associated with the inventory vectors U and $\tilde{U}$, respectively. Without loss of generality, we assume that these vectors stock precisely C units.

Our construction of $\tilde{U}$ implies that $v_k \leq \tilde{v}_k = \mathcal{V}(v_k)$. Now assume that, just before the

m-th arrival, the most preferred product remaining when stocking initially the vector $\tilde{U}$ corresponds to the j-th unit of the sequence, meaning that $\alpha_m(\tilde{U}) = \tilde{v}_j$, and that all units in $\{v_1, \dots, v_{j-1}\}$ have been consumed by previously arriving customers. Each of these units can be mapped to the arrival rank of the customer who purchases it, entailing the subsequence of arrivals $(m_1, \dots, m_{j-1})$. Clearly, each product v_k belongs to the preference list L_{m_k}. The key observation is that, since $v_k \leq \tilde{v}_k$, the preference list L_{m_k} also contains product v_k for each $k \in [j-1]$. Consequently, when initially stocking the vector U, after the first $m-1$ arrivals of customers, whereby the preference lists $(L_{m_1}, \dots, L_{m_{j-1}})$ occurred precisely in this order, the units $\{v_1, \dots v_{j-1}\}$ are consumed as well. Indeed, each unit v_k would necessarily be consumed by the list L_{m_k}, if it were not purchased by a previously arriving customer. We thus obtain that $v_j \leq \alpha_m(U)$, meaning that $\alpha_m(\tilde{U}) = \mathcal{V}(v_j) \leq \mathcal{V}(\alpha_m(U))$.

■

To conclude Lemma 4.2, recall that r_i is the expected revenue generated by a single arrival, conditional on product i being the most preferred one available. Therefore, the expected revenue generated by the m-th arrival satisfies

$$r_{\alpha_m(U)} \leq r_{\mathcal{V}(\alpha_m(U))} \leq r_{\mathcal{V}(\alpha_m(\tilde{U}))} = r_{\alpha_m(\tilde{U})} ,$$

where the first inequality follows from the definition of maximal products, and the second inequality is due to Claim 4.4 and the monotonicity of $(r_i)_{i \in \mathcal{V}}$. Therefore, based on the revenue decomposition given by equation (10), we conclude that

$$\mathbb{E}\left[\mathcal{R}(U)\right] = \sum_{m=1}^{\infty} \mathbb{E}\left[r_{\alpha_m(U)}\right] \leq \sum_{m=1}^{\infty} \mathbb{E}\left[r_{\alpha_m(\tilde{U})}\right] = \mathbb{E}\left[\mathcal{R}(\tilde{U})\right] .$$

4.3 Proof of Lemma 4.3

Notation. Following the previous section, for every subset $S \subseteq [N]$ we define $\alpha_m(S)$ as the most preferred product available at the m-th arrival, when initially stocking the set S. If all products have stocked out, or if the number of arrivals is smaller than m, the value of $\alpha_m(S)$ is set to ∞, which corresponds to a dummy product with price 0. Using these random variables, the expected revenue can be decomposed similar to equation (10), namely $f(S) = \mathbb{E}[\sum_{m=1}^{\infty} r_{\alpha_m(S)}]$.

Monotonicity. Consider a subset $S \subseteq [N]$ and some product $i \in [N] \setminus S$. For each realization of the consumption process, it is easy to verify that, just before each arrival, the most preferred product available under the initial set decision S, is larger or equal to the one under the initial set $S \cup \{i\}$. That is, $\alpha_m(S) \geq \alpha_m(S \cup \{i\})$ for any realization. Thus, the revenue function f is indeed monotone since

$$f(S) = \sum_{m=1}^{\infty} \mathbb{E}\left[r_{\alpha_m(S)}\right] \leq \sum_{m=1}^{\infty} \mathbb{E}\left[r_{\alpha_m(S \cup \{i\})}\right] = f(S \cup \{i\}) ,$$

where the above inequality follows from the monotonicity of $(r_i)_{i \in [N]}$.

Submodularity. To prove that f is also submodular, it is sufficient to show that for any subset S and distinct products $i, j \in [n] \setminus S$, the expected revenue function satisfies

$$f(S \cup \{i,j\}) - f(S \cup \{j\}) \leq f(S \cup \{i\}) - f(S) \ .$$

One important observation is that we can assume without loss of generality that $j < i$; otherwise, by permuting i and j we obtain an equivalent inequality. For ease of exposition, we introduce the subset $S_1 = S$, $S_2 = S \cup \{i\}$, $S_1' = S \cup \{j\}$, and $S_2' = S \cup \{i,j\}$. With this notation, the desired inequality is

$$f(S_2') - f(S_1') \leq f(S_2) - f(S_1) \ . \tag{11}$$

Note that, if inequality (11) is satisfied for any deterministic number of arrivals, it generalizes to the case where M is stochastic. Thus, we restrict our attention to a deterministic number of arrivals M. In this case, the expected revenue increment can be written as

$$f(S_2') - f(S_1') = \sum_{m=1}^{M} \mathbb{E}\left[r_{\alpha_m(S_2')}\right] - \sum_{m=1}^{M} \mathbb{E}\left[r_{\alpha_m(S_1')}\right] \ . \tag{12}$$

We define τ as the stopping time corresponding to the first arrival where the minimal product available is i, when initially stocking S_2. Namely, we have $\alpha_\tau(S_2) = i$, and $\alpha_{\tau-1}(S_2) < i$ or $\tau = 1$. (In case τ is not defined, the stopping time is set to ∞.) Note that the stopping time τ corresponds to the first arriving customer faced by distinct minimal products, when initially stocking S_1 and S_2. Hence, the revenue difference between these two sets is a function of the arrivals after τ. Formally, since $\alpha_m(S_2) = \alpha_m(S_1)$ with probability 1 for all arrivals $m < \tau$, we infer from an analog of equality (12) for S_1 and S_2 that

$$\begin{aligned} f(S_2) - f(S_1) &= \sum_{m=1}^{\tau-1} \mathbb{E}\left[r_{\alpha_m(S_2)} - r_{\alpha_m(S_1)}\right] + \sum_{m=\tau}^{M} \mathbb{E}\left[r_{\alpha_m(S_2)} - r_{\alpha_m(S_1)}\right] \\ &= \sum_{m=\tau}^{M} \mathbb{E}\left[r_{\alpha_m(S_2)} - r_{\alpha_m(S_1)}\right] \\ &= \sum_{k=0}^{M-\tau} \mathbb{E}\left[r_{\alpha_{\tau+k}(S_2)} - r_{\alpha_{\tau+k}(S_1)}\right] \ . \end{aligned} \tag{13}$$

This simplification is made intuitive by Figure 5. Similarly, by defining τ' as the minimal arrival m such that $\alpha_m(S_2') = i$, we obtain:

$$f(S_2') - f(S_1') = \sum_{k=0}^{M-\tau'} \mathbb{E}\left[r_{\alpha_{\tau'+k}(S_2')} - r_{\alpha_{\tau'+k}(S_1')}\right] \ . \tag{14}$$

Now observe that, by an inductive argument, it is not difficult to prove that, for any realization of the consumption process, $\tau \leq \tau'$, meaning that $M - \tau \geq M - \tau'$. As a result, combining

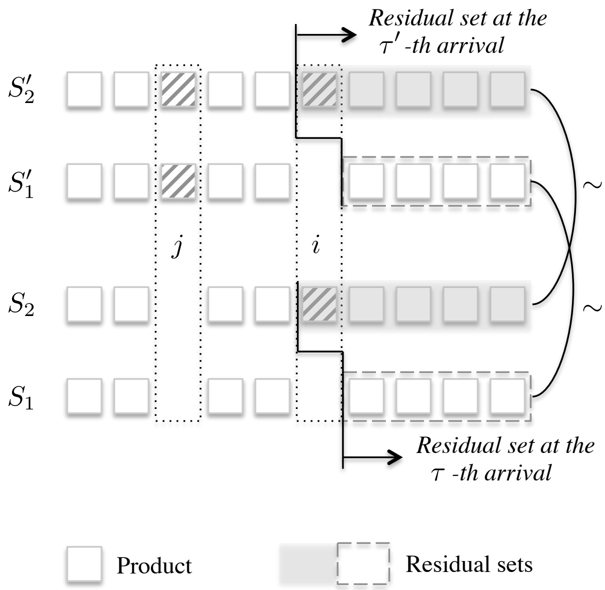

Figure 5: Equivalence between the residual sets at the stopping times τ and τ'.

inequalities (13) and (14), we have

$$(f(S_2) - f(S_1)) - (f(S_2') - f(S_1')) = \sum_{k=0}^{M-\tau'} \mathbb{E}\left[r_{\alpha_{\tau+k}(S_2)} - r_{\alpha_{\tau'+k}(S_2')}\right] - \sum_{k=0}^{M-\tau'} \mathbb{E}\left[r_{\alpha_{\tau+k}(S_1)} - r_{\alpha_{\tau'+k}(S_1')}\right] + \sum_{k=M-\tau'+1}^{M-\tau} \mathbb{E}\left[r_{\alpha_{\tau+k}(S_2)} - r_{\alpha_{\tau+k}(S_1)}\right] . \tag{15}$$

Now, the important observation is that the consumption process starting from the τ-th arrival only depends on the residual set of products available at that time. Hence, we can exploit the equivalence between the residual sets of products, reflected in Figure 5, to infer an equivalence between revenues:

- *Equivalence* $\alpha_{\tau'+k}(S_2') \sim \alpha_{\tau+k}(S_2)$. When initially stocking S_2', the residual set of products at the τ'-th arrival is the set $(\{i\} \cup S) \cap [i, N]$, which is equal to the residual set at the τ-th arrival when initially stocking S_2. Thus, we infer that the random variables $\alpha_{\tau'+k}(S_2')$ and $\alpha_{\tau+k}(S_2)$ are identically distributed for all $k \leq M - \tau'$.

- *Equivalence* $\alpha_{\tau'+k}(S_1') \sim \alpha_{\tau+k}(S_1)$. When initially stocking S_2', the residual set of products at the τ'-th arrival is the set $S \cap [i, N]$, which is exactly the residual set at the τ-th arrival when initially stocking S_2. Thus, we infer that the random variables $\alpha_{\tau'+k}(S_1')$ and $\alpha_{\tau+k}(S_1)$ are identically distributed for all $k \leq M - \tau'$.

As a result, equation (15) simplifies as follows:

$$(f(S_2) - f(S_1)) - (f(S_2') - f(S_1')) = \sum_{k=M-\tau'+1}^{M-\tau} \mathbb{E}\left[r_{\alpha_{\tau+k}(S_2)} - r_{\alpha_{\tau+k}(S_1)}\right] . \tag{16}$$

Since $S_1 \subseteq S_2$, it is easy to verify that $\alpha_m(S_2) \leq \alpha_m(S_1)$ with probability 1, for any arrival m.

Therefore, $r_{\alpha_m(S_2)} \geq r_{\alpha_m(S_1)}$, as our selection step guarantees that $r_1 \geq \cdots \geq r_n$. Combining this observation with inequality (16), we infer the desired inequality (11), proving that f is indeed submodular.

5 Computational Experiments

In this section, we demonstrate the practical merits of our algorithms against existing heuristics. In order to run such comparisons on instances of realistic scale, we take as a benchmark several efficient heuristics proposed in previous related work. Specifically, our algorithms are compared against the following: (i) a discrete-greedy algorithm; (ii) a local search heuristic; and (iii) a gradient-descent algorithm on a continuous extension of the expected revenue function. The latter two heuristics are directly inspired by the work of Mahajan and van Ryzin (2001) and Goyal et al. (2016). For the nested choice model, we have also implemented the enumeration-based algorithm of Goyal et al. (2016).

5.1 Heuristics and their implementation

In what follows, we summarize the different algorithms implemented for our computational experiments.

Discrete-greedy. The greedy algorithm starts with zero inventory levels for all products, and iteratively augments the current vector by a single unit of the product that incurs the largest marginal increase in the expected revenue, until reaching C units. As explained toward the end of this section, the expected revenue is evaluated by averaging random realizations of the revenue function, which are sampled by simulating the consumption process and choice behavior of arriving customers.

Local search. Starting from an initial inventory vector, the local search algorithm iteratively improves the expected revenue, by greedily transferring a single inventory unit from a one product to another. Formally, letting $U^{(k)}$ denote the inventory vector obtained at the beginning of step k, a swap is represented by an ordered pair of products (i, j) for which the current inventory level $u_i^{(k)}$ of product i is strictly positive. The inventory vector $U_{i \to j}^{(k)}$ resulting from this swap is derived from $U^{(k)}$ through decreasing $u_i^{(k)}$ by one unit and augmenting $u_j^{(k)}$ by one unit. With this definition, we either proceed to step $k+1$ with the inventory vector $U_{i \to j}^{(k)}$ that maximizes $\mathbb{E}[\mathcal{R}(U_{i \to j}^{(k)})]$ over all swaps (i, j), or terminate the algorithm when none of these swaps improves the expected revenue by at least 0.5%. Once again, the expected revenue function is estimated through sampling, while the initial inventory vector $U^{(1)}$ is defined by stocking C units of the best single-product assortment.

Gradient-descent approach. We consider an adaptation of the stochastic gradient-descent algorithm of Mahajan and van Ryzin (2001). In contrast to their setting, here the revenue function is defined only for integer-valued inventory vectors. Hence, similar to the approach of Goyal et al. (2016), we develop a continuous relaxation of the revenue function, defined through the Lovász extension of a discrete function. Letting $f : \mathbb{Z}^n \to \mathbb{R}$ denote the expected revenue function, its Lovász extension $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\hat{f}(U) = f(\lfloor U \rfloor) + \sum_{i=1}^{n} \left(u_{\pi(i)} - u_{\pi(i-1)}\right) \cdot \left[f\left(\lfloor U \rfloor + \sum_{k=1}^{i} e_{\pi(k)} \right) - f\left(\lfloor U \rfloor + \sum_{k=1}^{i-1} e_{\pi(k)} \right) \right] ,$$

where the permutation π sorts products by the increasing fractional part of their inventory, namely, $u_{\pi(1)} - \lfloor u_{\pi(1)} \rfloor \leq \cdots \leq u_{\pi(n)} - \lfloor u_{\pi(n)} \rfloor$. The Lovász extension is piecewise linear, and its gradient can be approximately computed through sampling.

Starting with the initial solution $U^{(0)} = 0$, and letting $U^{(k)}$ denote the solution obtained at the end of step k, each iteration consists of computing $U^{(k+1)} = \max\{0, U^{(k)} + \epsilon_{k+1} \nabla f(U^{(k)})\}$, where ϵ_{k+1} is the step size. When the latter vector does not lie in the feasible region $\{U \in \mathbb{R}^n : \|U\|_1 \leq C\}$, it is projected onto the boundary by linear rescaling. Through extensive trial and error, we chose an adaptive step size of

$$\epsilon_{k+1} = \max\left\{ \frac{C - \|U^{(k)}\|_1}{2}, 0.05 \cdot C \right\} .$$

Intuitively, the step size is larger when the vector $U^{(k)}$ is farther from the boundary, while still enforcing a minimal step size. The algorithm terminates when the objective value does not improve by a factor greater than 0.5%, or after 2000 iterations. Finally, it remains to round the resulting inventory vector to an integral one. Suppose that $U^{(k+1)}$ is the inventory vector obtained following the gradient-descent algorithm; then $\lfloor U^{(k+1)} \rfloor$ is augmented greedily, by stocking at each step a unit of the product with maximal marginal expected revenue, until reaching C units.

Approximation scheme for the nested choice model. We also implemented the approximation scheme developed by Goyal et al. Goyal et al. (2016) for the nested choice model. This algorithm begins by partitioning products into two classes, frequent and rare, the latter defined as the collection of products for which fewer than $\epsilon^2 C$ units are purchased in expectation when C units of that product are initially stocked. Next, all inventory vectors consisting of $O((1/\epsilon) \cdot \log(1/\epsilon))$ frequent products and a single rare product are approximately enumerated. However, even when we choose $\epsilon = 0.3$, corresponding to a guaranteed approximation ratio of only 0.1^5, this enumeration procedure is impractical and incurs exorbitant running times. In fact, time limits that still enable running the required experiments in practice (within less than an hour) resulted in only partial enumeration, even for the smallest instances considered. Therefore, to improve the observed performance, the enumeration order over inventory vectors is determined such that more expensive products are stocked with higher priority. After considering several different options, this heuristic was superior in balancing between performance and speed.

Revenue evaluation. To approximately evaluate the expected revenue function in each of the above-mentioned heuristics, we use a sample mean estimator over 100 realizations of the consumption process. The number of samples is uniform over all heuristics to provide "fair" comparisons. In contrast, our approximation algorithms, both for the general model and for intervals, make use of the newsvendor-based lower bound, that can be computed exactly and very efficiently.

5.2 Instance generation

We fix the number of products at $n = 20$, while the capacity parameter C is varied in $\{20, 40, 80, 150\}$. The prices of these products are generated randomly from a log-normal distribution with location $\mu = 0$ and scale $\sigma = 0.3$. The parameters underlying the consumption process are sampled randomly according to the following generative models.

Choice models. We run three series of experiments:

- *General choice model:* The number of preference lists (in the support of the distribution) is given by $|\mathcal{L}| \in \{50, 160\}$. The preference lists $L_j \in \mathcal{L}$ are generated independently and identically through the following random procedure. We first determine the set of products in L_j through a sequence of Bernoulli trials with $p = 0.35$, i.e., each of the 20 products is independently picked to appear in this list with probability 0.35. Next, for each list L_j, the ranking order over product alternatives in L_j is sampled uniformly over all $|L_j|!$ permutations.

- *Intervals choice model:* Similarly, we vary the number of interval preference lists $|\mathcal{L}|$ in $\{50, 200\}$. To create $\mathcal{L}$, we sample uniformly at random over all subsets formed by $|\mathcal{L}|$ distinct intervals (out of $n(n+1)/2$ intervals overall).

- *Nested choice model:* Here, we pick $|\mathcal{L}| = n$, meaning that every nested list is included in $\mathcal{L}$.

In each of the above models, the probability distribution over $\mathcal{L}$ is sampled uniformly over the unit simplex.

Number of customers M. We test several parametric distributions for the demand M. Specifically, letting $G \sim N(30, 40)$ and $P \sim \exp(0.02)$, we alternatively examine $M = \min\{[\lceil G \rceil | G \geq 0], 100\}$ and $M = \min\{\lceil P \rceil, 100\}$. For the nested choice model, we also test a nonparametric non-IFR demand distribution, where the support of M is constructed by randomly sampling 5 integers in the interval $[1, 100]$ (without replacement), and the corresponding probability measure is generated uniformly at random from the unit simplex.

5.3 Results

We implemented our algorithms, as well as the heuristics described in Section 5.1, using the Julia programming language. The experiments described in this section were conducted on a standard laptop with 2.7GHz Intel Core i5 processor and 8GB of RAM. In order to execute the required experiments within several days, we imposed a time limit of 10 minutes (per-instance) for all heuristics tested, noting that each of our algorithms terminates within a few seconds.

Relative performance. For each instance tested, in practice one cannot simply compute the optimal expected revenue through brute-force enumeration, as the latter involves considering all combinations of feasible inventory vectors, number of customers, and their choice preferences. Similarly, we found integer programming approaches to be impractical; for example, using a sample average approximation with 100 realizations of the consumption process, the running time exceeds 30 minutes for the smallest instance tested. Therefore, rather than estimating the exact optimality gaps, we compare the different algorithms on a relative basis. Specifically, for each instance, the benchmark is set as the expected revenue of the most profitable inventory vector obtained through all algorithms tested. Then, the relative performance of each algorithm (reported subsequently) is defined as the ratio between its expected revenue and the benchmark. For example, if our algorithm attains an expected revenue of 1, while all tested heuristics generate an expected revenue of 0.9, the relative performance is 100% for our algorithm, and 90% for the other heuristics.

Performance analysis. The results of our experiments are summarized in Tables 1, 2, and 3, where each entry is obtained by averaging over 30 random instances. Our algorithms dominate the other heuristics revenue-wise in all configurations. Under a general generative model, the discrete-greedy algorithm emerges as the most effective heuristic on average, but it still falls behind our general approximation by 7% to 36%. There is a single configuration where the gradient-descent algorithm outperforms the other heuristics, specifically when $C = 80$. Interesting, in the latter regime, the inventory capacity approximately matches the demand in expectation. Since this observation is consistent across all choice models and demand distributions, our experiments seem to indicate that the gradient-descent is particularly effective in such a regime (while still falling short of the algorithms we propose).

Table 1: Results under the general choice model ($|L|$=160)

parameters		average relative performance (%)				average running time (sec.)			
M	C	GA	DG	GD	LS	GA	DG	GD	LS
Gaussian	20	100	87.6	66.5	77.4	6.6	3.1	55.9	7.9
	40	100	70.6	77.2	62.1	7.1	6.4	216.2	16.9
	80	100	64.1	78.9	54.4	7.3	13	374.8	21.4
	150	100	66.8	62.2	50.6	8.1	24.3	384.2	19.7
Poisson	20	99.9	93.1	66.4	91.4	6.9	3.7	36.6	10.7
	40	100	77.3	80.7	65.6	6.9	7.4	238.8	17.1
	80	100	72.6	86.5	57.6	7.6	15.2	435.3	27.5
	150	100	71	67.8	53.4	8.3	28.4	442	20.4

Here, GA designates our general approximation algorithm, DG is the discrete-greedy algorithm, GD is the gradient-descent approach, and LS corresponds to the local search heuristic.

Surprisingly, even in the more specialized settings, our general-purpose approximation still enjoys strong practical performance. Specifically, its performance is comparable to the decomposition algorithm of Section 3 in the intervals case. Under the nested model, the selective-greedy algorithm developed in Section 4 improves revenue by 1% on average against our general approximation. It is worth mentioning that, as one would expect, the discrete-greedy and selective-greedy algorithms have identical performance under the nested model. However, our selection step restricts the incremental actions examined upon each iteration, as opposed to considering all possible stocking decisions at each iteration, and therefore reduces the running time significantly.

Table 2: Results under the intervals choice model ($|L|$=160)

parameters		average relative performance (%)					average running time (sec.)				
M	C	GA	IN	DG	GD	LS	GA	IN	DG	GD	LS
Gaussian	20	99.2	99.4	89.6	52.7	67.4	5.6	6.8	2.0	33.7	5.3
	40	99.2	99.0	62.1	60.6	25.6	6.0	10.7	4.0	136.2	8.8
	80	97.5	98.4	49.1	76.3	16.6	6.2	11.7	8.1	229.5	19.6
	150	99.5	97.7	48.3	27.6	15.7	6.7	7.1	15.1	244.8	13.5
Poisson	20	99.7	99.6	97.3	52.7	93.3	8.1	10.8	3.6	23.3	9.3
	40	99.4	99.2	76.8	63.4	37.2	8.5	9.0	7.3	232.7	16.6
	80	99.0	99.1	57.2	79.7	17.7	9.0	10.8	14.5	399.4	15
	150	98.4	99.3	57.3	37.1	18	9.9	10.4	27.2	430.1	51.9

Here, GA designates our general approximation algorithm, IN is our approximation for the intervals choice model, DG is the discrete-greedy algorithm, GD is the gradient-descent approach, and LS corresponds to the local search heuristic.

On the computational front, our algorithms tend to outperform the other heuristics while the gradient-descent algorithm is the slowest one. In contrast to other algorithms, the running time of the general approximation grows sub-linearly in the capacity value; hence, when $C = 150$, its running time is better by a factor of 3 to 40 against other heuristics. Under the nested choice model, the selective-greedy algorithm is an order of magnitude faster than any other algorithm for small instances ($C \leq 80$). In this setting, the approximation scheme of Goyal et al. (2016) is particularly inefficient, and reaches the time limit of 10 minutes prior to completing its full enumeration procedure. As a result, the revenue performance is not near-optimal, particularly for large instances ($C = 150$), where the average optimality gap can be as large as 18%.

6 Concluding Remarks

Robustness under a mixture of nested models. Given the very structured nature of nested preference lists, it is interesting to investigate whether our algorithms can be utilized in the case of a model misspecification, specifically, under a mixture of nested instances. Here, each segment of the mixture is uniquely described by its left endpoint product; this setting is a special case of the intervals choice model. It is not difficult to verify that, when the mixture is formed by K customer segments, we can derive an $O(\log K)$-approximation by adapting the recursive decomposition of Section 3.2. Specifically, at every step of the recursion, each segment of the partition is broken at the median left endpoint product of all remaining lists contained in that segment. As a result, the recursion depth is $O(\log K)$, thus leading to the above-mentioned approximation ratio.

Newsvendor-like models. The problem formulation considered in this paper incorporates a hard capacity constraint on the number of units stocked. A natural direction for future research

Table 3: Results under the nested choice model ($|L|$=20)

parameters		average relative performance (%)						average running time (sec.)					
M	C	GA	SG	GLS	DG	GD	LS	GA	SG	GLS	DG	GD	LS
Gaussian	20	96.4	99.8	93	99.8	78	99.7	0.8	< 0.1	600	0.1	1.4	0.3
	40	99.5	99.8	89.2	99.8	92.9	83.9	0.9	< 0.1	600	0.3	2	0.9
	80	99.9	99.7	91.6	99.7	99.3	69.9	0.9	0.3	600	1.3	3.8	0.9
	150	100	100	85.2	100	96.5	69.7	1	0.9	600	3.9	5.7	12.2
Poisson	20	95.6	99.4	92.3	99.4	83.2	99.9	0.8	< 0.1	600	< 0.1	1.5	0.3
	40	98.5	99.5	91.1	99.5	91.7	97.1	0.9	< 0.1	600	0.3	2.9	1
	80	99.5	99.7	89.8	99.7	99.4	83.9	0.9	0.3	600	1.3	4	2.6
	150	100	100	82	100	99.3	71.2	0.9	1	571	3.8	5	12.8
Non-IFR	20	97.1	99.2	92.1	99.2	79.8	99.5	0.8	< 0.1	600	< 0.1	1.5	0.3
	40	95.9	99	91.3	99	89.3	95.6	0.8	< 0.1	600	0.3	2.5	1.1
	80	99.9	98.9	89.8	98.9	97.7	84.4	0.9	0.2	600	1.2	4.8	1.7
	150	100	100	86.4	100	96.7	72.3	0.9	1.2	561.9	3.7	5	18.6

Here, GA designates our general approximation algorithm, SG is our selective-greedy approximation for the nested choice model, DG is the discrete-greedy algorithm, GD is the gradient-descent approach, LS corresponds to the local search heuristic, and GLS is the enumeration-based algorithm of Goyal et al. (2016).

is to study newsvendor-like models, where there is no capacity limitation, and instead, the salvage value of inventory has decreasing marginal gains. It would be interesting to investigate whether the technical ideas we developed can be leveraged to this setting.

Refined approximability results. Following the present work, one open question is that of determining whether the intervals model can be efficiently approximated within a constant factor. Although we were able to obtain a performance guarantee of $O(\log\log(p_{\max}/p_{\min}))$, even the simpler problem of evaluating the expected revenue of a given inventory vector in this model is still wide open. An interesting direction to consider would be to propose structural transformations, in the spirit of Section 4, in order to reveal certain submodularity properties. It is worth mentioning that, even with identical prices, the original revenue function is not submodular, as we demonstrate in Lemma A.4 (see Appendix A.3).

A Additional Proofs

A.1 Proof of Lemma 2.6

In what follows, we prove a sufficient condition for subadditivity, stating that with respect to any inventory vector, the deletion of units can only increase the consumption probability of any remaining unit. Formally, any inventory vector U is viewed as a collection of units, each of which is a separate copy of a given product; within units of the same product, a fixed (arbitrary) order is set, according to which they are consumed by customers. We denote by U^{-i} the inventory vector obtained from U by deleting the first unit of product i, and use $[\mathcal{E}_v|U]$ to designate the event where unit v is purchased during the consumption process with the initial inventory vector U.

Lemma A.1. *For any inventory vector U, any product $i \in [n]$ with $U_i \geq 1$, and any unit $v \in U^{-i}$,*

$$\Pr\left[\mathcal{E}_v \left| U^{-i}\right.\right] \geq \Pr\left[\mathcal{E}_v|U\right] \ .$$

Proof. Given an initial inventory vector U, we designate by $U_{(1)}, \ldots, U_{(M)}$ the random sequence of residual inventory levels facing each customer arrival, i.e., $U_{(m)}$ is the inventory vector facing the m-th arriving customer. To prove the desired claim, it suffices to show that $U_{(m)}(w) \geq U^{-i}_{(m)}(w)$ for any realization w of the consumption process. Here, any such realization corresponds to the specific outcomes of the number of arriving customers M and their choices of preference lists.

We now focus on a fixed realization w, and prove the latter claim by induction on the arrival rank m. For $m = 1$, we have by definition $U_{(1)}(w) = U > U^{-i} = U^{-i}_{(1)}(w)$. For the general case, by the induction hypothesis, we have $U_{(m-1)}(w) \geq U^{-i}_{(m-1)}(w)$. Now let k be the first product on the preference list of customer $m-1$ (picked according to w) that is stocked by $U_{(m-1)}(w)$; we define $k = \infty$ when not such product exists. As any product stocked by $U^{-i}_{(m-1)}(w)$ is necessarily stocked by $U_{(m-1)}(w)$, there are three cases:

1. $k = \infty$: Here, customer $m-1$ does not purchase any unit with respect to both $U_{(m-1)}(w)$ and $U^{-i}_{(m-1)}(w)$, meaning that $U_{(m)}(w) = U_{(m-1)}(w) \geq U^{-i}_{(m-1)}(w) = U^{-i}_{(m)}(w)$.

2. *$k < \infty$ and product k is stocked by both $U_{(m-1)}(w)$ and $U^{-i}_{(m-1)}(w)$*: In this case, customer $m-1$ purchases a single unit of product k in $U_{(m-1)}(w)$ and $U^{-i}_{(m-1)}(w)$, implying that $U_{(m)}(w) \geq U^{-i}_{(m)}(w)$.

3. *$k < \infty$ and product k is stocked by $U_{(m-1)}(w)$ but not by $U^{-i}_{(m-1)}(w)$*: Even though customer $m-1$ purchases a single unit of product k in $U_{(m-1)}(w)$, the remaining number of units stocked of this product is still greater or equal to the same quantity with respect to the residual vector after purchasing from $U^{-i}_{(m-1)}(w)$. For the latter, customer $m-1$ purchases a single unit of a product different than k or does not purchase at all; in either case, we have $U_{(m)}(w) \geq U^{-i}_{(m)}(w)$.

■

A.2 Proof of Lemma 2.4

The claim follows by observing that

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\min\{X, C\} + [X-C]^+\right] \\ &= \mathbb{E}\left[\bar{X}\right] + \Pr[X \geq C] \cdot \mathbb{E}[X - C | X \geq C] \\ &\leq \mathbb{E}\left[\bar{X}\right] + \delta \cdot \mathbb{E}[X] \ .\end{aligned}$$

The last inequality holds since X is IFR and since

$$\delta C \geq \mathbb{E}\left[\bar{X}\right] = \mathbb{E}[\min\{X, C\}] \geq C \cdot \Pr[X \geq C] \ .$$

A.3 Counter-Examples

Lemma A.2. *Under the nested choice model and a single arrival, the set function $f: \{0,1\}^N \to \mathbb{R}^+$ is neither monotone nor submodular .*

Proof. Consider the following instance: there are three products denoted by $\{1,2,3\}$ with respective prices $p_1 = 1, p_2 = 2$ and $p_3 = 3$, while the total capacity is $C = 1$. We model a single customer arrival, where the list $(1,2,3)$ occurs with probability 1. With a slight abuse of notation, where sets are used instead of binary sequences, it is easy to verify that for $S_1 = \{3\}$ and $S_2 = \{1,3\}$, we have $f(S_1) > f(S_2)$, and $f(S_1 \cup \{2\}) - f(S_1) = -1$ while $f(S_2 \cup \{2\}) - f(S_2) = 0$. ■

Lemma A.3. *Under the nested choice model and IFR demand distributions, the expected revenue function is not concave.*

Proof. Consider the following instance: there are two products denoted by $\{1,2\}$, with prices $p_1 = 0$ and $p_2 = 1$. There are two arriving customers, each of which draws the preference list $(1,2)$ with probability 1/2 and the empty list with probability 1/2. We consider the inventory vectors $(2,0)$, $(0,2)$ and $(1,1)$. We observe that $\mathbb{E}[\mathcal{R}(2,0)] = 0$, $\mathbb{E}[\mathcal{R}(0,2)] = 1$ and $\mathbb{E}[\mathcal{R}(1,1)] = 1/4$. As a result, we obtain that

$$\frac{1}{2} \cdot \mathbb{E}[\mathcal{R}(2,0)] + \frac{1}{2} \cdot \mathbb{E}[\mathcal{R}(0,2)] > \mathbb{E}[\mathcal{R}(1,1)] = \mathbb{E}\left[\mathcal{R}\left(\frac{1}{2} \cdot (2,0) + \frac{1}{2} \cdot (0,2)\right)\right] \ .$$

■

Lemma A.4. *Under the interval choice model, the revenue function is not submodular, even with deterministic arrivals and uniform prices.*

Proof. To construct a counterexample, we consider the collection of products $\{1, 2, 3\}$ such that $p_1 = p_2 = p_3 = 1$, and define the consumption process where there are exactly two arrivals. Each of these customers samples a preference list according to the following distribution: with probability $1 - \epsilon$, the customer chooses the list (2), and with probability ϵ she chooses the list $(1, 2, 3)$, where $\epsilon < 1/2$.

For this instance, focusing on the sets of products $S_1 = \{3\}$ and $S_2 = \{2, 3\}$, we have $\mathbb{E}[\mathcal{R}(S_1 \cup \{1\})] - \mathbb{E}[\mathcal{R}(S_1)] = \epsilon^2$ whereas $\mathbb{E}[\mathcal{R}(S_2 \cup \{1\})] - \mathbb{E}[\mathcal{R}(S_2)] = \epsilon \cdot (1 - \epsilon)$. Note that $\epsilon \cdot (1 - \epsilon) > \epsilon^2$ since $\epsilon < 1/2$, meaning that

$$\mathbb{E}\left[\mathcal{R}(S_2 \cup \{1\})\right] - \mathbb{E}\left[\mathcal{R}(S_2)\right] > \mathbb{E}\left[\mathcal{R}(S_1 \cup \{1\})\right] - \mathbb{E}\left[\mathcal{R}(S_1)\right] \ .$$

■

Acknowledgments.

The work of Ali Aouad and Retsef Levi was partially supported by NSF grant CMMI-1537536. Danny Segev's research on this project was supported by the ISRAEL SCIENCE FOUNDATION (grant no. 148/16).

References

Aouad, A., Farias, V., Levi, R. and Segev, D. (2015), The Approximability of Assortment Planning under Ranking Preferences. Operations Research (forthcoming), available as SSRN report 2612947.

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

Chen, F. and Bassok, Y. (2008), 'Variety and substitution'. Working paper.

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

Désir, A., Goyal, V., Segev, D. and Ye, C. (2015), 'Capacity constrained assortment optimization under the markov chain based choice model', Working paper, available as SSRN report 2626484.

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

Feige, U. (1998), 'A threshold of $\ln n$ for approximating set cover', *Journal of the ACM* **45**(4), 634–652.

Feldman, J. B. and Topaloglu, H. (2017), 'Revenue management under the markov chain choice model', *Operations Research* **65**(5), 1322–1342.

Fisher, M. (2011), 'Don't trust your gut with assortment planning'. Harvard Business Review.

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

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

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

Honhon, D., Gaur, V. and Seshadri, S. (2010), 'Assortment planning and inventory decisions under stockout-based substitution', *Operations research* **58**(5), 1364–1379.

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

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

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

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

Muckstadt, J. and Sapra, A. (2010), *Principles of inventory management: When you are down to four, order more*, Springer Science & Business Media.

Nagarajan, M. and Rajagopalan, S. (2008), 'Inventory models for substitutable products: optimal policies and heuristics', *Management Science* **54**(8), 1453–1466.

Nemhauser, G., Wolsey, L. and Fisher, M. (1978), 'An analysis of approximations for maximizing submodular set functions', *Mathematical Programming* **14**(1), 265–294.

Rusmevichiengtong, P., Roy, B. V. and Glynn, P. W. (2006), 'Nonparametric approach to multiproduct pricing', *Operations Research* **54**(1), 82–98.

Rusmevichientong, P. and Topaloglu, H. (2012), 'Robust assortment optimization in revenue management under the multinomial logit choice model', *Operations research* **60**(4), 865–882.

Ryzin, G. v. and Mahajan, S. (1999), 'On the relationship between inventory costs and variety benefits in retail assortments', *Management Science* **45**(11), 1496–1509.

Segev, D. (2015), Assortment planning with nested preferences: Dynamic programming with distributions as states? Working paper, available as SSRN report 2587440.

Shaked, M. and Shanthikumar, J. (1994), *Stochastic orders and their applications*, Academic Press, New York.

Smith, S. A. and Agrawal, N. (2000), 'Management of multi-item retail inventory systems with demand substitution', *Operations Research* **48**(1), 50–64.

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

Topaloglu, H. (2013), 'Joint stocking and product offer decisions under the multinomial logit model', *Production and Operations Management* **22**(5), 1182–1199.

Zeithalm, V. A. (1988), 'Consumer perception of price, quality and value: a means-end model and synthesis of evidence', *Journal of Marketing* **52**, 2–22.

Segev, D. [2015], Assortment planning with nested preferences: Dynamic programming with distributions as states? Working paper, available as SSRN report 2587440.

Shaked, M. and Shanthikumar, J. [1994], *Stochastic orders and their applications*, Academic Press, New York.

Smith, S. A. and Agrawal, N. [2000], 'Management of multi-item retail inventory systems with demand substitution', *Operations Research* **48**(1), 50–64.

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

Topaloglu, H. [2013], 'Joint stocking and product offer decisions under the multinomial logit model', *Production and Operations Management* **22**(5), 1182–1199.

Zeithalm, V. A. [1988], 'Consumer perception of price, quality and value: a means-end model and synthesis of evidence', *Journal of Marketing* **52**, 2–22.