Essays on Inventory Management and Object Allocation

Thiam Hui Lee

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

Abstract

Essays on Inventory Management and Object Allocation

Thiam Hui Lee

This dissertation consists of three essays. In the first, we establish a framework for proving equivalences between mechanisms that allocate indivisible objects to agents. In the second, we study a newsvendor model where the inventory manager has access to two experts that provide advice, and examine how and when an optimal algorithm can be efficiently computed. In the third, we study classical single-resource capacity allocation problem and investigate the relationship between data availability and performance guarantees.

We first study mechanisms that solve the problem of allocating indivisible objects to agents. We consider the class of mechanisms that utilize the Top Trading Cycles (TTC) algorithm (these may differ based on how they prioritize agents), and show a general approach to proving equivalences between mechanisms from this class. This approach is used to show alternative and simpler proofs for two recent equivalence results for mechanisms with linear priority structures. We also use the same approach to show that these equivalence results can be generalized to mechanisms where the agent priority structure is described by a tree.

Second, we study the newsvendor model where the manager has recourse to advice,

or decision recommendations, from two experts, and where the objective is to minimize worst-case regret from not following the advice of the better of the two agents. We show the model can be reduced to the class machine-learning problem of predicting binary sequences but with an asymmetric cost function, allowing us to obtain an optimal algorithm by modifying a well-known existing one. However, the algorithm we modify, and consequently the optimal algorithm we describe, is not known to be efficiently computable, because it requires evaluations of a function v which is the objective value of recursively defined optimization problems. We analyze v and show that when the two cost parameters of the newsvendor model are small multiples of a common factor, its evaluation is computationally efficient. We also provide a novel and direct asymptotic analysis of v that differs from previous approaches. Our asymptotic analysis gives us insight into the transient structure of v as its parameters scale, enabling us to formulate a heuristic for evaluating v generally. This, in turn, defines a heuristic for the optimal algorithm whose decisions we find in a numerical study to be close to optimal.

In the third essay, we study the classical single-resource capacity allocation problem. In particular, we analyze the relationship between data availability (in the form of demand samples) and performance guarantees for solutions derived from that data. This is done by describing a class of solutions called ε-backwards accurate policies and determining a suboptimality gap for this class of solutions. The suboptimality gap we find is in terms of ε and is also distribution-free. We then relate solutions generated by a Monte Carlo algorithm and ε-backwards accurate policies, showing a lower bound on the quantity of data necessary to ensure that the solution generated by the algorithm is ε-backwards accurate with a high probability. Combining the two results then allows us to give a lower bound on the data needed to generate an α-approximation with a given confidence probability $1 - \delta$. This lower bound is polynomial in the number of fares, M, and $1/\alpha$.

Contents

List of Figures

List of Tables

Acknowledgements

Throughout the course of my doctoral studies, I have been very lucky and fortunate to receive the support and encouragement of many people.

First of all, I would like to thank the professors who advised me, Prof. Tim Huh and Prof. Jay Sethuraman. Prof. Huh has always been one of the most approachable people I know. He would always be patient and willing to listen and have a discussion when I would spring a surprise visit to his office or accost him in the hallways of the department. What continually amazes me is his ceaseless energy and passion and his tremendous curiosity, which have an infectiousness that I can personally attest to. He instilled in me an open-mindedness toward learning about and tackling problems that are outside my zone of comfort, something I will always be grateful for. Prof. Sethuraman has been the other great influence and role-model for me. He has continually encouraged me to seek the right way to think about the problems we work on, to find approaches that grant insight into and unravel the intricacies of the problems. Moreover, the manner in which he analyses a situation or problem shows a lucidity of mind I have always sought to emulate. Both Profs. Huh and Sethuraman are two of the most inspiring and warmest people I know, and I will miss interacting with them and learning from them as I move forward from my doctoral studies. I am truly in their debt for what they have taught me and done for me.

Next, I would like to thank Prof. Retsef Levi. His intensity and work ethic had inspired me to push forward in my project with him and Prof. Huh when I had doubts on whether the results would bear out. His insightful comments were also very helpful in the completion of our work together.

Following, I would like to express my gratitude to Prof. Maria Chudnovsky and Adjunct Prof. Michael Lipkin. I have enjoyed and continue to enjoy discussions with them about art, literature, music and life. They have been wonderful mentors and friends.

I would also like to thank my fellow students in the IEOR department and friends from other departments at Columbia from the past and in the present. In particular, I would like to thank Rishi Talreja, Abhinav Verma, Margaret Pierson, Rae Yunzi Tan and Özge Sahin for their close friendships. They have enriched my life both intellectually and emotionally in so many ways over the years.

I also wish to thank my parents, Ah Bee Lee and See Bee Chew, and my siblings Jon and Cristal Lee for their constant love and support, without which this dissertation would not have been possible.

Last but not least, I would like to thank all members of my defense committee: Prof. Soulaymane Kachani and Prof. Vineet Goyal in addition to Prof. Tim Huh, Prof. Jay Sethuraman and Prof. Özge Sahin. I appreciate their taking time to read this dissertation carefully and also appreciate their comments which have been very constructive.

Chapter 1

Introduction

1.1 Overview

This dissertation studies facets of three major problems of Operations Research: the problem of allocating indivisible objects, the newsvendor problem, and the capacity allocation problem.

In Chapter 2, we study mechanisms for allocating indivisible objects to agents. These mechanisms are algorithms that find matchings between agents and objects given agents' (strict) preference orderings over the objects. In particular, we consider the class of mechanisms that use the Top Trading Cycles (TTC) algorithm, which we call TTC mechanisms. Part of the specification of any TTC mechanism is a prioritization over agents for each object that indicate which agent has ownership over that object under what conditions. Allocations are found by serially conducting exchanges of owned objects between unmatched agents, such that each agent involved in an exchange ends up with his favorite unmatched object at that time. Such exchanges are found are by forming graphs involving unmatched agents and finding cycles within these graphs. Members of this class of mechanisms differ on how agents are prioritized for each object, which influences what cycles may appear in the graph.

We establish an approach to proving equivalences between TTC mechanisms, i.e. a method of showing two mechanisms from this class generate the same matchings with the same probabilities. This approach describes the pair of mechanisms using recursion, and show that under certain conditions, the terms in the recursion due to cycles common to both mechanisms have the same contribution while terms due to cycles appearing in one mechanism but not the other have zero overall contribution. The latter is shown using a simple bijection.

Using our proof approach, we are able to show alternative and concise proofs for two recent equivalence results for TTC mechanisms with agent priorities specified by linear structures, i.e. agent priorities for each object is described by a linear ordering (which may not necessarily be complete). Additionally, we use the same framework to generalize these equivalence results to a class of mechanisms that use the TTC algorithm where agent priorities are specified by tree structures.

In Chapter 3, we study the newsvendor problem with expert advice. The newsvendor problem is one where an inventory manager has to decide a stocking level for goods he is selling, and, depending on the demand, incurs an overage cost if he stocks too much and an underage cost if he stocks too little. We look at a multi-period version of the problem where the goods are perishable and where the manager has access to advice about stocking levels given by experts. The manager wishes to minimize his regret of not having followed the advice of the best expert, i.e. to minimize the cost he would have incurred he had followed the best expert's advice throughout minus the cost he actually incurred. In particular, his objective to minimize the worst-case regret.

We show that this model can be reduced to a problem of predicting sequences with expert advice, a widely-studied problem in machine-learning and computer science, with an asymmetric linear cost function (i.e. with a different overage and underage cost). Moreover, we are able to modify an algorithm for the sequence prediction problem with linear costs so that it is optimal for the newsvendor context with asymmetric costs as well. However, computing decisions using this algorithm requires evaluating a function v described by optimal value of a recursively defined optimization problem, and the question whether v could be evaluated efficiently was previously unanswered. Analyzing v comprises the main focus of the chapter, where we attempt to answer this question. First, we find that v can be efficiently computed when the overage and underage costs are small multiples of some common factor, and even more efficiently if one is a multiple of the other. Second, we analyze the asymptotic behavior of v, which leads to a heuristic for v and consequently a heuristic algorithm for the manager's problem.

And in Chapter 4, we consider the problem of managing non-replenishable inventory of a single resource that is sold in stages with a different price in each stage, where the manager's problem is to maximize revenue by controlling in each stage how much inventory to reserve for future stages. This is a classical problem in revenue management typically called the single-resource capacity allocation problem. We study the relationship between the availability of data (in the form of demand samples) and performance for this problem. This study is done in two parts.

The first part examines a parametrized class of solutions to the problem, called "ε-backwards accurate" policies, and show a suboptimality bound on its performance in terms of ε that does not depend on the distribution of demands.

To connect this result to data availability, we show that a Monte-Carlo algorithm can generate an ε-backwards accurate policy with a given confidence probability if sufficiently many data samples are used as input to the algorithm. This allows us to connect a desired performance level, the confidence probability of achieving that performance level and the number of samples that the Monte Carlo algorithm needs to find such a solution.

1.2 Background

1.2.1 Allocation of indivisible objects

The problem of allocating indivisible objects is a problem where a number of indivisible objects are to be assigned to an equal number of agents, with each agent receiving exactly one object and without monetary exchange between the agents. Note that this problem is also commonly referred to as the *assignment* or *housing allocation* problem. This problem formulation has found use in several applications where allocation of resources should take place *without* the exchange of money, including allocating on-campus housing for students ([3],[20]), placing students in public schools ([1]), and designing kidney exchanges ([71]).

This problem is normally framed as one of how to ensure fairness and efficiency (in terms of social welfare). Agents are typically assumed to have strict individual preferences over the objects: for each agent, his utility of being assigned different objects is compared using a strict ordering of the objects (which may be unique to him) and not with a numerical scale; this is also called an *ordinal* preference ordering. For example, an agent has preferences $a \succ b \succ c$ over the objects $\{a, b, c\}$, so he prefers a to b to c. The objective is to find an algorithm, usually called a *mechanism* in this context, that finds a matching between the agents and the objects satisfying certain fairness criteria while being efficient. It is not difficult to see that deterministic mechanisms suffer from issues of fairness (for example, two differently labeled but otherwise identical agents must necessarily be treated differently by a deterministic mechanism), so it is common for mechanisms that are studied to use randomness. This problem is therefore called the *random assignment* problem.

Early literature considered two models: the *housing market* model introduced by Shapley and Scarf [78] in which each agent is already endowed with an object (and each agent is given top priority for the object he is endowed with); or the *house allocation* model introduced by Hylland and Zeckhauser [43] that had no endowments. There is a vast literature on both models at this point.

For the housing market model, Shapley and Scarf [78] propose an algorithm, called the *top trading cycles* (TTC) algorithm that computes a unique reallocation of objects to agents based on agent preferences and their endowments. The TTC algorithm does the following: at any juncture, it maintains a graph with agents as nodes, and has an arc from agent i to agent j if j is endowed with i's most preferred remaining object; this graph can be shown to contain a cycle (possibly a self-loop); a cycle is chosen and every agent in that cycle is matched with his most preferred remaining object, after which the matched agents and objects are removed, and the graph is updated and the procedure is repeated if any agents and objects remain. Shapley and Scarf showed that this algorithm computes a matching from the core of the associated cooperative game. Roth and Postlewaite [73] later showed that the core is unique. Additionally, under the scenario where agents are to required to report their preferences, Roth [72] showed that the TTC algorithm is strategyproof (each agent has no incentive to lie about his preferences) and later Bird [12] showed that it is in fact group-strategyproof (there is no incentive for any group of agents to collaborate and lie about their preferences).

For the housing allocation model, one natural and very common mechanism in practice is *random priority* (RP), also called *random serial dictatorship*, first studied in the literature by Zhou [89]. In this mechanism, a random ordering is drawn and each agent is allowed to choose the object they want assigned to them in this order. The outcome of this mechanism is Pareto efficient, strategy-proof and exhibits equal treatment of agents with identical preferences.

These two models were unified by Abdulkadiroglu and Sonmez [2], who described an algorithm called *random endowment* (RE) that utilized the TTC algorithm but generated endowments by choosing uniformly at random from all the possible ones. They were able to show that RE and RP output the same random matchings with the same probability, i.e. that RE and RP were *equivalent* mechanisms.

Since then, there have been several papers that showed the equivalence of mechanisms that are generalizations of RE and RP ([66],[28],[16]). We develop a proof technique in Chapter 2 that unifies and generalizes these results.

For a more in-depth overview of literature on allocating indivisible objects, we refer the readers to a recent survey by Sonmez and Unver [80]. Interested readers can also refer to Pathak [65] for applications of these ideas to student placement.

1.2.2 The newsvendor problem

The origins of the newsvendor problem can be dated to Edgeworth [26] who considered the problem of how a bank should set its level of cash reserves to cover withdrawals by its customers. The classic paper by Arrow et al. [4] later considered a more general model, which we describe here.

There is an inventory manager who has to decide how much to stock of a good he sells. There is a purchase cost, c, for each unit of inventory he decides to stock and a price, p, the customers pay for each unit of inventory they demand, and the manager wishes to maximize revenue. This can be equivalently modeled as the scenario where there is no cost for deciding how much to stock, but instead there is an *overage* cost, h, for each unit of inventory stocked in excess of demand and an *underage* cost, b, for each unit of demand that cannot be satisfied due to insufficient inventory, and where the manager's objective is to minimize cost. (It is not difficult to see that we can relate the two by writing $h = c$ and $b = p - c$.) We consider the second alternative.

Let the manager stocking decision be y and the customers' demand be D. Then the manager's incurred cost is

$$\lambda = h(y - D)^+ + b(D - y)^+.$$

Arrow et al. [4] and Dworetzky et al. [25] show that if the demand is stochastic with distribution function F, the stocking decision y that minimizes expected cost should satisfy

$$F(y) = \frac{b}{b + h}.$$

Instead of minimizing expected cost, the manager could try to minimize worst-case performance, typically called a *minimax* approach. Possible rationales for choosing such an objective include risk-aversion or lack of knowledge about the distribution of the demand. Scarf [74], and later Moon and Gallego [60] showed that if the demand has mean μ and standard deviation $\sigma > 0$, the stocking quantity y that minimizes the manager's cost against the worst-case demand distribution is

$$y = \mu + \frac{\sigma}{2}\left(\sqrt{b/h} - \sqrt{h/b}\right).$$

(This assumes that $h < b$, but the case where $h > b$ has a similar form.)

Time can be incorporated into the model by considering demand arriving over a time horizon with replenishments of inventory allowed. A common way to model this is to segment the time horizon into contiguous time periods and consider each period a separate newsvendor instance. Several variations are possible within this framework: unsatisfied demand in each period can either be lost or backlogged; goods can either be perishable (they are can be sold only in the period they are received by the manager) or non-perishable; the time taken between ordering and receiving goods, the *lead time*, can either be negligible (i.e. instantaneous delivery) or be non-negligible (i.e. more than one period or possibly stochastic); and there may be a fixed cost attached to each replenishment order made.

The case where goods are non-perishable and unsatisfied demand is backlogged is well-studied. Clark and Scarf [21] use dynamic programming to show that basestock policies, where in each period there is a target stocking level that the inventory is replenished to, are optimal. (Since then, basestock policies have been widely studied for many other inventory models, see e.g. [44], [85], [27].) There have also been recent studies that take a minimax approach to the multi-period problem ([31], [9], [10]).

In the case where goods are perishable, unsatisfied demand is lost, and there is no lead time or fixed costs, the newsvendor instances in each period decouple and may be optimized separately. We focus on this set up in Chapter 3, but with the following additions: there are two experts who provide advice to the inventory manager in each period; the manager's objective is to minimize his regret of not taking the better expert's advice throughout the periods, against the worst-case realization of expert advice and demand. Note that these additions prevent the newsvendor instances in each period from being decoupled.

For more discussion on additional models for inventory control, we refer the reader to Porteus [68] and Zipkin [90].

1.2.3 The single-resource capacity allocation problem

The capacity allocation problem is one selling identical capacity of a resource over a finite time horizon to various classes of demand who pay different prices (which we term fare classes). (The canonical example of this problem is the sale of tickets for a single-leg flight.) Demand for each fare class is stochastic and arrives sequentially, demand arrives individually and not in batches, and no overbooking is allowed (i.e. we cannot sell more than the capacity we have). The time horizon can thus be segmented into stages, with only fare class i demand arriving in stage i, and fare classes arriving in decreasing order (i.e. fare class $i+1$ arrives before fare class i). The manager seeks to maximize expected revenue by deciding at each stage how much of the remaining capacity to reserve (or protect) for sales in future stages. We call the vector of his decisions, $\mathbf{p} = (p_1, p_2, \ldots)$, a *protection level policy*, with the protection level p_i referring the how to much capacity to reserve for fare classes 1 through i.

As an example, we place this model within the context of a single-leg flight with two fare classes: leisure travelers and business travelers. Leisure travelers are willing to buy tickets earlier but are willing to pay less, while business travelers are apt to purchase tickets closer to the flight but are willing to pay more. The manager decides some protection level policy (p_1), which indicates how much remaining capacity to reserve for the high-fare class demand.

The earliest study of this problem dates to Littlewood [55], who examined the problem of managing sales of capacity on a single-leg flight with two independent fare classes, low- and high-fare and arriving in that order, and introduced a rule, commonly known as Littlewood's rule, for deciding at what capacity level to halt low-fare sales and only allow high-fare ones. Derivations of this rule were later provided by Richter [69]. Note that modeling low-fare before high-fare corresponds to a worst-case sequence of arrivals of fares.

The case with more than two fare classes (with fare classes arriving in ascending order of fare price) was first considered by Belobaba [7], who extended Littlewood's rule to the EMSR method, a suboptimal but easily implementable policy that performs reasonably well under certain distributions (see e.g. [87]) but not others (see e.g. [70]).

Characterizations of optimal policies and methods to compute them when the fare classes are monotonically increasing in fares were later described in Curry [23], Wollmer [87], and Brumelle and McGill [15]. In particular, they showed that protection level policies are optimal when they satisfy the following characterization: if we let D_i and f_i be fare class i demand and fare respectively and the number of classes be $M+1$, optimal protection policies $(p_1^*, p_2^*, \ldots, p_M)$ are specified by

$$P\left(D_1 > p_1^*, \ldots, \sum_{i=1}^{k} D_i > p_k^*\right) = \frac{f_{k+1}}{f_1} \qquad \text{for all } k = 1, \ldots, M.$$

This characterization of the optimal policy can be used to design a Monte Carlo algorithm for generating policies from demand samples (historical or otherwise). The same characterization was later exploited by Van Ryzin and McGill [83] to develop a stochastic gradient ascent algorithm, where the ascent direction for p_k in each iteration depended on whether the event $(D_1 > p_1, \ldots, \sum_{i=1}^{k} D_i > p_k)$ was realized in that iteration's demand samples. Another stochastic gradient ascent algorithm was later developed by Kunnamkal and Topaloglu [50], which utilized not the characterization above but arose instead from a dynamic programming formulation of the problem.

The case when the fare classes are non-monotonic in their fares (so low-fare demand need not necessarily arrive before high-fare demand) was studied by Robinson [70], who showed protection level policies were still optimal in that setting, characterized the optimal policy, and provided a Monte Carlo algorithm that computed an optimal policy from demand samples.

It is this last case of non-monotonic fare classes we consider in Chapter 4, where we characterize the relationship between the number of samples available and the performance achievable by a solution generated by the Monte Carlo algorithm described in Robinson [70]. We note that our result is distinct from those of Van Ryzin and McGill [83] and Kunnamkal and Topaloglu [50] because they obtain convergence rates (how fast solutions converge to an optimal one) for their algorithms while we obtain sampling bounds (how many demand samples are needed to achieve an approximation of a given performance level with a given confidence probability).

There is also a wide range of models in the revenue management literature related to the capacity allocation problem. Interested readers may refer to Talluri and van Ryzin [82] for a discussion of these models.

1.3 Guide to reading the dissertation

The chapters are each self-contained, so they may be read independently. Furthermore, each chapter has its own discussion or conclusion section.

Chapter 2

Allocating indivisible objects

In this chapter, we focus on the problem of allocating indivisible objects to agents, and examine mechanisms that solve this problem. In particular, we establish an approach to proving equivalences between mechanisms that allows us to give concise alternate proofs for results from recent literature and generalize them to more general mechanisms.

2.1 Introduction

The problem of allocating a finite number of indivisible objects to a set of agents has been studied extensively, since the pioneering work of Shapley and Scarf [78], and has served as a useful model in many real-world settings such as the assignment of schools to students [1] and the design of kidney exchanges [71]. The early literature was mostly on two distinct models: the *housing market* model of Shapley and Scarf in which each agent was endowed with an object; and the *house allocation* model, studied by Hylland and Zeckhauser [43] in which there were no endowments. There is by now a well-developed literature on each of these models as well as on a hybrid version, first proposed by Abdulkadiroğlu and Sönmez [3], in which some—but not all—agents are endowed with objects.

In this chapter our focus is exclusively on ordinal settings in which agents submit strict preference orderings over the objects. Shapley and Scarf [78] formulated their housing market model in this setting, and described the top-trading cycle (TTC) algorithm (attributed to Gale) that computes a unique reallocation of the objects to the agents. Roth [72] showed that the resulting mechanism is strategyproof (submitting true preference orderings is a dominant strategy for each agent) and Bird [12] showed that it is also group strategyproof (submitting true preference orderings is a dominant strategy for any group of agents); Roth and Postlewaite [73] showed that the core of the natural cooperative game associated with this problem is unique. Shapley and Scarf [78] had shown the solution computed by their algorithm is in the core of the associated cooperative game. For the house allocation problem, a fundamental mechanism in this setting is the *Random Priority* (RP) mechanism: a random ordering of the agents is drawn, and the agents are invited to choose objects in this order. The resulting allocation (for any given ordering) is Pareto efficient, so the outcome of the mechanism can be thought of as a lottery over Pareto efficient assignments. It is easy to see that RP is strategyproof and treats equals equally: agents with identical preference orderings receive identical (probabilistic) allocations. The literature on these problems evolved fairly independently until Abdulkadiroğlu and Sönmez [2] thought of the following mechanism for the house allocation problem: endow each agent with a random object, each possible endowment of the objects to the agents equally likely; and find the unique reallocation given by the TTC algorithm. They called this the *random endowment* (RE) mechanism. A natural question is to understand the relationship between RP and RE. Somewhat surprisingly, Abdulkadiroğlu and Sönmez [2] showed that these mechanisms are equivalent: given any preference structure for the agents, RP and RE lead to the same probability distribution over Pareto efficient assignments. Since their result, there have been a number of papers that establish the equivalence of seemingly different mechanisms for a variety of models assignment models: For example, Pathak and Sethuraman [66] show that in assigning students to schools (using the TTC mechanism), the mechanism in which all the schools use the same (random) priority ordering of the agents is equivalent to the mechanism in which every school generates its priority ordering randomly and independently.

Our work in this chapter is inspired by a recent paper of Carroll [16] that puts forth a general framework that subsumes many (but not all) of the known equivalence results in the literature. Specifically, Carroll introduces the notion of a priority framework in which the priority orderings of the objects are specified in terms of "roles," which are placeholders for the agents. By picking a bijection from the set of roles to the set of agents, we instantiate the priority framework, resulting in a priority ordering of the agents for each object. Carroll showed that if the bijection from roles to agents is chosen uniformly at random, and if the TTC algorithm is used to compute the allocation, then the outcome—a probability distribution over matchings—is independent of the priority framework! The equivalence of RP and RE follows because each of them can be modeled by a fixed priority framework. Carroll also introduced a more general priority structure for the objects by using an exogenous partition of the agents into groups. There is a fixed priority structure for the objects in terms of the given groups, but each group is free to order its roles any way it likes across the various objects. As before, the roles of each group are instantiated by a random bijection to the agents in that group, and this is done independently for each group. Carroll showed that the final outcome is the same regardless of how each group orders its roles across the various objects. In proving his result, Carroll observed that the straightforward counting approach of Pathak and Sethuraman [66] does not easily extend to this model. Nevertheless Carroll used a combination of techniques and gave essentially a bijective technique to prove his main result. In another recent paper, Ekici [28] proved a new equivalence result in the hybrid model where some (but not all) agents are endowed with objects. For this model, Ekici showed the equivalence of a natural random priority mechanism, first proposed by Abdulkadiroğlu and Sönmez [2], to a variant of the random endowment mechanism in which some agents may be endowed with multiple objects and others none; agents of the latter sort are nonetheless granted an "inheritance" right. We defer the details of this mechanism to §2.4.1, but note that this result is neither subsumed by Carroll's general model, nor by a similar looking equivalence result of Pathak and Sethuraman. Ekici's proof is again bijective.

Motivated by these recent equivalence results we describe a general technique to prove the equivalence of two mechanisms within a certain class (§2.3). Our approach is related to that of Pathak and Sethuraman [66]: like that approach we rely on induction (on the number of agents), and express the outcome of a mechanism in terms of its outcome on smaller instances. Given two mechanisms, we focus on those terms that appear in one but not the other, and argue that the overall contribution of such terms is zero. The key difference is in the way this last fact is established. Pathak and Sethuraman used a counting argument to do this, whereas we replace this counting argument with a simple bijection. The advantage is that one can now apply this argument more broadly. In particular, the results of Carroll [16], Ekici [28], and generalizations of these results can all be established easily using our technique, see §2.4. In §2.5 we show how the same technique can be used to prove equivalence results when the priority structures form an inheritance tree in the sense of Papai [64]. Our main contribution therefore is a unified approach that sheds light on all the equivalence results in the literature, in addition to suggesting new ones.

2.2 TTC Mechanisms

Let A denote the set of agents and S the set of objects, with $|A| = |S|$. Each agent has a strict preference ordering over the objects. Each agent wishes to be assigned exactly one object, and each object can be assigned to at most one agent. Our focus shall be on the family of algorithms called *top-trading cycles* (TTC) invented by Gale and first described by Shapley and Scarf [78]. This algorithm operates in phases; At the beginning of each phase, there is a set of (remaining) agents, a set of (remaining) objects, and each object has a top-priority agent (among the ones that remain). Consider the graph with a node for each (remaining) agent, and an arc from node i to node k if agent i's most preferred object is one for which agent k has top priority. Note that self-loops are possible: an arc (i, i) exists if agent i has the highest priority for his most-preferred (remaining) object. This graph, referred to as the TTC graph, must have a cycle c (which may be simply a self-loop), because every node has out-degree 1 and there are finitely many nodes. Every agent in c is matched with the object he most prefers among the ones that remain; the agents in c along with their matched objects are removed from the problem, and the top-priority agent for each remaining object is updated if necessary. We will usually refer to this process as *clearing the cycle* c. If any agents (equivalently objects) remain, the next phase starts in which the same algorithm is applied to the remaining objects and agents; Otherwise, all the agents have been matched and the algorithm terminates.

Note that we talk of the TTC *family* of algorithms rather than the TTC algorithm because the final allocation depends on how the top priority agent for each object is specified. Thus the mechanisms we discuss differ in how the agents are prioritized for each object. The following two examples, which are well-known mechanisms in the literature, give an idea of how mechanisms can differ within the framework just described. In each case, the agents are assigned objects using the TTC algorithm, but applied to different priority profiles for the objects.

Example 2.2.1. Random Priority (RP). *Let* $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ *be a permutation of the agents chosen uniformly at random, that is, every permutation of the agents is equally likely. Let* i^* *be the smallest* i *for which agent* σ_i *still remains in the problem. Then, the top-priority agent (in any phase) for* every *remaining object is* σ_{i^*}. *Equivalently, the top-priority agent for every remaining object in phase* k *is the agent* σ_k.

Example 2.2.2. Random Endowment (RE). *Let* $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ *be a permutation of the agents chosen uniformly at random, that is, every permutation of the agents is equally likely. For each* $j = 1, 2, \ldots, n$, *the top-priority agent for the* j*-th object is* σ_j. *Note that if the* j*-th object still remains in the problem, so will agent* σ_j, *so every remaining object will have a top-priority agent at every point in time.*

For both RP and RE, the priority orderings of the objects are determined randomly, so the outcome is a *random* matching, which can be described as a *probability distribution over perfect matchings*, where a perfect matching refers to a matching between the full set of agents and the full set of objects. This naturally leads us to the notion of *equivalence* of mechanisms, defined as follows.

Definition 2.2.1. Two mechanisms are *equivalent* if for any instance, every perfect matching occurs with the same probability under both mechanisms.

Although RP and RE look different and are described in different terms, they lead to the *same probability distribution over perfect matchings*, as was shown independently by Abdulkadiroğlu and Sönmez [2] and Knuth [49]. Thus, RP and RE are equivalent mechanisms.

2.3 Equivalence of Mechanisms: A general approach

Recent results have shown equivalences between mechanisms that allow for more complex priority structures [16, 28, 66]. We review in §2.4 two of these results and show how their proofs can be simplified. We unify and generalize these recent results in §2.5, where we formulate a model in which the priority structure for each object is given by an *inheritance* tree. The proof technique in all our proofs is essentially the same, and is discussed here.

Recall that different mechanisms in the TTC family differ only in how the priority structures for the various objects are determined. All the TTC mechanisms we consider satisfy the following *persistence* property:

(i) Once an agent has top priority for an object, he retains it until he is matched (to that object or a different one);

and the following *inheritance* property:

(ii) The top-priority agent for a remaining object at any time may depend only on the set of remaining agents and objects as well as the partial matching guaranteed by the all past cycles formed in the TTC graph.

Note that on a given priority profile for the objects it is possible for the corresponding TTC graph for $\mathbb{M}$ to have multiple cycles; by property (i), however, any cycle that is not cleared will *persist* in the TTC graph until it is cleared. This enables us to write the outcome of mechanism $\mathbb{M}$ as a simple recursion. To do this, we first set up some notation.

Let $\mathbb{M}(A, \pi)$ be the probability that mechanism $\mathbb{M}$ applied to the set of agents A results in the matching π.[1] For any $l \geq 1$, let $\mathcal{C}^l_{\mathbb{M}}$ be the collection of l disjoint top-

[1] Strictly speaking, we should also specify the set of objects that are "available." However the set of available objects will be clear from the context, justifying the notation.

trading cycles (also called a *cycle product* of size l) that could be simultaneously present initially when mechanism $\mathbb{M}$ is used. Each cycle can be written in a canonical way: the smallest-numbered agent in the cycle appears as the first agent in the cycle; for a cycle-product of size l, we arrange the cycles in ascending order of the first agents. Note that this representation is unique. For any cycle product c, let $\nu(c)$ be the induced matching, and $P_{\mathbb{M}}(c)$ be the probability that cycle product c is present in the TTC graph under mechanism $\mathbb{M}$. Finally, for matchings π and π', $\pi' \subseteq \pi$ if π' is a submatching of π (i.e., every matched pair in π' is also a matched pair in π); in that case, $\pi \setminus \pi'$ denotes the matching π restricted to the agents and objects that do not appear in π'. Then:

$$\mathbb{M}(A,\pi) = \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}}(c)\, \mathbb{M}(A\setminus c, \pi\setminus\nu(c)). \tag{2.1}$$

Eq. (2.1) can be justified in a straightforward way using the inclusion-exclusion principle and property (i).[2] Note that the information about the cycles that have been cleared is implicit in the definition of the residual problem.

As priority structures grow increasingly complex so does the behavior of the mechanism, so comparing different TTC mechanisms is not always easy. However, Eq. (2.1) suggests a simple way to think about the equivalence of two TTC mechanisms. Suppose there is another TTC mechanism $\mathbb{M}'$ satisfying properties (i) and (ii) such that

(iii) every collection of cycles that could be simultaneously present in the TTC graph of $\mathbb{M}$ can also be simultaneously present in the TTC graph of $\mathbb{M}'$. That is, $\mathcal{C}^l_{\mathbb{M}} \subseteq \mathcal{C}^l_{\mathbb{M}'}$ for any l; and

(iv) for any $c \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}}$, $P_{\mathbb{M}}(c) = P_{\mathbb{M}'}(c)$.

[2]We appeal to the inclusion-exclusion principle to avoid double-counting. Properties (i) and (ii) together imply that the outcome of the mechanism does not depend on which subset of cycles of the TTC graph is cleared during a phase of the TTC algorithm.

Then,

$$
\begin{aligned}
&\mathbb{M}'(A,\pi) \\
&\quad = \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}'}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)) \\
&\quad = \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\cap\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}'}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)) + \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\setminus\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}'}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)) \\
&\quad = \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)) + \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\setminus\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}'}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)), \quad (2.2)
\end{aligned}
$$

where the first term of the last equality follows from properties (iii) and (iv) of mechanism $\mathbb{M}'$.

Suppose we are able to show that

$$\sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\setminus\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}'}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)) \ = \ 0. \qquad (2.3)$$

Then expression (2.2) simplifies to

$$\mathbb{M}'(A,\pi) = \sum_{l\geq 1} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} (-1)^{l-1} P_{\mathbb{M}}(c)\ \mathbb{M}'(A\setminus c, \pi\setminus\nu(c)).$$

Notice that this is exactly the recursion defining $\mathbb{M}(A,\pi)$, assuming $\mathbb{M}(A',\pi') = \mathbb{M}'(A',\pi')$ for all $A' \subsetneq A$ and all matchings π' involving the agents in A'. This sets the stage for an inductive proof.[3]

We shall therefore focus on proving (2.3), which we do using the following general idea. For any $c \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$ pick two special agents x_1 and x_2 from c and see

[3] Any assumption on the priority structure for the problem involving the agents in A must be preserved for the reduced problem A'. This is generally easily verified.

how they appear in its canonical representation[4]: if x_1 and x_2 appear in the same cycle, say $(x_1a_0a_1 \ldots a_kx_2b_0b_1 \ldots b_l)$, then split the cycle into two cycles $(x_1a_0a_1 \ldots a_k)$ and $(x_2b_0b_1 \ldots b_l)$; if x_1 and x_2 appear in distinct cycles, say $(x_1a_0a_1 \ldots a_k)$ and $(x_2b_0b_1 \ldots b_l)$, then merge the two cycles as $(x_1a_0a_1 \ldots a_kx_2b_0b_1 \ldots b_l)$. We call this the *transformation* T applied to the cycle product c. Suppose that for any cycle product $c \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$, there is a method of picking the two agents x_1 and x_2 such that the following properties hold:

(a) the choice of x_1 and x_2 is depends only on the set of agents in the cycle product, and not on its cycle structure;

(b) $T(c) \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$, and $P_{\mathbb{M}'}(T(c)) = P_{\mathbb{M}}(c)$.

We record a few properties of this transformation T: first, $T(T(c)) = c$, so that T is self-inverse; second, $T(c)$ has one less or one more cycle than does c; and finally, the matchings induced by c and $T(c)$ are identical, therefore the residual problems are identical as well. These observations collectively imply that T is a bijection between $\{c \in \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}} : \nu(c) \subseteq \pi, l \text{ odd}\}$ and $\{c \in \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}} : \nu(c) \subseteq \pi, l \text{ even}\}$. Furthermore, $P_{\mathbb{M}'}(T(c)) = P_{\mathbb{M}'}(c)$ for any $c \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$. Rewriting Eq. (2.3) as

$$\sum_{l \text{ odd}} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\setminus\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} P_{\mathbb{M}'}(c)\, \mathbb{M}'(A \setminus c, \pi \setminus \nu(c)) \;=\; \sum_{l \text{ even}} \sum_{\substack{c\in\mathcal{C}^l_{\mathbb{M}'}\setminus\mathcal{C}^l_{\mathbb{M}} \\ \nu(c)\subseteq\pi}} P_{\mathbb{M}'}(c)\, \mathbb{M}'(A \setminus c, \pi \setminus \nu(c)),$$

we note that its validity is immediate from the preceding discussion: for each term involving a cycle product c on the left, there is a corresponding term involving $T(c)$ on the right such that the expressions are identical, and vice-versa.

[4]In all the applications of this idea, every cycle $c \in \bigcup_{l\geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$ has at least two candidates for the roles of x_1 and x_2, so this can always be done.

Example 2.3.1. *Consider an instance with* 4 *agents, with* $A = \{1, 2, 3, 4\}$ *and* $S = \{a, b, c, d\}$*, where the agents have strict preferences given by Table 2.1.*

1	2	3	4
a	b	c	a
b	a	d	c
c	c	a	b
d	d	b	d

Table 2.1: Agent preferences for Example 2.3.1, from most- to least-preferred.

Let RP be $\mathbb{M}$ *and RE be* $\mathbb{M}'$*. Consider the sets of cycle products under RP. These are simply*

$$\begin{aligned}\mathcal{C}^1_{\mathbb{M}} &= \{(1), (2), (3), (4)\}, \\ \mathcal{C}^2_{\mathbb{M}} &= \mathcal{C}^3_{\mathbb{M}} = \mathcal{C}^4_{\mathbb{M}} = \emptyset\end{aligned}$$

Similarly, the sets of cycle products under RE are

$$\begin{aligned}\mathcal{C}^1_{\mathbb{M}'} &= \mathcal{C}_{\mathbb{M}} \cup \{(12), (13), (23), (24), (34), (123), (132), (234), (243)\}; \\ \mathcal{C}^2_{\mathbb{M}'} &= \{(1)(2), (1)(3), (2)(3), (2)(4), (3)(4), (1)(23), (13)(2), (2)(34), \\ &\quad (12)(3), (24)(3), (23)(4)\}; \\ \mathcal{C}^3_{\mathbb{M}'} &= \{(1)(2)(3), (2)(3)(4)\}; \\ \mathcal{C}^4_{\mathbb{M}'} &= \emptyset\end{aligned}$$

Note that it is impossible for agents 1 *and* 4 *to be part of the same cycle product since they both have* a *as their most preferred object. Any cycle product in RE that is not in RP must have at least two agents; let* x_1 *and* x_2 *be, respectively, the lowest-numbered agent and the second lowest-numbered agent in such a cycle. It is easy to verify* *that the associated transformation* T *(with this choice of* x_1 *and* x_2*) satisfies the claimed properties. For the cycle* $c = (23) \in \mathcal{C}_{\mathbb{M}'} \setminus \mathcal{C}_{\mathbb{M}}$*, note that* $T(c) = (2)(3) \in \mathcal{C}_{\mathbb{M}'} \setminus \mathcal{C}_{\mathbb{M}}$*, and that* $P_{\mathbb{M}'}(c) = 1/12 = P_{\mathbb{M}'}(T(c))$*. If* $c' = (243)$*,* $T(c') = (24)(3)$*; and* $P_{\mathbb{M}'}(c') = 1/24$ *as is* $P_{\mathbb{M}'}(T(c'))$*. Finally, if* $\tilde{c} = (2)(3)(4)$*,* $T(\tilde{c}) = (23)(4)$*; again,* $P_{\mathbb{M}'}(\tilde{c}) = 1/24$ *as is* $P_{\mathbb{M}'}(T(\tilde{c}))$*. We leave it to the reader to verify the bijection for each of the other cycle products.*

We can summarize the discussion so far as follows: To prove that mechanisms $\mathbb{M}$ and $\mathbb{M}'$ satisfying properties (i)-(iv) are equivalent, it is *sufficient* to verify the equivalence of $\mathbb{M}$ and $\mathbb{M}'$ when there is a *single* agent and to prove Eq. (2.3). The latter can be done by showing how to choose x_1 and x_2 such that the associated transformation T satisfies properties (a) and (b). In the sections that follow, we shall carry out these (two) steps for various pairs of mechanisms.

2.4 Linear Priority Structures

2.4.1 RP and RE in a model with endowments

We consider an assignment model in which some of the agents are endowed with objects. The most prominent application is in the setting of house allocation where some of the houses already have existing tenants who are willing to move but only to other houses that are better for them. The RP and RE mechanisms can be generalized in many ways to this setting, and several recent papers show the equivalence between the generalized RP and RE mechanisms, see Sömnez and Ünver [79], Pathak and Sethuraman [66] and Ekici [28]. To illustrate our general approach, we provide a simple proof of Ekici's result, which is also the most recent.

Suppose the agents A are partitioned into N and E, and the objects are partitioned into V and O, such that $|E| = |O|$ and $|N| = |V|$. The interpretation is that every agent in E is an "existing tenant," and every object in O is an "occupied house." The agents in N are "new tenants", and the "houses" in V are "vacant". Each agent in E is endowed with an object in O, for which he has the highest priority. Ekici [28] examines the generalization of RP introduced by Abdulkadiroğlu and Sönmez [3] and proposes a natural generalization of the RE mechanism to this setting (we abuse notation and continue to call these RP and RE), which we describe next. (Recall that the matching is found by applying the TTC mechanism to a priority structure for the objects; the RP and RE mechanisms differ only in how the priority structure—to which the TTC algorithm is applied—is generated.)

RP. Let $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ be a permutation of the agents chosen uniformly at random, that is, every permutation of the agents is equally likely. Let i^* be the smallest $i \in \{1, 2, \ldots, n\}$ for which agent σ_i is still in the problem. The priority structure for the remaining objects is determined as follows: each object in V has σ_{i^*} as the top-priority agent; if the agent endowed with $o \in O$ still remains in the problem, he retains top priority for o, otherwise σ_{i^*} has top priority for o.

RE. Let $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ be a permutation of the agents chosen uniformly at random, that is, every permutation of the agents is equally likely. Suppose the objects are ordered in an arbitrary (but fixed) manner so that there is a first object, a second object, and so on. For each $j = 1, 2, \ldots, n$, if the j-th object is in V, then its initial top-priority agent is σ_j; otherwise, the j-th object is in O and is endowed to some agent $i \in E$. In the latter case ($j \in O$), agent σ_j becomes the "inheritor" of that agent i: when agent i departs, his priority for unassigned objects is passed on to σ_j, i.e. if i had top priority for an unassigned object just before he left, σ_j will now have top priority for it. Note that the inheritor of an agent $i \in E$ may be another agent $i' \in E$ who has already been assigned an object, triggering a chain of inheritances until an inheritor yet to be assigned an object is found. It is a simple matter to verify that each remaining object will have a top priority agent at any stage.

We now give an alternative proof that RP and RE are equivalent in this setting. First observe that both mechanisms satisfy properties (i)-(iv), with RP playing the role of $\mathbb{M}$ and RE the role of $\mathbb{M}'$ in the definitions of those properties. Additionally, when there is a single agent it is clear that these mechanisms are equivalent. Finally, under both RP and RE, after clearing a common cycle product c but with no further assumptions on realizations of randomness, the residual problem is simply a smaller instance of the original problem where we may treat an object $s \in O$ as an object from V if the agent in E endowed with s has already been matched. Consequently, we may use the induction argument from §2.3.

To complete the proof of equivalence, we need to specify the transformation T, which is the same as describing how to pick x_1 and x_2 given any cycle that is present in the TTC graph of RE, but not in that of RP. We call an agent *V-preferential* if his most preferred object is from V. Notice that each cycle product in the TTC graph for RP involves *at most* one V-preferential agent; the TTC graph for RE contains all of these cycle products but could contain other cycle products involving two or more V-preferential agents. Consequently, any cycle product $c \in \bigcup_{l \geq 1} \mathcal{C}^l_{\mathbb{M}'} \setminus \mathcal{C}^l_{\mathbb{M}}$ contains at least two V-preferential agents. Given any cycle product c that could be present in the TTC graph of RE, but not in that of RP, let x_1 be the smallest-indexed V-preferential agent in c and let x_2 be the second-smallest. This choice of x_1 and x_2 satisfies property (a). It is simple to verify that the cycle products c and $T(c)$ occur with the same probability in

RE, satisfying property (b).

2.4.2 Assignment models with agent groups

We turn now to a model, introduced by Carroll [16], in which we are given an exogenous partition of the agents into groups. Furthermore, for each object, we are given a (fixed) priority ordering of the groups. This is useful in modeling situations in which agents within a group should be treated equally, but agents across groups can be prioritized. Carroll showed that this model, described in more detail next, is rich enough to subsume most of the equivalence results in the literature, the only notable exception being the result of Ekici discussed in the previous section. In this model, Carroll proposed two mechanisms—the Random Serial Dictatorship in Groups mechanism (RSDIG) and the Within-Groups Top Trading Cycles mechanism (WGTTC)—and showed that they are equivalent.

As before, we are given a set A of agents, a set S of objects, with $|A| = |S|$. Agents have strict preferences over the objects. Suppose the agents are partitioned into groups $G_1, G_2, \ldots G_m$, for some $m \geq 1$. Each group G_i has an associated role group with the same number of members, $R_i = \{r^i_1, r^i_2, \ldots, r^i_{|G_i|}\}$. Each object's priority structure is given by an ordering of the roles, such that all the roles belonging to a given group occur consecutively (a property Carroll [16] calls *group-respect*); however the roles of a given group can be ordered differently in the priority orderings of different objects. One interpretation of these two features is as follows: each object is first endowed to some group and passes to a different group only when all the members of the original group have all been matched; however, a given group may choose to distribute the object it owns among its roles any way it pleases. (This latter feature allows constraints such as objects a and b cannot have the same top-priority agent.) The priority structure for the objects is specified in terms of roles, so following Carroll [16] we call this a *priority framework*. The two mechanisms studied by Carroll differ only in the priority frameworks to which the TTC algorithm is applied:

RSDIG. The roles R_i of each group G_i always appear in the order $r^i_1, r^i_2, \ldots, r^i_{|G_i|}$.

WGTTC. For each group G_i, there is no restriction on the ordering of roles, and the ordering may differ from object to object.

In each of these cases the roles have to be instantiated before applying the TTC algorithm. This is done by choosing, for each group G_i, a bijection uniformly at random from its set of roles R_i to the set of agents G_i. In such a case, Carroll showed that the RSDIG and WGTTC mechanisms are equivalent. We give an alternative proof of this result.

Observe that both mechanisms satisfy properties (i)-(iv), with RSDIG as $\mathbb{M}$ and WGTTC as $\mathbb{M}'$. Furthermore, it is clear the two mechanisms are equivalent when there is a single agent. Finally, in both RSDIG and WGTTC, after clearing a common cycle product c, the residual problem is a smaller instance of the original problem restricted to the remaining agents and objects. So we may use the induction argument of §2.3. It remains to show there is an appropriate way to pick x_1 and x_2 for transformation T so that $P_{\mathbb{M}'}(c) = P_{\mathbb{M}'}(T(c))$.

Observe that each cycle product in the TTC graph for RSDIG involves *at most* one agent from each group. The TTC graph for WGTTC contains all of these cycle products, but also includes those containing *two or more* agents from *some* group. Let c be a cycle product containing two or more agents from some group. Consider the smallest labeled group with two or more agents in c. Let x_1 be the smallest and x_2 the second-smallest labeled agents from that group in c, and consider the canonical transformation T described earlier. The resulting cycle product $T(c)$ corresponds to the cycle product induced by the same role-to-agent map as c except that the roles of x_1 and x_2 are swapped. Since x_1 and x_2 are in the same group, it follows that the role-to-agent maps inducing c and $T(c)$ both occur with the same probability. This verifies properties (a)-(b) so the equivalence result follows.

2.5 Tree-based inheritance with groups and endowments

We are ready to formulate a general object assignment model. As before we have a set A of agents and a set S of objects. Each agent wishes to be assigned exactly one object. Suppose the agents are partitioned into A and E and the objects are partitioned into V and O. Each object $o \in O$ is endowed to an agent $\rho(o) \in E$, and each agent $i \in E$ is endowed with *at least* one object in O. (Different objects in O may be endowed to the same agent.) Thus $|O| \geq |E|$ and $\rho(O) = E$. As in our earlier discussion, one can think of the agents in E as the existing tenants and the objects in O as the houses they occupy in a house allocation model; in that case, the agents in N and the houses in V are, respectively, the new tenants and the new houses. Agents have strict preferences over the objects. The priority structure of the objects—a distinguishing feature of our model—is specified by an *inheritance tree*, which generalizes the models of Carroll [16], as well as the earlier models of Papai [64] and Svensson & Larsson [81]. In particular, the priority structure allows for objects to be endowed to certain agents, admits group hierarchies, and additionally allows the inheritance of an object to depend on the partial matching at that stage.

The agents are also partitioned into groups $G_1, G_2, \ldots, G_m$ so that $G_i \cap G_j = \emptyset$ for all $i \neq j$ and $\bigcup G_i = A$. Note that composition of each of these groups is unrestricted: G_i may have a non-empty intersection with A or E or both. However the group membership of an agent may affect his priority for certain objects. Each agent group G_i is associated with a set of *roles* R_i with $|R_i| = |G_i|$, with $R_i \cap R_j = \emptyset$ for all $j \neq i$. The priority structure for each object is specified in terms of roles. When the mechanism is run, the roles, R_i, are instantiated with the agents in G_i, uniformly at random: each mapping of the roles R_i to the agents in G_i is equally likely. We say that an agent a or role r *owns* an object o if a or r has the topmost priority for o.

2.5.1 Inheritance Trees

The priority structure of each object is specified in terms of an inheritance *tree* as introduced by Papai [64], with one key difference: the nodes of the tree are populated by *roles* rather than the agents.

Every object $o \in S$ has an associated *inheritance tree* Γ_o, which is a rooted tree graph with directed arcs. If there are n agents in all, Γ_o has one root node (level 0), $n-1$ nodes at level 1, $(n-1)(n-2)$ nodes at level 2, etc., each of which is labeled with a role. Each node has exactly one incoming arc except the root (which has none), and each node at level k has $(n-k-1)$ outgoing arcs, each of which terminates at a node at level $k+1$ and is labeled with an object other than o and the k objects appearing as labels in the unique path from the root to that node. Moreover, in any path from the root to a leaf, each role label should appear exactly once. The agent who owns the object is the one whose role is the label of the root.

The tree Γ_o defines an inheritance plan for the object o in the following sense: consider the path from the root node v_0 to a node v_k, and suppose the labels of the nodes and arcs in the path are $r_0 - o_0 - r_1 - o_1 - r_2 - \ldots - r_{k-1} - o_{k-1} - r_k$. Then the agent with role r_k owns o if for all $i = 0, \ldots, k-1$ the agent with role r_i is assigned object o_i, and the agent with role r_k is still unassigned. (The assignment of agents with roles other than r_i, $i = 0, \ldots, k$ to objects other than $o, o_0, o_1, \ldots, o_{k-1}$ is immaterial.) Note that the definition of an inheritance tree ensures that in any partial assignment of objects to agents, either object o is assigned, or there is a unique maximal path originating from the root of Γ_o using only agents and objects in that partial assignment whose terminal node is the role of an unmatched agent, who next owns o.

We are now ready to specify the priority structure for each object $o \in S$. For $o \in V$, the priority structure is given by the inheritance tree Γ_o, with the roles instantiated uniformly at random from the corresponding groups; each object $o \in O$ is owned by the agent $\rho(o)$ until that agent is assigned an object; if $\rho(o)$ is assigned an object, but o is still unassigned, its ownership will be governed by its inheritance tree Γ_o.

Suppose the TTC mechanism is applied to a problem where the priority structure for each object is given by an inheritance tree. We describe how the trees are updated when a cycle c is cleared. Typically this process involves trimming branches and contracting arcs so that information that is no longer useful is discarded; the resulting updated tree will involve only those objects and roles that are still "unassigned". Let X be the set of objects that are assigned and let Y be the roles that these objects are assigned to when the TTC mechanism is applied. For each $x \in X$, let $\lambda(x)$ be the role in Y that x is assigned to. The inheritance trees are updated as follows:

- Discard the inheritance trees Γ_o for each $o \in X$;

- For each $x \in X$, and for each node labeled $\lambda(x)$ in every remaining inheritance tree, delete every outgoing arc from that node with a label other than x (and the subtree rooted at the other end of that arc). So for each $x \in X$, the only arc that emanates from any node labeled $\lambda(x)$ will be labeled x.

- Contract every arc with a label of x emanating from a node labeled $\lambda(x)$ (this is the same as deleting the arc x, and moving the subtree rooted at its head node to its tail node).

Observe that the updates to the inheritance trees removes all the paths that can no longer be realized and retains the paths that *may* still be realized. We end this section with an illustrative example.

Example 2.5.1. *Consider an instance of the problem with* 4 *agents* $\{1, 2, 3, 4\}$ *and* 4 *houses* $\{a, b, c, d\}$*, and suppose role* r_i *is mapped to agent* i *for each* i*. Focus on object* a*, whose inheritance tree is given by* Γ_a *in Figure 2.1(a). We illustrate in Figures 2.1(b)–(d) how ownership and the inheritance tree of object* a *evolves under some sequence of (sub-)matchings, say* $(1 \leftarrow b)$*,* $(3 \leftarrow d)$*,* $(2 \leftarrow c)$*, and* $(4 \leftarrow a)$ *in the given order. Initially,* a *is owned by agent 1, who is assigned* b *in phase 1. Following this,* a*'s inheritance tree is updated to* Γ'_a*, and agent 2 becomes the new owner of* a*. When agent 3 is assigned* d*, 2 continues to own* a*, but* a*'s inheritance tree is updated to* Γ''_a*. Then, agent 2 is assigned* c *at which point agent 4 becomes the owner of* a*; this is shown in the tree* Γ'''_a*. Finally* 4 *is matched with* a *so* a *is removed from the problem.*

2.5.2 Equivalence results

We generalize the equivalence results discussed earlier to settings in which the priority structure for each object is an inheritance tree. To state the equivalence result formally, we need the following definitions.

Given an inheritance tree Γ, define $G(\Gamma)$ to be the tree obtained by replacing each role at the nodes of Γ with the group to which that role belongs. Inheritance trees Γ and Γ'

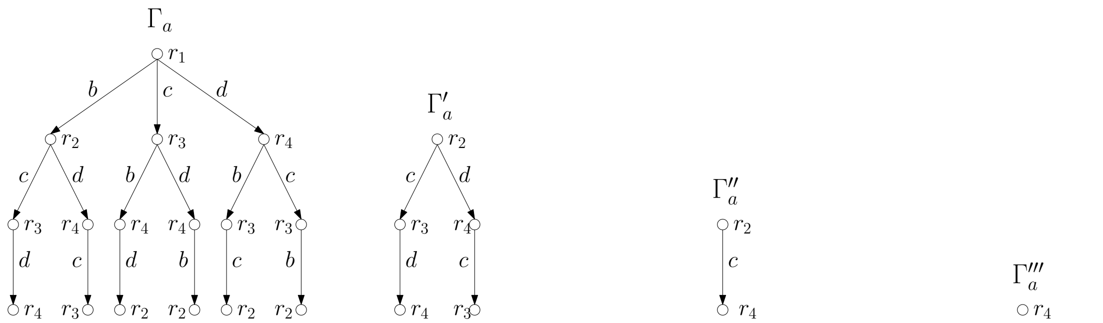
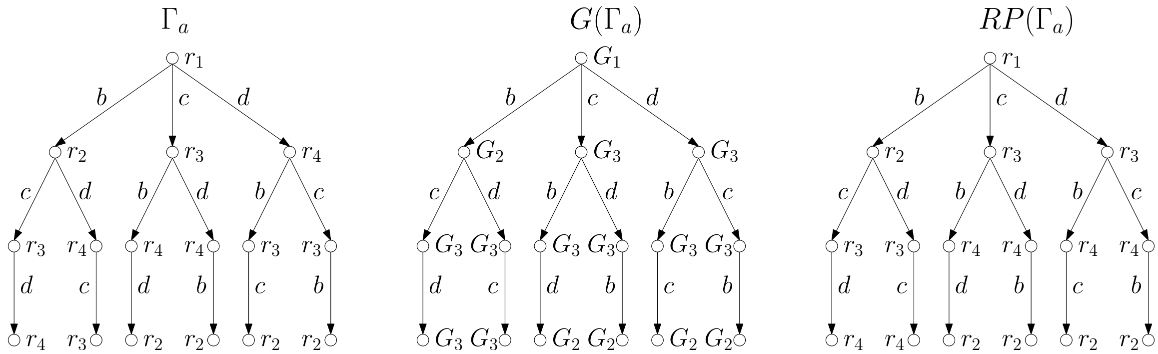


Figure 2.1: The evolution of Γ_a in Example 2.5.1.

are *G-similar* if $G(\Gamma)$ and $G(\Gamma')$ are identical.

An inheritance tree is a *random priority* inheritance tree if in every root-to-leaf path, the roles belonging to each group G_i appear in ascending order (from the lowest index to the highest index). Notice that when all the inheritance trees in the problem are random priority inheritance trees, each $o \in V$ owned by a member of a group G_i will be owned by the *same* role. Given an arbitrary inheritance tree Γ, there is a natural relabeling of the nodes that gives rise to the associated random priority inheritance tree $RP(\Gamma)$: find $G(\Gamma)$ and apply the random-priority labeling to it. It follows that if Γ and Γ' are G-similar, $RP(\Gamma) = RP(\Gamma')$.

For a partial assignment (or matching) π, and an object $s \in O$, let $g_\pi(s)$ be the group to which s is assigned; if s is not assigned to any agent in π, define $g_\pi(s)$ to be zero. Two partial assignments π and π' are *G-equivalent* if $g_\pi(s) = g_{\pi'}(s)$ for all $s \in O$. In other words, π and π' must assign the same objects, and each object assigned by them must be assigned to the same group under both π and π' (although the actual agents or roles to which s is assigned may differ). Given an inheritance tree Γ and a partial assignment π, let Γ^π be the updated inheritance tree after the objects and roles in π are removed from the problem. An inheritance tree Γ is *G-invariant* if $G(\Gamma^\pi) = G(\Gamma^{\pi'})$ for any G-equivalent partial assignments π and π'. Note that if Γ is G-invariant, then so is any inheritance tree Γ' that is G-similar to Γ.



(a) Inheritance tree (b) Replacing roles with groups (c) Induced random priority inheritance tree

Figure 2.2: Inheritance trees Γ_a, $G(\Gamma_a)$, and $RP(\Gamma_a)$ for Example 2.5.2.

Example 2.5.2. *Consider the inheritance tree Γ_a in Figure 2.2(a). Let the groups be $G_1 = \{1\}$, $G_2 = \{2\}$, $G_3 = \{3, 4\}$. After relabeling the roles with their associated groups, we get $G(\Gamma_a)$ as shown in Figure 2.2(b). The random-priority tree induced by Γ_a, $RP(\Gamma_a)$, is shown in Figure 2.2(c): in every path from the root to any leaf the roles of each group appear consecutively, and in ascending order. Moreover, both Γ_a and $RP(\Gamma_a)$ are G-invariant.*

Theorem 2.5.1. *Fix $A = (E, N)$, $S = (O, V)$ and ρ. Let $\{\Gamma_s\}$ be a set of inheritance trees on (A, S) that are G-invariant. Then the TTC mechanism using $\{\Gamma_s\}$ and the TTC mechanism using $\{RP(\Gamma_s)\}$ are equivalent.*

Proof. Let $\mathbb{M}$ be the TTC mechanism applied to $\{RP(\Gamma_s)\}$ and let $\mathbb{M}'$ be the TTC mechanism applied to $\{\Gamma_s\}$. With this interpretation, property (i) is immediate; property (ii) follows from the G-invariance of $\{\Gamma_s\}$ for $\mathbb{M}'$, and from the G-invariance of $\{RP(\Gamma_s)\}$ for

$\mathbb{M}$. (Note that G-invariance of $\{RP(\Gamma_s)\}$ is implied by G-invariance of $\{\Gamma_s\}$.) Properties (iii) and (iv) are also easily verified for the two mechanisms. Furthermore, it is clear that the two mechanisms are equivalent when there is a single agent.

Consider the TTC graph for $\mathbb{M}$ and $\mathbb{M}'$ in the residual problem under any sequence of cleared cycles, with no further assumptions on the realizations of randomness. Since the sequence of cycles cleared are the same, the inheritance trees are updated in both mechanisms similarly (the same arcs are deleted and contracted, though the role labels removed may differ). Moreover, the remaining agents and objects are the same in both mechanisms. It follows that the updated inheritance trees under $\mathbb{M}$ are the same as the random-priority trees induced by the updated inheritance trees under $\mathbb{M}'$ (a relabeling of roles may be necessary), so we can use the induction argument from §2.3.

It remains to show that there is an appropriate way to pick x_1 and x_2 for transformation T that satisfy properties (a)-(b). Consider the agents whose most preferred object is from V, whom we call V-preferential agents. Observe that each cycle product in the TTC graph for $\mathbb{M}$ has *at most* one V-preferential agent from G_k for each k. The TTC graph for $\mathbb{M}'$ contains these cycle products as well (in addition to zero or more cycle products containing at least two V-preferential agents from the same group). Consider the lowest labeled group with at least two V-preferential agents in c. Let x_1 and x_2 be, respectively, the lowest and second-lowest labeled V-preferential agents from this group. This satisfies property (a). Additionally, the resulting cycle product $T(c)$ is exactly the cycle product induced by the same role-to-agent map as c except that the roles mapping to x_1 and x_2 are swapped. Since x_1 and x_2 are in the same group, the role-to-agent maps inducing c and $T(c)$ both occur with the same probability, so property (b) is satisfied as well, and the equivalence result follows. □

Corollary 2.5.1. *Fix* $A = (E, N)$, $S = (O, V)$ *and* ρ. *Let* $\{\Gamma_s\}$ *and* $\{\Gamma'_s\}$ *be two sets of* *inheritance trees on* (A, S) *that are G-invariant, and let* Γ_s *be G-similar to* Γ'_s *for each* $s \in S$*. Then the TTC mechanism using* $\{\Gamma_s\}$ *and the TTC mechanism using* $\{\Gamma'_s\}$ *are equivalent.*

Proof. As $\{\Gamma_s\}$ and $\{\Gamma'_s\}$ are G-similar, $\{RP(\Gamma_s)\}$ and $\{RP(\Gamma'_s)\}$ are identical, and the result is immediate from Theorem 2.5.1 □

Remark. One can think of G-invariance as a natural generalization of the notion of a *group respecting* priority structure as defined by Carroll [16] (see §2.4.2). While this restriction is a technical necessity in the equivalence proof, it is not an unreasonable one: For example, in allocating public housing with agents grouped by income level, it may be undesirable for a house to pass from a low income group to a high income group and back to the low income group. This cannot occur if the inheritance structure is group respecting. Following Carroll, we could define an inheritance tree to be *group respecting* if in any path from the root to a leaf, all roles from the same group appear consecutively. It is easy to verify that every G-invariant tree is group respecting, but the converse is not true. Example 2.5.3 demonstrates the necessity of G-invariance for the equivalence result to hold. It also shows that equivalence may not hold when trees are group respecting but not G-invariant. If the inheritance tree is equivalent to a linear priority order as in Carroll [16], it can be shown that an inheritance tree is G-invariant if and only if it is group-respecting.

Example 2.5.3. *To demonstrate the necessity of G-invariance of all trees, consider the following* 4*-agent example. Let* $A = \{1, 2, 3, 4\}$*,* $S = \{a, b, c, d\}$*, with three groups* $G_1 = \{1, 2\}$ *(with corresponding roles* $\{r_1, r_2\}$*),* $G_2 = \{3\}$ *(with roles* $\{r_3\}$*) and* G_3 *(with roles* $\{r_4\}$*). Suppose the agents' preferences are given by Table 2.2, and that the inheritances for the objects are defined by the following trees: for* a*,* Γ_a *as given in 2.3(a); for* b*,* Γ_b *induced* *by the linear ordering of roles* $r_1 \succ r_2 \succ r_3 \succ r_4$*; for c,* Γ_c *induced by the linear ordering* $r_2 \succ r_1 \succ r_3 \succ r_4$*; and for d,* Γ_d *induced by the linear ordering* $r_1 \succ r_2 \succ r_3 \succ r_4$*. Notice that* Γ_b*,* Γ_c *and* Γ_d *are G-invariant. On the other hand, although* Γ_a *is group-respecting it is not G-invariant since the two matchings* $(1 \leftarrow b, 2 \leftarrow c)$ *(induced by the role mappings* $r_1 \rightarrow 1, r_2 \rightarrow 2$*) and* $(1 \leftarrow c, 2 \leftarrow b)$ *(induced by* $r_1 \rightarrow 2, r_2 \rightarrow 1$*) are G-equivalent but result in top-priority roles from different groups for a (*G_2 *and* G_3 *respectively).*

1	2	3	4
c	c	a	a
b	b	d	d
a	a	c	c
d	d	b	b

Table 2.2: Agent preferences for Example 2.5.3.

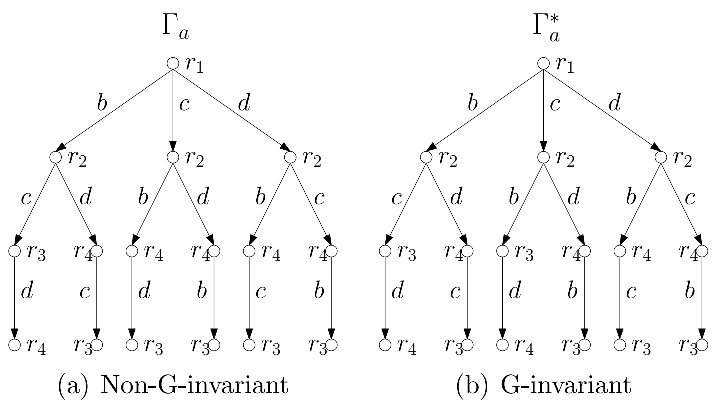


Figure 2.3: Two inheritance trees for object a in Example 2.5.3.

It is simple to verify the following: under $\{\Gamma_s\}$ *two matchings* $(1 \leftarrow b, 2 \leftarrow c, 3 \leftarrow a, 4 \leftarrow d)$ *and* $(1 \leftarrow c, 2 \leftarrow b, 3 \leftarrow a, 4 \leftarrow d)$ *are output with equal probability; under* $\{RP(\Gamma_s)\}$ *two matchings* $(1 \leftarrow b, 2 \leftarrow c, 3 \leftarrow d, 4 \leftarrow a)$ *and* $(1 \leftarrow c, 2 \leftarrow b, 3 \leftarrow d, 4 \leftarrow a)$ *are output with equal probability. Notice that under* $\{\Gamma_s\}$ 3 *is always assigned a and* 4 *is always assigned d, whereas under* $\{RP(\Gamma_s)\}$ *the reverse holds. So the two mechanisms are not equivalent.*

Now, consider if a has inheritance tree Γ_a^* *given in Figure 2.3(b) instead, while the inheritance trees of the other objects* Γ_s^* *are the same as* Γ_s*. It is not difficult to check that the mechanisms induced by* $\{\Gamma_s^*\}$ *and* $\{RP(\Gamma_s^*)\}$ *are equivalent, in particular both output two possible matchings* $(1 \leftarrow b, 2 \leftarrow c, 3 \leftarrow a, 4 \leftarrow d)$ *and* $(1 \leftarrow c, 2 \leftarrow b, 3 \leftarrow a, 4 \leftarrow d)$ *with equal probability.*

Remark. It is not difficult to see that inheritance trees allows for inheritance structure that cannot be captured by any linear ordering of roles. For example, consider a 4-agent instance where the agents $\{1,2,3,4\}$ are in three groups $G_1 = \{1,2\}$ (roles $\{r_1, r_2\}$), $G_2 = \{3\}$ (roles $\{r_3\}$) and $G_3 = \{4\}$ (roles $\{r_4\}$), and the objects are $\{a,b,c,d\}$. Suppose the inheritance tree a is given by Γ_a in Figure 2.4. If 1 is assigned b and 2 is assigned c, 3 becomes the owner of a. However, if instead 1 is assigned b and 2 is assigned d, 4 becomes the owner of a. It is not difficult to see that no group-respecting linear orderings can capture this sort of branching behavior.

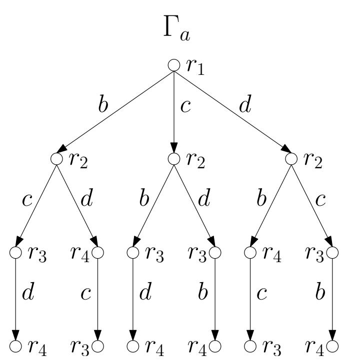


Figure 2.4: A G-invariant tree that cannot be represented by any linear inheritance.

2.6 Discussion

Our approach suggests the following general framework for thinking about allocating indivisible objects to a group of agents. Suppose we are given a set of agents with strict preferences over objects, and an exogenous partition of the agents into groups; suppose also that agents are partitioned into "existing tenants" and "new agents," and objects into "occupied" and "vacant" objects. Each occupied object has an obvious top-priority agent, namely, the agent who occupies it. Assume each vacant object is endowed to a unique group. Finally, each group can assign the top-priority role any way it pleases for all of the vacant objects endowed to it. Each group instantiates its roles uniformly at random, either without repetition (a given role is equally likely to be mapped to any remaining agent in that group who is not yet mapped) or with repetition (each role is equally likely to be mapped to any remaining agent in that group). Each group makes this decision independently, and can choose either type of mapping, regardless of what the other groups do. The TTC algorithm is used to clear (a subset of) cycles. If a vacant object remains, but its occupant does not, it is treated as a new object. If an object is endowed to a group, and its top-priority agent is no longer present, it is endowed to a different agent in the same group, assuming at least one such agent exists. If an object is endowed to a group, its top-priority agent is no longer present and no agent from that group is present, control of that object passes to another group (in a G-invariant way). Our general approach shows that the final outcome is the same regardless of how the groups assign the objects they control amongst themselves. It is easy to see that all the mechanisms considered in this chapter fit into this framework.

Consider the special case in which none of the agents are endowed with an object. Note that in this case all the mechanisms in this framework are strategyproof (SP), and always yield a Pareto efficient (PE) assignment. Furthermore, agents with identical preferences receive identical (probabilistic) allocations (ETE). The aforementioned result boils down to the statement that every mechanism in this framework is equivalent to the RP mechanism. An intriguing conjecture that suggests itself, in light of these equivalence results, is that the RP mechanism is *characterized* by these three properties. If we only require a subset of these properties, there are other mechanisms: giving an equal share of each object to each agent satisfies SP and ETE, but not PE; a serial dictatorship with an exogenous order of the agents satisfies SP and PE, but not ETE; and the probabilistic serial mechanism of Bogomolnaia and Moulin [14] satisfies PE and ETE, but not SP. The characterization result is, of course, of independent interest. We note, however, that such a characterization of RP may be an alternative approach to the proof of the equivalence results established here.

Chapter 3

Newsvendor with expert advice

In this chapter, we focus on the newsvendor problem under the presence of experts that provide stocking decisions advice to the inventory manager, with a minimax objective on the regret of the manager versus the best expert. In particular, we describe an optimal algorithm, and establish conditions under which the algorithm may be efficiently evaluated and how, as well as provide a novel asymptotic analysis that leads to the development of a heuristic for the optimal algorithm.

3.1 Introduction

The repeated newsvendor problem for systems facing unknown demand is an important classical problem in inventory management. Assuming that demand is i.i.d. across each period, demand can be learnt and used non-parametrically, i.e. without a priori assumptions on its form (see e.g. [33], [52]), or parametrically with demand that is uncensored (see e.g. [6], [45], [48], [56], [62], [75, 77]) or censored (see e.g. [22], [19], [24], [51], [59, 58] [41]). When the i.i.d. assumption on demand across the time horizon is relaxed, and dependencies and non-stationarity permitted, it is still possible to learn and use demand if the sequence of demands is predictable using time-series or other forecasting methods (see e.g. [76], [84], [46], [57], [35, 34], [39]).

However, when the demand is both non-i.i.d. and not amenable to standard forecasting techniques, the objective of minimizing the cumulative cost while learning demand becomes difficult to realize, especially in the case where demand is stock-dependent. In such scenarios, the minimax approach has found some measure of success (see e.g. [74], [60, 61], [32], [9], [10], [67]). In the minimax approach, it is assumed that some aspect of demand is fixed, e.g. the range of the demand or the mean and variance of the demand. The objective is to minimize the maximum cost across all possible demand realizations that satisfy the demand assumptions. This approach is a framework that is able to handle demand with complex dependencies or when the nature of dependencies is unknown, although it should be noted that it aims at optimizing worst-case performance and neglects average-case performance which can lead to overly conservative decisions.

We look at the following variation of the minimax approach. Consider the scenario where the inventory manager knows only that demand is bounded, but has recourse to advice on stocking level decisions provided by experts or oracles outside his control. In the real-world context, the stocking advice may be stocking levels generated by using different forecasting techniques, or provided by different pieces of software, to name a few examples. It is also not clear a priori which among the advice rendered would be the best in each period. We cannot expect the inventory manager to achieve minimal cost when all the experts provide poor advice, so the objective of the inventory manager is instead to minimize worst-case performance compared to the expert whose advice performs the best, i.e. to minimize his regret of not following the best expert. We concern ourselves mainly with the case where there are two experts that provide advice, but later extend our analysis to the case of multiple experts.

It turns out the scenario we are concerned is a variant within the class of *problems of predicting sequences with expert advice*. These problems of prediction are generally set up as follows: there is a series of trials, one in each period and with binary outcomes; in each period, the decision-maker has to predict that period's trial outcome and is made to pay a cost depending on how far off his prediction is from the outcome; however, before making his decision, the decision-maker has recourse to predictions made by experts outside his control, which he may utilize in making his prediction; the goal of the decision-maker is to make decisions to minimize the regret of not following the best expert. The decision-maker's prediction in each period is in the form of the probability that period's trial is successful. Furthermore, since the focus is on worst-case performance, the experts and trial outcomes may be assumed to be controlled by an adversary acting in opposition to the decision-maker. The first results for this set of models were due to Blackwell [13] and Hannan [37] and the model was subsequently re-discovered and widely studied by the machine learning and statistics community in recent years. Early works in the literature include the Weighted Majority algorithm of Littlestone and Warmuth [53] and the Aggregating algorithm of Vovk [86], where the decision-maker's prediction is a weighted sum of the predictions of the experts and the weights used are updated from period to period. Variants of the Weighted Majority algorithm were also studied in later years (see [30], [5], [88], among others). Alternative algorithms were also developed based on following the best-performing expert at any point in time, for example the Follow-the-Pertubed-Leader algorithm of Kalai and Vempala [47]. For extended discussion of these algorithms, interested readers may refer to Cesa-Bianchi and Lugosi [18].

This framework can be extended in a natural way to accommodate the newsvendor context by allowing the (demand) outcomes to be anywhere within some continuous interval (since demand may take values within some range) and allowing for an asymmetric linear cost function (since per-unit overage and underage costs and hence the costs of over- and under-prediction can differ). We call the resulting model the *newsvendor problem with expert advice.* Cesa-Bianchi et al. [17] give an algorithm *MM* that is optimal for the problem of predicting sequences with expert advice in the case where the cost of mis-prediction is linear. We show *MM* can be easily extended to give an optimal algorithm for the newsvendor problem with expert advice. The major issue with using either of these algorithms is that computing their decisions requires evaluating a function v that is described by recursively defined optimization problems. In fact, one question posed by Cesa-Bianchi et al. is whether v may be computed efficiently. Answering this question forms the core of this chapter, and we discover that the answer is in the affirmative. We find that when b and h are relatively integral, i.e. when they are both small multiples of some common factor or one is a multiple of the other, computing a value of v is computationally inexpensive. To handle cases where b and h are not relatively integral, we directly analyze the asymptotic behavior of v using a novel approach that models the function using martingales, in order to motivate a heuristic for calculating v. This is in contrast to existing approaches that draw conclusions about the limiting behavior of v from analyzing policies that are suboptimal but whose performance scales well with the number of periods. Our direct analysis of v uncovers its transient structure when scaling one of its parameters. This transient structure turns out to be very useful in formulating an approximation scheme for v. We use this to find an approximation for OPT that we call APX, whose decisions we show via a numerical experiment to be almost identical to OPT, and for which we give an example how it could extended to more than two experts.

To the best of our knowledge, OPT is the only optimal algorithm (and APX the only approximation to an optimal algorithm) for the newsvendor problem with expert advice. However, we note that there are several other algorithms, outside the scope of this chapter, that can be used for this problem either directly or with little modification. For example, Haussler, Kivinen and Warmuth [38] describes an algorithm that can handle general loss functions and continuous outcomes, and several of the other algorithms in Cesa-Bianchi et al. [17] can be modified to handle the newsvendor context, just as we had done with **MM** to get OPT. In contrast to OPT, these other algorithms typically have no assurances on working well when the number of periods is small, but are easy to compute in general and their performance can be shown to scale in the size of the problem (the number of periods) almost as well as an optimal algorithm when the size gets sufficiently large. We also note that while OPT and APX are tailored to the two expert case, these other algorithms are designed to use the advice of more than two experts.

Finally, there have also been attempts other than ours to utilize algorithms for prediction with expert advice models within the newsvendor context. O'Neil and Chaudhary [63] adapt the Weighted Majority algorithm of Littlestone and Warmuth [54] and the Follow the Perturbed Leader algorithm of Kalai and Vempala [47] for the newsvendor problem with expert advice. However, rather than allow the adversary to control the experts, they divide up the space of possible demands into buckets, assign each bucket to a different expert and have each expert recommend the minimax quantity within its bucket. Consequently, their analyses and their results pertain more to the newsvendor problem with a minimax objective rather than the newsvendor problem with expert advice.

The rest of this chapter is structured as follows: §3.2 describes the newsvendor problem with expert advice, for which we describe an optimal algorithm OPT in §3.3; §3.4 is the focus of this chapter, and in it we examine conditions under which computation of v is in polynomial time (§3.4.3) and analyze the asymptotic behavior of v directly (§3.4.4); finally, §3.5 describes an approximation scheme for v and the resulting the approximation algorithm APX (§3.5.1), compares its decisions with those of OPT via a numerical example

(§3.5.2) and gives a simple way to extend APX to use advice from more than two experts (§3.5.3).

3.2 Model description

In this section, we discuss a model for the newsvendor problem with expert advice. Models of the sort we will describe have been used extensively in problems of prediction. The particular one we use is a variant of the one discussed and analyzed in Cesa-Bianchi et al. [17], which looked at predicting binary sequences with recourse to expert advice. The main difference between the two variants is that the cost function they are concerned with is symmetric with respect to under- and over-prediction while the cost function we are concerned with can be asymmetric with respect to under- and over-prediction. We note that some recent papers have also considered similar models within the newsvendor context ([63]).

There is a finite planning horizon consisting of T periods, where time is indexed by t in a backward manner from T to 1; thus, T is the first period in the horizon and 1 is the last period. Denote the sequence of realized demands by $\mathbf{d} = (d_T, d_{T-1}, \ldots, d_1)$, where d_t represents the realized demand in period t. Assume that the range for possible values of demand is bounded and known; without loss of generality, the magnitude of demand is normalized to $d_t \in [0, 1]$.

In each period t, an inventory manager has to decide a stocking level decision $y_t \in [0, 1]$. He has recourse to two experts, $i = 1, 2$, each of whom provides him a reference stocking level decision ξ_t^i before he makes his decision. In addition, the inventory manager may base his decision on all other available historical data: the realized demand in previous periods, and the past decisions of both of the experts and himself. After the inventory manager makes his decision, demand d_t is realized and costs (both overage and underage) are incurred based on the stocking decision and the realized demand. The overage and underage costs are assumed to be linear with per-unit costs of $h \geq 0$ and $b \geq 0$ respectively, so the cost incurred by the inventory manager this period is

$$h(y_t - d_t)^+ + b(d_t - y_t)^+ \ . \tag{3.1}$$

Let λ_t be the cumulative cost incurred by the manager from periods T to $t+1$, and M_t^i be the cumulative cost through the same periods if the inventory manager had utilized decisions of expert i instead (i.e. if the inventory manager had decided $y_t = \xi_t^i$ for each $t = T, \ldots, t+1$). These quantities are recursively described by

$$\lambda_{t-1} = \lambda_t + h(y_t - d_t)^+ + b(d_t - y_t)^+ \ , \qquad \text{and} \tag{3.2}$$

$$M_{t-1}^i = M_t^i + h(\xi_t^i - d_t)^+ + b(d_t - \xi_t^i)^+ \qquad \text{for } i = 1, 2. \tag{3.3}$$

Additionally, $\lambda_T = M_T^1 = M_T^2 = 0$. Note that we will often abbreviate the vector (M_t^1, M_t^2) as $\mathbf{M}_t$.

The goal of the inventory manager is to minimize his regret of not having followed the advice of the better of the two experts, i.e. he seeks to minimize

$$\lambda_0 - \min\{M_0^1, M_0^2\} \ .$$

We will refer to the above expression as the manager's *regret*.

Remark. A greedy policy, where in each period t the inventory manager chooses $y_t = \xi_t^{i*}$ with $i^* = \text{argmin}\{M_t^i\}$, can perform badly. Hutter and Poland [42] give the following example applicable to the case of $b = h$:

$$\begin{pmatrix} \xi_1^1 & \xi_2^1 & \xi_3^1 & \xi_4^1 & \xi_5^1 & \dots \\ \xi_1^2 & \xi_2^2 & \xi_3^2 & \xi_4^2 & \xi_5^2 & \dots \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & 1 & 0 & \dots \end{pmatrix}$$

and $d_t = 0$ for all t; the regret using the greedy policy can be shown to be linear in the number of periods. Moreover, the same example can be extended to the case of $b \neq h$ by weighting all the adversary's decisions by $b/(b+h)$.

A worst-case approach is taken, and it is assumed that there is an adversary who controls the decisions of the both experts as well as the realized demand in each period and acts against the interest of the inventory manager.

Remark. The problem can viewed as a finitely-repeated sequential two-player game based on the newsvendor problem, where the first player is the inventory manager who decides the stocking level y_t, and the other player is an adversary who decides both the experts' choice of stocking levels (ξ_t^1, ξ_t^2) as well the realized demand d_t.

3.3 An optimal algorithm

In this section, we discuss a straightforward extension of the **MM** algorithm of Cesa-Bianchi et al. [17] to handle the newsvendor problem with expert advice. We first show that the newsvendor problem can be reduced to a problem of predicting binary sequences with an asymmetric cost function by examining the adversary's problem in §3.3.1. This implies that the **MM** algorithm, which is applicable and optimal for the newsvendor context when $b = h$, can be modified such that it is also optimal for the newsvendor context for general b and h. Describing this modified algorithm, which we call OPT, first requires us to define an auxiliary problem, which we do so in §3.4.1. OPT itself is outlined in 3.3.2.

3.3.1 The adversary's problem

Let $\mathcal{F}_t$ be the history up to the end of period t. (More formally, $\mathcal{F}_t = \{\xi_k^i, y_k, d_k : i = 1, 2, T \geq k \geq t\}$, with $\mathcal{F}_{T+1} = \emptyset$.) Fix the algorithm used by manager, which we denote by A. Let $A_t(\mathcal{F}_{t+1}, (\xi_t^1, \xi_t^2))$ be the manager's stocking decision in period t given the system history up to the end of period $t+1$ and both experts' stocking decisions in period t. The adversary's problem can be written as the following dynamic program. For $t \in \{T, T-1, \dots, 1\}$,

$$\begin{aligned} & J_t^A(\lambda_t, \mathbf{M}_t \mid \mathcal{F}_{t+1}) \\ & \quad = \max_{\xi_t^1,\ \xi_t^2,\ d_t} \ J_{t-1}^A\big(\lambda_{t-1}, \mathbf{M}_{t-1} \mid \mathcal{F}_{t+1} \cup \{d_t, y_t, (\xi_t^1, \xi_t^2)\}\big) \\ & \qquad \text{s. t.} \quad \text{Eq. (3.2)}, \quad \text{Eq. (3.3)}, \quad y_t = A_t(\mathcal{F}_{t+1}, (\xi_t^1, \xi_t^2)), \quad \text{and} \quad \xi_t^1, \xi_t^2, d_t \in [0,1] \ , \end{aligned} \tag{3.4}$$

and (for $t = 0$) $J_0^A(\lambda_0, \mathbf{M}_0 \mid \mathcal{F}_1) = \lambda_0 - \min\{M_0^1, M_0^2\}$. Under A, the inventory manager's regret is $J_T^A(0, \mathbf{0} \mid \mathcal{F}_{T+1})$, or simply $J_T^A(0, \mathbf{0})$. (Throughout this paper, we denote by $\mathbf{1}$ the vector of all 1's of an appropriate length, and by $\mathbf{0}$ the vector of all 0's.)

In the above definition of J_t^A, the maximization is taken over all possible values of d_t in the interval $[0, 1]$. The following proposition ensures that it suffices to consider the extreme values of d_t only. Interested readers may refer to §3.7.1 for the proof.

Proposition 3.3.1. *Fix the algorithm used by the manager, the set of optimal adversarial choices for the realized demand d_t in period t contains either 0 or 1 (or both).*

Proposition 3.3.1 allows us to assume that the adversary always chooses either 0 or 1 for the realized demand in every period. (This assumption may affect the incurred costs of the manager and the experts in each period, but the *difference* in the cost between the manager and that of each expert in every period is the same for any optimal adversarial choice.) More specifically, we may replace the constraint $d_t \in [0, 1]$ in the definition of J_t^A with $d_t \in \{0, 1\}$. This reduces the newsvendor problem with expert advice to a problem of predicting binary sequences with expert advice, but with a cost function that is asymmetric and linear with under- and over-prediction costs of b and h respectively.

3.3.2 The algorithm OPT and its optimality

An optimal algorithm for the problem of predicting binary sequences with expert advice under symmetric linear over- and under-prediction costs (i.e. the case where $b = h$) is the **MM** algorithm described by Cesa-Bianchi et al. [17]. Since we have reduced the newsvendor problem with expert advice to a prediction problem with an asymmetric linear cost function (where it is possible and probable that $b \neq h$), we would think that the **MM** algorithm could be modified to handle the more general cost function. Indeed, this turns out to be the case and we describe the result, which we call the algorithm OPT, in this section.

Before describing OPT, we define the following value function, which will be used to compute decisions in OPT and also plays an important role in its analysis:

$$\begin{aligned} v(\mathbf{M}, 0) &= \min_{i \in \{1,2\}} \{M^i\} , \\ v(\mathbf{M}, t) &= \min_{\mathbf{Z} \in [0,1]^2} \frac{b \cdot v(\mathbf{M} + h\mathbf{Z}, t-1) + h \cdot v(\mathbf{M} + b(\mathbf{1} - \mathbf{Z}), t-1)}{b+h} \qquad \text{for } t \geq 1, \end{aligned}$$

where $\mathbf{M} \in \mathbb{R}^2$. With this value function in mind, OPT is described in Figure 3.1. (Note that we can obtain **MM** from OPT by replacing both b and h by 1 in Eq. (3.5) and in the definition of v.)

Input: $T \in \{0, 1, 2, \ldots\}$; $b, h \in \mathbb{R}^+$
Output: Stocking level decisions $y_T, y_{T-1}, \ldots, y_1$.

1. $t \leftarrow T$; $M_i^T \leftarrow 0$ for $i = 1, 2$.
2. While $t > 0$, in the period t do:
3. Observe expert stocking decisions $\mathbf{Z} = (\xi_t^1, \xi_t^2)$.
4. Output the manager's stocking level decision in period t as

$$y_t = \frac{b}{b+h} + \frac{v(\mathbf{M}_t + h\mathbf{Z}, t-1) - v(\mathbf{M}_t + b(\mathbf{1} - \mathbf{Z}), t-1)}{b+h} \ , \quad (3.5)$$

where $\mathbf{M}_t = (M_t^1, M_t^2)$.

5. After demand d_t is realized, compute M_{t-1}^i, $i = 1, 2$, using Eq. (3.3).
6. $t \leftarrow t - 1$.

Figure 3.1: An optimal algorithm OPT.

The following two results show OPT is optimal and is stated without proof. They can be derived using arguments similar to those used in Cesa-Bianchi et al. [17] to prove the optimality of the **MM** algorithm. (Note that Proposition 3.3.1 is necessary to be able to make these arguments for OPT.)

Proposition 3.3.2. *For any algorithm A used by the manager,*

$$J_T^A(0, \mathbf{0}) \ \geq \ T \cdot \frac{bh}{b+h} - v(\mathbf{0}, T) \ .$$

Theorem 3.3.3.

$$J_T^{OPT}(0, \mathbf{0}) \ \leq \ T \cdot \frac{bh}{b+h} - v(\mathbf{0}, T) \ .$$

Observe that both the computation of and characterization of regret under OPT depend on v, effectively the objective value of a multi-stage optimization problem. We have no assurances that v is easily computable or easily characterizable, even in the case when $b = h$. Indeed, Cesa-Bianchi et al. [17] state that each evaluation of $v(M, t)$ can be computed via 2^{3T} recursive calls (when there are two experts and in the case of $b = h$) and pose the question of whether v can be computed efficiently. We address this question in the next section, the main focus of this chapter.

3.4 Computation and analysis of value function v

Since the decisions made by OPT involve the value function v, the computational efficiency of implementing this algorithm depends on the ready access to the values of v, which is recursively defined. We achieve two results regarding v in this section. The first is a partial answer to what the conditions are under which v may be efficiently computed, which is the subject of §3.4.3. The second is an analysis of the asymptotic behavior of v using a new approach using random-walks in §3.4.4, which will later motivate a heuristic for evaluating v and thus for OPT in §3.5.1. However, before proceeding, we state several basic properties of v in §3.4.1 that we will use throughout this section and explore how to reduce the decision space of the recursive optimization problem describing v in §3.4.2.

3.4.1 Properties of v

We will use the following properties of v, which can be proved in a straightforward manner by induction on t and using the definition of v.

A. For any $t \geq 0$, $m \in \mathbb{R}$, and $\mathbf{M} \in \mathbb{R}^2$, $v(\mathbf{M} + m\mathbf{1}, t) = v(\mathbf{M}, t) + m$.

B. For any $\mathbf{X} = (X^1, X^2) \in \mathbb{R}^2$, $\mathbf{Y} = (Y^1, Y^2) \in \mathbb{R}^2$, and $t \geq 0$,

$$\min_{i \in \{1,2\}} \left\{ X^i - Y^i \right\} \leq v(\mathbf{X}, t) - v(\mathbf{Y}, t) \leq \max_{i \in \{1,2\}} \left\{ X^i - Y^i \right\} .$$

C. $v(\mathbf{M}, t)$ is concave in $\mathbf{M}$, i.e., for all $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^2$ and $\alpha \in (0,1)$,

$$v(\alpha \mathbf{X} + (1-\alpha)\mathbf{Y}, t) \geq \alpha v(\mathbf{X}, t) + (1-\alpha) v(\mathbf{Y}, t) .$$

D. For all $\mathbf{M} \in \mathbb{R}^2$ and $\alpha \in (0,1)$, $v(\alpha \mathbf{M}, t) \geq \alpha v(\mathbf{M}, t)$.

Readers interested in specifics may refer to Cesa-Bianchi [17] which show properties (b) and (c) for the case of $b = h = 1$. (The approach for $b \neq h$ is similar.)

3.4.2 Reducing the decision space in v

Recall the definition of v, i.e.

$$v(\mathbf{M}, t) = \min_{\mathbf{Z} \in [0,1]^2} \frac{b \cdot v(\mathbf{M} + h\mathbf{Z}, t-1) + h \cdot v(\mathbf{M} + b(\mathbf{1} - \mathbf{Z}), t-1)}{b+h} \qquad \text{for } t \geq 1.$$

Each stage of recursion for $v(\mathbf{M}, t)$ requires minimizing a function over the unit square $[0,1]^2$. However, since $v(\mathbf{M}, t)$ is concave in $\mathbf{M}$ (property C), the expression in it over which minimization is performed is concave in $\mathbf{Z} \in [0,1]^2$ so the value of minimizer is one of the four corners of the unit square $[0,1]^2$: $(0,0)$, $(0,1)$, $(1,0)$, or $(1,1)$. This implies that $v(\cdot, t)$ may be evaluated via 4^t recursive calls, each corresponding to a specific sequence of minimizer choices for $\mathbf{Z}$. Note that this had already been observed in Cesa-Bianchi et al. [17] for the case where $b = h$.

To further reduce the search space, we introduce a function f that can be viewed as a transformation of the v function into single-dimensional space, at the cost of some loss of information. In this form, we are able to tighten our search space to two values at each stage, instead of four. Furthermore, function f will play an important role in the asymptotic analysis of v in §3.4.4.

Definition 3.4.1. Let $f : \mathbb{R} \times \{0, 1, 2, \ldots\} \to \mathbb{R}$ be a function given by

$$f(m,t) := v\left(\begin{bmatrix} 0 \\ m \end{bmatrix}, t\right) .$$

In §3.4.1, we mentioned several properties of v. It is not difficult to see that properties B-D are applicable to the function f as well (simply set the first entry of $\mathbf{M}$ to 0 in each of the properties). Moreover, property A immediately implies the following conversion between f and v:

Proposition 3.4.2. *Let* $(M^1, M^2) \in \mathbb{R}^2$ *and* $t \geq 0$. *Then*

$$v\left(\begin{bmatrix} M^1 \\ M^2 \end{bmatrix}, t\right) = f(M^2 - M^1, t) + M^1 .$$

We may characterize f via the following recursion. (The proof of this result may be found in §3.7.2.)

Proposition 3.4.3. *For* $m \in \mathbb{R}$ *and* $t \geq 1$,

$$f(m,t) = \frac{bh}{b+h} + \min_{\sigma \in \{1,-1\}} \left\{ \frac{b \cdot f(m + \sigma h, t-1) + h \cdot f(m - \sigma b, t-1)}{b+h} \right\} .$$

Also, $f(m, 0) = \min\{m, 0\}$.

Remark. Using Proposition 3.4.3, it can be easily shown by induction on t that $v(\mathbf{0}, t) = f(0, t) \leq t \cdot bh/(b + h)$.

Remark. As a corollary of Proposition 3.4.3, one could tighten the recursive expression for v such that the value of $\mathbf{Z}$ is restricted to a two-point set consisting of $(0, 1)$ or $(1, 0)$. In particular, the decision space $\mathbf{Z} \in [0, 1]^2$ in v can be replaced with $\mathbf{Z} \in \{(0, 1), (1, 0)\}$.

To compute one value of $v(\mathbf{M}, t)$, we have now reduced the number of recursive calls that need to be made from 4^t to 2^t. Even though it is a large improvement, computing v in this manner is still prohibitively slow when t is not very small. However, computing v can be significantly quicker in certain cases, which we will discuss next.

3.4.3 When computing v can be fast

From §3.4.2, we know that the task of computing $v(\mathbf{M}, t)$ is the same as the task of computing $f(M^2 - M^1, t)$. In this section, we focus on computing $f(m, t)$, and show that it can be done very efficiently in the case when both cost parameters b and h may be represented as small integral multiples of some common factor. In that case, we may evaluate any specific $f(m, t)$ in time polynomial in t but exponential in how small the common factor is relative to b and h. Stronger yet, if b is an integral multiple of h (or vice versa), we are able to compute full representations of $f(\cdot, t)$, $t = 1, 2, \ldots, T$ in time polynomial in (b/h) and T, after which each evaluation of $f(m, t)$ involves at most two lookups from a table of values.

Case 1: b and h are small integral multiples of a common factor

When both b and h can be represented by relatively small multiples of a common factor, i.e., there exists $\delta > 0$ such that $b = K\delta$ and $h = L\delta$ for small positive integers K and L, we show that a single evaluation of $f(m,t)$ for fixed m and t can be carried out efficiently. For this section, without loss of generality, we assume $b > h$, or equivalently $K \geq L$.

For a given value of m and t, we will describe a dynamic programming algorithm to compute $f(m,t)$ and prove its correctness. Define the set $\mathcal{H}_i := \{m - iK\delta, m - (iK - 1)\delta, \ldots, m - \delta, m, m + \delta, \ldots m + (iK - 1)\delta, m + iK\delta\}$. We begin by computing $f(x, 0)$ for $x \in \mathcal{H}_t$; note that $\mathcal{H}_t$ contains $2tK + 1$ points. What follows is the general schema: For $i \geq 0$ incrementally, presume that we have computed $f(x,i)$ for $x \in \mathcal{H}_{t-i}$. For each $x \in \mathcal{H}_{t-i-1}$, we compute $f(x, i+1)$. From Proposition 3.4.3,

$$f(x, i+1) \;=\; \frac{bh}{b+h} + \min_{\sigma \in \{1,-1\}} \left\{ \frac{b \cdot f(x + \sigma h, i) + h \cdot f(x - \sigma b, i)}{b+h} \right\} . \tag{3.6}$$

Since $x \in \mathcal{H}_{t-i-1}$ and also $h = L\delta$ and $b = K\delta$, it can be shown that each of $\{x - h, x + h, x - b, x + b\}$ appearing in the above expression belongs to $\mathcal{H}_{t-i}$. Hence, the computation of $f(x, i+1)$ requires only the values of $f(\cdot, i)$ which we have already computed. Finally, we note that when $i = t$, we compute $f(m,t)$, and the dynamic programming algorithm terminates.

The dynamic programming algorithm we have described takes $\mathcal{O}(t^2 K)$ to compute a single $f(m,t)$ value, a significant improvement over the 2^t recursive calls indicated in §3.4.2.

Case 2: b is an integral multiple of h

Computation of $f(\cdot, t)$ may be further improved in the special case where b is an integral multiple of h, i.e., $(b/h) \in \mathbb{Z}^+$. In many supply chain settings, b is significantly larger than h, and thus this assumption is a reasonable approximation. To improve computational efficiency in this case, we prove that, for each t, $f(\cdot, t)$ is a piecewise-linear function with all its breakpoints contained in a discrete set, whose size is linear in t. Furthermore, we show that the minimum operator in the recursive computation of $f(x,t)$ can be eliminated by dictating the optimal solution (refer to Eq. (3.6)). We state these results formally in the following theorem. The proof of this theorem is based on a transformation of the f function and is not straightforward. Interested readers may refer to §3.7.3 for the details.

Theorem 3.4.4. *Let b/h be a positive integer. Then, f has the following properties:*

(i) $f(\cdot,t)$ is a piece-wise linear function, where the breakpoints belong to the set $\{-tb,\dots,-2h,-h,0,h,2h,\dots,tb\}$.

(ii) The optimal choice of σ in Eq. (3.6) is -1 if $x \le 0$ and 1 if $x \ge 0$ for any t. I.e.,

$$f(x,t+1) = \frac{bh}{b+h} + \begin{cases} \frac{b}{b+h} f(x-h,t) + \frac{h}{b+h} f(x+b,t), & \textit{for } x \le 0 \\ \frac{h}{b+h} f(x-b,t) + \frac{b}{b+h} f(x+h,t), & \textit{for } x \ge 0. \end{cases}$$

As a consequence of Theorem 3.4.4, we can characterize $f(\cdot,t)$ as a single-variable function for each t, by evaluating $f(\cdot,t)$ only at breakpoints, and the number of such breakpoints is linear in t. Since an evaluation of $f(x,t)$ for any breakpoint x requires us evaluate $f(\cdot,t-1)$ for a constant number (only two) of different parameters, it follows that using dynamic programming we may compute complete representations of $f(\cdot,t)$ for $t = 1,2,\cdots,T$ in time polynomial in T, namely on the order of $\mathcal{O}((b/h)T^2)$. Part (ii) of Theorem 3.4.4 is a surprising result that eliminates a minimization operator in computation, and this result is consistently observed in numerical computations even when the integrality assumption of b/h is relaxed.

We contrast Theorem 3.4.4 with the results in the previous case where b and h share a common factor. We showed in the previous case that the problem of computing $f(m,t)$ for a given pair of m and t can be accomplished in time polynomial in t but exponential in how small the common factor is relative to b and h. In this case, we have shown that when b is a multiple of h we can pre-process and compute all breakpoints and corresponding values in time polynomial in (b/h) and T, after which any evaluation of $f(m,t)$ (equivalently $v(\mathbf{M},t)$) can be done by looking up the values of $f(\cdot,t)$ at the pair of breakpoints bracketing m and interpolating (or extrapolating if m is less than $-tb$ or greater than tb). This is a significant improvement in computational complexity over the previous case. We have also solved the problem of finding which value of σ minimizes the expression given in Eq. (3.6) based on the value of m only. The results of this section, however, require that b be an integral multiple of h, a stronger assumption than the previous case. (We remark that when h is a multiple of b, all the results of this section remain valid after obvious modification.)

3.4.4 Asymptotic analysis of v and f

We have studied in Sections 3.4.2 and 3.4.3 the conditions under which f (equivalently v) may be efficiently computed. In this section, we directly analyze the limiting behavior of $v(\mathbf{0},T)$ as $T \to \infty$ using a novel approach. The asymptotic behavior we find for $v(\mathbf{0},T)$ implies that the manager's optimal regret scales with $\sqrt{T}$. Note that the same observation about how optimal regret scales had already been made in Cesa-Bianchi et al. [17] for the case where $b = h$ and in Haussler, Kivinen and Warmuth [38] for more general loss functions (including the case of $b \neq h$), where they obtain these results not from directly analyzing v like we do but by analyzing the performance of non-optimal algorithms that nonetheless perform well in the long-run. Also note that our direct analysis of v will lead to a heuristic for computing v that we discuss in §3.5.1.

The outline of our asymptotic analysis is as follows. We define a two-dimensional stochastic process $\{S_{i,t}\}$, which can be thought of as an array of random walk-like single-dimensional processes, and show how $f(m,t)$ can be expressed in terms of $f(S_{t,t},0)$

(Proposition 3.4.5). This stochastic process under a central limit theorem-type scaling ($\{\tilde{S}_{i,t}\}$) falls under a class of martingales called *zero-mean square-integrable martingale arrays*, and we show that $\mathbb{E}[f(\tilde{S}_{t,t},0)]$ converges to $\mathbb{E}[f(N(0,1),0)]$, where $N(0,1)$ is a unit normal random variable (Proposition 3.4.7). Finally, we relate this back to the f and v functions as well as $J_T^A(0,\mathbf{0})$ (Proposition 3.4.8 and its corollaries).

We begin by defining the stochastic process that we will later relate to $f(m,r)$. Let $\{S_{t,k} \,:\, 0 \le k \le t, t \ge 0\}$ be a stochastic process given by

$$\begin{aligned} S_{t,0} &= m && \text{for } t \ge 0 \text{ and} \\ S_{t,k} &= S_{t,k-1} + X_{t,k}(S_{t,k-1}) && \text{for } k \in \{1,2,\ldots,t\}, \end{aligned}$$

where $\{X_{t,k}(S_{t,k-1})\}$ are independent random variables such that

$$X_{t,k}(m) = \begin{cases} \theta_{t-k+1}(m)h & \text{with probability } b/(b+h) \\ -\theta_{t-k+1}(m)b & \text{with probability } h/(b+h), \text{ and} \end{cases}$$

$$\theta_k(m) = \begin{cases} 1 & \text{if } bf(m+h,k-1)+hf(m-b,k-1) \\ & \qquad \le bf(m-h,k-1)+hf(m+b,k-1), \\ -1 & \text{otherwise.} \end{cases}$$

(Note that the condition in the definition of $\theta_k(m)$ is equivalent to $m \ge 0$ if b is an integer multiple of h; see Proposition 3.4.3 and Theorem 3.4.4.)

Observe that each $X_{t,k}(m)$ is a two-point random variable with the expected value of 0. This random variable takes either h with probability $b/(b+h)$ and $-b$ with probability $h/(b+h)$, or its negative counterpart that has $-h$ with probability $b/(b+h)$ and b with probability $h/(b+h)$. This choice is determined by the multiplier $\theta_{t-k+1}(m)$, which depends on the function $f(\cdot, k-1)$.

The following proposition relates the expected value of this process to $f(m,t)$. The proof of this proposition is based on induction and the definition of the $\{S_{t,k}\}$ process, and can be found in §3.7.4.

Proposition 3.4.5. *For all $t \geq 0$,*

$$\mathbb{E}[f(S_{t,t}, 0) \mid S_{t,0} = m] \;=\; f(m,t) - t \cdot \frac{bh}{b+h} \;.$$

While the quantity of our interest is $f(m,t)$, Proposition 3.4.5 shows that it is sufficient to evaluate the expected value of $f(S_{t,t}, 0)$. Thus, we focus our attention to the $\{S_{t,k}\}$ process. In particular, we proceed to examine the asymptotic behavior of this process under a central limit theorem-type scaling (i.e., we scale by the reciprocal of $\sqrt{t}$). As is typical for such results, we normalize the stochastic process such that it has zero-mean.

From the definition, the variance of $X_{t,k}$ is given by $\sigma^2 := \mathbb{E}[X_{t,k}^2 \mid S_{t,k-1}] = bh$, for any $1 \leq k \leq t < \infty$. Therefore,

$$\sum_{k=1}^{t} \mathbb{E}[X_{t,k}^2 \mid S_{t,k-1}] \;=\; \sigma^2 t \;.$$

We define a new stochastic process $\{\tilde{S}_{t,k} \;:\; 0 \leq k \leq t, t \geq 0\}$ based on $\{S_{t,k}\}$ by normalization and scaling: For all $1 \leq k \leq t < \infty$, define

$$\tilde{S}_{t,k} \;:=\; \frac{S_{t,k} - S_{t,0}}{\sigma\sqrt{t}} \;.$$

Also, define

$$\tilde{X}_{t,k}(\tilde{S}_{t,k-1}, S_{t,0}) \;:=\; \frac{X_{t,k}(S_{t,k-1})}{\sigma\sqrt{t}} \;.$$

It follows that

$$\tilde{S}_{t,k} = \sum_{j=1}^{k} \tilde{X}_{t,j}(\tilde{S}_{t,j}, S_{t,0}) \ .$$

We note that this process $\{\tilde{S}_{t,k}\}$ depends on $S_{t,0}$, but we suppress this dependency in our notation to simplify exposition. We also write $\tilde{X}_{t,k}$ in place of $\tilde{X}_{t,k}(\tilde{S}_{t,k}, S_{t,0})$ when there is no ambiguity.

To show that $\{\tilde{S}_{t,k}\}$ satisfies certain desirable properties, we remind the reader of the following definition.

Definition 3.4.6. (Hall and Heyde [36]) Let $(\Omega, \hat{\mathcal{F}}, \mathbb{P})$ be a probability space and $\{\hat{\mathcal{F}}_{t,k} \ : \ 1 \leq k \leq t, t \geq 1\}$ be an array of sub-σ-fields with $\hat{\mathcal{F}}_{t,k} \subseteq \hat{\mathcal{F}}_{t,k+1}$ for each $t \geq 1$. We say $\{\hat{S}_{t,k}, \hat{\mathcal{F}}_{t,k} \ : \ 1 \leq k \leq t, t \geq 1\}$ is a *zero-mean square-integrable martingale array with differences* $\{\hat{X}_{t,k} \ : \ 1 \leq k \leq t, t \geq 1\}$ if for all $t \geq 1$, $\hat{X}_{t,k}$ is $\hat{\mathcal{F}}_{t,k}$-measurable with $\mathbb{E}[\hat{X}_{t,k}^2] < \infty$ for each $1 \leq k \leq t$, $\mathbb{E}[\hat{X}_{t,k} \mid \hat{\mathcal{F}}_{t,k}] = 0$ for $1 \leq k \leq t$, and $\hat{S}_{t,k} = \sum_{1 \leq j \leq k} \hat{X}_{t,j}$ for $1 \leq k \leq t$.

We are now ready to obtain a series of results that characterizes the asymptotic behavior of $f(m,t)$ as $t \to \infty$ in terms of the standard normal distribution $N(0,1)$ and the $f(\cdot, 0)$ function, where we remind the reader that $f(m,0) = \min\{m, 0\}$. (Recall Definition 3.4.1 and Proposition 3.4.3.) These results are listed in Proposition 3.4.7 below. We first establish, in part (a), that $\{\tilde{S}_{t,k}\}$ satisfies the definition of a zero-mean square-integrable martingale array (Definition 3.4.6). Then, in part (b), we relate $f(\tilde{S}_{t,t}, 0)$ to $f\big(N(0,1), 0\big)$ by applying the Martingale Central Limit Theorem (Theorem 3.7.5 in §3.7.5). In part (c), we find an asymptotic expression of $f(m,t)$. The proof of part (c) is based on the relationship between $f(m,t)$ and $\mathbb{E}[f(S_{t,t}, 0)]$ (established earlier in Proposition 3.4.5), and uses the idea of how an appropriately-scaled version of $\mathbb{E}[f(S_{t,t}, 0)]$ can be sandwiched between two expressions involving $\mathbb{E}[f(\tilde{S}_{t,t}, 0)]$. In particular, we derive this by splitting

$S_{t,t}$ into two components: its starting initial condition $S_{t,0}$ and the normalized process $S_{t,t} - S_{t,0}$. When scaled by $1/(\sigma\sqrt{t})$, in the limit the initial condition vanishes and the scaled normalized process can be characterized using part (b). (This will have implications in formulating a heuristic in Section 3.5.1). The proof of Proposition 3.4.7 can be found in §3.7.5.

Proposition 3.4.7.

(a) The process $\{\tilde{S}_{t,k}, \mathcal{F}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$ *is a zero-mean square-integrable martingale with differences* $\{\tilde{X}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$.

(b) For any $m \in \mathbb{R}$, $\mathbb{E}\left[f(\tilde{S}_{t,t}, 0) \mid S_{t,0} = m\right] \ \to \ \mathbb{E}\left[f\big(N(0,1), 0\big)\right]$.

(c) For any $m \in \mathbb{R}$,

$$\lim_{t\to\infty} \frac{f(m,t) - t \cdot \frac{bh}{b+h}}{\sigma\sqrt{t}} \quad = \quad \lim_{t\to\infty} \mathbb{E}\left[f(\tilde{S}_{t,t}, 0) \mid S_{t,0} = m\right] \quad = \quad \mathbb{E}\left[f\big(N(0,1), 0\big)\right].$$

Now, having established the asymptotic behavior of $f(m,t)$ as $t \to \infty$, we can derive an asymptotically accurate expression for $v(\mathbf{M}, t)$, where $\mathbf{M} = (M^1, M^2) \in \mathbb{R}^2$. The relationship between $f(M^2 - M^1, t)$ and $v(\mathbf{M}, t)$ given in Proposition 3.4.2 shows that these expressions differ by M^1, a constant, that when scaled by $\big(1/\sqrt{t}\big)$ diminishes as $t \to \infty$. See §3.7.6 for the complete proof of Proposition 3.4.8.

Proposition 3.4.8. *For any* $\mathbf{M} = (M^1, M^2) \in \mathbb{R}^2$,

$$\lim_{t\to\infty} \frac{1}{\sqrt{bh}\cdot\sqrt{t}} \cdot \left(v\left(\mathbf{M}, t\right) - t \cdot \frac{bh}{b+h} \right) \quad = \quad -\,\frac{1}{\sqrt{2\pi}} \quad \approx \quad -\,0.399 \ .$$

For any vector $\mathbf{M}$, the above proposition shows that $v(\mathbf{M}, t) \approx [bh/(b+h)] \cdot t - \sqrt{bh/(2\pi)} \cdot \sqrt{t}$ when t is large. Note that this expression is a difference between a linear term in t and a square-root term in t. Furthermore, it is independent of $\mathbf{M}$.

As a corollary of Proposition 3.4.8, we can establish an asymptotic expression for the performance for OPT. The following result provides a lower bound on the limiting regret of any algorithm A as well as that of OPT. It is a direct consequence of Proposition 3.4.8, along with Proposition 3.3.2 and Theorem 3.3.3.

Theorem 3.4.9. *For any algorithm A of the manager,*

$$\lim_{T\to\infty} \frac{J_T^A(0,\mathbf{0})}{\sqrt{bh}\cdot\sqrt{T}} \;\geq\; \frac{1}{\sqrt{2\pi}} \;\approx\; 0.399\ .$$

Furthermore, the above inequality is replaced by equality for OPT.

The above theorem shows that under the optimal algorithm OPT, for large T the manager's regret satisfies $J_T^{OPT}(0,\mathbf{0}) \approx \sqrt{bh/(2\pi)}\sqrt{T}$ which is asymptotically accurate. This difference grows with the square-root of T. Therefore, the *per period* regret, $J_T^{OPT}(0,\mathbf{0})/T$, decreases to 0 as $T \to \infty$.

3.5 A heuristic for v

3.5.1 A heuristic for v, and the algorithm APX

So far, we have established that v is quickly computable if the costs satisfy certain integrality conditions (Section 3.4.3) as well as analyzed the asymptotic behavior of $v(\mathbf{M},t)$ as t grows very large (Section 3.4.4). One remaining issue is the computation of $v(\mathbf{M},t)$ when neither the aforementioned cost integrality conditions are satisfied nor is t very large. To address this, we derive here a heuristic for $v(\mathbf{M},t)$, where each evaluation of $v(\mathbf{M},t)$ is about as fast as evaluating a single value of the unit normal loss function. This heuristic value for $v(\mathbf{M},t)$ can be used as a replacement for the exact value for $v(\mathbf{M},t)$ in

OPT (specifically in Eq. (3.5)), and we term the algorithm that results from doing this APX.

As shown previously, the problem of computing $v(\mathbf{M}, t)$ is equivalent to the problem of computing $f(m, t)$. From Proposition 3.4.5, we decomposed f into two parts:

$$f(m,t) = t \cdot \frac{bh}{b+h} + \mathbb{E}[f(S_{t,t}, 0) \mid S_{t,0} = m] \ . \tag{3.7}$$

In Proposition 3.4.7(c), we showed that $\mathbb{E}[f(S_{t,t}, 0) \mid S_{t,0} = m]/(\sigma\sqrt{t})$ converges to $\mathbb{E}[f(N(0,1), 0)]$, and therefore showed that

$$\lim_{t\to\infty} f(m,t) = t \cdot \frac{bh}{b+h} + \sigma\sqrt{t} \cdot \mathbb{E}[f(N(0,1), 0)] \ .$$

This provides a reasonable approximation for very large values of t, but not necessarily for smaller values of t. In particular, we had derived that result by splitting $S_{t,t}/(\sigma\sqrt{t})$ into two components

$$\frac{S_{t,t} - S_{t,0}}{\sigma\sqrt{t}} + \frac{S_{t,0}}{\sigma\sqrt{t}} \ ,$$

and considering the asymptotic behavior of each component. The left component is a scaled version of $S_{t,t}$ initialized to begin at $S_{t,0} = 0$, and the right component is the actual initial condition $S_{t,0} = m$ at the same scaling. The left component converges asymptotically to the unit normal random variable, $N(0,1)$. And the right component goes to zero as t gets very large so we may disregard it in the limit. However, any reasonable approximation should work well when t is not very large, and it is possible for the right component to dominate the left one when $S_{t,0}$ is large and t is small.

Hence, in our heuristic we use the asymptotic limit for the left-hand term but retain the right-hand term. Since the left-hand term converges asymptotically to $N(0,1)$ and the right-hand term contributes $m/(\sigma\sqrt{t})$, we can "approximate" their sum by the normal random variable $N(m/(\sigma\sqrt{t}), 1)$. We thus approximate Eq. (3.7) by

$$t \cdot \frac{bh}{b+h} + \sigma\sqrt{t} \cdot \mathbb{E}\left[f\left(N\left(\frac{m}{\sigma\sqrt{t}}, 1\right), 0\right)\right] .$$

It remains to evaluate $\mathbb{E}[f(N(m/(\sigma\sqrt{t}), 1), 0)]$. Since $f(m, 0) = \min\{m, 0\} = -m^-$,

$$\mathbb{E}\left[f\left(N\left(\frac{m}{\sigma\sqrt{t}}, 1\right), 0\right)\right] = \mathbb{E}\left[-N\left(\frac{m}{\sigma\sqrt{t}}, 1\right)^-\right] = -\mathbb{E}\left[N\left(-\frac{m}{\sigma\sqrt{t}}, 1\right)^+\right] ,$$

where the last equality follows from the symmetry of the normal distribution about its mean. It is a well-established probability result that if a random variable X is normally distributed with mean λ and variance σ^2, we have that $\mathbb{E}\left[(X - x)^+\right] = \sigma L((x - \lambda)/\sigma)$, where L is the unit normal loss function $L(x) = \mathbb{E}\left[(N(0, 1) - x)^+\right]$. Hence, we may simplify the above expression to

$$\mathbb{E}\left[f\left(N\left(\frac{m}{\sigma\sqrt{t}}, 1\right), 0\right)\right] = -L\left(\frac{m}{\sigma\sqrt{t}}\right) .$$

This gives us the following approximation formula for $f(m, t)$,

$$f(m, t) \approx t \cdot \frac{bh}{b+h} - \sqrt{bht} \cdot L\left(m/\sqrt{bht}\right) .$$

Proposition 3.4.2 lets us to convert this into a heuristic for v, i.e.

$$v(\mathbf{M}, t) \approx M^1 + t \cdot \frac{bh}{b+h} - \sqrt{bht} \cdot L\left((M^2 - M^1)/\sqrt{bht}\right) .$$

It is easy to verify that as $t \to \infty$, this heuristic value and the actual $v(\mathbf{M}, t)$ have the same limiting value (the one given in Theorem 3.4.9). As stated earlier, APX is the variant of

OPT that computes the stocking level decision in each period t using this heuristic for $v(\mathbf{M}, t)$ instead of its exact value.

It is not difficult to see that the decisions of APX and OPT converge as $t \to \infty$, so one can show that

$$\lim_{T\to\infty} \frac{J_T^{APX}(0, \mathbf{0})}{\sqrt{bh} \cdot \sqrt{T}} = \lim_{T\to\infty} \frac{J_T^{OPT}(0, \mathbf{0})}{\sqrt{bh} \cdot \sqrt{T}}.$$

In other words, the regret using APX scales in T on the same order as that using OPT.

Additionally, in §3.5.2, we show numerically that the difference in the decisions made by OPT and APX diminishes as t increases, and are minute even for small t.

3.5.2 Numerical comparison decisions in APX and OPT

Here, we study numerically the decisions made by APX and by OPT for one instance and contrast how they vary across different values of t. We fix the following model parameters. Consider some time period t, with the backorder and holding costs being $b = 9$ and $h = 1$ respectively, and where expert 1's decision is $\xi_1^t = 0$ and expert 2's is $\xi_2^t = 1$. We consider three values of t: $t = 5$, $t = 50$ and $t = 500$, corresponding to short term, medium term and long term lengths of time until the end of the time horizon. From earlier analyses, for any fixed period t, in that period OPT mixes the decision of the two experts based on the difference of their accrued costs. Hence, we only vary $M_t^1 - M_t^2$, i.e. expert 1's accumulated cost minus that of expert 2's with t periods remaining in the time horizon. As $M_t^1 - M_t^2$ becomes more negative, expert 1's accrued cost becomes lower compared to expert 2's and vice versa. We consider how the decisions of APX and OPT vary with $M_t^1 - M_t^2$. The results are shown in Figure 3.2.

We see that APX is a very good approximation for OPT. When t is large ($t = 50$ or $t = 500$), their decisions are practically identical. Moreover, when t is small ($t = 5$), the difference between decisions of APX and OPT are minute.

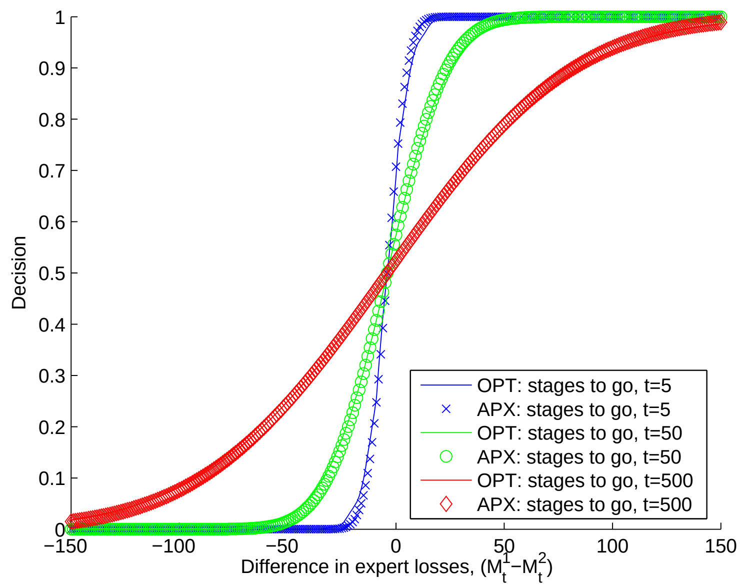


Figure 3.2: OPT and APX decisions against difference in expert costs, with expert decisions $\xi_1^t = 0$, $\xi_2^t = 1$ and costs $b = 9$, $h = 1$.

3.5.3 Extending APX to more than two experts

So far, we have assumed that the inventory manager has recourse to the advice of only two experts. We give a simple example of how APX can be extended to the case with $N = 2^k$ experts for some $k > 1$. (That N is a power of two is introduced for notational convenience, and the case where N is not a power of two can be analyzed using the same idea.) To extend APX to handle N experts, we decompose the problem with more than two experts into a collection of concurrent two-experts problems. To do so, we construct a "knockout" tournament tree of experts. In this tree, we begin with N experts, labeled $\Lambda^{0,1}, \Lambda^{0,2}, \ldots, \Lambda^{0,N}$, that mimic the decisions of the experts $1, 2, \ldots, N$ of the original problem. We pair up the experts in a consecutive manner, $\Lambda^{0,1}$ with $\Lambda^{0,2}$, $\Lambda^{0,3}$ with $\Lambda^{0,4}$, and so forth. For each of these pairs $\Lambda^{0,2j-1}$ and $\Lambda^{0,2j}$, we combine them into a new expert $\Lambda^{1,j}$ whose decisions are that of using APX on the two experts $\Lambda^{0,2j-1}$ and $\Lambda^{0,2j}$. We continually repeat this pairing and mixing procedure, using the decisions of $\Lambda^{i,2j-1}$

and $\Lambda^{i,2j}$ to form a decision for $\Lambda^{i+1,j}$ for each i, until we are at some round i^* with only one expert $\Lambda^{i^*,1}$. It is clear that each new round has only half the number of experts of the previous round, so for each i there are $N/2^i$ experts of the form $\Lambda^{i,j}$, and $i^* = \log_2 N$. The algorithm picks the decision of the inventory manager to be the same as $\Lambda^{\log_2 N,1}$. We maintain this tree of experts throughout the time horizon and keep track of each of the experts' ($\Lambda^{i,j}$'s) performance and decisions. As an example, a tournament tree for TRN when $N = 4$ is shown in Figure 3.3.

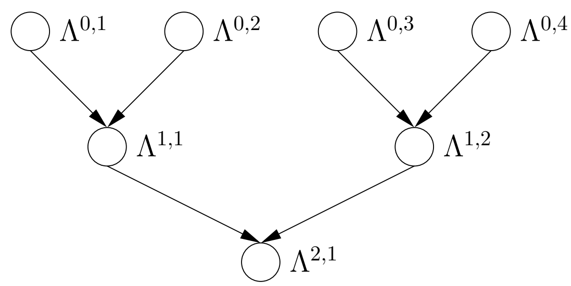


Figure 3.3: TRN tournament tree for $N = 4$ experts

Call this tournament tree algorithm TRN. Given that TRN has a tournament tree with $\log_2 N$ levels and uses APX whose regret scales in T the same as OPT (i.e. $\sqrt{T}$) in order to merge the decisions of each pair of experts in a level, it is not difficult to show that

$$\lim_{T\to\infty} \frac{J_T^{TRN}(0,\mathbf{0})}{\sqrt{T}\cdot\sqrt{bh}\cdot\log_2 N} \;\leq\; \frac{1}{\sqrt{2\pi}} \;\approx\; 0.399\ .$$

In other words, the regret of TRN is of order $\log_2 N\sqrt{T}$, which is worse than several existing algorithms to mix N experts for general loss functions (see for example [38]) that have regret of order $\sqrt{\log_2 N}\sqrt{T}$. However, this is unsurprising since our goal was to describe a simple extension of APX to handle more than two agents, and TRN is the simplest way to do so where the asymptotic behavior can be established immediately from the previous results in this chapter.

3.6 Conclusion

There are two main contributions of this chapter. The first is establishing polynomial time algorithms for computing v in the cases where b and h are small multiples of a common factor or when one is a multiple of the other. These show that OPT can run in polynomial time under those conditions. The second is a direct asymptotic analysis of v that distinguishes itself from prior analyses that rely on examining the limiting behavior of suboptimal algorithms. Our direct analysis gives us greater insight into the structure of $v(\cdot, t)$ as t increases, insight that allows us to formulate a heuristic for v and hence a heuristic version of OPT we call APX. Moreover, a numerical experiment indicates the decisions of APX are almost identical to those of OPT.

An interesting question is whether the analysis of v in this chapter can be extended easily to more than two experts. A positive answer would allow for a version of APX with better performance bounds than the tournament tree version we described, and would also allow a characterization of when the more-than-two-expert version OPT can be computed efficiently.

3.7 Additional Proofs

3.7.1 Proof of Proposition 3.3.1

Suppose that in period t, the expert choices are (ξ_t^1, ξ_t^2), and the manager decision is y_t. Let $\Delta^i(x)$, where $i = 1, 2$, denote the difference in cost between the manager and expert i in period t when that period's demand is x, i.e.,

$$\Delta^i(x) = \Big(b(x - y_t)^+ + h(y_t - x)^+\Big) - \Big(b(x - \xi_t^i)^+ + h(\xi_t^i - x)^+\Big).$$

To show the required result, it suffices to show that for any adversarial choice of the realized demand $d_t \in (0,1)$ one of the following holds: (i) $\Delta^i(0) \geq \Delta^i(d_t)$ for both $i = 1, 2$; or (ii) $\Delta^i(1) \geq \Delta^i(d_t)$ for both $i = 1, 2$. (The proposition follows since the choice of t was arbitrary; thus, it would also hold for $t-1, t-2, \ldots, 1$.)

Without loss of generality, suppose that $\xi_t^1 \leq \xi_t^2$. We consider the following two cases separately: $d_t < y_t$ and $d_t \geq y_t$. In the first case $d_t < y_t$, from the form of $\Delta^i(x)$, it follows that

$$\begin{aligned}
\Delta^i(0) &= (h(y_t - 0)) - \left(b(0 - \xi_t^i)^+ + h(\xi_t^i - 0)^+\right), \\
\Delta^i(d_t) &= (h(y_t - d_t)) - \left(b(d_t - \xi_t^i)^+ + h(\xi_t^i - d_t)^+\right), \text{ and} \\
\Delta^i(0) - \Delta^i(d_t) &= (hd_t) + \left(b(d_t - \xi_t^i)^+ - b(0 - \xi_t^i)^+\right) + \left(h(\xi_t^i - d_t)^+ - h(\xi_t^i - 0)^+\right) \\
&\geq 0 .
\end{aligned}$$

Thus, (i) holds in the first case. A similar analysis shows that (ii) holds in the second case $d_t \geq y_t$. □

3.7.2 Proof of Proposition 3.4.3

If $t = 0$, then the result that $f(m, 0) = \min\{m, 0\}$ follows from the definition of $v((0, m), 0)$. Suppose $t \geq 1$. From the definitions of f and v,

$$\begin{aligned}
&f(m,t) \\
&= \min_{(Z^1,Z^2)\in[0,1]^2} \frac{b \cdot v((hZ^1, m + hZ^2), t-1) \ + \ h \cdot v((b(1-Z^1), m + b(1-Z^2)), t-1)}{b+h} \\
&= \min_{(Z^1,Z^2)\in[0,1]^2} \frac{bf(m + h(Z^2 - Z^1), t-1) \ + \ bhZ^2 \ + \ hf(m + b(Z^1 - Z^2), t-1) \ + \ bh(1 - Z^1)}{b+h} \\
&= \frac{bh}{b+h} + \min_{(Z^1,Z^2)\in[0,1]^2} \frac{b \cdot f(m + h(Z^2 - Z^1), t-1) + h \cdot f(m + b(Z^1 - Z^2), t-1)}{b+h} .
\end{aligned}$$

Since $f(m,t-1)$ is a concave with respect m for any t (property C), it follows that we can reduce the feasible set of minimization from $[0,1]^2$ to its four corner points: $\{(0,0),(0,1),(1,0),(1,1)\}$.

We claim that we can further reduce the feasible set to two points: $\{(0,1),(1,0)\}$. To prove this claim, observe that the objective value inside the minimization operator is simplified to $f(m,t-1)$ if (Z^1,Z^2) is either $(0,0)$ or $(1,1)$. By the concavity of f with respect to its first argument (Property C), since m can be written as a convex combination of $m-b$ and $m+h$, it follows that

$$f(m,t-1) \;\geq\; f\left(\frac{h}{b+h}\cdot(m-b),t-1\right)+f\left(\frac{b}{b+h}\cdot(m+h),t-1\right) \;.$$

By property D,

$$\begin{aligned} &f(m,t-1)\\ &\quad\geq \frac{b}{b+h}\cdot f(m+h,t-1)+\frac{h}{b+h}\cdot f(m-b,t-1)\\ &\quad= \left.\frac{b\cdot f(m+h(Z^2-Z^1),t-1)+h\cdot f(m+b(Z^1-Z^2),t-1)}{b+h}\right|_{(Z^1,Z^2)=(0,1)} . \end{aligned}$$

Therefore, in the minimization operator above, the choice of $(0,1)$ dominates both $(0,0)$ and $(1,1)$. It follows that

$$\begin{aligned} f(m,t) &= \frac{bh}{b+h}\\ &\quad+\min_{(Z^1,Z^2)\in\{(0,1),(1,0)\}}\frac{b\cdot f(m+h(Z^2-Z^1),t-1)+h\cdot f(m+b(Z^1-Z^2),t-1)}{b+h} , \end{aligned}$$

which is equivalent to the required result. ∎

3.7.3 Proof of Theorem 3.4.4

To provide a simple proof of Theorem 3.4.4, we find it useful to introduce a sequence of functions $\{g_t : t \geq 0\}$, that have one-to-one correspondence with $\{f(\cdot, t) : t \geq 0\}$. Each g_t is symmetric around 0 and has a simpler recursive formulation. We define g_t, $t \geq 0$, as follows:

Definition 3.7.1. Let $\alpha > 0$. Define $g_0(x) = \alpha x$ for $x \leq 0$ and $g_0(x) = -\alpha x$ for $x \geq 0$. For any $t \geq 0$, recursively define

$$g_{t+1}(x) \quad = \quad \min\{hg_t(x-b) + bg_t(x+h),\ bg_t(x-h) + hg_t(x+b)\}\ . \tag{3.8}$$

In the following proposition, we prove the one-to-one correspondence between g_t and $f(\cdot, t)$.

Proposition 3.7.2. *Suppose $\alpha = 1/2$ in the definition of g_t. Then,*

$$g_t(x) \quad = \quad (b+h)^t \left[f(x,t) - t \cdot \frac{bh}{b+h} - \alpha x \right]\ . \tag{3.9}$$

Proof. It is easy to verify the statement for $t = 0$. We assume the result for $t - 1$ and prove the result for t. From Eq. (3.8),

$$g_t(x) \quad = \quad \min\{hg_{t-1}(x-b) + bg_{t-1}(x+h),\ bg_{t-1}(x-h) + hg_{t-1}(x+b)\}\ .$$

Applying the induction hypothesis to the g_{t-1} terms,

$$
\begin{aligned}
&hg_{t-1}(x-b)+bg_{t-1}(x+h)\\
&=(b+h)^{t-1}\left[hf(x-b,t-1)+bf(x+h,t-1)-(t-1)bh-h\alpha(x-b)-b\alpha(x+h)\right]\\
&=(b+h)^{t}\left[\left(\frac{hf(x-b,t-1)+bf(x+h,t-1)}{b+h}+\frac{bh}{b+h}\right)-t\cdot\frac{bh}{b+h}-\alpha x\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&bg_{t-1}(x-h)+hg_{t-1}(x+b)\\
&=(b+h)^{t-1}\left[hf(x-b,t-1)+bf(x+h,t-1)-(t-1)bh-h\alpha(x-b)-b\alpha(x+h)\right]\\
&=(b+h)^{t}\left[\left(\frac{bf(x-h,t-1)+hf(x+b,t-1)}{b+h}+\frac{bh}{b+h}\right)-t\cdot\frac{bh}{b+h}-\alpha x\right].
\end{aligned}
$$

Therefore, from the definition of $f(x,t)$ in Eq. (3.6), we obtain Eq. (3.9), which completes the induction step. □

Before proceeding further, we note the following properties of g_t. The proofs of these results are simple, and thus we omit them here.

Assertion 1. *For any t, g_t is symmetric, i.e., $g_t(x)=g_t(-x)$.*

Assertion 2. *For any t, the two arguments of the minimum operator in Eq. (3.8) of Definition 3.7.1 are the same if $x=0$, i.e., $hg_t(0-b)+bg_t(0+h)=bg_t(0-h)+hg_t(0+b)$, for any t.*

The following proposition shows some properties of g_t, and these properties are similar to those we want to show for $f(\cdot,t)$.

Proposition 3.7.3. *Let b/h be a positive integer. Then, g_t has the following properties for all $t\geq 0$.*

(i) g_t *is a piece-wise linear function, where the breakpoints belong to the set* $\{\dots, -2h, -h, 0, h, 2h, \dots\}$.

(ii) The minimization operator in Eq. (3.8) in the definition of g_{t+1} *can be solved in the following manner based on the value of* x*:*

$$g_{t+1}(x) \quad = \quad \begin{cases} bg_t(x-h) + hg_t(x+b), & \text{for } x \leq 0 \\ hg_t(x-b) + bg_t(x+h), & \text{for } x \geq 0. \end{cases}$$

Proof. The base cases of $t = 0$ and $t = 1$ can be verified easily. We assume the result for $t - 1$ and t and prove the result for $t + 1$.

We first consider the case of $x \geq h$. Then, since $x - h \geq 0$, by the induction hypothesis,

$$\begin{aligned} & bg_t(x-h) + hg_t(x+b) \\ = \; & bhg_{t-1}(x-b-h) + b^2 g_{t-1}(x) + h^2 g_{t-1}(x) + bhg_{t-1}(x+b+h) \\ = \; & h[bg_{t-1}(x-b-h) + hg_{t-1}(x)] + b[bg_{t-1}(x) + hg_{t-1}(x+b+h)] \\ \geq \; & h[g_t(x-b)] + b[g_t(x+h)] \; , \end{aligned}$$

where the last inequality follows from the definition of g_{t+1} according to Eq. (3.8). Thus, we prove (ii) for $x \geq h$. By a similar argument (or by Assertion 1 above), we prove (ii) also for $x \leq -h$.

From the above induction argument, it can be verified that (i) also follows for $x \in (-\infty, -h] \cup [h, \infty)$. Furthermore, from the fact that each of $hg_t(x - b) + bg_t(x + h)$ and $bg_t(x-h)+hg_t(x+b)$ is piecewise-line with the breakpoints in $\{\dots, -2h, -h, 0, h, 2h, \dots\}$ and that these two functions are equal at $x = 0$, we can obtain that g_{t+1} is linear in each of $[-h, 0]$ and $[0, h]$. Thus, we prove (i) for any $x \in (-\infty, \infty)$.

It remains to show (ii) for x in the interval $(-h, h)$ only. Since g_{t+1} is piecewise-linear, i.e., (i), it suffices to verify this result for $x = 0$ only. If $x = 0$, then Assertion 2 shows the required result for (ii). This completes the induction step. □

By the one-to-one correspondence of g_t and $f(\cdot, t)$, the properties of g_t just shown in Proposition 3.7.3 have their counterparts for $f(\cdot, t)$. Part (ii) of Proposition 3.7.3 and Proposition 3.7.2 together imply part (ii) of the required result (stated in Theorem 3.4.4). Furthermore, since breakpoints of a piecewise linear function are preserved under scaling and under addition of an affine term, part (i) of Proposition 3.7.3 and Proposition 3.7.2 imply that $f(\cdot, t)$ is a piece-wise linear function, where the breakpoints belong to the set $\{\ldots, -2h, -h, 0, h, 2h, \ldots\}$. However, this set is infinite in size. Thus, to prove part (i) of the required result, it remains to show that this set can be truncated on both sides to $\{-tb, -tb + h, \ldots, -h, 0, h, \ldots, tb - h, tb\}$. We accomplish this by showing that, for any fixed t, $f(m, t)$ is linear in the region $m \leq -tb$, and also linear in the region $m \geq tb$. This result, stated in Proposition 3.7.4, completes the proof of Theorem 3.4.4.

Proposition 3.7.4. *For any fixed t, $f(m, t)$ is linear in m for $m \leq -tb$ with gradient 1, and linear in m for $m \geq tb$ with gradient 0, i.e.*

$$f(m_2, t) - f(m_1, t) = \begin{cases} m_2 - m_1 & \text{for } m_1, m_2 \leq -tb \\ 0 & \text{for } m_1, m_2 \geq tb \ . \end{cases}$$

Proof. It is easy to check that the statement holds for $t = 0$. We assume the statement holds for $t-1$, i.e., $f(m, t-1)$ is linear in m with gradient 1 for $m \leq (t-1)b$ and is linear in m with gradient 0 for $m \geq (t-1)b$.

We first consider the interval $(-\infty, -tb)$. For any m_1 and m_2 satisfying $m_1 \leq m_2 \leq$

$-tb$,

$$f(m_2,t) - f(m_1,t) = \min_{\sigma\in\{1,-1\}} \left\{ \frac{bf(m_2+\sigma h, t-1) + hf(m_2 - \sigma b, t-1)}{b+h} \right\} - \min_{\sigma'\in\{1,-1\}} \left\{ \frac{bf(m_1+\sigma' h, t-1) + hf(m_1 - \sigma' b, t-1)}{b+h} \right\} .$$

Note that each of $\{m_1+h, m_1+b, m_2+h, m_2+b\}$ is bounded above by $-tb+b = -(t-1)b$. Thus, from the induction hypothesis, we may replace all the $f(x, t-1)$ terms by a linear function $x + \beta$, for some constant β. Hence, the above expression can be written as

$$f(m_2,t) - f(m_1,t) = \frac{(b+h)(m_2 - m_1 + \beta - \beta)}{b+h} = m_2 - m_1 ,$$

proving the first part of the result. The case for $m_1, m_2 \geq tb$ is similar, and we complete the induction step. □

3.7.4 Proof of Proposition 3.4.5

The required result is clearly true for $t = 0$ since both sides are $f(m, 0)$. As an induction hypothesis, we assume the result for $t-1$, and show that it is also true for $t \geq 1$.

From Proposition 3.4.3 and the definition of θ_t,

$$\begin{aligned} & f(m,t) - \frac{bh}{b+h} \\ &= \min_{\sigma\in\{1,-1\}} \left\{ \frac{b \cdot f(m+\sigma h, t-1) + h \cdot f(m - \sigma b, t-1)}{b+h} \right\} \\ &= \frac{b}{b+h} \cdot f(m + \theta_t(m)h, t-1) + \frac{h}{b+h} \cdot f(m - \theta_t(m)b, t-1) . \end{aligned} \tag{3.10}$$

We turn our attention to $\mathbb{E}[f(S_{t,t},0) \mid S_0 = m]$. Conditioning it on on $X_{t,1}$, we obtain

$$\begin{aligned}
&\mathbb{E}[f(S_{t,t},0) \mid S_0 = m] \\
&= \mathbb{P}\left(X_{t,1} = \theta_t(m)h\right) \cdot \mathbb{E}\left[f(S_{t,t},0) \mid S_{t,0} = m, S_{t,1} = m + \theta_t(m)h\right] \\
&\quad + \mathbb{P}\left(X_{t,1} = -\theta_t(m)b\right) \cdot \mathbb{E}\left[f(S_{t,t},0) \mid S_{t,0} = m, S_{t,1} = m - \theta_t(m)b\right]. \quad (3.11)
\end{aligned}$$

We simplify the above expression. First, notice that

$$\mathbb{P}\left(X_{t,1} = \theta_t(m)h\right) = \frac{b}{b+h} \qquad \text{and} \qquad \mathbb{P}\left(X_{t,1} = -\theta_t(m)b\right) = \frac{h}{b+h}.$$

Now, for any $0 \le j \le k \le t$, we have $(S_{t,k} \mid S_{t,0}, S_{t,1}, \ldots, S_{t,j}) = (S_{t,k} \mid S_{t,j})$ from the definition of $S_{t,k}$. Therefore, for any α and β, we obtain

$$\begin{aligned}
\mathbb{E}\left[f(S_{t,t},0) \mid S_{t,0} = \alpha, S_{t,1} = \beta\right] &= \mathbb{E}\left[f(S_{t,t},0) \mid S_{t,1} = \beta\right] \\
&= \mathbb{E}\left[f(S_{t-1,t-1},0) \mid S_{t-1,0} = \beta\right] \\
&= f(\beta, t-1) - (t-1) \cdot \frac{bh}{b+h}. \quad (3.12)
\end{aligned}$$

Above, the second equality follows since the definition of $S_{t,k}$ implies $(S_{t,k} \mid S_{t,j} = \alpha)$ has the same distribution as $(S_{t-j,k-j} \mid S_{t-j,0} = \alpha)$ for any α and $0 \le j \le k \le t$. The last equality follows from the induction hypothesis.

Applying Eq. (3.12) to Eq. (3.11) gives

$$\begin{aligned}
&\mathbb{E}[f(S_{t,t},0) \mid S_0 = m] \\
&= \frac{b}{b+h}\left\{f(m+\theta_t(m)h, t-1) - (t-1)\cdot\frac{bh}{b+h}\right\} \\
&\quad + \frac{h}{b+h}\left\{f(m-\theta_t(m)b, t-1) - (t-1)\cdot\frac{bh}{b+h}\right\} \\
&= \frac{b}{b+h}\cdot f(m+\theta_t(m)h, t-1) + \frac{h}{b+h}\cdot f(m-\theta_t(m)b, t-1) - (t-1)\cdot\frac{bh}{b+h}\ .
\end{aligned}$$

This equation combined with Eq. (3.10) completes the induction step. □

3.7.5 Proof of Proposition 3.4.7

We first establish some properties of the zero-mean square-integrable martingale array.

Let $A_t \Rightarrow B$ indicate convergence in distribution of some stochastic process $\{A_t\}$ to some random variable B. We state the following theorem, which is adapted from Corollary 3.2 of Hall and Heyde [36].

Theorem 3.7.5 (Martingale Central Limit)**.** *Let* $\{\hat{S}_{t,k}, \hat{\mathcal{F}}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$ *be a zero-mean square-integrable martingale array with differences* $\{\hat{X}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$*. For any* t*, let* $\hat{\sigma}^2_{t,k} = \mathbb{E}\left[\hat{X}^2_{t,k} |\ \hat{\mathcal{F}}_{t,k-1}\right] \in (0,\infty)$ *for all* k*, and* $\hat{v}_t = \sum_{k=1}^t \hat{\sigma}^2_{t,k}$*. Suppose*

(i) for any $\varepsilon > 0$*,* $\sum_{k=1}^t \mathbb{E}\left[\hat{X}^2_{t,k}\cdot \mathbf{I}\left\{\left|\hat{X}_{t,k}\right| > \varepsilon\right\} \mid \hat{\mathcal{F}}_{t,k}\right]$ *converges to* 0 *in probability, and*

(ii) $\hat{v}_t$ *converges to* $\hat{\eta}^2$ *in probability, where* $\hat{\eta}^2$ *is an almost surely finite random variable satisfying* $\mathbb{P}(\hat{\eta}^2 > 0) = 1$*.*

Then, $\hat{S}_{t,t}/\sqrt{\hat{v}_t} \Rightarrow N(0,1)$*.*

We now prove the following corollary of Theorem 3.7.5, which we shall require.

Proposition 3.7.6. *Let* $\{\hat{S}_{t,k}, \hat{\mathcal{F}}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$ *be a zero-mean square-integrable martingale array with differences* $\{\hat{X}_{t,k} \ : \ 1 \le k \le t, t \ge 1\}$*. Suppose that the conditions of Theorem 3.7.5 is satisfied. Furthermore, suppose that* $\mathbb{E}[\hat{X}_{t,k}\hat{X}_{t,j}] = 0$ *for any* $1 \le k < j \le t$*. Then, for any continuous function* $h : \mathbb{R} \to \mathbb{R}$ *satisfying* $|h(x)| \le |x|$*,*

$$\mathbb{E}\left[h\left(\hat{S}_{t,t}/\sqrt{\hat{v}_t}\right)\right] \longrightarrow \mathbb{E}\left[h\big(N(0,1)\big)\right] \ . \tag{3.13}$$

Proof. From Theorem 3.7.5,

$$\hat{S}_{t,t}/\sqrt{\hat{v}_t} \ \Rightarrow \ N(0,1).$$

Since h is continuous, by the Continuous Mapping Theorem (see, for example, the corollaries to Theorem 25.7 in Billingsley [11]),

$$h\left(\hat{S}_{t,t}/\sqrt{\hat{v}_t}\right) \ \Rightarrow \ h\big(N(0,1)\big) \ .$$

Thus, if we can prove that $\{h(\hat{S}_{t,t}/\sqrt{\hat{v}_t}) \ : \ t \ge 1\}$ is uniformly integrable, then we obtain the required result Eq. (3.13). (See, for example, Theorem 25.12 of Billingsley [11].)

For the remainder of the proof, we show the uniform integrability of $\{h(\hat{S}_{t,t}/\sqrt{\hat{v}_t}) \ : \ t \ge 1\}$. Since $|h(x)| \le |x|$ implies $h(x)^2 \le x^2$, it follows that, for any t,

$$\mathbb{E}\left[h\left(\hat{S}_{t,t}/\sqrt{\hat{v}_t}\right)^2\right] \ \le \ \mathbb{E}\left[\hat{S}_{t,t}^2/\hat{v}_t\right] \ = \ \frac{1}{\hat{v}_t} \cdot \mathbb{E}\left[\hat{S}_{t,t}^2\right] \ .$$

Recall

$$\hat{S}_{t,t}^2 \ = \ \left(\sum_{i=k}^{t} \hat{X}_{t,k}\right)^2 \ = \ 2 \sum_{1 \le k < j \le t} (\hat{X}_{t,k}\hat{X}_{t,j}) + \sum_{k=1}^{t} \hat{X}_{t,k}^2 \ .$$

Since $\mathbb{E}[\hat{X}_{t,i}\hat{X}_{t,j}] = 0$ and $\hat{v}_t = \sum_{i=1}^{t} \mathbb{E}[\hat{X}_{t,i}^2]$, we obtain that

$$\mathbb{E}\left[h\left(\hat{S}_{t,t}/\sqrt{\hat{v}_t}\right)^2\right] \;\le\; \frac{1}{\hat{v}_t}\cdot\mathbb{E}\left[\hat{S}_{t,t}^2\right] \;=\; 1\;,$$

which is a sufficient condition for the uniform integrability of $\{h(\hat{S}_{t,t}/\sqrt{\hat{v}_t}) \;:\; t \ge 1\}$. □

We now prove Proposition 3.4.7.

Proof of Part (a). Observe that

$$\begin{aligned}
\mathbb{E}[\tilde{X}_{t,k} \mid \mathcal{F}_{t,k}] &= \frac{\mathbb{E}[X_{t,k}(S_{t,k-1})]}{\sigma\sqrt{t}} = 0 \\
\mathbb{E}[\tilde{X}_{t,k}^2 \mid \mathcal{F}_{t,k}] &= \frac{\mathbb{E}[X_{t,k}(S_{t,k-1})^2]}{\sigma^2 t} = \frac{1}{t} < \infty\;.
\end{aligned}$$

Therefore, from Definition 3.4.6, $\{\tilde{S}_{t,k}, \mathcal{F}_{t,k} \;:\; 1 \le k \le t, t \ge 1\}$ is a zero-mean square-integrable martingale with differences $\{\tilde{X}_{t,k} \;:\; 1 \le k \le t, t \ge 1\}$. □

Proof of Part (b). We first claim that $\{\tilde{S}_{t,k}, \mathcal{F}_{t,k} \;:\; 1 \le k \le t, t \ge 1\}$ satisfies the conditions of Proposition 3.7.6. To prove this claim, observe that $\sum_{i=1}^{t} \mathbb{E}[\tilde{X}_{t,k}^2] = t \cdot \frac{1}{t} = 1$, and that for any $\epsilon > 0$,

$$\mathbb{E}[\tilde{X}_{t,k}^2 \cdot \mathbf{1}\{|\tilde{X}_{t,k}| > \varepsilon\}] \;\le\; \mathbb{E}[\tilde{X}_{t,k}^2] \;=\; \frac{1}{t} \longrightarrow 0 \qquad \text{as } t \to \infty.$$

Furthermore, for any $1 \le i \le j \le t < \infty$,

$$\mathbb{E}[\tilde{X}_{t,k}\tilde{X}_{t,j}] \;=\; \frac{1}{\sigma^2 t}\mathbb{E}[X_{t,k}(S_{t,k-1})X_{t,j}(S_{t,j-1})] \;=\; 0.$$

Hence, we complete the proof of the claim.

From part (a) and the above claim, we may apply the results of Proposition 3.7.6 to

$\{\tilde{S}_{t,k}\}$. Recall that $f(\cdot, 0)$ satisfies $|f(m,0)| = |\min\{m,0\}| \leq |m|$. Since $\sum_{i=1}^{t} \mathbb{E}[\tilde{X}_{t,k}^2] = t \cdot \frac{1}{t} = 1$, we obtain

$$\mathbb{E}\left[f(\tilde{S}_{t,t}, 0)\right] = \mathbb{E}\left[f\left(\frac{\tilde{S}_{t,t}}{\sqrt{\sum_{i=1}^{t} \mathbb{E}[\tilde{X}_{t,k}^2]}}, 0\right)\right] \longrightarrow \mathbb{E}\left[f\big(N(0,1), 0\big)\right],$$

which is the required result. □

Proof of Part (c). Since Proposition 3.4.5 implies $f(m,t) - t \cdot \frac{bh}{b+h} = \mathbb{E}[f(S_{t,t}, 0)|S_{t,0} = m]$, it suffices to show that

$$\lim_{t\to\infty} \frac{\mathbb{E}[f(S_{t,t}, 0)|S_{t,0} = m]}{\sigma\sqrt{t}} = \mathbb{E}\left[f\big(N(0,1), 0\big)\right] . \tag{3.14}$$

Since $f(m,0) = \min\{m,0\}$ (Proposition 3.4.3), it follows that $f(x,0) - |y| \leq f(x+y, 0) \leq f(x,0) + |y|$ for any pair of scalars x and y. Thus,

$$\begin{aligned} &\mathbb{E}[f(S_{t,t} - S_{t,0}, 0)|S_{t,0} = m] - |m| \\ &\quad \leq \mathbb{E}[f(S_{t,t}, 0)|S_{t,0} = m] \\ &\qquad \leq \mathbb{E}[f(S_{t,t} - S_{t,0}, 0)|S_{t,0} = m] + |m| . \end{aligned}$$

Furthermore, since $f(\alpha x, 0) = \alpha f(x,0)$ for any $\alpha \geq 0$,

$$\begin{aligned} \frac{\mathbb{E}[f(S_{t,t} - S_{t,0}, 0)|S_{t,0} = m]}{\sigma\sqrt{t}} &= \mathbb{E}\left[f\left(\frac{S_{t,t} - S_{t,0}}{\sigma\sqrt{t}}, 0\right)|S_{t,0} = m\right] \\ &= \mathbb{E}\left[f\left(\tilde{S}_{t,t}, 0\right)|S_{t,0} = m\right] . \end{aligned}$$

Combining these two results, we obtain

$$\begin{aligned}
\mathbb{E}\left[f\left(\tilde{S}_{t,t},0\right)|S_{t,0}=m\right]-\frac{|m|}{\sigma\sqrt{t}} \\
\leq \ \frac{\mathbb{E}[f(S_{t,t},0)|S_{t,0}=m]}{\sigma\sqrt{t}} \\
\leq \ \mathbb{E}\left[f\left(\tilde{S}_{t,t},0\right)|S_{t,0}=m\right]+\frac{|m|}{\sigma\sqrt{t}} \ .
\end{aligned}$$

Now, we take the limit of the above inequalities as $t \to \infty$. Part (b) implies $\mathbb{E}[f(\tilde{S}_{t,t},0)|S_{t,0} = m] \longrightarrow \mathbb{E}[f\big(N(0,1),0\big)]$, and also $|m|/(\sigma\sqrt{t}) \to 0$. Thus, we obtain Eq. (3.14) and complete the proof. □

3.7.6 Proof of Proposition 3.4.8

Proposition 3.4.2 implies $v(\mathbf{M},t) = f(m) + M^1$, where $m = M^2 - M^1$. Since $M^1/\sqrt{t} \to 0$ as $t \to \infty$,

$$\lim_{t\to\infty}\frac{1}{\sqrt{bh}\cdot\sqrt{t}}\cdot\left[v\left(\mathbf{M},t\right)-t\cdot\frac{bh}{b+h}\right] \ = \ \lim_{t\to\infty}\frac{f(m,t)-t\cdot\frac{bh}{b+h}}{\sigma\sqrt{t}} \ = \ -\,\mathbb{E}\left[N(0,1)^+\right] \ .$$

where the last equality follows from Proposition 3.4.7, as well as the fact that $f(m,0) = \min\{m,0\}$ (Proposition 3.4.3) and the symmetry of the uniform distribution. It is easy to verify that $\mathbb{E}\left[N(0,1)^+\right]$ is $1/\sqrt{2\pi}$, from which the required result follows. □

Chapter 4

Single-item capacity allocation

In this chapter, we focus on performance guarantees for a classic problem in revenue management: the single-resource capacity allocation problem under non-monotonic fare arrivals. We are interested in the relationship between performance (the expected revenue generated) and the amount of information about demand. In particular, we derive an explicit lower bound on the amount of demand data required so that a solution policy derived from that data achieves a specified performance with high probability.

4.1 Introduction

The single-resource capacity allocation problem is one of maximizing expected sales revenue of a limited resource whose price and demand changes with time. The literature's resource of choice has traditionally been single-leg flight tickets, but equally applicable are single-night stays in a hotel and tickets for cruises, sporting and music events. We use the context of single-leg flight tickets and talk about selling capacity (on the flight) for different fares.

The time horizon in this problem can be modeled in one of two ways, either time is continuous (where price may evolve continuously as well, see e.g. [29]) or time is segmented into contiguous intervals (where the price is static within each interval).

We focus on the latter model, where the sales horizon is divided into a finite number of stages. Capacity is sold at a different prescribed fare in each stage. The revenue manager controls the amount of capacity available for purchase in each stage by deciding the quantity remaining capacity to reserve for future sales, which is called a *protection level*. In each stage, capacity is sold to satisfy demand in each stage until demand is completely satiated or the remaining capacity is less than or equal to the protection level. Demand is assumed to be independent across the stages, and the sequence of fares across the stages can be arbitrary but is known a priori. The goal of the revenue manager is to set these protection levels in order to maximize expected revenue.

The problem we described is well-understood when demand distributions are known explicitly. Littlewood [55], Belobaba [7, 8], Curry [23], Wollmer [87], and Brumelle and McGill [15] analyzed the problem with the assumption that the fares across stages are monotonically increasing, and found that protection level policies are optimal. Robinson [70] later showed that even when fares change non-monotonically across stages, protection level policies were optimal.

What is less understood is the relationship between how much is known about the demand and the level of performance that knowledge guarantees. One common measure of information is the number of demand samples available. This measure of information makes sense when demand patterns are consistent across instances of the problem, say Friday evening flights on non-holiday weekends between Boston and New York, so samples would be demand data from ticket sales of these flights. The main results of relevance are those of Van Ryzin and McGill [83] and Kunnamkal and Topaloglu [50], which describe stochastic gradient descent algorithms and analyze how quickly the decision parameters converge to optimal ones.[1] These algorithms iteratively choose a policy, obtain a sample of the demand distributions, and update the policy according to the outcome. An issue with the analyses in these studies is that while they relate how quickly the decision parameters converge to optimal ones, they do not relate how quickly performance converges. In particular, in instances where performance is insensitive to decision parameters, a wide range of policies perform close to optimal and a well-performing policy can be found quickly, even if decision parameters converge slowly.

We attempt to address this gap in understanding using analytical approaches inspired by Levi, Roundy and Shmoys [52], who analyze multi-stage stochastic inventory problems and derive bounds on the number of data samples required to generate a provably-good solution with high probability. The dynamic program they analyze has the feature that the cost function in each stage is convex in the decision variable if the decisions made in later stages are optimal. They show that with a sufficient number of data samples, it is possible to estimate with high probability decisions that preserve this convexity while performing almost as well as under the optimal decisions. They furthermore extend their approach to stochastic dynamic programs where the cost function in each stage is convex or concave in the decision variable. However, their results cannot be directly applied to the capacity allocation problem since the expected revenue function in each stage of its dynamic program is not concave in the decision variable. We refer interested readers to Kunnamkal and Topaloglu [50] for an example that demonstrates this non-concavity.

In order to handle the non-concavity inherent in the capacity allocation problem, we develop an approach different from that of Levi, Roundy and Shmoys. Instead of attempting to estimate decisions that enforce concavity of the revenue function in the decision

[1] One important distinction between their results and ours is that they are able to work with censored demand samples, while our analysis requires demand samples be uncensored. Another distinction is that while their results assume fares are monotonically increasing across the stages, we make no such assumption.

variable in each stage of the dynamic program, we show it is sufficient to estimate decisions that almost satisfy known optimality conditions for the capacity allocation problem One interesting consequence of the approach that we take versus that of Levi, Roundy and Shmoys is that the estimated decisions in their approach must always be greater than the corresponding optimal decisions, while the estimated decisions in our approach have no such restrictions.

Our approach to this problem is as follows: we define a class of policies called ε-backwards accurate policies (for any arbitrary $\varepsilon > 0$), analyze their performance relative to an optimal policy, and show an explicit bound on number of data samples required to generated an instance from this class of policies.

ε-backwards accurate policies are approximations of optimal policies in the following sense: optimal policies can be characterized by a series of equations involving the probabilities of certain events, and ε-backwards accurate policies allow each of these equations to be off by up to ε.

Characterizing how this class of policies performs is the subject of the first half of this chapter. We show an upper bound for the maximum regret under an ε-backwards accurate policy (the absolute difference in expected revenue using it versus using an optimal policy). We achieve this by comparing a ε-backwards accurate policy to an optimal one by doing the following period-by-period: start with some remaining capacity, consider any demand realization in the current period, and the remaining capacities under both policies given the realized demand, match the capacities under both policies by adjusting the remaining capacity under the ε-backwards accurate policy, and do some cost accounting for the adjustment. The total cost accounting gives a bound on the suboptimality of the ε-backwards accurate policy. We also derive a lower bound for the expected revenue of an optimal policy by considering alternative policies that only protect a given amount of capacity until a certain period and makes all capacity available for sale thereafter. The upper bound on the absolute gap divided by the lower bound on optimal expected revenue in turn yields a upper bound for the suboptimality of ε-backwards accurate policies expressed as a ratio of the optimal performance that is distribution-free.

The second half of the chapter focuses on deriving the number of demand samples required to generate an ε-backwards accurate policy. We show that a straightforward Monte Carlo integration algorithm suffices to generate such a policy if it is provided enough demand samples. In such an algorithm, an empirical probability measure for the demand distributions is generated from the demand samples, and an optimal solution is found for the problem using this empirical measure in place of the true one. We lower bound the number of samples required for this algorithm to compute an ε-backwards accurate policy to an arbitrary confidence probability.

The structure of the chapter is as follows: we outline the capacity allocation problem in §4.2 and describe its known optimality results in §4.3. In §4.4, we define ε-backwards accurate policies and analyze their optimality gap. Finally, in §4.5, we describe a well-known Monte Carlo sampling algorithm and show when it is given a sufficient number of samples as inputs, it constructs from samples a ε-backwards accurate policy with high probability. Additionally, we bound the minimum number of samples required for a policy computed via this Monte Carlo algorithm to achieve some performance level with a desired confidence probability.

4.2 Model description

We now describe the model formally. There is x_0 capacity of a flight to be sold in $M+1$ stages, beginning from stage $M+1$ and ending after stage 1. In each stage i, the fare charged for a unit of capacity is f_i. The fares f_i, $i = 1, \ldots, M+1$, are pre-determined but need not be monotonic in i. We will refer to capacity sold in stage i (at fare f_i) as *fare class* i. In each stage i, the revenue manager decides some protection level p_{i-1} which is the amount of capacity to reserve for future stages (i.e. all stages $j < i$). The demand for capacity in each stage i is D_i (which is stochastic in nature) which, depending on the remaining capacity x and the protection level p_{i-1}, can be either fully or only partially satiated. In particular, in each stage i, $\min\{D_i, x - p_{i-1}\}$ capacity is sold for a revenue of $f_i \cdot \min\{D_i, x - p_{i-1}\}$ and the remaining capacity for future stages is $x - \min\{D_i, x - p_{i-1}\}$. We assume that each D_i has continuous support, finite mean, and cdf F_i (and ccdf $\overline{F}_i$). Moreover, the D_i's are mutually independent.

Recall that the manager's role is to decide the protection levels $p_0, p_1, \ldots, p_M$. Define $\mathbf{p}^i$ to be the vector $(p_0, p_1, \ldots, p_i)$. We will call these vectors of protection levels *protection level policies*, since for any i, $\mathbf{p}^i$ is a complete description of the decisions that need to made from stages $i+1$ through 1. We will typically define some $\mathbf{p}^M = (p_0, \ldots, p_M)$ and use $\mathbf{p}^i$ to refer to the first $i+1$ entries $(p_0, \ldots, p_i)$. (Similarly we might define $\hat{\mathbf{p}}^M$ and refer to its first $i+1$ entries by $\hat{\mathbf{p}}^i$).

For $i = 1, \ldots, M+1$, let $W_i(\mathbf{p}^{i-1}, x)$ to be the expected revenue to go from stage i onwards with x capacity remaining at the beginning of the stage and using the protection level policy $\mathbf{p}^{i-1}$. We may write W_i as

$$W_i(\mathbf{p}^{i-1}, x) = \begin{cases} \int_0^{x-p_{i-1}} \left[f_i r_i + W_{i-1}(\mathbf{p}^{i-2}, x - r_i)\right] dF_i(r_i) & \\ \quad + \overline{F}_i(x - p_{i-1}) \cdot \left(f_i \cdot (x - p_{i-1}) + W_{i-1}(\mathbf{p}^{i-2}, p_{i-1})\right), & \text{if } x > p_{i-1}, \\ W_{i-1}(\mathbf{p}^{i-2}, x), & \text{if } x \le p_{i-1}, \end{cases} \tag{4.1}$$

$$W_i(\mathbf{p}^{i-1}, 0) = 0,$$

$$W_0(\cdot, \cdot) = 0.$$

The revenue manager's objective is to select $\mathbf{p}^M$ to maximize $W_{M+1}(\mathbf{p}^M, x)$.

4.3 Characterizing optimal policies

The structure of optimal policies to the capacity allocation problem is well studied, and in this section we review the characterization due to Robinson [70]. (For a characterization of the model where the fares can only increase across stages, see Brumelle and McGill [15].) We first define some notation that will be used throughout the rest of this chapter.

Definition 4.3.1. Let $[i]$ be the fare class that has the highest fare amongst classes $1, 2, \ldots, i$, i.e. $[i] := \text{argmax}_j \{f_j : 1 \le j \le i\}$, so that $f_{[i]} = \max\{f_i, f_{i-1}, \ldots, f_1\}$. We term $[i]$ a *highest remaining fare class* and $f_{[i]}$ a *highest remaining fare*. Additionally, let J be the set of highest remaining fare classes, i.e. $J := \{j : j = [j]\}$, and J_i be the set of entries in J that are less than or equal to i, i.e. $J_i := \{j \in J : j \le i\}$.

Definition 4.3.2. For $j \le i$ and $j \in J$, $\pi_i^j(\mathbf{p}^{i-1}, x)$ is the probability of a *positive stockout* in stage j, i.e. that by the end of stage j all capacity has been sold and there is unsatisfied demand in that stage, given that at the beginning of stage i there is x capacity remaining and the protection level policy $\mathbf{p}^{i-1}$ is used.

We can express $\pi_i^j(\mathbf{p}^{i-1}, x)$ recursively. For any i and $j \in J$,

$$\pi_i^j(\mathbf{p}^{i-1}, x) = \int_0^{x-p_{i-1}} \pi_{i-1}^j(\mathbf{p}^{i-2}, x - r_i) dF_i(r_i) + \overline{F}_i(x - p_{i-1})\pi_{i-1}^j(\mathbf{p}^{i-2}, p_{i-1}), \quad \text{if } i > j \text{ and } x > p_{i-1} \tag{4.2}$$

$$\pi_i^j(\mathbf{p}^{i-1}, x) = \pi_{i-1}^j(\mathbf{p}^{i-2}, x) \quad \text{if } i > j \text{ and } x \le p_{i-1}$$

$$\pi_j^j(\mathbf{p}^{i-1}, x) = \overline{F}_j(x) \tag{4.3}$$

$$\pi_i^j(\mathbf{p}^{i-1}, x) = 0 \quad \text{if } i < j.$$

The characterization of optimal policies involves non-trivial convex combinations of different $\pi_i^j(\cdot, \cdot)$ terms, so we define the following shorthand.

Definition 4.3.3.

$$\pi_i(\mathbf{p}^{i-1}, x) := \sum_{j \in J_i} \alpha_i^j \pi_i^j(\mathbf{p}^{i-1}, x) \tag{4.4}$$

where

$$\alpha_i^j := \begin{cases} f_j / f_{[i]}, & \text{if } j = 1, \\ \left(f_j - f_{[j-1]}\right) / f_{[i]}, & \text{if } j \in J_i,\ j \neq 1. \end{cases} \tag{4.5}$$

Note that $\sum_{j \in J_i} \alpha_i^j = 1$. Moreover, both $\pi_i^j(\mathbf{p}^M, x)$ and $\pi_i(\mathbf{p}^M, x)$ are non-increasing in x.

We describe several properties of α_i^j that we will use. These properties follow in a straightforward manner from the definitions.

Proposition 4.3.4. *The coefficients α_i^j have the following properties:*

(a) If $i \notin J$, then $\alpha^j_{i-1} = \alpha^j_i$.

(b) $f_{[i-1]} \leq f_{[i]}$ and $\alpha^j_{i-1} \leq \alpha^j_i$.

(c) $f_{[i-1]}\alpha^j_{i-1} = f_{[i]}\alpha^j_i$.

We now characterize optimal policies. Consider the first-order derivatives of Eq. (4.1).

$$\frac{\partial W_i(\mathbf{p}^{i-1}, x)}{\partial p_{i-1}} = \begin{cases} \overline{F}_i(x - p_{i-1}) \cdot \left(\frac{\partial W_{i-1}}{\partial x}\Big|_{(\mathbf{p}^{i-2}, p_{i-1})} - f_i \right), & \text{if } x \geq p_{i-1}, \\ 0, & \text{if } x < p_{i-1}, \end{cases} \tag{4.6}$$

$$\frac{\partial W_i(\mathbf{p}^{i-1}, x)}{\partial x} = \begin{cases} \int_0^{x-p_{i-1}} \frac{\partial W_{i-1}}{\partial x}\Big|_{(\mathbf{p}^{i-2}, x-r_i)} dF_i(r_i) + \overline{F}_i(x - p_{i-1}) \cdot f_i, & \text{if } x \geq p_{i-1}, \\ \frac{\partial W_{i-1}(\mathbf{p}^{i-2}, x)}{\partial x}, & \text{if } x < p_{i-1}. \end{cases} \tag{4.7}$$

A necessary condition for some policy $\mathbf{p}$ to be optimal is for Eq. (4.6) to evaluate to zero for any x, for each i; in turn, Eq. (4.7) must evaluate to f_{i+1} for $x = p_i$, for each i. By induction, this requirement can simplified to the following lemma:

Lemma 4.3.5. *Let $\mathbf{p}^{*,M}$ be an optimal policy. For all $i = 1, \ldots, M+1$,*

$$\frac{\partial W_i(\mathbf{p}^{*,i-1}, x)}{\partial x} = f_{[i]}\pi_i(\mathbf{p}^{*,i-1}, x). \tag{4.8}$$

One implication of the above lemma is that under an optimal policy, the expected revenue-to-go at each stage, W_i, is concave with respect to remaining capacity, since $\pi_i(\mathbf{p}^{*,i-1}, x)$ is non-increasing in x.

Corollary 4.3.6. *Let $\mathbf{p}^{*,M}$ be an optimal policy. For all i, $W_i(\mathbf{p}^{*,i-1}, x)$ is a concave function in x.*

Moreover, applying Eq. (4.8) to Eq. (4.6) implies $W_i(\mathbf{p}^{*,i-2}, p_{i-1}, x)$ is quasiconcave in

p_{i-1}. It follows that the necessary condition is also sufficient, which gives us a characterization of an optimal policy.

Corollary 4.3.7. *Let* $\mathbf{p}^{*,M}$ *be an optimal protection level policy. Then*

$$f_i = f_{[i-1]}\pi_{i-1}(\mathbf{p}^{*,i-2}, p^*_{i-1}) \tag{4.9}$$

for all $i \notin J$*. Furthermore,* $p^*_{i-1} = 0$ *for all* $i \in J$*.*

Such a characterization of the optimal protection levels provides a recursive definition of an optimal protection level policy.

4.4 Approximately optimal policies

In this section, we define and analyze an approximately optimal class of policies we call ε-backwards accurate policies. These are policies that satisfy a relaxed version of the optimality conditions given by Eq. (4.9). To be more exact, for each $i \notin J$, the convex combinations of stockout probabilities ($\pi_i(\cdot,\cdot)$, see Definition 4.3.3) need not be exactly $f_{i+1}/f_{[i-1]}$ but may deviate by some small $\varepsilon > 0$. However, like an optimal policy, for all $i \in J$, ε-backwards accurate policies should have $p_{i-1} = 0$.

Definition 4.4.1. Let $0 < \varepsilon < \min_k(f_k/f_{[k]})$. A policy $\mathbf{p}^M$ is *ε-backwards accurate* if

$$\frac{f_i}{f_{[i-1]}} - \varepsilon \le \pi_{i-1}(\mathbf{p}^{i-2}, p_{i-1}) \le \frac{f_i}{f_{[i-1]}} + \varepsilon \tag{4.10}$$

for all $i \notin J$, and $p_{i-1} = 0$ for all $i \in J$.

Remark. We also call some partial policy $\mathbf{p}^{i-1}$ ε-backwards accurate if it is ε-backwards accurate for a smaller instance of the capacity allocation problem restricted to stages 1 through i.

The goal of our analysis is to characterize the suboptimality gap of an ε-backwards accurate policy. To do this, we derive an upper bound for the difference in expected revenue earned by an optimal policy and an ε-optimal policy (§4.4.2) as well as a lower bound for the expected revenue earned by an optimal policy (§4.4.3). Combining these two results (§4.4.4), we can obtain an upper bound on the suboptimality gap as a ratio of the optimal expected revenue:

Theorem 4.4.2. *Let $\mathbf{p}^{*,M}$ be an optimal policy and $\hat{\mathbf{p}}^M$ be an ε-backwards accurate policy. The relative gap between the performances of the two policies is*

$$\frac{W_{M+1}(\mathbf{p}^{*,M}, x_0) - W_{M+1}(\hat{\mathbf{p}}^M, x_0)}{W_{M+1}(\mathbf{p}^{*,M}, x_0)} \leq \frac{2M f_{[M]}\varepsilon}{\min_{j \notin J}\{f_j - 2f_{[j-1]}\varepsilon\}}.$$

Note that this bound is distribution-free, depending only on the fares and ε.

4.4.1 Subgradient bounds

In this section, we discuss some results concerning the subgradient of the revenue-to-go function when using ε-backwards accurate policies. These will be used in §4.4.2 and §4.4.3.

We can show that under a ε-backwards accurate policy, the subgradient of the revenue-to-go functionis the same as that under an optimal policy (see Lemma 4.3.5), but with an additional error that is homogeneous with respect to ε. This is summarized in the next Lemma, whose proof is discussed in §4.7.1.

Lemma 4.4.3. *Let $\hat{\mathbf{p}}^M$ be an ε-backwards accurate policy. For each i and $x \geq 0$,*

$$f_{[i]}\left(\pi_i(\hat{\mathbf{p}}^{i-1}, x) - \varepsilon\right) \leq \frac{\partial W_i(\hat{\mathbf{p}}^{i-1}, x)}{\partial x} \leq f_{[i]}\left(\pi_i(\hat{\mathbf{p}}^{i-1}, x) + \varepsilon\right). \tag{4.11}$$

From Lemma 4.3.5, $\partial W_i(\mathbf{p}^{*,i-1}, x)/\partial x$ is concave in x and has value $f_{[i]}\pi_i(\hat{\mathbf{p}}^{i-1}, x)$. The key implication of Lemma 4.4.3 is that we can use $\partial W_i(\mathbf{p}^{*,i-1}, x)/\partial x$ in place of $\partial W_i(\hat{\mathbf{p}}^{i-1}, x)/\partial x$ and the error is at most $f_{[i]} \cdot \varepsilon$. This in turns allows us to upper- and lower-bound the differences in expected revenue-to-go under the ε-backwards accurate policies at different remaining capacities by using the property that concave functions have a decreasing subgradient and correcting for the aforementioned error along the way, which is the subject of the next Proposition.

Proposition 4.4.4. *Let $\hat{\mathbf{p}}^M$ be an ε-backwards accurate policy. For all $i \notin J$:*

$$W_{i-1}(\hat{\mathbf{p}}^{i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a) \begin{cases} \geq (b-a)f_i - 2(b-a)f_{[i-1]}\varepsilon, & \text{if } a \leq b \leq \hat{p}_{i-1}, \\ \leq (b-a)f_i + 2(b-a)f_{[i-1]}\varepsilon, & \text{if } \hat{p}_{i-1} \leq a \leq b. \end{cases}$$

Proof. We begin by showing the first part. For all $x \leq \hat{p}_{i-1}$, Lemma 4.4.3 implies that

$$\begin{aligned} \left.\frac{\partial W_{i-1}}{\partial x}\right|_{(\hat{\mathbf{p}}^{i-2},x)} &\geq f_{[i-1]}\left(\pi_{i-1}(\hat{\mathbf{p}}^{i-2}, x) - \varepsilon\right) \geq f_{[i]}\left(\pi_{i-1}(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) - \varepsilon\right) \\ &\geq f_{[i-1]}\left(\frac{f_i}{f_{[i-1]}} - \varepsilon - \varepsilon\right) = f_i - 2f_{[i-1]}\varepsilon, \end{aligned}$$

where the second inequality follows from the non-increasing property of π_{i-1} (see the remark following Definition 4.3.3) and the third inequality follows from the definition of an ε-backwards accurate policy. Therefore, since $b \leq \hat{p}_{i-1}$ we have that

$$\begin{aligned} W_{i-1}(\hat{\mathbf{p}}^{i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a) &\geq \left(\min_{a \leq x \leq b} \left.\frac{\partial W_{i-1}}{\partial x}\right|_{(\hat{\mathbf{p}}^{i-2},x)}\right) \cdot (b-a) \\ &\geq \left(f_i - 2f_{[i-1]}\varepsilon\right) \cdot (b-a). \end{aligned}$$

We now show the second part. For all $x \geq \hat{p}_{i-1}$,

$$\begin{aligned}\left.\frac{\partial W_{i-1}}{\partial x}\right|_{(\hat{\mathbf{p}}^{i-2},x)} &\leq f_{[i-1]} \cdot (\pi_{i-1}(\hat{\mathbf{p}}^{i-2},x)+\varepsilon) \leq f_{[i-1]} \cdot \big(\pi_{i-1}(\hat{\mathbf{p}}^{i-2},\hat{p}_{i-1})+\varepsilon\big) \\ &\leq f_{[i-1]}\left(\frac{f_i}{f_{[i-1]}}+\varepsilon+\varepsilon\right) = f_i + 2f_{[i-1]}\varepsilon.\end{aligned}$$

Since $\hat{p}_{i-1} \leq a$, it follows that

$$\begin{aligned}W_{i-1}(\hat{\mathbf{p}}^{i-2},b) - W_{i-1}(\hat{\mathbf{p}}^{i-2},a) &\leq \left(\max_{a\leq x\leq b} \left.\frac{\partial W_{i-1}}{\partial x}\right|_{(\hat{\mathbf{p}}^{i-2},x)}\right)\cdot(b-a) \\ &\leq \big(f_i + 2f_{[i-1]}\varepsilon\big)\cdot(b-a),\end{aligned}$$

as required. □

4.4.2 Upper bounding the performance gap

In this section, we upper bound the difference in expected revenue between using an optimal and an ε-backwards accurate policy. We motivate our approach as follows. Imagine that we are able to implement the two policies side-by-side for the same demand. In each stage i from $M+1$ to 1, the sales quantities to fare class i under the two policies can either be the same or differ. If they are the same, we take no additional action in this stage. If they are different, we balance the remaining capacities under the two policies by "draining" or "filling" the remaining capacity under the ε-backwards accurate policy as needed to bring it to parity with the remaining capacity under the optimal policy. Cost accounting is done for this action by crediting or charging the ε-backwards accurate policy the expected revenue associated with the capacity drained or filled respectively. The suboptimality gap between the two policies in absolute terms will then be the net cost incurred via this cost accounting across all stages.

For concision and ease of exposition, we define a function $\mathcal{H}_i$ that is easily validated to be an upper bound on the difference in capacity sold in stage i by the two policies.

Definition 4.4.5. Let $\hat{\mathbf{p}}^M$ be an ε-backwards accurate policy and $\mathbf{p}^{*,M}$ be an optimal policy. Then

$$H_i(\mathbf{p}^{*,M}, \hat{\mathbf{p}}^M, x) := \begin{cases} \min\{p^*_{i-1}, x\} - \hat{p}_{i-1}, & \text{if } x > \hat{p}_{i-1} \text{ and } \hat{p}_{i-1} \leq p^*_{i-1} \\ \min\{\hat{p}_{i-1}, x\} - p^*_{i-1}, & \text{if } x > p^*_{i-1} \text{ and } \hat{p}_{i-1} > p^*_{i-1} \\ 0, & \text{if } x \leq \min\{p^*_{i-1}, \hat{p}_{i-1}\}. \end{cases}$$

Moreover, we write $\mathcal{H}_i(x)$ in place of $H_i(\mathbf{p}^{*,M}, \hat{\mathbf{p}}_{i-1}, x)$ when the context is clear, omitting the two parameters indicating the protection levels of the two policies.

Note that $H_i(\mathbf{p}^{*,M}, \hat{\mathbf{p}}^M, x)$ (or equivalently $\mathcal{H}_i(x)$) is increasing in x, a fact we will utilize. We now state and prove an upper bound on the revenue difference between using the two policies.

Theorem 4.4.6. *Let $\hat{\mathbf{p}}^M$ be an ε-backwards accurate policy and $\mathbf{p}^{*,M}$ be an optimal policy. Then for all i and $x \geq 0$,*

$$W_i(\mathbf{p}^{*,i-1}, x) - W_i(\hat{\mathbf{p}}^{i-1}, x) \leq 2\varepsilon \sum_{k=2}^{i} \mathcal{H}_k(x) f_{[k-1]}.$$

Proof. We prove this by induction. For $i = 1$, since $1 \in J$, $p^*_0 = \hat{p}_0 = 0$, the statement is trivially true. Suppose that the statement is true for $k \leq i - 1$. We show the statement is true for i.

There are two main cases to consider, and these cases are distinguished by whether both policies will sell the same amount of capacity to class i demand, or that the two policies will sell a differing amount of capacity.

Case 1. *(A. $x < \min\{p^*_{i-1}, \hat{p}_{i-1}\}$; B. $D_i < x - \max\{p^*_{i-1}, \hat{p}_{i-1}\}$ and $p^*_{i-1} \neq \hat{p}_{i-1}$; or C. $p^*_{i-1} = \hat{p}_{i-1}$.)* In this case, both policies sell the same amount of inventory to class i demand. Since both policies sell the same amount of inventory (possibly zero) in stage i, in the next stage $i-1$ both policies will begin with the same remaining capacity. It follows that no difference in revenue is accrued in stage i and the difference in expected revenue-to-go, $W_i(\mathbf{p}^{*,i-1}, x) - W_i(\hat{\mathbf{p}}^{i-1}, x)$, is that of stage $i-1$ onwards, i.e.

$$W_{i-1}(\mathbf{p}^{*,i-2}, y) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, y)$$

for some $0 \leq y \leq x$. By the induction hypothesis, this amount is less than or equal to

$$2\varepsilon \sum_{k=2}^{i-1} \mathcal{H}_k(y) f_{[k-1]}.$$

Since $\mathcal{H}_k(y) \leq \mathcal{H}_k(x)$ for all k and $\mathcal{H}_i(x) \geq 0$, we obtain the required upper bound to complete the induction step.

Case 2. *($p^*_{i-1} \neq \hat{p}_{i-1}, D_i > x - \max\{p^*_{i-1}, \hat{p}_{i-1}\}$ and $x > \min\{p^*_{i-1}, \hat{p}_{i-1}\}$.)* In this case, the two policies sell differing amounts of capacity to class i demand. We consider separately the subcases where $\hat{p}_{i-1} < p^*_{i-1}$ and $\hat{p}_{i-1} > p^*_{i-1}$.

Subcase 1. ($\hat{p}_{i-1} < p^*_{i-1}$.) Using $\hat{p}_{i-1}$ as the protection level for this stage, $\min\{(x - D_i)^+, (x - \hat{p}_{i-1})^+\}$ capacity will be sold. Alternatively, using p^*_{i-1} as the protection level for this stage, $\min\{(x - D_i)^+, (x - p^*_{i-1})^+\}$ will be sold. The difference in sales under the two differing protection levels (the former sales quantity minus the latter) is some strictly positive value between zero and $(\max\{x, p^*_{i-1}\} - \hat{p}_{i-1})$. At the beginning of stage $i-1$, the remaining capacity is $b := \min\{p^*_{i-1}, x\}$ if p^*_{i-1} were used in stage i, or $a := \max\{\hat{p}_{i-1}, x - D_i\}$ if $\hat{p}_{i-1}$ were used in stage i instead. Note that $\hat{p}_{i-1} \leq a < \min\{p^*_{i-1}, x\} = b$. Thus, using $\hat{p}_{i-1}$ instead of p^*_{i-1} for the protection level in stage i would earn $(b - a) \cdot f_i$ more in revenue but would result in $b-a$ less remaining capacity in stage $i-1$. The difference in the (expected) revenue-to-go function from stage i onwards can thus be expressed as

$$W_i(\mathbf{p}^{*,i-1}, x) - W_i(\hat{\mathbf{p}}^{i-1}, x) = W_{i-1}(\mathbf{p}^{*,i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a) - (b-a)f_i \ .$$

Subtracting and adding $W_{i-1}(\hat{\mathbf{p}}^{i-2}, b)$ to and from the first and second terms on the right-hand side respectively, the above expression can be re-written as

$$\big(W_{i-1}(\mathbf{p}^{*,i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, b)\big) + \big(W_{i-1}(\hat{\mathbf{p}}^{i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a)\big) - (b-a)f_i.$$

By Proposition 4.4.4, the second and third terms can be bounded above by $2\varepsilon(b-a)f_{[i-1]}$. From the induction hypothesis, the first bracketed term is bounded above by $2\varepsilon\sum_{k=2}^{i}\mathcal{H}_k(b)f_{[k-1]}$. Combining these results, we obtain that the above expression is bounded from above by

$$2\varepsilon\sum_{k=2}^{i}\mathcal{H}_k(b)f_{[k-1]} + 2\varepsilon(b-a)f_{[i-1]}.$$

To complete the induction step, it suffices to show that $\mathcal{H}_k(b) \leq \mathcal{H}_k(x)$ and $(b-a) \leq \mathcal{H}_i(x) := \max\{x, p^*_{i-1}\} - \hat{p}_{i-1}$. The first follows from the definition of $H_i(\cdot,\cdot,\cdot)$. The second follows from $\hat{p}_{i-1} \leq a < b = \min\{x, p^*_{i-1}\}$, completing the induction step.

Subcase 2. ($\hat{p}_{i-1} < p^*_{i-1}$.) This subcase is similar to the first subcase except that the policies are reversed. Following the same line of reasoning, we have that the difference in expected-revenue-to-go function can be characterized by

$$W_i(\mathbf{p}^{*,i-1}, x) - W_i(\hat{\mathbf{p}}^{i-1}, x) = W_{i-1}(\mathbf{p}^{*,i-2}, a) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, b) + (b-a)f_i,$$

for some value of a satisfying $p^*_{i-1} \leq a < b$, where $b := \min\{\hat{p}_{i-1}, x\}$. Subtracting and adding $W_{i-1}(\hat{\mathbf{p}}^{i-2}, a)$ from the first and second terms of the right-hand side respectively, we have

$$\big(W_{i-1}(\mathbf{p}^{*,i-2}, a) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a)\big) - \big(W_{i-1}(\hat{\mathbf{p}}^{i-2}, b) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, a)\big) + (b-a)f_i.$$

By Proposition 4.4.4, the second and third terms are upper bounded by $2\varepsilon(b-a)f_{[i-1]}$. From the induction hypothesis, we can upper bound first term by $2\varepsilon \sum_{k=2}^{i} \mathcal{H}_k(a) f_{[k-1]}$. So the above expression is at most

$$2\varepsilon \sum_{k=2}^{i-1} \mathcal{H}_k(a) f_{[k-1]} + 2\varepsilon(b-a) f_{[i-1]}.$$

To complete the induction step, it suffices to show that $\mathcal{H}_k(a) \leq \mathcal{H}_k(x)$ and $(b-a) \leq \mathcal{H}_i(x) := \min\{\hat{p}_{i-1}, x\} - p^*_{i-1}$. The former requirement follows from $H_i(\cdot,\cdot,y)$ being non-decreasing in y, and the latter requirement follows from $p^*_{i-1} \leq a < b = \min\{\hat{p}_{i-1}, x\}$, which completes the induction step. □

4.4.3 Lower bounding the optimal performance

In this section, we develop a lower bound on the expected revenue under an optimal policy, in particular one with a form that may be utilized together with Theorem 4.4.6 to give a suboptimality gap that is distribution-free bound. To find such a lower bound, we construct a sequence of alternative policies. Each policy in the sequence corresponds to each stage i, and protects an amount equivalent to the higher of p^*_{i-1} and $\hat{p}_{i-1}$ until stage $i-1$ and protects nothing thereafter. Each of these policies is feasible, so its expected revenue is a lower bound for that of an optimal policy. By extension, the maximum of these lower bounds is also a lower bound for the optimal expected revenue. Note that this bound is weak, but is necessary for our purposes.

Lemma 4.4.7. *Let* $\hat{\mathbf{p}}^M$ *be an* ε*-backwards accurate policy and* $\mathbf{p}^{*,M}$ *be an optimal policy. The expected revenue under the optimal policy,*

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(\min_{j \notin J} f_j - 2f_{[j-1]}\varepsilon \right) \left(\max_{j \notin J} \mathcal{H}_j(x) \right)$$

Proof. Let $\mathcal{R}^*(z) = \{j \in \{1, \ldots, M+1\} : j \notin J, z > p^*_{j-1}\}$ and $\hat{\mathcal{R}}(z) = \{j \in \{1, \ldots, M+1\} : j \notin J, z > \hat{p}_{j-1}\}$. Recall that $\mathcal{H}_i(x) = 0$ if $x \leq \min\{p^*_{i-1}, \hat{p}_{i-1}\}$, and that $\mathcal{H}_i(x) = 0$ for all $i \in J$ since $p^*_{i-1} = \hat{p}_{i-1} = 0$ (see Definition 4.4.5). It follows that for any i not in $\mathcal{R}^*(x) \cup \hat{\mathcal{R}}(x)$, $\mathcal{H}_i(x) = 0$. Hence, to prove the required result it is sufficient to show for any $i \in \mathcal{R}^*(x) \cup \hat{\mathcal{R}}(x)$,

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(f_i - 2f_{[i-1]}\varepsilon \right) \mathcal{H}_i(x)\ .$$

We consider three cases: when $i \in \mathcal{R}^*(x) \cap \hat{\mathcal{R}}(x)$; when $i \in \mathcal{R}^*(x) \setminus \hat{\mathcal{R}}(x)$; and when $i \in \hat{\mathcal{R}}(x) \setminus \mathcal{R}^*(x)$.

Case 1. ($i \in \mathcal{R}^*(x) \cap \hat{\mathcal{R}}(x)$.) In this case, $x > \max\{p^*_{i-1}, \hat{p}_{i-1}\}$. Recall from Corollary 4.3.6 that $W_{i-1}(\mathbf{p}^{*,i-2}, z)$ is concave in z. Thus for any $z \leq p^*_{i-1}$, it follows that

$$\left.\frac{\partial W_{i-1}}{\partial z}\right|_{(\mathbf{p}^{*,i-2}, z)} \geq \left.\frac{\partial W_{i-1}}{\partial z}\right|_{(\mathbf{p}^{*,i-2}, p^*_{i-1})} = f_{[i-1]}\pi_{i-1}(\mathbf{p}^{*,i-2}, p^*_{i-1}) = f_i$$

where the first equality is due to Lemma 4.3.5, and the last equality is due to Corollary 4.3.7. This implies that $W_{i-1}(\mathbf{p}^{*,i-2}, p^*_{i-1}) \geq f_i p^*_{i-1}$, which we will transform into an appropriate bound. Notice that $W_{i-1}(\mathbf{p}^{*,i-2}, p^*_{i-1}) = W_i(\mathbf{p}^{*,i-1}, p^*_{i-1}) \leq W_{M+1}(\mathbf{p}^{*,M}, x)$.

Moreover, $f_i \geq f_i - 2f_{[i-1]}\varepsilon$ and $p^*_{i-1} \geq (p^*_{i-1} - \hat{p}_{i-1})$. Hence,

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(f_i - 2f_{[i-1]}\varepsilon\right)\left(p^*_{i-1} - \hat{p}_{i-1}\right). \tag{4.12}$$

On the other hand, we have that

$$W_i(\hat{\mathbf{p}}^{i-1}, \hat{p}_{i-1}) \;=\; W_{i-1}(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) \;=\; W_{i-1}(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, 0).$$

Applying Proposition 4.4.4 to the right-hand side gives us that

$$\begin{aligned} &W_i(\hat{\mathbf{p}}^{i-1}, \hat{p}_{i-1}) \\ &\quad\geq\; \hat{p}_{i-1}f_i - 2\hat{p}_{i-1}f_{[i-1]}\varepsilon \;=\; \left(f_i - 2f_{[i-1]}\varepsilon\right)\hat{p}_{i-1} \;\geq\; \left(f_i - 2f_{[i-1]}\varepsilon\right)\left(\hat{p}_{i-1} - p^*_{i-1}\right). \end{aligned}$$

Furthermore, the expected revenue using the optimal policy satisfies $W_{M+1}(\mathbf{p}^{*,M}, x) \geq W_i(\mathbf{p}^{*,i-1}, x) \geq W_i(\hat{\mathbf{p}}^{i-1}, \hat{p}_{i-1})$, so it follows that

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(f_i - 2f_{[i-1]}\varepsilon\right)\left(\hat{p}_{i-1} - p^*_{i-1}\right). \tag{4.13}$$

Combining Eq. (4.12) and Eq. (4.13) completes the proof for this case, i.e.

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(f_i - 2f_{[i-1]}\varepsilon\right)\max\{p^*_{i-1} - \hat{p}_{i-1}, \hat{p}_{i-1} - p^*_{i-1}\} = \left(f_i - 2f_{[i-1]}\varepsilon\right)\mathcal{H}_i(x).$$

Case 2. ($i \in \mathcal{R}^*(x) \setminus \hat{\mathcal{R}}(x)$.) In this case, $p^*_{i-1} \leq x < \hat{p}_{i-1}$. From the definition of W_i and Proposition 4.4.4,

$$W_i(\hat{\mathbf{p}}^{i-1}, x) = W_{i-1}(\hat{\mathbf{p}}^{i-2}, x) = W_{i-1}(\hat{\mathbf{p}}^{i-2}, x) - W_{i-1}(\hat{\mathbf{p}}^{i-2}, 0) \geq xf_i - 2xf_{[i-1]}\varepsilon.$$

Since $x \geq (x - p^*_{i-1}) = \mathcal{H}_i(x)$, we obtain

$$W_i(\hat{\mathbf{p}}^{i-1}, x) \geq (f_i - 2f_{[i-1]}\varepsilon)\mathcal{H}_i(x).$$

The optimal policy's expected revenue satisfies $W_{M+1}(\mathbf{p}^{*,M}, x) \geq W_i(\mathbf{p}^{*,i-1}, x) \geq W_i(\hat{\mathbf{p}}^{i-1}, x)$, so it follows that $W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(f_i - 2f_{[i-1]}\varepsilon\right)\mathcal{H}_i(x)$.

Case 3. $(i \in \hat{\mathcal{R}}(x) \backslash \mathcal{R}^*(x).)$ Then $\hat{p}_{i-1} \leq x < p^*_{i-1}$. As similar to Case 1, the combination of using Corollary 4.3.6, Lemma 4.3.5 and Corollary 4.3.7 gives us that for any $z \leq p^*_{i-1}$,

$$\left.\frac{\partial W_{i-1}}{\partial z}\right|_{(\mathbf{p}^{*,i-2}, z)} = f_i.$$

It follows that $W_{i-1}(\mathbf{p}^{*,i-2}, x) \geq f_i x$. Note that $W_{M+1}(\mathbf{p}^{*,M}, x) \geq W_i(\mathbf{p}^{*,i-1}, x) = W_{i-1}(\mathbf{p}^{*,i-2}, x)$, $f_i \geq (f_i - 2f_{[i-1]}\varepsilon)$ and $x \geq x - \hat{p}_{i-1} = \mathcal{H}_i(x)$, so we conclude that $W_{M+1}(\mathbf{p}^{*,M}, x) \geq (f_i - 2f_{[i-1]}\varepsilon)\mathcal{H}_i(x)$ as required. $\square$

4.4.4 Proving the distribution-free suboptimality gap

We may now prove Theorem 4.4.2. We divide the upper bound for the absolute suboptimality gap (Theorem 4.4.6) by the lower bound for the optimal expected revenue (Lemma 4.4.7) to yield a bound for the suboptimality gap as a ratio of optimal expected revenue. The details are as follows:

From Theorem 4.4.6,

$$\begin{aligned} W_{M+1}(\mathbf{p}^{*,M}, x_0) - W_{M+1}(\hat{\mathbf{p}}^M, x_0) &\leq 2\varepsilon \sum_{j=2}^{M+1} \mathcal{H}_j(x) f_{[j-1]} \\ &\leq 2\varepsilon M f_{[M]} \max_{2\leq j\leq M+1} \mathcal{H}_j(x) \\ &= 2\varepsilon M f_{[M]} \max_{j\notin J} \mathcal{H}_j(x), \end{aligned}$$

where the equality follows from $\mathcal{H}_j(x) = 0$ for all $j \in J$ (see Definition 4.4.5) and $1 \in J$. Additionally, from Lemma 4.4.7,

$$W_{M+1}(\mathbf{p}^{*,M}, x) \geq \left(\min_{j\notin J} \left[f_j - 2f_{[j-1]}\varepsilon\right]\right) \left(\max_{j\notin J} \mathcal{H}_j(x)\right).$$

It follows that

$$\begin{aligned} \frac{W_{M+1}(\mathbf{p}^{*,M}, x_0) - W_{M+1}(\hat{\mathbf{p}}^M, x_0)}{W_{M+1}(\mathbf{p}^{*,M}, x_0)} &\leq \frac{2\varepsilon f_{[M]} M \max_{j\notin J} \mathcal{H}_j(x)}{\left(\min_{j\notin J} \left[f_j - 2f_{[j-1]}\varepsilon\right]\right) \left(\max_{j\notin J} \mathcal{H}_j(x)\right)} \\ &= \frac{2M f_{[M]}\varepsilon}{\min_{j\notin J} \left[f_j - 2f_{[j-1]}\varepsilon\right]}\ , \end{aligned}$$

which completes the proof.

4.5 Generating an approximately optimal policy

In this section, we show that a straightforward Monte Carlo algorithm suffices to compute an ε-backwards accurate policy with high probability, as long as it is provided sufficiently many data samples. Such an algorithm is described in Robinson [70], and is simple enough that we reproduce it in Figure 4.1.

We can show that if a sufficient number of samples is used, there is only a small

Input: N demand samples, with N_i samples for each stage $i \notin J$, where $\sum_{i \notin J} N_i = N$. Each demand sample is of the form $d^t = (d_1^t, d_2^t, \ldots, d_{M+1}^t)$ where d_k^t is the t-th sample for demand in the k-th stage.

Output: A protection level policy $\hat{\mathbf{p}}^M = (\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_M)$

1. For $i = 1$ through $M+1$ in increasing order:
2. If $i \in J$, set $\hat{p}_{i-1} \leftarrow 0$. Otherwise do:
3. Initialize $z_{i-1,j}^t \leftarrow 0$ for $t = 1$ to N_i and $j \in J_{i-1}$. For each $j \in J_{i-1}$:
4. For $n = j$ to $i-1$ in increasing order:
5. If $z_{i-1,j}^t \geq \hat{p}_{n-1}$, set $z_{i-1,j}^t \leftarrow z_{i-1,j}^t + d_n^t$.
6. Set $\hat{p}_{i-1} \leftarrow \inf\{x \geq 0 : (1/N_i) \sum_{t=1}^{N_i} \sum_{j \in J_{i-1}} \alpha_{i-1}^j I\{x \leq z_{i-1,j}^t\} \leq (f_i / f_{[i-1]})\}$

Figure 4.1: A Monte-Carlo algorithm for computing protection levels.

probability that each of the protection levels found in step 6 of Monte-Carlo algorithm is outside the acceptable interval of values that correspond to an ε-backwards accurate policy. The proof for the result uses standard concentration inequality arguments and interested readers may refer to §4.7.2.

Theorem 4.5.1. *For any $i \notin J$, $\varepsilon > 0$, and $\delta \in (0,1)$, let $\overline{N}_i(\varepsilon, \delta) := \ln(2/\delta)/(2\varepsilon^2)$. Furthermore, let a given set of protection levels $\mathbf{p}^{i-2} := (p_1, \ldots, p_{i-2})$ be ε-backwards accurate. Suppose the number of samples N_i used to compute $\hat{p}_{i-1}$ in steps 3–6 of the Monte Carlo algorithm is at least $\overline{N}_i(\varepsilon, \delta)$. The policy with $\hat{p}_{i-1}$ appended, i.e. $(p_1, \ldots, p_{i-2}, \hat{p}_{i-1})$, is ε-backwards accurate with probability at least $1 - \delta$, i.e.,*

$$P\left(\frac{f_i}{f_{[i-1]}} - \varepsilon \leq \pi_{i-1}(\mathbf{p}^{i-2}, \hat{p}_{i-1}) \leq \frac{f_i}{f_{[i-1]}} + \varepsilon\right) \geq 1 - \delta.$$

We will use its corollary, given as follows, shortly.

Corollary 4.5.2. *The algorithm outputs a protection level policy that is ε-backwards accurate with probability at least $1-\delta$, where $\varepsilon > 0$ and $\delta \in (0,1)$, if we provide at least $N_{total}(\varepsilon, \delta)$ demand samples to the algorithm, where*

$$N_{total}(\varepsilon, \delta) = M \ln(2M/\delta)/(2\varepsilon^2).$$

Proof. From Theorem 4.5.1, if we use $\ln(2M/\delta)/(2\varepsilon^2)$ samples to estimate each protection level (for a total of $M \ln(2M/\delta)/(2\varepsilon^2)$ samples used), then with probability $(1-\delta/M)^M$, we will have an ε-backwards accurate policy. The corollary then follows since

$$\left(1 - \frac{\delta}{M}\right)^M > 1 - \delta$$

where the inequality holds because for any real number $y \in (0,1)$ and strictly positive integer n, $(1-y/n)^n > 1-y$. □

An implication of the above corollary and Theorem 4.4.2 is that we can bound the number of samples needed to ensure the computed solution is an α-approximation (i.e. the solution has suboptimality gap as a ratio of optimal performance of at most α) with a given confidence probability.

Lemma 4.5.3. *Consider some $\alpha > 0$ and $\delta \in (0,1)$. Suppose we wish to construct a protection level policy under whose expected revenue is at least $(1-\alpha)$ of that of an optimal policy with probability $1-\delta$. The Monte-Carlo algorithm outputs such a policy if we use at least*

$$\frac{2f_{[M]}^2 M\left(M+\alpha\right)^2 \cdot (\ln(2M) - \ln\delta)}{\alpha^2\left(\min_{j\notin J} f_j\right)^2} = \mathcal{O}\left(\left(\frac{f_{[M]}}{\min_{j\notin J} f_j}\right)^2 \cdot \frac{M^3(\log(2M) - \log\delta)}{\alpha^2}\right)$$

demand samples.

Proof. From Theorem 4.4.2, we see that for an ε-backwards optimal policy $\hat{\mathbf{p}}^M$ to have an optimality gap less than or equal to α, we require that

$$\frac{2Mf_{[M]}\varepsilon}{\min_{j\notin J} f_j - 2f_{[j-1]}\varepsilon} \leq \alpha. \tag{4.14}$$

We claim that choosing any ε such that

$$\varepsilon \leq \frac{\alpha \min_{i\notin J} f_j}{2(M+\alpha)f_{[M]}} \tag{4.15}$$

is sufficient to ensure this. To see this this true, substituting the right-hand side of inequality Eq. (4.15) for the ε terms in Eq. (4.14), we have that

$$\begin{aligned} &\left(2Mf_{[M]} \cdot \frac{\alpha \min_{i\notin J} f_i}{2(M+\alpha)f_{[M]}}\right) \Bigg/ \left(\min_{j\notin J} f_j - 2f_{[j-1]} \cdot \frac{\alpha \min_{k\notin J} f_k}{2(M+\alpha)f_{[M]}}\right) \\ = \ &\frac{2Mf_{[M]} \cdot \alpha \min_{i\notin J} f_i}{\min_{j\notin J} 2(M+\alpha)f_j f_{[M]} - 2f_{[j-1]} \cdot \alpha \min_{k\notin J} f_k} \end{aligned}$$

where the equality follows by multiplying the numerator and denominator by $2(M+\alpha)f_{[M]}$. In the denominator term, we upper bound the $f_{[j-1]}$ term by $f_{[M]}$ to get that the above equation is less than or equal to

$$\frac{2Mf_{[M]} \cdot \alpha \min_{i\notin J} f_i}{\min_{j\notin J} 2(M+\alpha)f_j f_{[M]} - 2\alpha f_{[M]} \min_{k\notin J} f_k} = \frac{\alpha \cdot 2Mf_{[M]} \min_{i\notin J} f_i}{2Mf_{[M]} \min_{j\notin J} f_j} = \alpha$$

which proves our claim.

Choose $\varepsilon = (\alpha \min_{j\notin J} f_j)/(2(M+\alpha)f_{[M]})$. Applying this choice of ε and the required failure probability δ to Corollary 4.5.2 yields the desired result. $\square$

4.6 Conclusion

In this chapter, we analyzed a class of near-optimal solutions to the capacity allocation problem called ε-backwards accurate policies, proved a distribution-free bound on its suboptimality in terms of ε, and showed that a member of this class can be generated by a Monte Carlo algorithm given sufficiently many data samples. As a result, we were able to characterize the relationship between data availability and guaranteed performance, in the form of the minimum number of data samples required to ensure that solution generated by the Monte Carlo algorithm is an approximation of a given degree with a given confidence probability.

We note that in order to derive a sampling bound that is distribution-free, we had to find a lower bound for the optimal expected revenue in §4.4.3 that is homogeneous in the difference between the protection levels of the optimal and ε-backwards accurate policies. Since these differences can be very small, this lower bound is weak. As the sampling bounds we derive are inversely proportional to this lower bound, we may expect that in most applications, the number of samples needed to achieve a given performance level can be much less than the bound we find.

Finally, we had highly leveraged the closed form characterization of the optimal policy, so it is an interesting question if this approach may also be utilized for other stochastic dynamic programs with value functions that are non-convex (non-concave) in the decision variables but have closed form characterizations of their optimal solutions.

4.7 Additional Proofs

4.7.1 Proof of Lemma 4.4.3

Before we begin, define $\Lambda_i^j(\mathbf{p}^{i-1}, x)$ to be the probability that under using the policy $\mathbf{p}^{i-1}$, if we had an additional infinitesimal unit at the beginning of stage i, it would be sold at stage j. Or more formally,

$$\Lambda_i^j(\mathbf{p}^{i-1}, x) :=$$
$$\begin{cases} \int_0^{x-\max\{p_{i-1},p_{j-1}\}} \Lambda_{i-1}^j(\mathbf{p}^{i-2}, x - r_i) dF_i(r_i), & \text{if } i > j \text{ and } x > \max\{p_{i-1}, p_{j-1}\} \\ \Lambda_{i-1}^j(\mathbf{p}^{i-2}, x), & \text{if } i > j \text{ and } p_{j-1} < x \leq p_{i-1} \\ \bar{F}_i(x - p_{i-1}), & \text{if } i = j \text{ and } x \geq p_{i-1} \\ 0, & \text{otherwise.} \end{cases}$$

Since $\hat{\mathbf{p}}^M$ is ε-backwards accurate, it follows from the definition that for all $j \notin J$, there exists some $\varepsilon_j \in [-\varepsilon, \varepsilon]$ such that

$$\pi_{j-1}(\hat{\mathbf{p}}^{j-2}, \hat{p}_{j-1}) = (f_j / f_{[j-1]}) + \varepsilon_j; \tag{4.16}$$

and for all $j \in J$, $\hat{p}_{j-1} = 0$. We prove the following result:

$$\frac{\partial W_i(\hat{\mathbf{p}}^{i-1}, x)}{\partial x} = f_{[i]} \pi_i(\hat{\mathbf{p}}^{i-1}, x) + \sum_{j \in \mathcal{S}_i(x)} f_{[j]} \varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x), \tag{4.17}$$

where $\Lambda_i^j(\cdot)$ is as defined earlier and $\mathcal{S}_i(x) = \{j : j \leq i, j \notin J_i, x > \hat{p}_{j-1}\}$. Note that $\mathcal{S}_i(x)$ is the set of stages for which some sales could be made given that we start stage i with remaining capacity x and which are also stages that do not correspond to highest remaining fares. Since $\Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x)$ are probability measures of mutually exclusive events for different j, it follows that $\sum_{j \in \mathcal{S}_i(x)} \Lambda_i^j(\hat{\mathbf{p}^{i-1}}, x) \le 1$. Applying this together with $f_{[j]} \le f_{[i]}$ for $j \le i$ and $-\varepsilon \le \varepsilon_j \le \varepsilon$, Eq. (4.17) is sufficient to give us the lemma statement.

We prove Eq. (4.17) by induction. For $i = 1$, we have $1 \in J$ so by the definition of an ε-backwards accurate policy, $\hat{p}_0 = 0$. Thus from Eq. (4.7)

$$\frac{\partial W_1(\mathbf{p}^0, x)}{\partial x} = f_1 x \cdot dF_1(x) - f_1 x \cdot dF_1(x) + \bar{F}_1(x) f_1 = \bar{F}_1(x) f_1 = f_{[1]} \pi_1^1(x - \hat{p}_0).$$

Suppose that the proposition is true for all $j \le i - 1$. We will show the proposition is also true for i. There are three cases to consider: $x < \hat{p}_{i-1}$ and $i \notin J$ (case 1), $x \ge \hat{p}_{i-1}$ and $i \notin J$ (case 2), and $i \in J$ (case 3).

Case 1. ($i \notin J, x < \hat{p}_{i-1}$.) Then $W_i(\hat{\mathbf{p}}^{i-1}, x) = W_{i-1}(\hat{\mathbf{p}}^{i-2}, x)$ from the definition of W_i and $\Lambda_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x) = \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x)$ for all $j < i$ from the definition of Λ_i^j. Thus, from Eq. (4.7) and the induction hypothesis applied to $i - 1$,

$$\begin{aligned}
\frac{\partial W_i(\hat{\mathbf{p}}^{i-1}, x)}{\partial x} &= \frac{\partial W_{i-1}(\hat{\mathbf{p}}^{i-2}, x)}{\partial x} \\
&= f_{[i-1]} \pi_{i-1}(\hat{\mathbf{p}}^{i-2}, x) + \sum_{j \in \mathcal{S}_{i-1}(x)} f_{[j]} \varepsilon_j \Lambda_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x) \\
&= f_{[i-1]} \left(\sum_{j \in J_{i-1}} \alpha_{i-1}^j \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x) \right) + \sum_{j \in \mathcal{S}_{i-1}(x)} f_{[j]} \varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x), \qquad (4.18)
\end{aligned}$$

where the last equality follows from the definition of π_{i-1}. Since $i \notin J$, we have that $\alpha_{i-1}^j = \alpha_i^j$ from Proposition 4.3.4. Moreover, since $x \le \hat{p}_{i-1}$, we have that $\pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x) = \pi_i^j(\hat{\mathbf{p}}^{i-1}, x)$, which follows from the definition of π_i^j. Note that $[i] = [i-1]$ so $J_i = J_{i-1}$.

As a result, we may rewrite the first term of Eq. (4.18) as

$$f_{[i-1]}\left(\sum_{j\in J_{i-1}}\alpha_{i-1}^{j}\pi_{i-1}^{j}(\hat{\mathbf{p}}^{i-2},x)\right)=f_{[i]}\left(\sum_{j\in J_{i}}\alpha_{i}^{j}\pi_{i}^{j}(\hat{\mathbf{p}}^{i-1},x)\right)=f_{[i]}\pi_i(\hat{\mathbf{p}}^{i-1},x).$$

Additionally, since $x<\hat{p}_{i-1}$, $\mathcal{S}_i(x)=\{j: j\leq i, j\notin J, x>\hat{p}_{j-1}\}=\{j: j\leq i-1, j\notin J, x>\hat{p}_{j-1}\}=\mathcal{S}_{i-1}(x)$ which we apply to the second term of Eq. (4.18), thereby yielding

$$\frac{\partial W_i(\hat{\mathbf{p}}^{i-1},x)}{\partial}=f_{[i]}\pi_i(\hat{\mathbf{p}}^{i-1},x)+\sum_{j\in\mathcal{S}_i(x)}f_{[j]}\varepsilon\Lambda_i^j(\hat{\mathbf{p}}^{i-1},x),$$

completing the induction step for i.

For the remaining cases 2 and 3, we have that $x\geq\hat{p}_{i-1}$. For case 2, this is implicit in the case statement. For case 3, this is because for all $i\in J$, $\hat{p}_{i-1}=0$ from the definition of an ε-backwards accurate policy. Hence for these cases, from Eq. (4.7),

$$\begin{aligned}\frac{\partial W_i(\hat{\mathbf{p}}^{i-1},x)}{\partial x}&=\int_0^{x-\hat{p}_{i-1}}\left.\frac{\partial W_{i-1}}{\partial x}\right|_{(\hat{\mathbf{p}}^{i-2},x-r_i)}dF_i(r_i)+\bar{F}_i(x-\hat{p}_{i-1})f_i\\&=\int_0^{x-\hat{p}_{i-1}}f_{[i-1]}\pi_{i-1}(\hat{\mathbf{p}}^{i-2},x-r_i)\\&\qquad+\sum_{j\in\mathcal{S}_{i-1}(x-r_i)}f_{[j]}\varepsilon_j\Lambda_{i-1}^j(\mathbf{p}^{i-2},x-r_i)dF_i(r_i)\\&\qquad+\bar{F}_i(x-\hat{p}_{i-1})f_i\\&=\int_0^{x-\hat{p}_{i-1}}f_{[i-1]}\pi_{i-1}(\hat{\mathbf{p}}^{i-2},x-r_i)dF_i(r_i)\\&\qquad+\int_0^{x-\hat{p}_{i-1}}\sum_{j\in\mathcal{S}_{i-1}(x-r_i)}f_{[j]}\varepsilon_j\Lambda_{i-1}^j(\mathbf{p}^{i-2},x-r_i)dF_i(r_i)\\&\qquad+\bar{F}_i(x-\hat{p}_{i-1})f_i,\end{aligned}$$

where the second equality is from substituting the first-order derivative term using the induction hypothesis for $i-1$ and the third equality follows from a rearrangement of the terms. Using the property that $j \in \mathcal{S}_{i-1}(x-r_i)$ if and only if $j \in \mathcal{S}_{i-1}(x)$ and $x-r_i > \hat{p}_{j-1}$, we rewrite the second term of the last line above as

$$\begin{aligned}
&\int_0^{x-\hat{p}_{i-1}} \sum_{j\in\mathcal{S}_{i-1}(x-r_i)} f_{[j]}\varepsilon_j \Lambda_{i-1}^j(\mathbf{p}^{i-2}, x-r_i)dF_i(r_i) \\
&= \int_0^{x-\hat{p}_{i-1}} \sum_{j\in\mathcal{S}_{i-1}(x)} I\{x-r_i > \hat{p}_{j-1}\} f_{[j]}\varepsilon_j \Lambda_{i-1}^j(\mathbf{p}^{i-2}, x-r_i)dF_i(r_i) \\
&= \sum_{j\in\mathcal{S}_{i-1}(x)} f_{[j]}\varepsilon_j \int_0^{x-\max\{\hat{p}_{i-1},\hat{p}_{j-1}\}} \Lambda_{i-1}^j(\mathbf{p}^{i-2}, x-r_i)dF_i(r_i) \\
&= \sum_{j\in\mathcal{S}_{i-1}(x)} f_{[j]}\varepsilon_j \Lambda_i^j(\mathbf{p}^{i-1}, x),
\end{aligned}$$

where the last equality follows from the definition of Λ_i^j. As a result, we are able to express $\partial W_i(\hat{\mathbf{p}}^{i-1}, x)/\partial x$ as

$$\begin{aligned}
\frac{\partial W_i(\hat{\mathbf{p}}^{i-1}, x)}{\partial x} &= \int_0^{x-\hat{p}_{i-1}} f_{[i-1]}\pi_{i-1}(\hat{\mathbf{p}}^{i-2}, x-r_i)dF_i(r_i) && \text{(A1)} \\
&\quad + \sum_{j\in\mathcal{S}_{i-1}(x)} f_{[j]}\varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x) && \text{(A2)} \\
&\quad + \bar{F}_i(x-\hat{p}_{i-1})f_i. && \text{(B)}
\end{aligned}$$

Case 2. ($i \notin J, x \geq \hat{p}_{i-1}$.) If $i \notin J$, $[i] = [i-1]$, which implies that $J_{i-1} = J_i$ and $\alpha_{i-1}^j = \alpha_i^j$. Additionally, from Eq. (4.16),

$$f_i = f_{[i-1]}\pi_{i-1}(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) + f_{[i-1]}\varepsilon_i = f_{[i]} \sum_{j\in J_{i-1}} \alpha_i^j \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) + f_{[i-1]}\varepsilon_i,$$

where the last equality follows from the definition of π_{i-1} and $[i] = [i-1]$. Thus Eq. (B) satisfies

$$\bar{F}_i(x-\hat{p}_{i-1})f_i = f_{[i]}\sum_{j\in J_{i-1}} \alpha_i^j \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2},\hat{p}_{i-1})\bar{F}_i(x-\hat{p}_{i-1}) + f_{[i-1]}\varepsilon_i\bar{F}_i(x-\hat{p}_{i-1}).$$

From the definition of π_{i-1} and since $f_{[i-1]} = f_{[i]}, J_{i-1} = J_i$ and $f_{[i-1]}\alpha_{i-1}^j = f_{[i]}\alpha_i^j$ by Proposition 4.3.4, Eq. (A1) satisfies

$$\begin{aligned}
&\int_0^{x-\hat{p}_{i-1}} f_{[i-1]}\pi_{i-1}(\hat{\mathbf{p}}^{i-2}, x-r_i)dF_i(r_i)\\
&= \int_0^{x-\hat{p}_{i-1}} f_{[i-1]}\left[\sum_{j\in J_{i-1}} \alpha_{i-1}^j\pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x-r_i)\right] dF_i(r_i)\\
&= \int_0^{x-\hat{p}_{i-1}} f_{[i]}\left[\sum_{j\in J_i} \alpha_i^j\pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x-r_i)\right] dF_i(r_i)\\
&= f_{[i]}\left(\left[\sum_{j\in J_i}\alpha_i^j\int_0^{x-\hat{p}_{i-1}}\pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x-r_i)dF_i(r_i)\right]\right).
\end{aligned}$$

Now, for Eq. (A2), note that $\mathcal{S}_{i-1}(x)\cup\{i\} = \mathcal{S}_i(x)$ since $i \notin J$ and $x > \hat{p}_i$. Then

$$\begin{aligned}
\sum_{j\in\mathcal{S}_{i-1}(x)} f_{[j]}\varepsilon_j\Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x) &= -f_{[i]}\varepsilon_i\Lambda_i^i(\hat{\mathbf{p}}^{i-1}, x) + \sum_{j\in\mathcal{S}_i(x)} f_{[j]}\varepsilon_j\Lambda_i^j\\
&= -f_{[i]}\varepsilon_i\bar{F}_i(x-\hat{p}_{i-1}) + \sum_{j\in\mathcal{S}_i(x)} f_{[j]}\varepsilon_j\Lambda_i^j
\end{aligned}$$

from the definition of Λ_i^i. Now we combine the above expressions for Eq. (A1), Eq. (A2) and Eq. (B) to rewrite the expression for $\partial W_i(\hat{\mathbf{p}}^{i-1}, x)/\partial x$:

$$\begin{aligned}\frac{\partial W_i(\hat{\mathbf{p}}^{i-1}, x)}{\partial x} &= f_{[i]} \left(\sum_{j \in J_i} \alpha_i^j \left[\int_0^{x-\hat{p}_{i-1}} \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x - r_i) dF_i(r_i) + \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, \hat{p}_{i-1}) \right] \right) \\ &\quad + \sum_{j \in \mathcal{S}_i(x)} f_{[j]} \varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x) \\ &= f_{[i]} \pi_i(\hat{\mathbf{p}}^{i-1}, x) + \sum_{j \in \mathcal{S}_i(x)} f_{[j]} \varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x),\end{aligned}$$

where the last equality follows from the recursive definition of $\pi_i(\hat{\mathbf{p}}^{i-1}, x)$. This completes the induction step.

Case 3. ($i \in J$.) If $i \in J$, $\hat{p}_{i-1} = 0$. This allows us to rewrite Eq. (A1)+Eq. (B) as

$$\int_0^x f_{[i-1]} \pi_{i-1}(\hat{\mathbf{p}}^{i-2}, x - r_i) dF_i(r_i) + \bar{F}_i(x) f_i \tag{4.19}$$

$$= \int_0^x f_{[i-1]} \left[\sum_{j \in J_{i-1}} \alpha_{i-1}^j \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x - r_i) \right] dF_i(r_i) + f_{[i]} \bar{F}_i(x) \tag{4.20}$$

$$= \left(\left[\sum_{j \in J_{i-1}} f_{[i-1]} \alpha_{i-1}^j \int_0^x \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x - r_i) dF_i(r_i) \right] + f_{[i]} \bar{F}_i(x) \right). \tag{4.21}$$

Since $f_{[i-1]}\alpha_{i-1} = f_{[i]}\alpha_i$ from Proposition 4.3.4 and $\sum_{j \in J_{i-1}} \alpha_i^j + \alpha_i^i = 1$ from the definition of α_i^j in Eq. (4.3.3), it follows that the above expression can be written as

$$\begin{aligned}
&\left(\sum_{j \in J_{i-1}} f_{[i]} \alpha_i^j \left[\int_0^x \pi_{i-1}^j(\hat{\mathbf{p}}^{i-2}, x - r_i) dF_i(r_i) + \bar{F}_i(x) \right] + \alpha_i^i f_{[i]} \bar{F}_i(x) \right) \\
&= f_{[i]} \left(\sum_{j \in J_{i-1}} \alpha_i^j \pi_i^j(\hat{\mathbf{p}}^{i-1}, x) + \alpha_i^i \pi_i^i(\hat{\mathbf{p}}^{i-1}, x) \right) \\
&= f_{[i]} \left(\sum_{j \in J_i} \alpha_i^j \pi_i^j(\hat{\mathbf{p}}^{i-1}, x) \right) \\
&= f_{[i]} \pi_i(\hat{\mathbf{p}}^{i-1}, x),
\end{aligned}$$

where the first equality follows from the definition of $\pi_i^i = \bar{F}_i(x)$, and the last equality follows from the definition of π_i. Furthermore, since $i \in J$, we have that $\mathcal{S}_i(x) = \{j : j \leq i, j \notin J, x > p_{j-1}\} = \{j : j \leq i - 1, j \notin J, x > p_{j-1}\} = \mathcal{S}_{i-1}(x)$, so we may rewrite Eq. (A2) as

$$\sum_{j \in \mathcal{S}_i(x)} f_{[j]} \varepsilon_j \Lambda_i^j(\hat{\mathbf{p}}^{i-1}, x),$$

which completes the induction step for i. □

4.7.2 Proof of Theorem 4.5.1

In the proof, we will use Hoeffding's inequality, a standard concentration inequality.

Lemma 4.7.1 (Hoeffding Inequality, [40])**.** *Let $X^1, \ldots, X^N$ be i.i.d. one-dimension random variables such that $P(X^1 \in [\alpha, \beta]) = 1$ for some $\alpha < \beta$. Then, for any $\varepsilon > 0$, we* *have*

$$P\left(\frac{1}{N}\sum_{i=1}^{N} X^i - E[X^1] \leq -\varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2 N}{(\beta-\alpha)^2}\right), \text{ and}$$
$$P\left(\frac{1}{N}\sum_{i=1}^{N} X^i - E[X^1] \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2 N}{(\beta-\alpha)^2}\right).$$

Proof of Theorem 4.5.1. Since $\mathbf{p}_{i-2}$ is fixed, we omit that parameter in all references to $\pi_{i-1}(\cdot,\cdot)$ in this proof. We show that with sufficiently high probability, that protection level $\hat{p}_{i-1}$ falls within an interval $[q_1, q_2]$ that preserves ε-backwards accurateness for the policy $(\mathbf{p}^{i-2}, \hat{p}_{i-1})$. More formally, these interval end-points q_1 and q_2 are given by

$$q_1 := \inf\left\{x \geq 0 : \pi_{i-1}(x) \leq (f_i/f_{[i-1]}) + \varepsilon\right\}, \text{ and}$$
$$q_2 := \sup(\left\{x \geq 0 : \pi_{i-1}(x) \geq (f_i/f_{[i-1]}) - \varepsilon\right\} \bigcup \{0\}).$$

It is straightforward to verify that both quantities are well-defined and $q_1 \leq q_2$.

Let $S(i-1, j, x)$ be an indicator for the event that there is a positive stockout by stage j when there is x capacity remaining at the beginning of stage $i-1$ while using the protection level policy $(\hat{p}_0, \ldots, \hat{p}_{i-2})$. Let the random variable

$$G(x) := \sum_{j \in J_{i-1}} \alpha_{i-1}^j S(i-1, j, x).$$

From the definition of π_i, π_i^j and S, we see that $\pi_i(x) = \mathbb{E}[G(x)]$. Consider step 6 of the algorithm. Each $z_{i-1,j}^t$, where $j \in J_{i-1}$ and $t = 1, \ldots, N_i$, gives the largest remaining capacity at the start of $i-1$ that results in a positive stockout by stage j when the demand is that of the t-th sample. Hence $I\{x \leq z_{i-1,j}^t\}$ is a realization of $S(i-1, j, x)$ and $\sum_{j \in J_{i-1}} \alpha_{i-1}^j I\{x \leq z_{i-1,j}^t\}$ is a realization of the random variable $G(x)$, corresponding to t-th demand sample. It follows that

$$\hat{\pi}_{i-1}(x) := (1/N_i) \sum_{t=1}^{N_i} \sum_{j \in J_{i-1}} \alpha_{i-1}^{j} I\{x \leq z_{i-1,j}^{t}\}$$

is the function $\pi_{i-1}(x)$ evaluated under the empirical probability measure induced by the N_i demand samples d_i^t, $t = 1, \ldots, N_i$. Moreover,

$$\hat{p}_{i-1} = \inf \left\{ x \geq 0 : \hat{\pi}_{i-1}(x) \leq \frac{f_i}{f_{[i-1]}} \right\},$$

and $\hat{\pi}_{i-1}(\cdot)$ and $\pi_{i-1}(\cdot)$ satisfy the conditions Hoeffding's Inequality.

Before proceeding, note that the complement of the event $[q_1 \leq \hat{p}_{i-1} \leq q_2]$ is $[q_1 > \hat{p}_{i-1}] \bigcup [q_2 < \hat{p}_{i-1}]$. We show that the probability of the event $[q_1 > \hat{p}_{i-1}] \bigcup [q_2 < \hat{p}_{i-1}]$ is upper bounded by δ, by showing this event is contained within a larger one, i.e.

$$\left[\left[\hat{\pi}_{i-1}(q_1) \leq f_i/f_{[i-1]} \right] \bigcup \left[\hat{\pi}_{i-1}(q_2) \geq f_i/f_{[i-1]} \right] \right],$$

and that this larger event occurs with probability at most δ.

Since $\hat{\pi}_{i-1}$ is a decreasing function, we have that $q_1 > \hat{p}_{i-1}$ implies $\hat{\pi}_{i-1}(q_1) \leq \hat{\pi}_{i-1}(\hat{p}_{i-1}) \leq f_i/f_{[i-1]}$. Therefore, the event $\left[\hat{\pi}_{i-1}(q_1) \leq f_i/f_{[i-1]}\right]$ must include the event $[q_1 > \hat{p}_{i-1}]$. Now, let $q_2 < \hat{p}_{i-1}$. Since $\hat{p}_{i-i}$ is the smallest number such that any other number x strictly smaller than it must satisfy $\hat{\pi}_{i-1}(x) > f_i/f_{[i-1]}$, it follows that $q_2 < \hat{p}_{i-1}$ implies $\hat{\pi}_{i-1}(q_2) > f_i/f_{[i-1]}$ which in turn implies $\hat{\pi}_{i-1}(q_2) \geq f_i/f_{[i-1]}$. Therefore the event $\left[\hat{\pi}_{i-1}(q_2) \geq f_i/f_{[i-1]}\right]$ must include the event $[q_2 < \hat{p}_{i-1}]$. Thus, we can conclude that $\left[\left[\hat{\pi}_{i-1}(q_1) \leq f_i/f_{[i-1]}\right] \bigcup \left[\hat{\pi}_{i-1}(q_2) \geq f_i/f_{[i-1]}\right]\right]$ contains the event $[q_1 > \hat{p}_{i-1}] \bigcup [q_2 < \hat{p}_{i-1}]$. It remains to show the larger event occurs with probability at most δ.

As $\hat{\pi}_{i-1}(y) \in [0, 1]$ for any y, applying the lower bound and upper bound portion of

Hoeffding's Inequality to the random variables $G(q_1)$ and $G(q_2)$ respectively, we have that

$$P\left(\hat{\pi}_{i-1}(q_1) - \pi_{i-1}(q_1) \leq -\varepsilon\right) \leq \exp\left(-2\varepsilon^2 N\right) \leq \frac{\delta}{2}, \text{ and}$$
$$P\left(\hat{\pi}_{i-1}(q_2) - \pi_{i-1}(q_2) \geq \varepsilon\right) \leq \exp\left(-2\varepsilon^2 N\right) \leq \frac{\delta}{2}.$$

For any one-dimensional real random variable Y, the event $\left[Y \leq f_i/f_{[i-1]}\right]$ implies $Y - \pi_{i-1}(q_1) \leq f_i/f_{[i-1]} - \pi_{i-1}(q_1) \leq -\varepsilon$, where the second inequality follows from the definition of q_1. So

$$\begin{aligned} P\left(\pi_{i-1}(q_1) \leq \frac{f_i}{f_{[i-1]}}\right) &\leq P\left(\hat{\pi}_{i-1}(q_1) \leq \pi_{i-1}(q_1) - \varepsilon\right) \\ &= P\left(\hat{\pi}_{i-1}(q_1) - \pi_{i-1}(q_1) \leq -\varepsilon\right) \leq \frac{\delta}{2}. \end{aligned}$$

Similarly,

$$P\left(\hat{\pi}_{i-1}(q_2) \geq \frac{f_i}{f_{[i-1]}}\right) \leq P\left(\hat{\pi}_{i-1}(q_2) - \pi_{i-1}(q_2) \geq \varepsilon\right) \leq \frac{\delta}{2}.$$

Thus,

$$\begin{aligned} P\left(q_1 \leq \hat{p}_{i-1} \leq q_2\right) &= 1 - P\left([q_1 > \hat{p}_{i-1}] \bigcup [q_1 < \hat{p}_{i-1}]\right) \\ &\geq 1 - P\left(\left[\hat{\pi}_{i-1}(q_1) \leq \frac{f_i}{f_{[i-1]}}\right] \bigcup \left[\hat{\pi}_{i-1}(q_2) \geq \frac{f_i}{f_{[i-1]}}\right]\right) \\ &\geq 1 - P\left(\hat{\pi}_{i-1}(q_1) \leq \frac{f_i}{f_{[i-1]}}\right) - P\left(\hat{\pi}_{i-1}(q_2) \geq \frac{f_i}{f_{[i-1]}}\right) \\ &\geq 1 - \delta. \end{aligned}$$

□

Bibliography

[1] A. Abdulkadiroğlu, P. Pathak, and A. E. Roth. The New York City high school match. *American Economic Review, Papers and Proceedings*, 95(2):364–367, 2005.

[2] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.

[3] A. Abdulkadiroğlu and T. Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88:233–260, October 1999.

[4] K. Arrow, T. Harris, and J. Marshack. Optimal inventory policy. *Econometrica*, 19:250–272, 1951.

[5] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 2002.

[6] K. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31(9):1150–1160, 1985.

[7] P. P. Belobaba. Airline yield management: An overview of seat inventory control. *Transportation Science*, 21:63–73, 1987.

[8] P. P. Belobaba. Application of a probabilistic decision model to airline seat inventory control. *Operations Research*, 37:183–197, 1989.

[9] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[10] D. Bienstock and N. Özbay. Computing robust basestock levels. *Discrete Optimization*, 5(2):389 – 414, 2008. In Memory of George B. Dantzig.

[11] P. Billingsley. *Probability and Measure*. Wiley, New York, NY, 1995. Third Edition.

[12] C. G. Bird. Group incentive compatibility in a market with indivisible goods. *Economic Letters*, 14(4):309–313, 1984.

[13] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[14] A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100:295–328, 2001.

[15] S. L. Brumelle and J. I. McGill. Airline seat allocation with multiple nested fare classes. *Operations Research*, 41(1):127–137, 1993.

[16] G. Carroll. A general equivalence theorem for allocation of indivisible objects. Working Paper.

[17] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Hembold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the Association for Computing Machinery*, 44(3):427–485, 1997.

[18] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, Cambridge, MA, 2006.

[19] L. Chen and E. Plambeck. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing and Service Operations Management*, 10(2):236–256, 2008.

[20] Y. Chen and T. Sönmez. Improving efficiency of on-campus housing. *American Economic Review*, 92(5):1669–1686, 2002.

[21] A. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6:475–490, 1960.

[22] S. A. Conrad. Sales data and estimation of demand. *Operations Research Quarterly*, 27(1):123–127, 1976.

[23] R.E. Curry. Optimal airline seat allocation with multiple nested fare classes. *Transportation Science*, 24:193–204, 1990.

[24] X. Ding, M. L. Puterman, and A. Bisi. The censored newsvendor and the optimal acquisition of information. *Operations Research*, 50(3):517–527, 2002.

[25] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. The inventory problem. *Econometrica*, 20:187–222, 1952.

[26] F. Edgeworth. The mathematical theory of banking. *Journal of the Royal Statistical Society*, 51:113–127, 1888.

[27] R. Ehrhart. (s,S) policies for a dynamic inventory model with stochastic lead times. *Operations Research*, 32:121–132, 1984.

[28] Ö. Ekici. Fair and efficient discrete resource allocation: A market approach. Job Market Paper.

[29] Y. Feng and G. Gallego. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40:999–1020, 1994.

[30] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[31] G. Gallego, J. Ryan, and D. Simchi-Levi. Minimax analysis for finite horizon inventory models. *IIE Transactions*, 33:861–874, 2001.

[32] G. Gallego, J. Ryan, and D. Simchi-Levi. Minimax analysis for finite-horizon inventory models. *IIE Transactions*, 33:861–874, 2001.

[33] G. Godfrey and W. B. Powell. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with application to inventory and distribution problems. *Management Science*, 47(8):1101–1112, 2001.

[34] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research*, Supplement to 46:S35–S49, 1998.

[35] S. C. Graves, H. C. Meal, S. Dasu, and Y. Qiu. Two-stage production planning in a dynamic environment. In S. Axsäter, C. Shneeweiss, and E. Silver, editors, *Multi-Stage Production Planning and Control: Lecture Notes in Economics and Mathematical Systems*, volume 266, pages 9–43. Springer-Verlag, Berlin, 1986.

[36] P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Applications.* Academic Press, New York, NY, 1980.

[37] J. Hannan. Approximation to Bayes risk in repeated plays. In M. Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, pages 97–139. Princeton University Press, Princeton, NJ, 1957.

[38] D. Haussler, J. Kivinen, and M. K. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44(5):1906–1925, September 1998.

[39] D. C. Heath and P. L. Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, 26(3):17–30, 1994.

[40] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[41] W. T. Huh, R. Levi, P. Rusmevichientong, and J. Orlin. Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research*, To appear.

[42] M. Hutter and J. Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6:639–660, Apr 2005.

[43] A. Hylland and R. J. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political Economy*, 87(2):293–314, April 1979.

[44] D. Iglehart. Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. *Management Science*, 9(2):259–267, 1963.

[45] D. L. Iglehart. The dynamic inventory problem with unknown demand distribution. *Management Science*, 10(3):429–440, 1964.

[46] G. D. Johnson and H. E. Thompson. Optimality of myopic inventory policies for certain dependent demand processes. *Management Science*, 21:1303–1307, 1975.

[47] A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[48] S. Karlin. Dynamic inventory policy with varying stochastic demands. *Management Science*, 6(3):231–258, 1960.

[49] D. E. Knuth. An exact analysis of stable allocation. *Journal of Algorithms*, 20(2):431–442, 1996.

[50] S. Kunnumkal and H. Topaloglu. A stochastic approximation method for the single-leg revenue management problem with discrete demand distributions. *Mathematical Methods of Operations Research*, 70:477–504, 2009.

[51] M. A. Lariviere and E. L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45(3):346–363, 1999.

[52] R. Levi, R. Roundy, and D. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32(4):821–839, 2007.

[53] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[54] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[55] K. Littlewood. Forecasting and control of passenger bookings. In *Proceedings of the 12th AGIFORS Symposium*, pages 95–117, 1972.

[56] W. S. Lovejoy. Myopic policies for some inventory models with uncertain demand distributions. *Management Science*, 36(6):724–738, 1990.

[57] W. S. Lovejoy. Stopped myopic policies in some inventory models with generalized demand processes. *Management Science*, 38:688–707, 1992.

[58] X. Lu, J.-S. Song, and K. Zhu. Inventory control with unobservable lost sales and bayesian updates. Working paper, 2006.

[59] X. Lu, J.-S. Song, and K. Zhu. Analysis of perishable-inventory systems with censored data. *Operations Research*, 58(4):1034–1038, 2008.

[60] I. Moon and G. Gallego. The distribution-free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, 44(8):825–834, 1993.

[61] I. Moon and G. Gallego. Distribution free procedures for some inventory models. *Journal of the Operational Research Society*, 45(6):651–658, 1994.

[62] G. R. Murray and E. A. Silver. A Bayesian analysis of the style goods inventory problem. *Management Science*, 12(11):785–797, 1966.

[63] S. T. O'Neil and A. Chaudhary. Comparing online learning algorithms to stochastic approaches for the multi-period newsvendor problem. In *Proceedings of the 9th Workshop on Algorithm Engineering and Experiments (ALENEX)*, 2008.

[64] S. Pápai. Strategy-proof assignment by hierachical exchange. *Econometrica*, 68:1403–1433, 2000.

[65] P. Pathak. The mechanism design approach to student assignment. *Annual Reviews of Economics*, 3, Forthcoming.

[66] P. Pathak and J. Sethuraman. Lotteries in student assignment: An equivalence result. *Theoretical Economics*, 6:1–17, 2011.

[67] G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

[68] E. L. Porteus. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA, 2002.

[69] H. Richter. The differential revenue method to determine optimal seat allotments by fare type. *AGIFORS Symposium Proceedings*, 22, 1982. Lagonissi, Greece.

[70] L. W. Robinson. Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes. *Operations Research*, 43(2):252–263, 1995.

[71] A. Roth, T. Sönmez, and M. U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119(2):457–488, May 2004.

[72] A. E. Roth. Incentive compatibility in a market with indivisible goods. *Economic Letters*, 9:127–132, 1982.

[73] A. E. Roth and A. Postlewaite. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4:131–137, 1977.

[74] H. Scarf. A min-max solution to an inventory problem. In K. J. Arrow, S. Karlin, and Scarf H., editors, *Studies in Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.

[75] H. Scarf. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics*, 30(2):490–508, 1959.

[76] H. Scarf. The optimality of (s,S) policies in dynamic inventory problems. In K. Arrow, S. Karlin, and P. Suppes, editors, *Mathematical Models in the Social Sciences*. Stanford University Press, Stanford, 1960.

[77] H. Scarf. Some remarks on Bayes solutions to the inventory problem. *Naval Research Logistics Quarterly*, 7(4):591–596, 1960.

[78] L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–28, 1974.

[79] T. Sönmez and M. U. Ünver. House allocation with existing tenants: An equivalence. *Games and Economic Behavior*, 52(1):153–185, 2005.

[80] T. Sönmez and M. U. Ünver. Matching, allocation, and exchange of discrete resources. In J. Benhabib, A. Bisin, and M. Jackson, editors, *Handbook of Social Economics*, volume 1A, pages 781–852. North-Holland, The Netherlands, 2011.

[81] L-G. Svensson and B. Larsson. Strategy-proofness, core, and sequential trade. *Review of Economic Design*, 9(2):167–190, 2005.

[82] K. T. Talluri and G. Van Ryzin. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, 2004.

[83] G. Van Ryzin and J. I. McGill. Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science*, 46(6):760–775, 2000.

[84] A. Veinott. Optimal policy for a multi-product, dynamic, nonstationary inventory problem. *Management Science*, 12:206–222, 1965.

[85] A. Veinott. On the optimality of (s,S) inventory policies: New conditions and a new proof. *SIAM Journal on Applied Mathematics*, 14:1067–1083, 1966.

[86] V. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 372–383. Morgan Kaufmann, Rochester, NY, 1990.

[87] R. D. Wollmer. An airline seat management model for a single leg route when lower fare classes book first. *Operations Research*, 40:831–844, 1992.

[88] R. Yaroshinsky, R. El-Yaniv, and S. Seidan. How to better use expert advice. *Machine Learning*, 55(3):271–309, 2004.

[89] L. Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52:123–135, 1990.

[90] P. H. Zipkin. *Foundations of Inventory Management.* McGraw-Hill, New York, NY, 2000.