# Essays on Online Learning and Resource Allocation 

Steven Yin

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

## COLUMBIA UNIVERSITY

© 2022

Steven Yin

All Rights Reserved


#### Abstract

Essays on Online Learning and Resource Allocation Steven Yin

This thesis studies four independent resource allocation problems with different assumptions on information available to the central planner, and strategic considerations of the agents present in the system.

We start off with an online, non-strategic agents setting in Chapter 1, where we study the dynamic pricing and learning problem under the Bass demand model. The main objective in the field of dynamic pricing and learning is to study how a seller can maximize revenue by adjusting price over time based on sequentially realized demand. Unlike most existing literature on dynamic pricing and learning, where the price only affects the demand in the current period, under the Bass model, price also influences the future evolution of demand. Finding a revenue-maximizing dynamic pricing policy in this model is non-trivial even in the full information case, where model parameters are known. We consider the more challenging incomplete information problem where dynamic pricing is applied in conjunction with learning the unknown model parameters, with the objective of optimizing the cumulative revenues over a given selling horizon of length $T$. Our main contribution is an algorithm that satisfies a high probability regret guarantee of order $\mathrm{m}^{2 / 3}$; where the market size $m$ is known a priori. Moreover, we show that no algorithm can incur smaller order of loss by deriving a matching lower bound.

We then switch our attention to a single round, strategic agents setting in Chapter 2, where we study a multi-resource allocation problem with heterogeneous demands and Leontief utilities.


Leontief utility function captures the idea that for certain resource allocation settings, the utility of marginal increase in one resource depends on the availabilities of other resources. We generalize the existing literature on this model formulation to incorporate more constraints faced in real applications, which in turn requires new algorithm design and analysis techniques. The main contribution of this chapter is an allocation algorithm that satisfies Pareto optimality, envy-freenss, strategy-proofness, and a notion of sharing incentive.

In Chapter 3, we study a single round, non-strategic agent setting, where the central planner tries to allocate a pool of items to a set of agents who each has to receive a prespecified fraction of all items. Additionally, we want to ensure fairness by controlling the amount of envy that agents have with the final allocations. We make the observation that this resource allocation setting can be formulated as an Optimal Transport problem, and that the solution structure displays a surprisingly simple structure. Using this insight, we are able to design an allocation algorithm that achieves the optimal trade-off between efficiency and envy.

Finally, in Chapter 4 we study an online, strategic agent setting, where similar to the previous chapter, the central planner needs to allocate a pool of items to a set of agents who each has to receive a prespecified fraction of all items. Unlike in the previous chapter, the central planner has no a priori information on the distribution of items. Instead, the central planner needs to implicitly learn these distributions from the observed values in order to pick a good allocation policy. Additionally, an added challenge here is that the agents are strategic with incentives to misreport their valuations in order to receive better allocations. This sets our work apart both from the online auction mechanism design settings which typically assume known valuation distributions and/or involve payments, and from the online learning settings that do not consider strategic agents. To that end, our main contribution is an online learning based allocation mechanism that is approximately Bayesian incentive compatible, and when all agents are truthful, guarantees a sublinear regret for individual agents' utility compared to that under the optimal offline allocation policy.

## Table of Contents

Acknowledgments ..... ix
Dedication ..... x
Introduction ..... 1
Chapter 1: Dynamic Pricing and Learning under the Bass Model ..... 6
1.1 Background and Motivation ..... 6
1.2 Main contributions ..... 9
1.3 Literature Review ..... 10
1.4 The Bass model and problem formulation ..... 13
1.5 Algorithm description ..... 17
1.5.1 Algorithm outline ..... 17
1.5.2 Estimation and price computation details ..... 18
1.6 Regret upper bound ..... 20
1.7 Regret lower bound ..... 24
1.8 Conclusion and Future Directions ..... 29
Chapter 2: Dominant Resource Fairness with Meta-Types ..... 31
2.1 Background and Motivation ..... 31
2.2 Literature Review ..... 32
2.3 Problem Formulation ..... 34
2.3.1 Desirable Properties ..... 37
2.4 Dominant Resource Fairness with Meta-Types ..... 38
2.4.1 Integral Allocation from Rounding ..... 43
2.4.2 Connection to Previous Dominant Resource Fairness Algorithms ..... 44
2.4.3 Alternative Fair Allocation Mechanisms ..... 44
2.4.4 Extension to Arbitrary Group Structure ..... 45
2.5 Numerical Experiments ..... 45
Chapter 3: Optimal Efficiency Envy Trade-Off via Optimal Transport ..... 47
3.1 Background and Motivation ..... 47
3.2 Review of Optimal Transport ..... 49
3.3 Literature Review ..... 50
3.4 Problem Formulation ..... 52
3.5 Optimal Solution Structure ..... 53
3.6 Stochastic Optimization ..... 55
3.7 Learning from Samples ..... 56
3.8 Experiments ..... 61
3.9 Conclusion and Future Directions ..... 62
Chapter 4: Online Allocation and Learning in the Presence of Strategic Agents ..... 64
4.1 Background and Motivation ..... 64
4.2 Literature Review ..... 67
4.3 Problem formulation ..... 68
4.3.1 The offline problem ..... 68
4.3.2 The online problem: approximate Bayesian incentive compatibility and regret ..... 70
4.4 Algorithm and main results ..... 72
4.5 Proof ideas ..... 74
4.6 Proof details ..... 77
4.6.1 Individual Regret Bound (Theorem 6) ..... 77
4.6.2 Approximate-Bayesian Incentive Compatibility (Theorem 5) ..... 78
4.7 Conclusion and Future Directions ..... 80
References ..... 82
Appendix A: Appendices for Chapter1 ..... 89
A. 1 A Concentration Result on Exponential Random Variables ..... 89
A. 2 Some Properties of the Deterministic Optimal Price Curve ..... 90
A.2.1 Optimal pricing policy expression ..... 90
A.2.2 Price Lipschitz Bound ..... 92
A. 3 Proof of Lemma 1 and Lemma 2 ..... 95
A. 4 Upper Bound Proofs ..... 102
A.4.1 Step 1: Bounding the estimation errors (Proof of Lemma 3, Lemma 4) ..... 102
A.4.2 Step 2: Proof of Lemma 5 ..... 106
A.4.3 Step 3: Proof of Lemma 6 and Lemma 7 ..... 109
A.4.4 Step 4: Putting it all together for proof of Theorem 1 ..... 111
A. 5 Lower Bound Proofs ..... 111
A.5.1 Step 1: missing lemmas and proofs ..... 111
A.5.2 Step 2: missing lemmas and proofs ..... 117
A.5.3 Step 3: Lipschitz bound on the optimal pricing policy ..... 121
A.5.4 Step 4: Putting it all together for proof of Theorem 2 ..... 122
A. 6 Auxiliary Results ..... 124
Appendix B: Appendices for Chapter2 ..... 126
B. 1 Missing Proofs of Results ..... 126
B.1.1 Proof of Equivalence in Definition 2 ..... 126
B.1.2 Proof of Claim 2 ..... 126
B.1.3 Proof of Lemma 9 and Fact 1 ..... 127
B.1.4 Proof of Lemma 10 ..... 128
B.1.5 Proof of Lemma 11 ..... 129
B.1.6 Proof of Lemma 12 ..... 132
B. 2 Experimental Setup and Additional Experiments ..... 134
B. 3 Beyond Sharing Incentive: Proportionality ..... 137
B. 4 Alternative Design of DRF-MT ..... 141
Appendix C: Appendices for Chapter3 ..... 143
C. 1 Proof of Theorem 3 ..... 143
Appendix D: Appendices for Chapter4 ..... 145
D. 1 Concentration Results ..... 145
D.1.1 Proof of Lemma 14 ..... 145
D. 2 Proof of Theorem 6 ..... 146
D.2.1 Proof of Lemma 15 ..... 146
D.2.2 Proof of Theorem 6 ..... 153
D. 3 Proof of Theorem 5 ..... 156
D.3.1 Proof of Lemma 16 ..... 156
D.3.2 Proof of Lemma 17 ..... 157
D.3.3 Proof of Lemma 18 ..... 161
D.3.4 Proof of Lemma 19 ..... 163
D. 4 Auxiliary Proofs ..... 166
D.4.1 Proof of Claim 8 ..... 166

## List of Figures

1.1 Left: Bass model's daily sales predictions over time have the same bell shape as the realized daily sales, whereas the other i.i.d. models do not. Right: The Bass model is able to capture the S -shaped adoption curve while the other models are
not. The faint green line in the background is the price history.
2.1 All three hospitals can accept both types of doctors. However, hospitals I and II can only accept Nurse type C, while hospital III accepts only Nurse type D.36
2.2 Left: Running time comparison. Middle: Normalized max envy comparison. Right: Distribution of normalized difference in social welfare between Discrete MNW and DRF-MT over all trials. Normalized difference is calculated by subtracting the social welfare of Discrete MNW from that of (rounded) DRF-MT and then dividing by the social welfare of Discrete MNW.44
3.1 Left: An illustration of what Laguerre cells look like when $n=2$. Consider any distribution on the support $[0,1]^{2}$. The optimal division of the space is to move the diagonal line up or down until the probability mass contained in the orange region is equal to $p^{*}$. Right: A pictorial proof of the optimality of such partition. Suppose one can find an $\epsilon$ mass above this line that is matched to $A$, and an $\epsilon$ mass below the line that is matched to $B$, then switching the assignments of these two regions increases the matched weights.54
3.2 Allocation policy for artificial data under different envy constraints. From left to right: $\epsilon=0.2,0.1,0.0$. When the envy constraint is loose (large $\epsilon$ ), $B$ envies $A$, since both agents prefer the items on the top right, but most of them are allocated to $A$. As the envy constraint tightens, the allocation boundary tilts in the direction that makes the allocations more even between the two agents.
3.3 The trade-off curve between envy and welfare for both data-sets. The shaded region is between 25th and 75th percentile of the trials. The non-monotonicity in the plot for the simulator data is due to the stochasticity in the SGD algorithm.60
3.4 Approximation gap with respect to sample size. Both $x$ and $y$ axis are in $\log$ scale. The solid line is the median and the shaded region is between the 25 th and 75 th percentile. The dashed lines show what the theoretical $1 / \sqrt{m}$ rate would look like.
B. 1 Running times comparison with meta-types: $\Omega_{1}=\{0\}, \Omega_{2}=\{1,2\}, \Omega_{3}=\{3,4,5\}, \Omega_{4}=$ $\{6,7,8,9\}, \Omega_{5}=\{10,11,12,13,14\} \ldots . . . . . . . . . . . . . . . . . . .$.
B. 2 Running time comparison with respect to number of resources . . . . . . . . . . . 135
B. 3 Running time comparison between MNW and DRF-MT. The meta types are $\Omega_{1}=$ $\{0\}, \Omega_{2}=\{1,2\}, \Omega_{3}=\{3,4,5\}, \Omega_{4}=\{6,7,8,9\} \ldots \ldots . . . . . . . . .$.
B. 4 Running time comparison between MNW and DRF-MT with respect to number of resources. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 136
B. 5 Normalized difference in social welfare between Discrete MNW and DRF-MT, grouped by number of agents.137
B. 6 Normalized difference in social welfare between Discrete MNW and MNW over all trials. Normalized difference is calculated by subtracting the social welfare of Discrete MNW from that of DRF-MT and then dividing by the social welfare of Discrete MNW.

## List of Tables

1 The four chapters of this thesis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
B. 1 Allocations from DRF-MT in Example 1 and the comparison of the resulting utilities with utilities of proportional allocation. . . . . . . . . . . . . . . . . . . . . . 140

## Acknowledgements

First and foremost, I want to thank my advisor Shipra Agrawal, who has provided constant guidance and wisdom for the past five years. I would also like to thank my collaborators over the years, Assaf Zeevi, Christian Kroer, Shatian Wang, Lingyi Zhang, Ruizhe Jia. Working with you all will be the one thing I miss the most about graduate school. I also want to thank my undergrad advisor Professor Lieven Vandenberghe. Your EE 133 class at UCLA was the sole reason why I decided to apply to graduate school in the first place.

I want thank my friends: Muye, James, Sasha, Mitch, Mali, Yi, Yuan, Oussama, Agathe, Tugce, Ahmet, Mike, Shawn, Jacob, Wenbo, just to name a few. Without whom this journey would have been way less colorful.

As always, none of this would have been possible without the love and support from Mom and Dad. Thank you for providing me everything I need to pursue my interests. And finally, to Yining, thank you for inspiring me with your wisdom and passion every day.

## Dedication

To Mom and Dad.

## Introduction

A surprisingly large number of applications can be viewed as resource allocation problems. From hospitals and federal agencies, to Ad auction platforms and online retailers, many organizations faces the problem of allocating some scarce resource (e.g. equipment, funding, user attention, and inventory) to a group of agents (e.g. patients, research groups, advertisers, and customers). Depending on the specific application, agents can sometimes act strategically in order to maximize their own benefits, often at the expense of other participants in the system. This poses interesting challenges to the central planner that needs to design allocation policies that perform well in the presence of strategic agents. Additionally, with the advent of data driven decision making, there has been increasing interests in building systems that can continuously improve based on past observations. Designing online learning algorithms that can adapt to new information presents another dimension of challenge for the central planner. In this thesis, we explore four independent resource allocation problems, each with a different combination of the above properties (online vs single round setting, strategic vs non-strategic agents). Table 1 provides an overview of the problems we study.

Table 1: The four chapters of this thesis.

|  | Non-strategic agents | Strategic agents |
| :---: | :---: | :---: |
| Online/Multiple rounds | Chapter 1 | Chapter 4 |
| Single round | Chapter 3 | Chapter 2 |

In the remainder of this section, We will give an overview of the motivations behind each of
the problems we studied, and a summary of our contribution to each problem.

Chapter 1 In this chapter, we study a revenue maximization problem, where a seller tries to maximize its revenue by the end of a fixed time period. The main observation from the dynamic pricing and learning literature is that by dynamically adjusting the pricing strategies based on the observations collected over time, sellers can achieve higher revenues than they would have otherwise. Demands are typically realized following a "demand model", which describes the rate at which customers will purchase the product at a given price. Most of the work however has been focused on i.i.d. and contextual demand models, where in each time step, the price that the seller picks only affects demand in that time step. Departing from such relatively simple demand models, we consider a novel formulation of the dynamic pricing and demand learning problem, where the evolution of demand in response to posted prices is governed by a stochastic variant of the popular Bass model with parameters $(\alpha, \beta)$ that are linked to the so-called "innovation" and "imitation" effects. With these two parameters, the Bass model is able to capture the so called "word of mouth" effect, where an item that is able to garner significant popularity becomes even more popular as a result. Unlike the more commonly used i.i.d. and contextual demand models, in this model the posted price not only affects the demand and the revenue in the current round but also the future evolution of demand, and hence the fraction of potential market size $m$ that can be ultimately captured. This long term dependence of demand on price introduces many challenges to designing good pricing algorithms. In fact, finding a revenue-maximizing dynamic pricing policy in this model is non-trivial even in the full information case, where model parameters are known. In this paper, we consider the more challenging incomplete information problem where dynamic pricing is applied in conjunction with learning the unknown model parameters, with the objective of optimizing the cumulative revenues over a given selling horizon of length $T$.

Our main contribution is an algorithm that satisfies a high probability regret guarantee of order $m^{2 / 3}$; where the market size $m$ is known a priori. Moreover, we show that no algorithm can incur smaller order of loss by deriving a matching lower bound. Unlike most regret analysis results,
in the present problem the market size $m$ is the fundamental driver of the complexity; our lower bound in fact, indicates that for any fixed $\alpha, \beta$, most non-trivial instances of the problem have constant $T$ and large $m$. We believe that this insight sets the problem of dynamic pricing under the Bass model apart from the typical i.i.d. setting and multi-armed bandit based models for dynamic pricing, which typically focus only on the asymptotic with respect to time horizon $T$.

Chapter 2 Conceptually, the challenges we faced in Chapter 1 stem from the fact the central planner has no a priori information on the demand model parameters, and that the underlying demand model has intricate state dependencies. In this chapter, we turn our attention to a different resource allocation problem setting where the challenge to designing a good algorithm stems from a completely different source. In this setting, the central planner interacts with the agents only once, but the agents can potentially behave strategically to benefit themselves at the expense of others. The main challenge here is that the agents could misreport their valuations for the different resources. As such, we have to take advantage of the structure in the agents' utility function in order to design an algorithm that performs well in the presence of strategic agents.

The problem was initially motivated by the COVID-19 pandemic, where we saw many hospitals were in need of different medical supplies, and that the federal and local governments were in a position to distribute/reallocate some of the supplies. We study a generalization of the multiresource allocation problem with heterogeneous demands and Leontief utilities. Unlike existing settings, we allow each agent to specify requirements to only accept allocations from a subset of the total supply for each resource. These requirements can take form in location constraints (e.g. A hospital can only accept volunteers who live nearby due to commute limitations). This can also model a type of substitution effect where some agents need 1 unit of resource A or B , both belonging to the same meta-type. But some agents specifically want A , and others specifically want B . We propose a new mechanism called Dominant Resource Fairness with Meta Types which determines the allocations by solving a small number of linear programs. The proposed method satisfies Pareto optimality, envy-freeness, strategy-proofness, and a notion of sharing incentive for our setting. To
the best of our knowledge, we are the first to study this problem formulation, which improved upon existing work by capturing more constraints that often arise in real life situations. Finally, we show numerically that our method scales better to large problems than alternative approaches.

Chapter 3 The challenges in the first two chapters stem from 1) unknown demand distribution and 2) strategic agents. One might be under the impression that if the central planner has perfect information on the demand distribution and agents' individual utility for the items, then the allocation problem would be easy. This is not the case, and we study one such instance in this chapter. We consider the problem of allocating a distribution of items to $n$ receivers where each receiver has to be allocated a fixed, pre-specified fraction of all items, while ensuring that each receiver does not experience too much envy. Intuitively, the main challenge here is a computational one: with items represented as a distribution, naive formulation of the resource allocation problem lead to an infinite dimensional optimization problem. We show however, that in this specific setting, the problem can be formulated as a variant of the semi-discrete optimal transport (OT) problem, whose solution structure in this case has a concise representation and a simple geometric interpretation. Unlike existing literature that treats envy-freeness as a hard constraint, our formulation allows us to optimally trade off efficiency and envy continuously. Additionally, we study the statistical properties of the space of our OT based allocation policies by showing a polynomial bound on the number of samples needed to approximate the optimal solution from samples. Our approach is suitable for large-scale fair allocation problems such as the blood donation matching problem, and we show numerically that it performs well on a prior realistic data simulator.

Chapter 4 Finally, in the last chapter, we study a resource allocation problem that combines all of the challenges we have encountered so far. We study the problem of allocating $T$ sequentially arriving items among $n$ homogenous agents under the constraint that each agent must receive a prespecified fraction of all items, with the objective of maximizing the agents' total valuation of items allocated to them. The agents' valuations for the item in each round are assumed to be i.i.d. but their distribution is a priori unknown to the central planner. Therefore, the central planner needs to
implicitly learn these distributions from the observed values in order to pick a good allocation policy. However, an added challenge here is that the agents are strategic with incentives to misreport their valuations in order to receive better allocations. This sets our work apart both from the online auction mechanism design settings which typically assume known valuation distributions and/or involve payments, and from the online learning settings that do not consider strategic agents. To that end, our main contribution is an online learning based allocation mechanism that is approximately Bayesian incentive compatible, and when all agents are truthful, guarantees a sub-linear regret for individual agents' utility compared to that under the optimal offline allocation policy.

# Chapter 1: Dynamic Pricing and Learning under the Bass Model 

### 1.1 Background and Motivation

The dynamic pricing and learning literature, often referred to as "learning and earning," has at its focal point the objective of maximizing revenues jointly with inferring the structure of a demand model that is a priori not known to the decision maker. It is an extremely active area of research that can essentially be traced back to two strands of work. Within the computer science community, the first paper on the topic is [1] that studies a posted price auction with infinite inventory in which the seller does not know the willingness-to-pay of buyers and must learn it over repeated interactions. The problem is stateless, and demand is independent from period to period. Various refinements and improvements have been proposed since in what has become a very active field of study in economics, computer science and operations research. The second strand of work originates in the operations research community [2] which focuses on the same finite horizon regret criteria in the posted-price auction problem but with limited inventory. This problem is sometimes referenced as "bandits with knapsacks" due to the follow up work of [3] and subsequent papers. In that problem, the learning objective is more subtle as the system state (i.e., the remaining inventory and time) is changing over time. For further references and historical notes on the development of the topic see, e.g., the recent survey [4].

Most of the literature that has evolved from the inception points identified above has focused on a relatively simple setting where given current pricing decision, demand is independent of past actions and demand values. In addition, much of the literature has focused on the "stateless" problem setting, which provides further simplification and tractability. With the evolution of online platforms and marketplaces, the focus on such homogeneous modeling environments is becoming increasingly less realistic. For example, platforms now rely more and more on online reviews and
ratings to inform and guide consumers. Product quality information is also increasingly available on online blogs, discussion forums, and social networks, that create further word-of-mouth effects. One clear implication on the dynamic pricing and learning problem is that the demand environment can no longer be assumed to be static; for example, in the context of online reviews, sales of the product trigger reviews/ratings, and these in turn influence subsequent demand behavior etc.

While it is possible to model demand shifts and non-stationarities in a flexible (nonparametric) manner within the dynamic pricing and learning problem (see, e.g., [5], and [6] for a general MAB formulation), this approach can be too broad and unstructured to be effective in practical dynamic pricing settings. To that end, product diffusion models, such as the popular Bass model [7, 8], are known to be extremely robust and parsimonious, capturing aforementioned word-of-mouth and imitation effects on the growth in sales of a new product. The Bass diffusion model, originally proposed by Frank Bass in 1969 [7] has been extremely influential in marketing and management science, often described as one of the most celebrated empirical generalizations in marketing. It describes the process by which new products get adopted as an interaction between existing users and potential new users.

The Bass diffusion model [7, 8] has two parameters: the "coefficient of innovation" representing external influence; and the "coefficient of imitation" representing internal influence or word-of-mouth effect. Let $m$ be the number of potential buyers, i.e., the market size, and let $X_{t}$ be fraction of customers who has already adopted the product until time $t$. Then, $m X_{t}$ represents the cumulative sales (i.e., adoptions) up until time $t$, and $m\left(1-X_{t}\right)$ is the size of remaining market yet to be captured. The instantaneous sales at time $t, m \frac{d X_{t}}{d t}$ can then be expressed as

$$
\begin{equation*}
m \frac{d X_{t}}{d t}=\underbrace{m \alpha\left(1-X_{t}\right)}_{\text {sales due to external influence }}+\underbrace{m \beta X_{t}\left(1-X_{t}\right)}_{\text {sales due to internal influence or imitation }} \tag{1.1}
\end{equation*}
$$

A generalization of the Bass diffusion model that can be harnessed for the dynamic pricing context was proposed by Robinson and Lakhani [9]. In the latter model, $p_{t}$ denotes the price


Figure 1.1: Left: Bass model's daily sales predictions over time have the same bell shape as the realized daily sales, whereas the other i.i.d. models do not. Right: The Bass model is able to capture the $S$-shaped adoption curve while the other models are not. The faint green line in the background is the price history.
posted at time $t$. Then, the number of new adoptions at time instant $t$ is given by

$$
\begin{equation*}
m \frac{d X_{t}}{d t}=m e^{-p_{t}}\left(\alpha+\beta X_{t}\right)\left(1-X_{t}\right) \tag{1.2}
\end{equation*}
$$

Thus, the current price affects not only the immediate new adoptions and revenue, but also future adoptions due to its dependence on the adoption level $X_{t}$.

The Bass model therefore creates a state-dependent evolution of market response which is well aligned with the impact of recent technological developments, such as online review platforms, on the customer purchase behavior. To that end, several recent empirical studies in marketing science and econometrics utilize abundant social data from online platforms to quantify the impact of word-of-mouth effect on consumer purchase behaviors and a new product diffusion process (e.g., [10, 11, 12, 13, 14, 15], also see [16, 17] for literature surveys).

In Figure 1.1, we present some empirical motivation for this model using the UCI online retail dataset [18]. This dataset contains information on roughly half a million transactions from a UKbased online retail, including price and time of each transaction, so that the price curve and demand evolution over time can be observed. In Figure 1.1, we fit the Bass model (as given by (1.2)) to
this data and compare it to the best fitting i.i.d./stationary models with linear and exponential price response functions. Figure 1.1(a) shows the curve of new adoptions, i.e., sales per day, and Figure 1.1(b) shows the typical S-shaped curve of cumulative sales. Clearly, the Bass model is able to nicely capture the S -shape in the cumulative adoption curve, whereas the i.i.d./stationary models fail to capture this phenomenon.

Motivated by these observations, the objective of this chapter is to investigate a novel formulation of the the dynamic pricing and demand learning problem, where the evolution of demand is governed by the Bass diffusion model, and where the parameters of this model are unknown a priori and need to be learned from repeated interactions with the market.

### 1.2 Main contributions

The proposed model and main problem studied We consider a stochastic variant of the above Bass model, where customer arrivals at time $t$ is governed by a non-homogeneous Poisson process with rate $\lambda_{t}$ given by the right hand side of (1.2). More details on the stochastic model are provided in Section 1.4 where we describe the full problem formulation and performance objectives.

The problem of dynamic pricing under demand learning can then be described, informally, as follows: the learner (seller) is required to dynamically choose prices $\left\{p_{t}\right\}$ to be posted at time $t \in[0, T]$, where $p_{t}$ is chosen based on the past observations that include the number of customers arrived so far, their times of arrival (which stand for the inter-arrival times between customers) and the price decisions made in the past. The number of customers $d_{t}$ arriving until any time $t$ is given by the stochastic Bass model. Note that we use the term "customer arrival" and "customer adoption" interchangeably to mean the same thing, i.e., every customer arriving at time $t$ adopts the product and pays the price $p_{t}$. We assume that the size of the market $m$ and the horizon $T$ are known to the learner, but the Bass model parameters $\alpha, \beta$ (i.e., coefficient of innovation and coefficient of imitation) are unknown. The aim is to maximize the cumulative revenue over the horizon $T$, i.e., $\sum_{d=1}^{d_{T}} p_{d}$, via a suitable sequence of prices, where $p_{d}$ is the price paid by the $d^{\text {th }}$ customer and $d_{T}$ is the total number of adopters until time $T$. Finding a revenue-maximizing optimal pricing
trajectory in such a dynamic demand model is non-trivial even in a deterministic setting and when the model parameters are known, e.g., see [19] for some characterizations. In the learning problem considered here, we will not be directly maximizing this quantity but rather, and much in line with the online learning and multi-armed bandits literature, will focus on evaluating a suitable notion of regret and establishing "good" regret performance of our proposed pricing and learning algorithm.

Main contributions The paper makes two significant advances in the study of the aforementioned Bass model learning and earning problem. First, we present a learning and pricing algorithm that achieves a high probability $\tilde{O}\left(m^{2 / 3}\right)$ regret bound. (Here, and in what follow we make rather standard use of the big-Oh notation convention, delaying further clarifying remarks to the subsequent section.) Second, under certain mild restrictions on algorithm design, we provide a matching lower bound, showing that any algorithm must incur order $\Omega\left(m^{2 / 3}\right)$ regret for this problem. The precise statements of these results are provided as Theorem 1 and Theorem 2, in Section 1.6 and 1.7 respectively. Hence the "price" of incomplete information and the "cost" of learning it on the fly, are reasonably small. Unlike most regret analysis results, in the present setting the market size $m$ is the fundamental driver of the problem complexity; our lower bound, in fact, indicates that for any fixed $\alpha, \beta$, most non-trivial instances of the problem have constant $T$ and large $m$. This is also reflected in the regret of our algorithm which depends sub-linearly on market size $m$, and only logarithmically on the time horizon $T$. This insight sets the problem of dynamic pricing under Bass model uniquely apart from the typical i.i.d. and multi-armed bandit-based models for dynamic pricing.

### 1.3 Literature Review

Several other models of non-stationary demand have been considered in recent literature on dynamic pricing and demand learning. Boer [20] studies an additive demand model where the market condition is determined by the sum of an unknown market process and an adjustment term that is a known function of the price. The paper numerically studies several sliding window and
weight decay based algorithms, including one experiment where their pricing strategy is tested on the Bass diffusion model. However, they do not provide any regret guarantees under the Bass model. Besbes and Zeevi [21] and Chen et al. [22] consider the pricing problem under a piecewise stationary demand model. Keskin and Zeevi [5] consider a two-parametric linear demand model whose parameters are indexed by time, and the total quadratic variation of the parameters are bounded. Many of the existing dynamic pricing and learning algorithms borrow ideas from related settings in multi-armed bandit literature. Garivier and Moulines [23] and Cao et al. [24] study piecewise stationary bandits based on sliding window and change point detection techniques. Besbes et al. [6] and Russac et al. [25] and Luo et al. [26] study regret based on an appropriate measure of reward variation. In rested and restless bandits literature, Tekin and Liu [27] and Bertsimas and Niño-Mora [28] also study a related problem where the reward distribution depends on the state which follows an unknown underlying stochastic process. However, these models of non-stationarity are very broad and unstructured, and fundamentally different from the stateful structured Bass diffusion model studied here.

There is also much recent work on learning and regret minimization in stateful models using general MDP and reinforcement learning frameworks (for example [29, 30, 31]). However, these results typically rely on having settings where each state can be visited many times over the learning process. This is ensured either because of an episodic MDP setting (e.g., [29]), or through an assumption of communicating or weakly communicating MDP with bounded diameter, i.e., a bound on the number of steps to visit any state from any starting state under the optimal policy (e.g., see [31, 30]). Our setting is non-episodic, and every state is transient - the state is described by the number of cumulative adopters so far and the remaining time, both of which can only move in one direction. Therefore, the results in the above and related papers on learning general MDPs are not applicable to our problem setting.

In the marketing literature, there is significant work on using the Bass model to forecast demand before launching a product. Stochastic variants of Bass model have also been considered previously, e.g., in Grasman and Kornelis [32] and Niu [33]. In addition to the work mentioned
in the introduction, Lee et al. [34], Fan et al. [35], Yin et al. [36], and Grasman and Kornelis [32] present empirical methods for using historical data from similar products or from past years to estimate Bass model parameters, in order to obtain reliable forecasts. In this context, we believe our proposed dynamic pricing and learning algorithm may provide an efficient method for adaptively estimating the Bass model parameters while optimizing revenue. However, the focus of this paper is on providing theoretical performance guarantees; the empirical performance of the proposed method has not been studied.

The work most closely related to ours is a parallel recent paper by Zhang et al. [37]. In [37], the authors consider a dynamic pricing problem under a similar stochastic Bass model setting as the one studied here. They propose an algorithm based on Maximum Likelihood Estimation (MLE) that is claimed to achieve a logarithmic regret bound of $\tilde{O}(\log (m T))$. At first glance this regret bound seems to contradict our lower bound ${ }^{1}$ of $\Omega\left(m^{2 / 3}\right)$. However, it appears the results in [37] may have some mistakes (see, in particular, the current statement and proofs of Lemma 3, Theorem 5 and Theorem 6 in [37] which we believe contain the aforementioned inconsistencies). To the best of our understanding, these inconsistencies cannot be fixed without changing the current results in a significant way. Further evidence for the $m^{2 / 3}$ lower bound can be found in a recent parallel work in [38], where the authors showed that one needs at least $m^{2 / 3}$ observations to estimate Bass model parameters up to a constant error. However, their paper does not consider the adaptive learning setting where different actions (e.g., dynamically setting prices) can be used to potentially learn more efficiently. Our paper shows that even with adaptive learning, $m^{2 / 3}$ regret is unavoidable, and we also provide an online learning algorithm with a regret upper bound that closely matches that lower bound.

[^0]
### 1.4 The Bass model and problem formulation

The stochastic model and problem primitives There is an unlimited supply of durable goods of a single type to be sold in a market with $m$ customers. We consider a dynamic pricing problem with sequential customer arrivals over a time interval $[0, T]$. We denote by $d_{t}$, the number of arrivals by time $t$; with $d_{0}=0$. At any given time $t$, the seller observes $d_{t}$, the number of arrivals so far as well as their arrival times $\left\{\tau_{1}, \tau_{2}, \ldots, \tau_{d_{t}}\right\}$, and chooses a new price $p_{t}$ to be posted for times $t^{\prime}>t$. The seller can use the past information to update the price any number of times until the end of time horizon $T$. As mentioned earlier, in our formulation a customer arriving at time $t$ immediately adopts the product and pays the posted price $p_{t}$, and therefore the terms "adoption" and "arrival" are used interchangeably throughout the text.

The customer arrival process for any pricing policy is given by a stochastic Bass diffusion model with (unknown) parameters $\alpha, \beta$. In this model, number of arrivals $d_{t}$ by time $t$ is a nonhomogeneous Poisson point process ${ }^{2}$ with rate $\lambda_{t}$ given by (1.2), the adoption rate in the deterministic Bass diffusion model [9]. That is,

$$
\lambda_{t}=m e^{-p_{t}}\left(\alpha+\beta X_{t}\right)\left(1-X_{t}\right)
$$

where $X_{t}=d_{t} / m$. For convenience of notation, we define the function

$$
\lambda(p, x):=m e^{-p}(\alpha+\beta x)(1-x),
$$

so that $\lambda_{t}=\lambda\left(p_{t}, X_{t}\right)$, with $X_{t}=d_{t} / \mathrm{m}$.
The seller's total revenue is simply the sum of the prices paid by the customers who arrived before time $T$, under the dynamic pricing algorithm used by the seller. We denote by $p_{d}$, the price

[^1]paid by $d^{t h}$ customer, i.e., $p_{d}=p_{\tau_{d}}$. Then, the revenue over $[0, T]$ is given by:
$$
\operatorname{Rev}(T)=\sum_{d=1}^{d_{T}} p_{d}
$$

The optimal dynamic pricing policy is defined as the one that maximizes the total expected revenue $E[\operatorname{Rev}(T)]$. We denote by $V^{\text {stoch }}(T)$, the total expected revenue under the optimal dynamic pricing policy. Then, regret is defined as the difference between the optimal expected revenue $V^{\text {stoch }}(T)$ and seller's revenue, i.e.,

$$
\begin{equation*}
\operatorname{Regret}(T)=V^{\text {stoch }}(T)-\operatorname{Rev}(T) \tag{1.3}
\end{equation*}
$$

In this chapter, we aim to provide a dynamic learning and pricing algorithm with a high probability upper bound of $\tilde{O}\left(\mathrm{~m}^{2 / 3}\right)$ on regret, as well as a closely matching lower bound. Instead of directly analyzing the regret, we define a notion of "pseudo-regret," which measures regret against the optimal revenue under the deterministic Bass model $\left(V^{\operatorname{det}}(T)\right)$. This is useful because as we will show later in (1.4), there is a simple expression for the optimal prices in the deterministic Bass model. This can be leveraged using a fluid model approach, widely used in the analysis of stochastic systems, which targets a deterministic "skeleton" as a stepping stone toward developing insights for policy design and complexity drivers. Later, we show more formally that the pseudoregret upper bounds the regret measure $\operatorname{Regret}(T)$, and is also within $\tilde{O}(\sqrt{m})$ of $\operatorname{Regret}(T)$. This justifies our focus on the pseudo-regret for both the purpose of developing lower bounds, as well as analyzing upper bounds on policy performance. Next, we discuss the expressions for optimal pricing policy and revenue under the deterministic Bass model, and use those to define the notion of pseudo-regret.

Optimal revenue and pricing policy under the deterministic Bass model Recall (refer to introduction) that the adoption process under the deterministic Bass model is described as follows: given the current adoption level $X_{t}$ and price $p_{t}$ at time $t$, the new adoptions are generated deter-
ministically with rate given in $(1.2)^{3}$.
The optimal price curve for the deterministic Bass model is then defined as the price trajectory $\left\{p_{t}\right\}$ that maximizes the total revenue $m \int_{0}^{T} p_{t} d X_{t}$ under the above adoption process. We denote by $V^{\operatorname{det}}(T)$ the total revenue under the optimal price curve.

An analytic expression for the optimal price curve under the deterministic Bass model is in fact known. It can be derived using optimal control theory (see expression (8) in [19]). For a Bass model with parameters $\alpha, \beta$, horizon $T$, and initial adoption level 0 , the price at adoption level $x$ in the optimal price curve is given by the following expression:

$$
\begin{equation*}
p^{*}(x, \alpha, \beta)=1+\log \left(\frac{(\alpha+\beta x)(1-x)}{\left(\alpha+\beta X_{T}^{*}\right)\left(1-X_{T}^{*}\right)}\right), \tag{1.4}
\end{equation*}
$$

where $X_{T}^{*}$, the adoption level at the end of horizon $T$, is given by the following equations:

$$
\begin{equation*}
X_{T}^{*}=\frac{1}{e}\left(\alpha+\beta X_{T}^{*}\right)\left(1-X_{T}^{*}\right) T \tag{1.5}
\end{equation*}
$$

or, more explicitly

$$
\begin{equation*}
X_{T}^{*}=\frac{T(\beta-\alpha)-e+\sqrt{[T(\beta-\alpha)-e]^{2}+4 \alpha \beta T^{2}}}{2 \beta T} . \tag{1.6}
\end{equation*}
$$

For completeness, a derivation of the above expression of $X_{T}^{*}$ is included in Appendix A.2. Using the above notation for optimal price curve, we can write $V^{\operatorname{det}}(T)$, the optimal total revenue, as:

$$
V^{\operatorname{det}}(T)=m \int_{0}^{X_{T}^{*}} p^{*}(x, \alpha, \beta) d x .
$$

[^2]Pseudo-Regret We define pseudo-regret as the difference between $V^{\operatorname{det}}(T)$, the optimal total revenue in the deterministic Bass model, and the algorithm's total revenue $\operatorname{Rev}(T)$. That is,

$$
\begin{equation*}
\operatorname{Pseudo}-\operatorname{Regret}(T)=V^{\operatorname{det}}(T)-\operatorname{Rev}(T) \tag{1.7}
\end{equation*}
$$

Essentially, pseudo-regret replaces the benchmark of stochastic optimal revenue $V^{\text {stoch }}(T)$ used in the regret definition by deterministic optimal revenue $V^{\text {det }}(T)$. We show that the deterministic optimal revenue is a stronger benchmark, in the sense that it is always larger than the stochastic optimal revenue. Furthermore, we show that it is within $\tilde{O}(\sqrt{m})$ of the stochastic optimal revenue. To prove this relation between the two benchmarks we demonstrate a concavity property of deterministic optimal revenue which is crucial for our results. Specifically, we define an expanded notation $V^{\text {det }}(x, T)$ as the deterministic optimal revenue starting from adoption level $x$ and remaining time $T$. Note that $V^{\operatorname{det}}(T)=V^{\text {det }}(0, T)$. Then, we show that $V^{\operatorname{det}}(x, T)$ is concave in $x$ for any $T$, and any market parameters $m, \alpha, \beta$. More precisely, we prove the following key lemma.

Lemma 1 (Concavity of deterministic optimal revenue). For any deterministic Bass model, $V^{\text {det }}(x, T)$, defined as the optimal revenue starting from adoption level $x$ and remaining time $T$, is concave in $x$, for all $T \geq 0$, and all adoption levels $x \in[0,1]$.

Using these observations, we can prove the following relation between Pseudo-Regret $(T)$ and $\operatorname{Regret}(T)$; all proofs can be found in the appendix.

Lemma 2 (Pseduo-regret is close to regret). With probability 1 , for any $T \geq 0$,

$$
\begin{gathered}
\text { Pseudo-Regret }(T) \geq \text { Regret }(T) \\
\text { Pseudo-Regret }(T) \leq \operatorname{Regret}(T)+O\left(\sqrt{m} \log ^{2}(m)+\log ((\alpha+\beta) T) \log ^{2}(m \log ((\alpha+\beta) T))\right) .
\end{gathered}
$$

A simpler summary of this result can be stated as the following almost sure inequality:

$$
\operatorname{Regret}(T) \leq \operatorname{Pseudo} \text {-Regret }(T) \leq \operatorname{Regret}(T)+\tilde{O}(\sqrt{m})
$$

Therefore, an upper bound on Pseudo-Regret $(T)$ implies the same upper bound on $\operatorname{Regret}(T)$. And, a lower bound on Pseudo-Regret $(T)$ implies a lower bound on $\operatorname{Regret}(T)$ within $\tilde{O}(\sqrt{m})$. In the rest of the paper, we therefore only derive bounds on the pseudo-regret.

Notation conventions Throughout this Chapter, if a potentially fractional number (like $m X_{T}^{*}$, $m^{2 / 3}, \gamma m$ etc.) is used as a whole number (for example as number of customers or as boundary of a discrete sum) without a floor or ceiling operation, the reader should assume that it is rounded down to its nearest integer. As indicated earlier, we use the conventional big-Oh $O(\cdot)$ to mean a quantity is "of this order," and $\tilde{O}(\cdot)$ when ignoring poly-logarithmic terms. When such an order statement is applied to stochastic quantities it may be interpreted either to hold in the almost sure sense, or with suitably high probability. To that end, $\delta \in(0,1)$ will frequently be used to define events that hold with $1-\delta$ probability. The $\Omega(\cdot)$ notation is the equivalent to $O(\cdot)$ for lower bound purposes and will be used following similar convention.

### 1.5 Algorithm description

The concavity property of deterministic optimal revenue and the implied relation between pseudo-regret and regret derived in Lemma 2 suggests that deterministic optimal revenue provides a benchmark that is comparable to the stochastic optimal revenue. Further, this benchmark is more tractable than the stochastic optimal due to the known and simple analytical expressions for optimal pricing policy, as stated in Section 1.4. Using these insights, our algorithm is designed to essentially follow (an estimate of) the optimal price curve for the deterministic model at every time. We believe this approach could be applied to other finite horizon MDP problems where such concavity property holds, which may be of independent interest. We now describe our algorithm.

### 1.5.1 Algorithm outline

Our algorithm is provided, as input, the market size $m$, and a constant upper bound $\phi$ on $\alpha+\beta$. For many applications, it is reasonable to assume that $\alpha+\beta \leq 1$, i.e., $\phi=1$, but we allow for
more general settings. The market parameters $\alpha, \beta$ are unknown to the algorithm. The algorithm alternates between using a low "exploratory" price aimed at observing demand in order to improve the estimates of model parameters, and using (a lower confidence bound on) the deterministic optimal prices for the estimated market parameters. Setting the exploratory price $p_{0}$ as 0 suffices for our analysis, but more generally it can be set as any lower bound on the deterministic optimal prices, i.e., we just need $0 \leq p_{0} \leq p^{*}(x, \alpha, \beta), \forall x$. Using a non-zero exploratory price could be better in practice.

Our algorithm changes price only on arrival of a new customer, and holds the price constant in between two arrivals. The prices are set as follows. The algorithm starts with using an exploratory price $p_{0}$ for the first $\gamma m$ customers, where $\gamma=m^{-1 / 3}$. The prices $p_{1}, \ldots, p_{\gamma m}$ and the observed arrival times $\tau_{1}, \ldots, \tau_{\gamma m}$ for the first $\gamma m$ customers are then used to obtain an estimation of $\alpha$, and a high probability error bound $A$ on this estimate. The estimate of $\alpha$ is not updated in subsequent time steps. The algorithm then proceeds in epochs $i=1,2, \ldots K$. Let $\gamma_{i}=2^{i} \gamma$. Epoch $i$ starts right after customer $\gamma_{i} m$ arrives and ends either when customer $2 \gamma_{i} m$ arrives, or when we reach the end of the planning horizon $T$. In the beginning of each epoch $i$, the exploratory price $p_{0}$ is again offered for the first $\gamma m$ customers in that epoch. The arrival times observed from these customers are used to update the estimate of $\beta$ and its' high probability error bound $B_{i}$. For the remaining customers in epoch $i$, the algorithm offers a lower confidence bound $\underline{p_{i}}$ on the deterministic optimal price $p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right)$ computed using the current estimates of $\alpha, \beta$ and their error bounds.

### 1.5.2 Estimation and price computation details

Since our algorithm fixes prices between two arrivals, the arrival rate of the (Poisson) adoption process is constant in between arrivals, which in turn implies that the inter-arrival times are exponential random variables. This greatly simplifies the estimation procedure for $\alpha, \beta$ : the estimates $\hat{\alpha}$, and $\hat{\beta}_{i}$ for every epoch $i$ are calculated using the following equations which match the observed

```
ALGORITHM 1: Dynamic Learning and Pricing under Bass Model
    Input: Horizon \(T\), market size \(m, \delta \in(0,1)\), a constant upper bound \(\phi\) on \(\alpha+\beta\).
    Initialize: \(X_{0}=0, \gamma=m^{-1 / 3}, \gamma_{i}=2^{i-1} \gamma\) for \(i=1,2, \ldots\), and exploratory price \(p_{0}=0\)
    Offer \(p_{0}\) for the first \(\gamma m\) customers.
    Use observed arrival times of the first \(\gamma m\) customers to compute \(\hat{\alpha}\) according to (1.8).
    Also, obtain a high probability estimation bound on \(|\alpha-\hat{\alpha}| \leq A\) where \(A\) is as defined in
        (1.10).
    for \(i \leftarrow 1,2,3, \ldots\) do
        Post price \(p_{0}\) for customers \(d=\gamma_{i} m+1, \ldots, \gamma_{i} m+\gamma m\).
        Use \(\hat{\alpha}\) along with the observed arrival times of the first \(\gamma m\) customers in the current
        epoch to calculate a new estimate \(\hat{\beta}_{i}\) according to (1.9). Also update the high
        probability bound \(\left|\beta-\hat{\beta}_{i}\right| \leq B_{i}\), with \(B_{i}\) as defined in (1.10).
        Use \(\hat{\alpha}, A, \hat{\beta}_{i}, B_{i}\) to compute price \(\underline{p_{i}}\left(\frac{d-1}{m}\right)\) according to (1.11). Post price \(\underline{p_{i}}\left(\frac{d-1}{m}\right)\) for
            \(d=\gamma_{i} m+\gamma m+1, \ldots, 2 \gamma_{i} m\), or until end of time horizon \(T\).
        If \(\gamma_{i} \geq \frac{1}{3}\) then extend the current epoch until the end of the planning horizon.
    end
```

inter-arrival times to the expected value of the corresponding exponential random variables:

$$
\begin{align*}
\frac{1}{e^{-p_{0}} \hat{\alpha} m}\lfloor\gamma m\rfloor & =\tau_{\gamma m}  \tag{1.8}\\
\frac{\lfloor\gamma m\rfloor}{e^{-p_{0}}\left(\hat{\alpha}+\hat{\beta}_{i} \gamma_{i}\right)\left(1-\gamma_{i}\right) m} & =\tau_{\gamma_{i} m+\gamma m}-\tau_{\gamma_{i} m} . \tag{1.9}
\end{align*}
$$

Also, we define

$$
\begin{equation*}
A:=\frac{8 \phi}{(1-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right)} m^{-1 / 3}, B_{i}:=\frac{16 \phi}{\gamma_{i}(1 / 3-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right)} m^{-1 / 3} . \tag{1.10}
\end{equation*}
$$

Then, using concentration results for exponential random variables, we can show that the estimation errors in $\hat{\alpha}, \hat{\beta}_{i}$ are bounded by $A$ and $B_{i}$ respectively. Specifically, we show the following result. The proof is in Appendix A.4.

Lemma 3. Assume $m^{1 / 3} \geq 64 \frac{(\alpha+\beta)^{2}}{\alpha^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$. Then with probability $1-\delta,|\alpha-\hat{\alpha}| \leq A$. And for any $i=1, \ldots, K$, with probability $1-\delta,\left|\beta-\hat{\beta}_{i}\right| \leq B_{i}$.

Note that only the estimate and error bounds ( $\hat{\beta}_{i}$ and $B_{i}$ ) for the parameter $\beta$ are updated in every epoch. The estimate and error bound ( $\hat{\alpha}$ and $A$ ) for $\alpha$ is computed once in the beginning and
remains fixed throughout the algorithm. Given $\hat{\alpha}, A, \hat{\beta}_{i}, B_{i}$, we define the price $\underline{p_{i}}(d / m)$ to be used by the algorithm in epoch $i$ as follows. For any $d \leq m$,

$$
\begin{array}{r}
\underline{p_{i}}\left(\frac{d}{m}\right)=\operatorname{clamp}\left(p^{*}\left(\frac{d}{m}, \hat{\alpha}, \hat{\beta}_{i}\right)-L_{\alpha} A-L_{\beta i} B_{i},[0, \log (e+\phi T)]\right), \\
\quad \text { where } \quad L_{\alpha}=\frac{2}{\hat{\alpha}-A}+\frac{\hat{\beta}_{i}+B_{i}}{(\hat{\alpha}-A)^{2}}, \quad L_{\beta i}=\frac{3}{\hat{\alpha}-A}+\frac{3}{\hat{\beta}_{i}-B_{i}} . \tag{1.11}
\end{array}
$$

Here $p^{*}(\cdot, \cdot, \cdot)$ is the optimal deterministic price curve as given by (1.4). In above, we clamped the price to the range $[0, \log (e+\phi T)]$. That is, if the computed price is less than 0 we set $\underline{p_{i}}$ to 0 ; and if it is above $\log (e+\phi T)$, which is an upper bound on the optimal price (proven later in Lemma 7), we set $\underline{p_{i}}$ equal to this upper bound.

Later in Lemma 4, we show that the quantities $L_{\alpha}, L_{\beta i}$ play the role of Lipschitz constants for the optimal price curve: they provide high probability upper bounds on the derivatives of the optimal price curve with respect to $\alpha, \beta$ respectively. Consequently (in Corollary 1), we can show that with high probability the price defined in (1.11) will be lower than the corresponding deterministic optimal price. The reason that the algorithm uses a lower confidence bound on the optimal price is that we want to acquire at least as many customers in time $T$ as the optimal trajectory. The intuition here is that losing customers (i.e., the revenue associated with those customers, as well as the potential word-of-mouth effect that they bring) cost a lot more than losing a little bit of revenue from each customer. Our algorithm is described in detail as Algorithm 1.

### 1.6 Regret upper bound

The main result from this section is the following upper bound on the pseudo-regret of the algorithm proposed in the previous section. Since pseudo-regret upper bounds regret (refer to Lemma 2) it directly implies the same upper bound on $\operatorname{Regret}(T)$.

Theorem 1 (Regret upper bound). For any market with parameters $\alpha, \beta, \alpha+\beta \leq \phi$, market size
$m$, and time horizon T, Algorithm 1 achieves the following regret bound with probability $1-\delta$,

$$
\text { Pseudo-Regret }(T) \leq O\left(m^{2 / 3} \log (m) \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right)=\tilde{O}\left(m^{2 / 3}\right)
$$

Proof Intuition We first give an intuitive explanation for why our algorithm works well. As mentioned earlier in the introductory sections, there is a simple closed form expression for the optimal prices in the deterministic model for any $\alpha, \beta, T$ (see (1.4)-(1.6)). Moreover, the definition of pseudo-regret and Lemma 2 allows us to replace the stochastic optimal revenue $V^{\text {stoch }}(T)$ with the deterministic optimal revenue $V^{\text {det }}(T)$. Our algorithmic strategy is then to follow (an estimate of) the deterministic optimal price trajectory, and show that the resulting revenue is close to the deterministic optimal revenue with high probability.

To prove this, we show that the prices $\underline{p_{i}}$ used by our algorithm were set so that they lower bound the deterministic optimal price with high probability. Intuitively, using a lower price would ensure that the algorithm sees at least as many as optimal number of customer adoptions in horizon $T$, so that the gap in revenue can be bounded simply by the gap in prices paid by those customers. The final piece of the puzzle is to show that we can learn or estimate $\alpha, \beta$ at a fast enough rate so that the estimated prices are increasingly close to the optimal price and we do not lose too much revenue from learning.

Proof Outline In the remainder of this section we outline the proof of Theorem 1 in four steps. All the missing proofs from this section can be found in Appendix A.4.

Step 1 (Bounding the estimation errors) In Lemma 3 we provided a high probability upper bound on the estimation errors in $\alpha, \beta$ in each epoch of the algorithm. Using these error bounds and the definition of price $\underline{p_{i}}$ in (1.11), we show that the prices offered by the algorithm are, with high probability, close lower bounds of the optimal prices. Specifically, we prove the following result.

Lemma 4 (Error bounds for estimated prices). Given any market parameters $\alpha, \beta$ and their es-
timators $\hat{\alpha}, \hat{\beta}_{i}$ that satisfy $|\alpha-\hat{\alpha}| \leq A,\left|\beta-\hat{\beta}_{i}\right| \leq B_{i}, \hat{\alpha}-A>0, \hat{\beta}_{i}-B_{i}>0$, then for every $d=0, \ldots, m-1$,

$$
\left|p^{*}\left(\frac{d}{m}, \hat{\alpha}, \hat{\beta}_{i}\right)-p^{*}\left(\frac{d}{m}, \alpha, \beta\right)\right| \leq L_{\alpha} A+L_{\beta i} B_{i},
$$

where $L_{\alpha}, L_{\beta i}$ are as defined in (1.11).
The proof of above results can be found in Appendix A.4.1. From the expressions of $A$ (error bound for $\hat{\alpha}$ ) vs. $B_{i}$ (error bound for $\hat{\beta}_{i}$ in epoch $i$ ) in (1.10), observe that $A=\tilde{O}\left(m^{-1 / 3}\right)$ and $B_{i}=\tilde{O}\left(\frac{1}{\gamma_{i}} m^{-1 / 3}\right)$. Therefore, the estimation of $\beta$ is the bottleneck here: it has an extra $\frac{1}{\gamma_{i}}$ factor in the error bound. This is because the imitation factor $\beta$ is multiplied with the current adoption level in the definition of the Bass model (see (1.1)). This means that the estimation error on $\beta$ is likely to be large when the adoption level is low. This is why the algorithm needs to keep updating the estimate of $\beta$ in each epoch but not $\alpha$.

The above lemma and the definition of $\underline{p_{i}}$ immediately implies the following.
Corollary 1 (Lower confidence bound on optimal price). Under the same assumptions as in Lemma 4, and given the definition of $p_{i}$ in (1.11), we have that for every $d=0, \ldots, m-1$, $\underline{p_{i}}\left(\frac{d}{m}\right) \leq p^{*}\left(\frac{d}{m}, \alpha, \beta\right)$.

Step 2 (Bounding the revenue loss due to lower price) Using the fact that there are $\gamma_{i} m$ customers in epoch $i$, and the price difference bound in Lemma 4 in Step 1, we can show that we lose at most $\tilde{O}\left(\gamma_{i} m\left(L_{\alpha} m^{-1 / 3}+L_{\beta i} \frac{1}{\gamma_{i}} m^{-1 / 3}\right)\right)=\tilde{O}\left(m^{2 / 3}\right)$ revenue per epoch. This loss bound per epoch is benchmarked against the optimal revenue earned from the same $\gamma_{i} m$ customers, assuming that the stochastic trajectory is given enough time to reach the same number of customers. More precisely, let $\operatorname{Rev}_{i}$ denote the algorithm's revenue from the $\left(2 \gamma_{i} m \wedge m X_{T}^{*}\right)-\gamma_{i} m$ customers in an epoch $i$ completed by the algorithm, and $V_{i}^{\text {det }}$ denote the potential revenue from those customers if they paid the deterministic optimal price instead. Then, we prove the following bound.

Lemma 5. For any epoch $i$ in the algorithm, with probability $1-\delta$,

$$
V_{i}^{d e t}-\operatorname{Rev}_{i} \leq O\left(m^{2 / 3} \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right)=\tilde{O}\left(m^{2 / 3}\right)
$$

Step 3 (Bounding the revenue loss due to fewer adoptions) In the previous step, we upper bounded the potential revenue loss for the $\gamma_{i} m$ customers that arrive in any epoch $i$ of the algorithm. However it is possible that the algorithm reaches the end of time horizon $T$ early and never even get to observe some customers, i.e. $d_{T}<X_{T}^{*} m$. In that case the algorithm would incur an additional regret due to fewer adoptions. Therefore, we need to show that our total number of adoptions in time $T$ cannot be much lower than that in the optimal trajectory. To show this, recall in Step 1 we showed that the algorithm always offer a lower confidence bound on the optimal prices (note that the exploration price $p_{0}=0$ is always a lower bound on the optimal prices). Due to the use of lower prices, we can show that with high probability our final number of adoptions can be at most $\tilde{O}(\sqrt{m})$ below the optimal number of adoptions in the deterministic Bass model. Further, we can prove an $O(\log (T))$ upper bound on the optimal price, which allows us to easily bound the revenue loss that may result from the small gap in number of adoptions. Lemma 6 and Lemma 7 make these results precise.

Lemma 6. If the seller follows Algorithm 1, then with probability at least $1-\delta \log (m)$, the number of adoptions at the end of time horizon $T$ is lower bounded as:

$$
d_{T} \geq m X_{T}^{*}-\sqrt{8 m X_{T}^{*} \log \left(\frac{4}{\delta}\right)}
$$

Lemma 7 (Upper bound on optimal prices). All prices in the optimal price curve for deterministic Bass model are upper bounded as:

$$
p^{*}(x, \alpha, \beta) \leq \log (e+(\alpha+\beta) T)
$$

These two results combined tell us that, compared to the optimal, we lose at most $\tilde{O}(\sqrt{m} \log (T))$ revenue due to fewer adoptions. The dominant term in regret comes from the seller losing roughly $\tilde{O}\left(m^{-1 / 3}\right)$ revenue on each customer, which led to $\tilde{O}\left(m^{2 / 3}\right)$ revenue loss per epoch in Step 2.

Step 4 (Putting it all together). Finally, we put the previous steps together to prove Theorem 1. Since the number of customers in each epoch grows geometrically and there are at most $m$ customers in total, the number of epochs is bounded by $O(\log (m))$. By Step 2, the algorithm loses at most $\approx m^{2 / 3}$ revenue on adoptions in each epoch. By Step 3, it loses at most $\approx \sqrt{m} \log (T)$ revenue due to missed adoptions. The total regret is therefore bounded by $\approx\left(\log (m) m^{2 / 3}+\sqrt{m} \log (T)\right)$.

### 1.7 Regret lower bound

We prove a lower bound on the regret of any dynamic pricing and learning algorithm under the following mild assumptions on algorithm design. Through out our discussion of the lower bound, we also make two additional assumptions on the pricing policy:

Assumption 1. The algorithm can change price only on arrival of a new customer. The price is held constant between two arrivals.

Assumption 2. Given a planning horizon T, the price offered by the algorithm at any time $t \in$ $[0, T]$ is upper bounded by $p_{\max }:=\log (T)+C(\alpha, \beta)$, for some function $C$ of $\alpha, \beta$.

The above assumptions are indeed satisfied by Algorithm 1 since it changes prices only on arrival of a new customer, and the prices offered are clamped to the range $[0, \log (e+\phi T)]$ (refer to (1.11)) where $\phi$ is a constant upper bound on $\alpha+\beta$. These assumptions are also satisfied by the optimal dynamic pricing policy for the deterministic Bass model since the optimal prices are bounded by $\log (e+(\alpha+\beta) T)$ (refer to Lemma 7). Note that since customer arrivals are continuous in the deterministic Bass model, Assumption 1 is vacuous in that setting.

We argue that these assumptions do not significantly handicap an algorithm and preserve the difficulty of the problem. Assuming an upper bound on the price is a common practice in dynamic pricing literature. Indeed, unlike most existing literature which assumes a constant upper bound, Assumption 2 allows the price to potentially grow with the planning horizon. Intuitively, given enough time to sell the product, the seller should be able to potentially increase prices in exchange
for a slower adoption rate. ${ }^{4}$ Such a dynamics is observed in the optimal dynamic pricing policy for deterministic Bass model, where the price can grow with the time horizon. However, the optimal prices are still uniformly upper bounded by $\log (e+(\alpha+\beta) T)$.

Furthermore, in the proof of Lemma 2 (specifically, refer to Lemma 24 in Appendix A.3), we show that there exist pricing policies in the stochastic Bass model that satisfy both the above assumptions and achieve an expected revenue that is at most $\tilde{O}(\sqrt{m})$ additive factor away from the deterministic optimal revenue $V^{\text {det }}(T)$. Since the lower bound provided in this section is of the order $m^{2 / 3}$, this indicates that removing these assumptions from algorithm design is unlikely to provide an advantage significant enough to overcome the current lower bound.

Our main result in this section is the following regret lower bound on any algorithm under the above assumptions.

Theorem 2 (Regret lower bound). Fix any $\alpha>0, \beta>0$, and $T=\frac{2(1+\sqrt{2}) e}{\alpha+\beta}$. Then, given any pricing algorithm satisfying Assumption 1 and 2, there exist Bass model parameters ( $\alpha, \beta^{\prime}$ ) with $\beta^{\prime} \in\left[\beta, \beta+\frac{(\alpha+\beta)^{2}}{\alpha}\right]$ such that the expected pseudo-regret of the algorithm in this market has the following lower bound:

$$
\mathbb{E}[\text { Pseudo-Regret }(T)] \geq \Omega\left(\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right) .
$$

Here the expectation is taken with respect to the stochasticity in the Bass model as well as any randomness in the algorithm.

Note that given the relation between pseudo-regret and regret in Lemma 2, the above theorem directly implies an $\Omega\left(m^{2 / 3}\right)$ lower bound on $\mathbb{E}[\operatorname{Regret}(T)]$.

Lower bound implications Theorem 2 highlights the regime of horizon $T$ where this learning problem is the most challenging. In this problem, we observe that if $T$ is large, the seller can simply offer a high price for the entire time horizon and still capture most of the market, making

[^3]the problem trivial. Specifically, for $T \geq \Omega(\log (m))$, under an additional assumption that there is a constant upper bound on price, we can show that an $O(\log (m))$ upper bound on regret can be trivially achieved by offering the maximum price at all times (see Lemma 32). On the other hand, when $T=o(1)$, then we can show that achieving a sub-linear (in $m$ ) regret is trivial for any algorithm. This is because from (1.5) we have that in this case the optimal number of adoptions $X_{T}^{*} \leq \frac{\alpha+\beta}{e} T=o(1)$. Furthermore, from Lemma 7 we know that all prices in the optimal curve can be bounded by $p_{\text {max }}^{*}=O(1)$ in this case. This means that the optimal revenue is at most $m X_{T}^{*} p_{\max }^{*}=o(m)$, i.e., sublinear in $m$. Intuitively, for such a small $T$, the word-out-mouth effect never comes into play (i.e., the $\alpha+\beta x$ term in the Bass model is dominated by $\alpha$ ), making the problem uninteresting. Our lower bound result therefore pinpoints the exact order of $T$, i.e., $T=$ $\Theta(1 /(\alpha+\beta))$, where the difficult and interesting instances of this problem come from.

In the rest of this section, we describe the intuition and an outline for proving 2.

Proof Intuition We start with showing that the pseudo-regret of any algorithm can be lower bounded in terms of its cumulative pricing errors. Therefore, in order to achieve low regret, any algorithm must be able to estimate optimal prices accurately. Next, we observe that in this problem, the main difficulty for any algorithm is in estimating the market parameter $\beta$. Since $\beta$ 's observed influence on the arrival rate of customers is proportional to the current adoption level $x$, one cannot estimate $\beta$ accurately when $x$ is small. This makes intuitive sense because when the number of adopters is small, we do not expect to be able to measure the word of mouth effect accurately. In fact, we demonstrate that for any $\varepsilon$, before the adoption level exceeds $\left(\frac{m}{\varepsilon}\right)^{2 / 3}$, no algorithm can distinguish two markets with Bass model parameters $(\alpha, \beta)$ vs. $(\alpha, \beta+\varepsilon)$. Further, we can show that in some problem instances (specifically for instances with $T=\Theta(1 /(\alpha+\beta))$ ), the optimal prices for the first $m^{2 / 3}$ customers are very sensitive to the value of $\beta$. Therefore, if an algorithm's estimation of $\beta$ is not accurate, it cannot accurately compute the optimal price for these customers. This presents an impossible challenge for any algorithm: it needs an accurate estimate of $\beta$ in order to compute an accurate enough optimal price for the first $m^{2 / 3}$ customers, but it cannot possibly
obtain such an accurate estimate while the adoption level is that low; thus, it must incur pricing errors resulting in the given lower bound on regret.

Proof Outline A formal proof of Theorem 2 is obtained through the following four steps. All the missing proofs from this section can be found in Appendix A.5.

Step 1. First we show that the pseudo-regret of an algorithm can be lower bounded in terms of cumulative pricing errors. Note that this result is not a priori obvious because as we have discussed earlier, prices have long term effects on the adoption curve: the immediate loss of revenue in the current round by offering too low of a price might be offset by the fact that we saved some time for the future rounds (lower price means faster arrival rate). On the other hand, if we offer a price that is higher than the optimal price, the resulting delay in customer arrival (higher price means slower arrival rate) may lead to less remaining time and fewer adoptions, which could be more harmful than the immediate extra revenue. The result proven here is crucial for precisely quantifying these tradeoffs and lower bounding the regret in terms of pricing errors.

To obtain this result, we first lower bound the impact of offering a suboptimal price at any time, on the remaining value in the deterministic Bass model. Given any current adoption level $x$ and remaining time $T^{\prime}$, the instantaneous impact or 'disAdvantage' $A^{\text {det }}\left(x, T^{\prime}, p\right)$ of price $p$ is defined as the overall decrease in value over the remaining time when $p$ is offered for an infinitesimal time and then optimal policy is followed. That is,

$$
A^{\operatorname{det}}\left(x, T^{\prime}, p\right):=\lim _{\delta \rightarrow 0} \frac{V^{\operatorname{det}}\left(x, T^{\prime}\right)-p \lambda(p, x) \delta-V^{\operatorname{det}}\left(x+\lambda(p, x) \delta / m, T^{\prime}-\delta\right)}{\delta}
$$

We prove the following lemma lower bounding this quantity.

Lemma 8. At any adoption level $x$ and remaining time $T^{\prime}$, the disadvantage of offering a suboptimal price $p$ in the deterministic Bass model is lower bounded as:

$$
A^{d e t}\left(x, T^{\prime}, p\right) \geq \lambda(p, x) \min \left(\frac{1}{4}\left(\pi^{*}\left(x, T^{\prime}\right)-p\right)^{2}, \frac{1}{10}\right)
$$

where $\pi^{*}\left(x, T^{\prime}\right)=\arg \min _{p} A^{\text {det }}\left(x, T^{\prime}, p\right)$ denotes the optimal price at $x, T^{\prime}$.

Now recall that pseudo-regret is defined as the total difference in revenue of the algorithm, which is potentially offering suboptimal prices, compared to the deterministic optimal revenue $V^{\text {det }}(T)$. We use the above result to quantify the impact of offering suboptimal prices for the first $n \approx m^{2 / 3}$ customers, along with a bound on difference in stochastic vs. deterministic optimal revenue, to obtain the following lower bound on the pseudo-regret:

$$
\mathbb{E}[\text { Pseudo-Regret }(T)] \geq \mathbb{E}\left[m \int_{0}^{n / m} \min \left(\frac{1}{4}\left(\pi_{x}^{*}-p_{x}\right)^{2}, \frac{1}{10}\right) d x\right]-\tilde{O}\left(m^{1 / 3}\right) .
$$

Here, $p_{x}$ denotes the price trajectory obtained on using the algorithm's prices in the deterministic Bass model, and $\pi_{x}^{*}$ denote the prices that would minimize the disadvantage at each point in this trajectory. A more precise statement of this result is in Lemma 29 in Appendix A.5.

The remaining proof focuses on lower bounding the cumulative difference in pricing errors $\left(\pi_{x}^{*}-p_{x}\right)^{2}$ that any algorithm must make for the first $n \approx m^{2 / 3}$ customers.

Step 2. Consider two markets with parameters $(\alpha, \beta)$, and $(\alpha, \beta+\epsilon)$ for some constant $\epsilon$. We show that for the first $n \approx\left(\frac{m}{\epsilon}\right)^{2 / 3}$ customers, any pricing algorithm will be "wrong" with a constant probability. Here by being wrong we mean that the algorithm will set a price that is closer to the optimal price of the other market. Lemma 30 and Corollary 4 formalize this idea. The proof is based on a standard information theoretic analysis using KL-divergence.

Step 3. Next we show that for $T=\Theta\left(\frac{1}{\alpha+\beta}\right)$, the difference between the optimal prices for the two markets $((\alpha, \beta)$ and $(\alpha, \beta+\varepsilon))$ is large $\left(\approx \frac{\alpha \varepsilon}{(\alpha+\beta)^{2}}\right.$. We show this by proving a bound on the derivative of optimal price with respect to $\beta$. Lemma 31 in Appendix A. 5 gives the precise statement.

Step 4. Finally, we put together the observations made in the previous three steps to prove Theorem 2 . We have shown that with constant probability any algorithm will make a pricing mistake for
the first $\approx\left(\frac{m}{\epsilon}\right)^{2 / 3}$ customers [Step 2] and that this mistake will be large (on the order of $\epsilon$ ) [Step 3] under the condition that $T=\Theta(1 /(\alpha+\beta))$. We also have a lower bound on pseudo-regret in terms of total (square of) pricing errors made over the first $n \approx m^{2 / 3}$ customers [Step 1]. Combining these observations with an appropriately chosen $\epsilon$ gives us that the pseudo-regret is lower bounded by $O\left(\epsilon^{2}\left(\frac{m}{\epsilon}\right)^{2 / 3}\right) \approx O\left(m^{2 / 3}\right)$.

All the missing details of this proof can be found in Appendix A.5.

### 1.8 Conclusion and Future Directions

In this Chapter we investigated a novel formulation of dynamic pricing and learning, with a non-stationary demand process governed by an unknown stochastic Bass model. In particular, we presented an online algorithm that learns to price from past observations without a priori knowledge of the model parameters. A key insight that we derive and utilize in our algorithm design is the concavity of the optimal value in the deterministic Bass model setting. Using this concavity property, we can show that the optimal value in the deterministic Bass model is always higher than in the stochastic model, and therefore can be used as a stronger benchmark to compete with. Based on this insight, the main algorithmic idea is to follow the optimal price curve for the deterministic model but with estimated model parameters substituted in place of their true values.

Our main technical result is an upper bound of $\tilde{O}\left(\mathrm{~m}^{2 / 3}\right)$ on the regret of our algorithm in a market of size $m$, along with a matching lower bound of $\Omega\left(m^{2 / 3}\right)$ under mild restrictions on algorithm design. Thus, our algorithm has close to the best performance achievable by any algorithm for this problem. The derivation of our lower bound is especially involved, and requires deriving novel dynamic-programming based inequalities. These allow for lower bounding the loss in long-term revenue in terms of instantaneous pricing errors made by any non-anticipating algorithm.

An interesting and perhaps surprising aspect of our bounds is the role of the horizon $T$ vs. market size $m$. Our upper bound depends sublinearly on $m$ but only logarithmically on the horizon $T$. And in fact our lower bound indicates that for any fixed $\alpha, \beta$ the most "interesting" (and challenging) instances of this problem are characterized by $T$ which is of constant order, and large $m$.

This highlights the distinct nature of pricing under state-dependent models, like the Bass model, when compared to the independent demand models and multi-armed bandit based formulations where asymptotics with respect to $T$ form the main focus of the analysis. Interesting directions for future research include investigation of other state-dependent demand models where the concavity property and other new dynamic programming based insights derived here may be useful.

## Chapter 2: Dominant Resource Fairness with Meta-Types

Taking a step back from the specific dynamic pricing application that we studied in Chapter 1, we can view the problem from the broader perspective of a central planner having to make sequential decisions while learning the underlying dynamics of the system over time. In order to achieve good performance, the central planner has to take actions that not only give him good short and long term reward, but also allow him to estimate the unknown parameters quickly. In other words, the challenge mainly came from the "learning" component. In this chapter, we turn our attention to a resource allocation problem where the main challenge stem from the potentially strategic agents. Agents report how much they value the items and the central planner decides on an allocation based on the reported values. This creates obvious incentives for the agents to potentially misreport their valuations so that they could get more or better items. As we will see later in the chapter, we are able to take advantage of the structure of the agents' utility function in order to design an allocation algorithm that is robust to strategic behaviors.

### 2.1 Background and Motivation

The recent COVID-19 pandemic has brought forward many problems that are particularly relevant to the operations research and computer science communities. Among them, an often overlooked problem is the effective and fair allocation of resources, such as volunteer medical workers, ventilators, and emergency field hospital beds.

There are several key challenges to the medical resource allocation problem in the face of an infectious disease outbreak. First, utilities from different types of resources are not additive nor linear. For example, when there are enough nurses but not enough doctors, the marginal utility of having one additional nurse on staff is very low. Second, not all resources are accessible to all
hospitals (referred to henceforth as accessibility constraints). For instance, the home location of each volunteer medical worker determines where she can commute to work; thus, she can only be assigned to hospitals within her commutable radius. Third, hospitals have different capacities (big medical centers versus small hospitals) and are in different stress levels (hospitals in an epicenter versus the ones in rural areas with few cases), so they should naturally be prioritized differently.

Another setting that has the above characteristics is the compute resource sharing problem with sub-types. For example, suppose a compute server has several compute nodes, and there are different types of GPU/CPU on the various nodes (e.g. NVIDIA vs. AMD GPU, Intel vs. AMD CPU). Some users look for specific hardware configurations (e.g. accept only Intel CPU) while others might be less selective (e.g. accept all CPU brands).

In this paper, we propose a new market mechanism that tackles the three challenges outlined above and achieves desirable properties including Pareto optimality, envy-freeness, strategyproofness, and sharing incentive. In our numerical experiments, we demonstrate that compared to the Maximum Nash Welfare (MNW), and the Discrete MNW approach, our mechanism is cheaper to compute and enjoys theoretical properties that MNW approaches do not have.

### 2.2 Literature Review

Recently, a flurry of papers have come out of the operations research, statistics, and computer science communities addressing resource allocations in the aftermath of a pandemic. For instance, Jalota et al. [39] proposed a mechanism for allocating public goods that are capacity constrained due to social distancing protocols, focusing on achieving a market clearing outcome. Mehrotra et al. [40] studied the allocation of ventilators under a stochastic optimization framework, minimizing the expected number of shortages in ventilators while also considering the cost of transporting ventilators. Zenteno [41] combined influenza modeling techniques with robust optimization to handle workforce shortfall in a pandemic. These papers differ from our work in that they do not explicitly address any fairness considerations that we study in this paper.

There has also been a growing interest in developing resource allocation mechanisms with
fairness properties. Under a fairly general class of utility functions including the Leontief utility, computing market equilibrium under the fisher market setting (divisible goods) can be done using an Eisenberg-Gale (EG) convex program Eisenberg and Gale [42]. Market equilibrium solutions satisfy Pareto optimality, proportionality, and envy-freeness. It is also known that an EG convex program implicitly maximizes Nash welfare, which is the product of all agents' utilities. However, MNW is generally not strategy-proof, and can be computationally expensive for large problems. Chen et al. [43] studied the computation and approximation of market equilibria under the socalled "hybrid Linear-Leontief" utilities. They assume that there are $m$ disjoint groups and each group contains several types of resources. An agent's utility is a linear combination of $m$ Leontief utilities, each associated with a group. Note that although we use the terms "meta-type" and "group" in our problem formulation, our setting is different from theirs because there is only one Leontief utility per user in our setting.

For Leontief utilities, Ghodsi et al. [44] introduced the Dominant Resource Fairness allocation mechanism (DRF) which in addition to the three properties satisfied by market equilibrium solutions, is also strategy-proof. Later Parkes et al. [45] extended the setting to allow agents to be weighted and have zero demand over some resources while maintaining all four desiderata. Our paper generalizes the setting further by allowing accessibility constraints, which as mentioned in the introduction, arise naturally in many practical settings.

For indivisible resources, Caragiannis et al. [46] showed that maximizing Nash welfare with integer constraints (Discrete MNW) satisfies envy-freeness up to one resource unit and has nice guarantees on the Max-Min Share ratio. However, similar to MNW, it is not strategy-proof and as we will see in the numerical section, it does not scale well to large number of agents and resource types. Although exact market equilibrium might not exist in indivisible settings, Budish [47] showed that a close approximation of it exists in the unweighted, binary allocation case. This was later put into practice for course allocation in Budish et al. [48]. However, the theory does not provide useful approximation bounds when assignments are not binary (e.g., each student only needs one seat from each class, but each hospital may require hundreds of doctors), and therefore
it is not well suited to our setting.

### 2.3 Problem Formulation

For the remainder of the paper we use local medical personnel allocation as a running example, even though other resource allocation problems can be formulated in a similar fashion. We group resources into meta-types: doctors, nurses, ventilators, emergency field hospital beds, etc. Within each meta-type (e.g. doctors), we have types (e.g. doctors from the Bronx, doctors from Brooklyn, doctors from Manhattan, etc. ${ }^{1}$ ). We assume that demands are given over meta-types, but each agent sometimes can only receive allocation from a subset of the resources in a meta-type because of constraints such as location (e.g. a hospital is indifferent to where doctors assigned to it come from as long as they are within commutable radius). We refer to such subsets of resource types in each meta-type as the agents' demand groups.

Let $\Omega_{1}, \Omega_{2}, \ldots, \Omega_{L}$ denote the meta-types. Each meta-type $\Omega_{l}$ is a collection of resource types. We assume that $\Omega_{i} \cap \Omega_{j}=\emptyset$ for any two different meta-types $i, j$, which means that each resource type belongs to exactly one meta-type. Let $R$ denote the set of all resource types: $R=\cup_{l \in[L]} \Omega_{l}$, and $N$ denote the set of agents. We use $m=|R|$ and $n=|N|$ to denote the total number of resource types and agents. Each type of resource $r$ has a finite supply of $S_{r}$. We assume that the supplies are normalized within each meta-type:

$$
\sum_{r \in \Omega_{l}} S_{r}=1 \quad \forall l \in[L] .
$$

Each agent $i \in N$ submits a demand vector $\left[d_{i 1}, \ldots, d_{i L}\right]$ where $d_{i l}$ denotes the fraction of metatype $l$ that agent $i$ needs in order to get one unit of utility (one can think of this as each agent trying to complete as many units of work as possible, where each unit of work requires $d_{i l}$ units of meta-type $l$ ). Additionally, each agent has access to only a subset of resource types within each meta-type. We represent this accessibility constraint in the form of a set of demand groups. Let

[^4]$G_{i}=\left\{g_{l}^{i} \subseteq \Omega_{l}: l \in[L], d_{i l}>0\right\}$, be the set of demand groups for agent $i$, where $g_{l}^{i} \subseteq \Omega_{l}$ is agent $i$ 's demand group for $l$, specifying the subset of resource types belonging to meta-type $l$ that agent $i$ can access. Note that we only include in $G_{i}$ meta-types that agent $i$ has non-zero demand of. This is to simplify notation in the later analysis. Intuitively, the introduction of meta-types models the substitution effects, and the introduction of demand groups models the accessibility constraints. When $i$ is clear from the context, we sometimes use $g_{l}$ instead of $g_{l}^{i}$ to simplify the notation.

Following the setup in Parkes et al. [45], we also allow agents to be weighted differently for each meta-type and we denote the weight of agent $i$ for meta-type $l$ as $w_{i l}$. Having different weights for different meta-types makes the model more expressive: if we let $w_{i 1}=\ldots=w_{i L}$, then this reduces to having a single priority weight for each agent. This weight can depend on factors such as agent $i$ 's contribution to the resource pool of meta-type $l$, as well as the size and stress level of agent $i$. In the case of medical supplies allocation, weights can represent how much each hospital is in need of extra resources. In the cloud compute setting, weights can represent how much money each user has paid for each meta-type of resource. Note that weights are fixed apriori, not self reported by the agents, nor determined by the allocation algorithm. We assume that weights are normalized within each meta-type: $\sum_{i \in N} w_{i l}=1$ for $l \in[L]$. Note that weights represent agents' priorities over the meta-types, not agents' preferences. Therefore they do not appear in the agents' utility functions, as we will define next.

Let $x_{i}$ be the allocation vector of agent $i: x_{i r}$ represents the assignment of resource type $r$ to $i$. For each meta-type $l, \sum_{r \in g_{l}^{i}} x_{i r}$ is the fraction of the total supply of meta-type $l$ that is assigned to agent $i$. The utility of agent $i$ is then defined as:

$$
\begin{equation*}
u_{i}\left(x_{i}\right):=\min _{g_{l} \in G_{i}}\left\{\frac{1}{d_{i l}} \sum_{r \in g_{l}} x_{i r}\right\} \tag{2.1}
\end{equation*}
$$

Since agent $i$ needs $d_{i l}$ fraction of each meta-type $l$ to finish one unit of work, $u_{i}\left(x_{i}\right)$ is the total units of work that agent $i$ can finish given allocation vector $x_{i}$. This form of utility measure is called the Leontief utility.


Figure 2.1: All three hospitals can accept both types of doctors. However, hospitals I and II can only accept Nurse type C, while hospital III accepts only Nurse type D.

We now give a concrete example, which is also illustrated in Figure 2.1. To keep the example simple we assume that each agent has the same weight over all meta-types: only one weight $w_{i}$ is defined for each agent $i$.

Example 1. Consider a case of three hospitals (agents) $\{1,2,3\}$ and two resource meta-types. The first meta-type consists of two types of doctors (resource $A, B$ ), and the second consists of two types of nurses (resource $C, D): \Omega_{1}=\{A, B\}, \Omega_{2}=\{C, D\}$. The normalized weights for the three hospitals are: $w_{1}=w_{2}=\frac{1}{4}, w_{3}=\frac{1}{2}$. The supply for each type of doctor and nurse is 500. Thus, the total available units of each meta-type is $500+500=1000$, and $S_{r}=\frac{500}{1000}=$ $\frac{1}{2} \forall r \in\{A, B, C, D\}$. All three hospitals have access to both types of doctors but hospitals 1,2 only have access to nurse type $C$ while the third hospital only has access to nurse type $D$ : $G_{1}=\left\{g_{1}^{1}=\{A, B\}, g_{2}^{1}=\{C\}\right\}, G_{2}=\left\{g_{1}^{2}=\{A, B\}, g_{2}^{2}=\{C\}\right\}, G_{3}=\left\{g_{1}^{3}=\{A, B\}, g_{2}^{3}=\{D\}\right\}$. For each unit of utility, hospital 1 demands 4 doctors and 1 nurse, hospital 2 demands 1 doctor and 4 nurses, and hospital 3 demands 1 doctor and 1 nurse. Since the total units of supply for each meta-type is $1000, d_{1}=\left[\frac{4}{1000}, \frac{1}{1000}\right], d_{2}=\left[\frac{1}{1000}, \frac{4}{1000}\right], d_{3}=\left[\frac{1}{1000}, \frac{1}{1000}\right]$.

### 2.3.1 Desirable Properties

Now we formally define the properties studied in this paper.

Pareto optimality. An allocation mechanism is Pareto optimal if compared to the output allocation $x$, there does not exist another allocation $x^{\prime}$ where some agent is strictly better off without some other agent being strictly worse off: $\exists i$ s.t. $u_{i}\left(x_{i}^{\prime}\right)>u_{i}\left(x_{i}\right) \Longrightarrow \exists j$ s.t. $u_{j}\left(x_{j}^{\prime}\right)<u_{j}\left(x_{j}\right)$.

Weighted envy-freeness. Given an allocation $x_{j}$ for agent $j$, let $\tilde{x}_{i}$ be the same allocation adjusted to agent $i$ 's weights and demand groups, i.e., $\tilde{x}_{i r}=x_{j r} \frac{w_{i l}}{w_{j l}}$ for all $r \in g_{l}^{i}, l \in[L]$, and $\tilde{x}_{i r}=0$ otherwise. $u_{i}\left(\tilde{x}_{i}\right)-u_{i}\left(x_{i}\right)$ is how much $i$ envies $j$. An allocation is weighted envy free if for any $i, j \in N$ this quantity is non-positive, i.e.,

$$
u_{i}\left(\tilde{x}_{i}\right)-u_{i}\left(x_{i}\right) \leq 0 .
$$

Intuitively, this means an agent prefers her allocation over the allocation of any other agent scaled by the weight ratios of the two agents. Note that since there is a separate weight for every metatype $l$, the allocations for each resource type $r$ is scaled according to the corresponding weight for the meta-type that it belongs to.

Strategy-proofness. In the existing literature, agents can only be strategic by misreporting their demand vector. In our setting however, agents have the additional possibility of misreporting their accessibility constraints for the meta-types (e.g. One can report that she accepts both Intel and AMD CPUs but in fact her program only runs on Intel CPU). Our definition of strategy-proofness guards against both types of misreporting. Let $x$ be the allocation returned by the mechanism under truthful reporting from all agents. Let $x^{\prime}$ be an allocation returned by the mechanism when all agents report truthfully except agent $i$ reports an alternative demand vector and/or alternative demand groups. The mechanism is strategy-proof if $u_{i}\left(x_{i}\right) \geq u_{i}\left(x_{i}^{\prime}\right)$.

Sharing incentive. In settings where the supplies for each resource come from the participants' contribution, sharing incentive is satisfied when the resulting allocation gives each participant at least as much utility as she originally had. More specifically, for each $i \in N$ and $l \in[L]$, let $s_{i l}$ be the proportion of meta-type $l$ contributed by agent $i$ that she can also access. We can also think of $s_{i l}$ as the amount of "useful" resource agent $i$ originally possessed of meta-type $l$ (she might contribute more than $s_{i l}$ to the pool). Prior to reallocation of resources, agent $i$ 's utility would be

$$
u_{i}^{o}:=\min _{g_{l} \in G_{i}}\left\{\frac{s_{i l}}{d_{i l}}\right\} .
$$

Sharing incentive says that $u_{i}\left(x_{i}\right) \geq u_{i}^{o} \forall i \in N$, where $x_{i}$ is the output allocation of the algorithm (i.e., all agents have incentives to share (pool) their individual resources for reallocation). ${ }^{2}$

### 2.4 Dominant Resource Fairness with Meta-Types

Before describing the algorithm we first define some key concepts used in the algorithm:

$$
l_{i}^{*}:=\arg \min _{l \in[L]} \frac{w_{i l}}{d_{i l}} \quad d_{i *}:=d_{i l_{i}^{*}} \quad w_{i *}:=w_{i l_{i}^{*}}
$$

Namely, $l_{i}^{*}$ is the meta-type from which agent $i$ demands the biggest proportional share, adjusted by her priority weights. We refer to $l_{i}^{*}$ as the dominant resource meta-type for agent $i$. $d_{i *}$ is the proportional share demanded by agent $i$ from its dominant resource meta-type to finish one unit of work.

We now present our fair allocation mechanism, which we call Dominant Resource Fairness with Meta-Types (DRF-MT). The mechanism proceeds in rounds and agents are gradually "eliminated". In each round $t$, we use the linear program described in (2.2) to maximize a fractional value $y_{t}$ so that each remaining agent $i\left(i \in N_{t}\right)$ receives at least $y_{t} \times w_{i *}$ fraction of the total supply from its' dominant resource meta-type $l_{i}^{*}$, and more generally $y_{t} \times w_{i *} \times d_{i l} / d_{i *}$ of each demanded

[^5]meta-type $l .{ }^{3}$ Based on this solution, we eliminate at least one resource and one agent using Definition 1 and 2 (although the algorithm only needs to explicitly maintain a list of active/eliminated agents, not resources). For each agent $i$ eliminated in round $t$, we set $\gamma_{i}=y_{t}$. We fix the fraction of dominant meta-type $l_{i}^{*}$ assigned to agent $i$ to $\gamma_{i} \times w_{i *}$, without fixing the specific allocations of the resources. We first observe the following fact (proof is in the Appendix):

Fact 1. In any round $t$ of Algorithm 2, the allocation constraints in Equation 2.2 for $i \notin N_{t}$ are tight for optimal solutions.

This fact implies that when an agent is eliminated, her utility in the final allocation is fixed, even though the exact allocation is not. Not fixing the allocation is a deliberate algorithmic design choice because agents who are flexible with their demand groups should accommodate agents who are more restrictive (e.g. if agent 1 accepts both type A and B, and agent 2 only accepts type A, then we should allocate agent 1 mostly type B resource, and leave type A resource for agent 2). When the number of agents and resource types is large, it is difficult to characterize such dynamics explicitly. So it is crucial to not fix the allocation to the agents until the last iteration.

We will show that there is at least one new resource and one agent being eliminated in each round. Thus our algorithm requires at most $\min (m, n)$ rounds (in practice it often terminates in 2-3 rounds even with a large number of resources types and agents). Since each round requires solving a polynomial-sized linear program, the overall procedure can be run in polynomial time.

Let $N_{t}, R_{t}$ be the set of active agents and resources at the beginning of round $t$. The LP for round $t$ is defined in (2.2). Note that the ratio $\frac{d_{i l}}{d_{i *}}$ is simply making sure that there is no waste in the allocation. For an agent who has been eliminated, $\frac{\gamma_{i} w_{i *}}{d_{i *}}$ is her final utility. If agent $i$ is not yet eliminated after round $t$, then $\frac{y_{t} w_{i *}}{d_{i *}}$ represents how much utility she is currently guaranteed to

[^6]receive (it will never decrease in later rounds, see Fact 2).
$\max y_{t}$
s.t. (active agents allocation constraints)
$y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}} \leq \sum_{r \in g_{l}} x_{i r} \quad \forall i \in N_{t}, g_{l} \in G_{i}$
(eliminated agents allocation constraints)
$\gamma_{i} \times w_{i *} \times \frac{d_{i l}}{d_{i *}} \leq \sum_{r \in g_{l}} x_{i r} \quad \forall i \notin N_{t}, g_{l} \in G_{i}$
(supply constraints)
$\sum_{i \in N} x_{i r} \leq S_{r} \quad \forall r \in R$
(non-negativity constraints)
$x_{i r} \geq 0 \quad \forall i \in N, r \in R$

Fact 2. The optimal value for Equation 2.2 is non-decreasing over rounds: $y_{1}^{*} \leq y_{2}^{*} \leq \ldots$, where $y_{t}^{*}$ is the optimal objective function value of the LP in round $t$.

This follows because the constraints on eliminated agents are less restrictive than the constraints on active agents, and the set of active agents is decreasing over time.

Definition 1. Resource $r$ is eliminated in round $t$ if $t$ is the first round in Algorithm 2 in which $\sum_{i \in N} x_{i r}=S_{r}$ for every optimal $x$.

By Fact 2 it is also easy to see that the set of remaining resources $R_{t}$ decreases over time.

Definition 2. We give two equivalent definitions for eliminating agents:

- Agent $i$ is eliminated in round $t$ when there exists $g_{l} \in G_{i}$ such that $g_{l} \cap R_{t+1}=\emptyset$.
- Agent $i$ is eliminated in round $t$ when there exists $g_{l} \in G_{i}$ such that $y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}}=\sum_{r \in g_{l}} x_{i r}$ for every optimal $x, y_{t}$.

Intuitively, both definitions are saying that agent $i$ can not improve her utility further in later rounds. Due to space constraint we defer the proof of their equivalence, and most of the other results to the Appendix. We include the proof of Claim 1 here because it is a good representation of the flavor of arguments used in other proofs. First we address the question of whether the DRF-MT can be efficiently implemented.

```
ALGORITHM 2: Dominant Resource Fairness with Meta-Types (DRF-MT)
    Input: Agents \(N\), resources \(R\), supplies \(S_{r} \forall r \in R\), demand groups \(G_{i} \forall i \in N\), normalized
    demands \(d_{i l} \forall i \in N, g_{l} \in G_{i}\), priority weights \(w_{i l} \forall i \in N, l \in[L]\)
    Initialize \(N_{0}=N\)
    for \(t \leftarrow 0,1,2, \ldots\) do
        \(y_{t}^{*} \leftarrow\) Solve (2.2)
        Update the remaining active agents \(N_{t+1}\) (using Claim 2)
        for agent i eliminated in this round do
                \(\gamma_{i} \leftarrow y_{t}^{*}\)
        end
        if \(N_{t+1}=\emptyset\) then
            Solve Equation 2.2 and assign resources according to \(x_{i r}\) with rounding
            break
        end
    end
```

Claim 1. In each round t of Algorithm 2, at least one remaining resource $r \in R_{t}$ and one remaining agent $i \in N_{t}$ is eliminated.

Proof. Suppose no resource is eliminated in round $t$, then for each $r \in R_{t}$, there exists an optimal solution such that $\sum_{i \in N} x_{i r}<S_{r}$. Then the convex combination of these solutions gives us an optimal solution $x^{*}$ that satisfies $\sum_{i \in N} x_{i, r}^{*}<S_{r} \forall r \in R_{t}$. However, by Definition 2, for every remaining agent $i \in N_{t}, g_{l} \cap R_{t} \neq \emptyset \forall g_{l} \in G_{i}$. So if we assign a little more of every active resource to every active agent, then the overall objective value would be higher. This contradicts the optimality of the LP.

Now suppose some resource $r \in R_{t}$ is eliminated in round $t$ but no agent is eliminated. Suppose this resource type is part of meta-type $l$. By the first definition in Definition 2, this means that for every $i$ such that $x_{i r}>0$, there exists $r^{\prime} \in g_{l}^{i}$ such that $r^{\prime}$ is not eliminated. By the same convex
combination argument above, we know that there is an optimal solution such that $\sum_{i} x_{i r^{\prime}}<S_{r^{\prime}}$ for every such $r^{\prime}$. Then for every such agent we can remove $\epsilon$ allocation of $r$ from her and replace it with $\epsilon$ allocation of the corresponding $r^{\prime}$. This gives us an allocation that has the same objective as before without using up the entire supply of $r$, contradicting $r$ being eliminated.

This result shows that DRF-MT can be implemented efficiently by solving at most $\min (m, n)$ number of polynomial-size linear programs. However, it does not tell us how to find the eliminated agents. The following theorem says that we can do so by looking at the dual variables of the LP. Note that the algorithm does not need to explicitly maintain a list of active resources (Equation 2.2 does not depend on $R_{t}$ ).

Claim 2. This claim has two parts:

- If the shadow price of an allocation constraint of an active agent in round tis positive, then its corresponding agent is eliminated in round $t$.
- In each round $t$, at least one allocation constraint corresponding to an agent in $N_{t}$ has a positive shadow price.

Now we state our main results.

Lemma 9. DRF-MT is Pareto optimal.

Lemma 10. DRF-MT is weighted envy-free.

Lemma 11. DRF-MT is strategy-proof.

The proofs for these three lemmas all involve a case analysis of different scenarios and showing that the undesirable outcomes violate either the optimality of the LP or the definition of eliminated resources/agents, similar to the arguments presented in the proof of Claim 1.

Lemma 12. Assume that demands, weights and supplies are all rational numbers. If priority weights of the algorithm are set according to the each agent's accessible contribution to the resource pool (for each meta-type), then DRF-MT satisfies sharing incentive.

In resource pooling settings, having a separate weight for each meta-type is crucial in proving this result (e.g. contributing a ton of hard drive space but no GPU should not give the agent high priority if GPU is its dominant/bottleneck resource type). The proof constructs a bipartite graph of supplies and demands of the resources, then uses Hall's Theorem [49] to show that there exists a feasible solution to the first round's LP that already gives every agent at least as much utility as they could get without participating in the pool. Since agents' utilities only increase in later rounds, the final allocation must also satisfy sharing incentive.

### 2.4.1 Integral Allocation from Rounding

So far we have implicitly assumed that the resources are divisible, and all fairness results are stated with respect to the fractional assignment output of Algorithm 2. In practice we round down the output to obtain the final assignment, since resources such as ventilators are indivisible. Each agent loses at most 1 unit of each type of resource through rounding. Since we focus on problems where each agent receives hundreds of units of each resource, the performance loss due to rounding is small. For example, starting with an envy-free fractional allocation, one agent can envy another by at most $2 m$ items after rounding. In Example 1, $m$ is 4, while the total allocations each agent receives are in the hundreds. So an envy of $2 m$ items is not significant. Note that such divisibility assumption is also standard in existing DRF literature, which often focus on the compute resource sharing problem: even though CPU cores are discrete, it's common to treat the problem as a continuous problem since there is a large quantity of cores in a compute cluster.

There are many existing algorithms that focus on fair allocation of indivisible goods (e.g. Discrete MNW from Caragiannis et al. [46]). Indivisible resource allocation is particularly important when the quantities of the resources are small (e.g. fairly assigning a car, a house, and a ring to two people). However, as is the case with most discrete optimization problems, these algorithms do not scale well to the sizes that we deal with in a pandemic with hundreds of hospitals and many types of resources. In settings where each agent receives hundreds of units of each resource, the performance loss due to rounding is often small compared to the dramatic increase in computational cost


Figure 2.2: Left: Running time comparison. Middle: Normalized max envy comparison. Right: Distribution of normalized difference in social welfare between Discrete MNW and DRF-MT over all trials. Normalized difference is calculated by subtracting the social welfare of Discrete MNW from that of (rounded) DRF-MT and then dividing by the social welfare of Discrete MNW.
for solving Mixed Integer Programs (see Section 2.5 for a numerical comparison).

### 2.4.2 Connection to Previous Dominant Resource Fairness Algorithms

The core difference between our problem setup and the existing DRF settings ([44], [45]) is the addition of accessibility constraints. When $\left|\Omega_{l}\right|=1$ for each $l \in[L]$, both our problem formulation and the DRF-MT algorithm reduce to the problem and algorithm studied in those papers. Note that in this simplified setting one can write out the closed form solutions to the LPs, so no actual optimization needs to be performed. However, it is natural that resources come in different "flavors" and that agents have different constraints/preferences over these variations. So our formulation captures a much wider range of problems encountered in practice.

### 2.4.3 Alternative Fair Allocation Mechanisms

As discussed in Section 2.2 and Section 2.4.1, the other most suitable approaches in our setting are MNW and Discrete MNW. When the weights are equal, MNW is also commonly referred to as the Competitive Equilibrium with Equal Income (CEEI) approach. Without the accessibility constraints, MNW is known to be Pareto optimal, envy-free and satisfy sharing incentive. However, unlike DRF-MT, it is known that MNW is not strategy-proof (see Section 6 of [44] for an example). We show in Section 2.5 that our DRF-MT mechanism achieves almost as much social welfare (i.e. sum of utilities of all agents) as MNW, and also runs faster in practice.

### 2.4.4 Extension to Arbitrary Group Structure

We currently assume that resources and demands follow a meta-type/group/type structure. One might be interested in a general group structure where a demand group can contain any subset of all possible resources (not necessarily from a single meta-type). The problem with this kind of flexible group structure is that it opens up possibilities for people to cheat the system by misreporting their true demand structure (e.g. instead of reporting that they are indifferent to resource A and B , and that they only need one unit of either one to finish a unit of work, agents can claim that they need one unit each from both A and B to finish one unit of work). In particular, Dominant Resource Fairness based approaches will likely not work, since it is unclear how one would even define the dominant resource under such a general setting. We leave this as an open question for future work.

### 2.5 Numerical Experiments

We compare the algorithms on running time, normalized max envy, and social welfare. Normalized max envy is the maximum envy (see Section 2.3.1) between any pair of agents normalized as a fraction of each agent's allocated utility. Social welfare is the sum of utilities of all agents. We fix a meta-type structure $\left(\Omega_{1}=\{0\}, \Omega_{2}=\{1,2\}, \Omega_{3}=\{3,4,5\}, \Omega_{4}=\{6,7,8,9\}\right)$ and randomly generate the demands, group structures, and weights for the agents. For each choice of number of agents, we ran 16 trials. All three algorithms Allocations are rounded down for MNW and DRF-MT. All three algorithms allow specifying different agent weights and also observe the accessibility constraints. Gurobi[50] and Mosek [51] are used to implement the algorithms. More details on the experimental setup and additional experiments can be found in Appendix B .

First we investigate scalability. As shown in Figure 2.2 (left), the running time for Discrete MNW quickly explodes while MNW and DRF-MT are much more scalable. DRF-MT runtime in particular grows very slowly. The error region represents one standard deviation from the mean.

Recall that DRF-MT is envy-free before rounding. We now investigate envy when the solution is rounded. Without accessibility constraints, MNW is also envy-free before rounding, and Dis-
crete MNW satisfies envy-free up to one good. Figure 2.2 (middle) shows that all three algorithms have small max envy after rounding in practice ( $<4 \%$ in most trials).

Finally, we compare the social welfare obtained under DRT-MT and Discrete MNW. Figure 2.2 (right) shows that in roughly $95 \%$ of the trials, DRF-MT obtained at least $90 \%$ as much social welfare as Discrete MNW. In Appendix B we show that MNW has slightly lower social welfare than Discrete MNW, so the above conclusion holds when DRF-MT is compared to rounded MNW as well.

In conclusion, compared to both Discrete MNW and MNW, DRF-MT 1) achieves almost as much social welfare, 2) has comparable level of max envy, 3) has the additional property of strategy-proofness (in the fractional case), and 4) is more scalable. An interesting avenue for future work is to determine the properties of the rounded variant of DRF-MT. A particularly interesting question would be whether one can show that approximate strategy-proofness holds when there is large supply of each item.

# Chapter 3: Optimal Efficiency Envy Trade-Off via Optimal Transport 

So far we have seen how having partial information of the system (Chapter 1) and dealing with potentially strategic agents (Chapter 2) can lead to different challenges for the central planner who wants to optimize some objective. In this chapter, we study a problem where despite the central planner having full information of the system (which means that the central planner does not need to rely on any input from the agents' themselves) finding a solution is still non-trivial. In particular, we assume that the central planner knows exactly which items are being allocated, and how much each agent values each and every item. As we will show later, the challenge in this setting is a computational one. Instead of relying on statistical tools (which were helpful for tackling the learning challenge in Chapter 1), or the structure of the utility function (which were crucial in solving the problem in Chapter 2), in this chapter we will use tools from optimization and duality theory to design tractable algorithms.

### 3.1 Background and Motivation

In this chapter, we focus on the problem of finding a resource allocation policy that divides a pool of items, represented by a distribution $\mathcal{D}$, to $n$ receivers, under the constraint that each receiver $i$ must be allocated a prespecified fraction $p_{i}^{*}$ of the items, where $p_{i}^{*} \in(0,1), 1^{\top} p^{*}=1$ is an input to the problem that characterizes the priority of each receiver. We refer to $\left\{p_{i}^{*}\right\}_{i=1}^{n}$ as the target matching distribution. In addition to this matching distribution constraint, we also require that the receiver's envy, which we will define formally later, be bounded. Unlike the existing resource allocation literature, where envy-free is either treated as a hard constraint or not considered at all, we allow the central planner to specify the level of envy that is tolerated, and find the most efficient allocation given the amount of envy budget. This allows us to reduce the envy significantly without
paying the full price of fairness with respect to efficiency.
To concretely motivate our model, let us consider the blood donor matching problem that was first studied in [52]. The Meta platform has a tool called Facebook Blood Donations, where users who opt in to receive notifications are notified about blood donation opportunities near them. Depending on the user's and the blood bank's specific characteristics (e.g., age, occupation for the user, and locations, hours for the blood bank), notifications about different donation opportunities have different probabilities of resulting in an actual blood donation. The platform would like to send each user the most relevant notifications (to maximize the total number of potential blood donations), while maintaining certain fairness criteria for all the blood banks that participate in this program. Although the platform can theoretically send each user multiple notifications about multiple blood banks, for user experience and other practical reasons, this is not done, at least in the model introduced in [52]. Therefore in this problem, users' attention is the scarce resource that the platforms needs to allocate to different blood banks. The most natural type of fairness criteria in this setting is perhaps the number of users that received notifications about each of the blood banks. For example, it is not desirable to match zero users to a particular, potentially inconveniently-located, blood bank, even if matching zero users to this blood bank results in more blood donations in expectation.

Another fairness desideratum commonly studied in the literature is called envy-freeness. We say that receiver $A$ envies receiver $B$ if $A$ values $B$ 's allocation more than her own. Intuitively, an allocation that is envy-free-where no agent envies another agent-is perceived to be fair. This chapter considers both of the two fairness criteria mentioned above: we study a setting where the goal is to maximize social welfare under a matching distribution constraint, while ensuring that each receiver has bounded envy. We make the following contributions in regards to this problem:

1. We formulate it as a constrained version of a semi-discrete optimal transport problem and show that the optimal allocation policy has a concise representation and a simple geometric structure. This is particularly attractive for large-scale allocation problems, due to the fast computation of a match given an item. This insight also shines new light on the question of
when envy arises, and when the welfare price on envy-freeness is large.
2. We propose an efficient stochastic optimization algorithm for this problem and show that it has a provable convergence rate of $O(1 / \sqrt{T})$.
3. We investigate the statistical properties of the space of our optimal transport based allocation policies by showing a Probably Approximately Correct (PAC)-like sample complexity bound for approximating the optimal solution given finite samples.

In Section 3.4 we formally define the problem we are interested in. In Section 3.5, we show that this problem can be formulated as a semi-discrete optimal transport problem, whose solution has a simple structure with a nice geometric interpretation. Section 3.6 develops a stochastic optimization algorithm. In Section 3.7, we show that an $\epsilon$-approximate solution can be found with high probability given $\tilde{O}\left(\frac{n}{\epsilon^{2}}\right)$ samples, where $n$ is the number of receivers. Finally, in Section 3.8 we demonstrate the effectiveness of our approach using both artificial and a semi-real data.

### 3.2 Review of Optimal Transport

Since our work draws an explicit connection to optimal transport (OT), we provide a summary of key OT results here. The results in this Section will also be useful later in Chapter 4 where we discuss a different but related resource allocation problem. Let $\alpha, \beta$ be two probability measures on the metric spaces $\mathcal{X}, \mathcal{Y}$ respectively. We define $\Pi(\alpha, \beta)$ as the set of joint probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals $\alpha$ and $\beta$. The Kantorovich formulation of the optimal transport problem [53] can be written as

$$
\begin{equation*}
L(\alpha, \beta):=\min _{\pi \in \Pi(\alpha, \beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d \pi(x, y) \tag{3.1}
\end{equation*}
$$

where $c(x, y)$ is the cost associated with "moving" $x$ to $y$. This is called a transportation problem because the conditional probability $\pi(y \mid x)$ specifies a transportation plan for moving probability mass from $\mathcal{X}$ to $\mathcal{Y}$. If $\beta$ is a discrete measure, i.e. $\mathcal{Y}$ is finite, then it is known from [54] that
the dual to (3.1) can be written as (here we abuse the notation $\beta$ to also represent the vector of probability masses, where $\beta_{i}$ is the probability mass on point $y_{i}$ ):

$$
\begin{equation*}
\max _{g \in \mathbb{R}^{n}} \mathcal{E}(g):=\sum_{i \in[n]} \int_{\mathbb{L}_{y_{i}}(g)} c\left(x, y_{i}\right)-g_{i} d \alpha(x)+g^{\top} \beta \tag{3.2}
\end{equation*}
$$

where $n=|\mathcal{Y}|$, and $\mathbb{L}_{y_{i}}$ is what is sometimes referred to as the Laguerre cell:

$$
\begin{equation*}
\mathbb{L}_{y_{i}}(g)=\left\{x \in \mathcal{X}: \forall i \neq j, c\left(x, y_{i}\right)-g_{i} \leq c\left(x, y_{j}\right)-g_{j}\right\} \tag{3.3}
\end{equation*}
$$

Proposition 1 (Proposition 2.1 [54]). If $\alpha$ is a continuous measure, and $\beta$ a discrete measure, then $L(\alpha, \beta)=\max _{g} \mathcal{E}(g)$, and the optimal solution $\pi$ of (3.1) is given by the partition $\left\{\mathbb{L}_{y_{i}}\left(g^{*}\right), i \in[n]\right\}$, i.e. $d \pi\left(x, y_{i}\right)=d \alpha(x)$ if $x \in \mathbb{L}_{y_{i}}\left(g^{*}\right)$, 0 otherwise.

### 3.3 Literature Review

Blood donation matching. McElfresh et al. [52] introduced this problem and modeled it as an online matching problem, where the matching quality between an user and a blood bank is assumed to be known to the platform. The model formulation there is complex, as it takes into account the fact that not every blood bank is in need of blood every day, a dynamic which we do not consider here. Furthermore, their matching policy is rather cumbersome, requiring a separate parameter for each (donor, receiver) pair. Compared to their paper, we are able to provide better structural insights to the problem by utilizing a simpler model that still captures the most salient part of the problem.

Online Resource Allocation. Another strand of work that our paper is closely related to is that of online resource allocation, especially those with i.i.d. or random permutation input models [55, 56, 57]. Agrawal et al. [58] studied the setting with linear objective and gave competitive ratio
bounds. Then, Agrawal and Devanur [59] generalized the results to concave objectives and convex constraints. Later, Devanur et al. [57] improved the approximation ratio bounds and relaxed the input assumptions on the budgets. Balseiro et al. [60] show that online mirror descent on the dual multpliers does well under both i.i.d. adversarial, and certain non-stationary input settings. However, none of theses papers study the envy-free criterion. Recently, Balseiro et al. [61] studied an online resource allocation problem with fairness regularization. Although the authors did not explicitly study envy regularization, their regularization framework can be modified to accommodate envy regularization. However, like all the other papers mentioned in this paragraph, the offline solution is used as the benchmark to measure regret, but no explicit solution is given to the offline problem. Our analysis focuses on the offline problem, and draws an explicit connection to optimal transport, which allowed us to provide a novel PAC-like analysis on the sample complexity of the problem. In another recent paper, Sinclair et al. [62] studied the trade-off between minimizing envy and minimizing waste, which refers to un-allocated resources. Despite close similarity between our titles, their offline benchmark is the standard Eisenberg-Gale program, which is envy-free, but does not address the welfare cost of achieving envy-freenes.

Fair Division. Offline resource allocation, or commonly known as fair division, is also a popular research area. A large body of these papers are formulated as a cake-cutting problem ([63, 64, 65]) where the resources are modeled as an interval and the agents' valuations are represented as functions on this interval. However most existing results are based on relatively simple (e.g., piece wise uniform, or piece wise linear) valuation functions, where as our problem can be thought of as a cake cutting problem with arbitrarily complex valuation functions. Other fair division literature that studies envy-freeness as a fairness criterea usually treat it as a hard constraint, and tries to maximize social welfare subject to that constraint $([66,67])$. Others have also characterized the worst-case loss in social welfare due to the requirement of envy-freeness [68], and other fairness notions [69]. Unlike these papers that focus on achieving zero envy, we treat the allowable envy as a parameter, and find the most efficient solutions subject to the desired amount of envy.

### 3.4 Problem Formulation

There is a set of $n$ receivers $\mathcal{Y}$. There is a "pool" of items, represented by a distribution $\alpha$ over $\mathcal{X} \subseteq[0, \bar{x}]^{n}$. Each random draw from this distribution $X \sim \alpha$ is a vector representing the $n$ receivers' valuations of this item. The goal is to maximize the expected matched utilities of the recipients, while maintaining the constraint that the receiver $y_{i}$ is matched $p_{i}^{*}$ fraction of the times in expectation. Here $\left\{p_{i}^{*}\right\}_{i=1}^{n}$ is called the target matching distribution which intuitively represents receivers' importance. A matching policy $\pi$ takes a valuation vector and maps it (potentially with randomness) to one of the $n$ receivers. Let $\pi(y \mid x)$ denote the probability of matching the item to $y$ given valuation vector $x$. The basic problem formulation is to solve the following optimization problem:

$$
\begin{align*}
& \left.\max _{\pi} \mathbb{E}_{X \sim \alpha}\left[\sum_{i=1}^{n} X_{i} \pi\left(y_{i} \mid X\right)\right]\right]  \tag{3.4}\\
& \text { s.t. } \quad \mathbb{P}\left[\pi\left(y_{i} \mid X\right)\right]=p_{i}^{*} \quad \forall i \in[n]
\end{align*}
$$

We assume that $p_{i}^{*}>0$ for all $i$. This is WLOG, because we can always pretend that a receiver does not exist, if the target fraction for that receiver is 0 . An example of such problem can be see in Figure 3.1, where $\alpha$ is a distribution over the unit square, and the goal is to partition the square into blue and orange regions (given to $A$ and $B$ respectively) such that each region covers the desired $p_{A}^{*}, p_{B}^{*}$ probability mass. Note that the orange and blue regions are allowed to over lap (probabilitistic partition), and that the boundary does not have to be linear as illustrated in the figure. As we will show later in Section 3.5, despite the large design space permitted by the formulation in (3.4), we can in fact focus on a much smaller design space.

In resource allocation problems, it is often the case that we care not just about efficiency (maximizing social welfare, or the sum of all receivers' utilities), but also other fairness criteria. One of the most commonly studied fairness criteria is envy-freeness. Agent $y_{i}$ envies another agent $y_{j}$ if agent $y_{i}$ values the allocation given to $y_{j}$ more (after adjusting for their priority weights). We can
formally define agent $y_{i}$ 's envy as

$$
\begin{equation*}
\operatorname{Envy}\left(y_{i}\right)=\max _{j} \mathbb{E}_{\alpha}\left[X_{i} \pi\left(y_{j} \mid X\right) \frac{p_{i}^{*}}{p_{j}^{*}}-X_{i} \pi\left(y_{i} \mid X\right)\right] \tag{3.5}
\end{equation*}
$$

Instead of the vanilla formulation in (3.4), we consider the following more general formulation:

$$
\begin{align*}
& \left.\max _{\pi} \mathbb{E}_{X \sim \alpha}\left[\sum_{i=1}^{n} X_{i} \pi\left(y_{i} \mid X\right)\right]\right]  \tag{3.6}\\
& \text { s.t. } \quad \mathbb{P}\left[\pi\left(y_{i} \mid X\right)\right]=p_{i}^{*} \quad \forall i \in[n] \\
& \\
& E n v y\left(y_{i}\right) \leq \lambda_{i} \quad \forall i
\end{align*}
$$

In existing fair resource allocation literature, people focus on finding allocations such that $\operatorname{Envy}\left(y_{i}\right)$ is at most 0 for every $y_{i}$. This can be a very restrictive constraint, often satisfied at the cost of reducing efficiency by a significant amount (This reduction is sometimes referred to as the Cost-of-Fairness). We take a different approach, and allow the central planner to set non-negative constraints on envy.

### 3.5 Optimal Solution Structure

The space of feasible solutions for (3.6) is large, which makes the problem difficult to optimize directly. However we can use the tools from OT to reduce the search space to something with much more structure. The key observation is that (3.6) can be formulated as variation of the semi-discrete optimal transport problem given in Equation (3.1).

Let's first consider the simpler case in (3.4) where there are no envy constraints. In this case, the problem can be stated in the form of (3.1) as follows: the cost function is the negative utility of the matched receiver $c\left(x, y_{i}\right)=-x_{i}$, the $\beta$ measure is the discrete measure $\sum_{i=1}^{n} p_{i}^{*} \delta_{y_{i}}$, and the matching policy $\pi(y \mid x)$ in (3.6) is exactly the conditional probability of the joint distribution in (3.1). From Theorem 1 it follows that the optimal matching policy is represented by Laguerre cells given in (3.3): $x$ is matched to $y_{i}$ if $i=\arg \min _{k}-x_{k}-g_{k}$. Note that the dual variables $g \in \mathbb{R}^{n}$ serve


Figure 3.1: Left: An illustration of what Laguerre cells look like when $n=2$. Consider any distribution on the support $[0,1]^{2}$. The optimal division of the space is to move the diagonal line up or down until the probability mass contained in the orange region is equal to $p^{*}$. Right: A pictorial proof of the optimality of such partition. Suppose one can find an $\epsilon$ mass above this line that is matched to $A$, and an $\epsilon$ mass below the line that is matched to $B$, then switching the assignments of these two regions increases the matched weights.
as an "adjustment" over the agents' reported utilities, and the resulting matching policy is simply a greedy policy over this adjusted valuation vector.

Geometrically, each Laguerre cell is simply the intersection of half-spaces: $\mathbb{L}_{i}(g)=\cap_{k}\{x:$ $\left.x_{i}+g_{i} \geq x_{k}+g_{k}\right\}$. To visualize this better, consider the simple setting with two receivers $A, B$ where their valuations for an item is a joint distribution supported on $[0,1]^{2}$. Suppose we want to match $p^{*}$ fraction of the items to receiver $B$. Figure 3.1 gives a proof-by-picture that the optimal strategy is to divide the space up with a slope- 1 diagonal line such that the probability mass lying above the line is equal to $p^{*}$. This geometric interpretation of the matching policy plays a crucial role in getting us a sample complexity bound later in Section 3.7.

With this geometric interpretation of the solution space in mind, let us consider the more general case with envy constraints as formulated in (3.6). The envy constraints can be added to the OT problem in (3.1) like so:

$$
\begin{align*}
L(\alpha, \beta, \lambda) & =\min _{\pi \in \Pi(\alpha, \beta)} \int_{X \times y} c(x, y) d \pi(x, y)  \tag{3.7}\\
\text { s.t. } & \int_{\mathcal{X}} c\left(x, y_{j}\right) d \pi\left(x, y_{j}\right)-\int_{\mathcal{X}} c\left(x, y_{j}\right) d \pi\left(x, y_{k}\right) \frac{\beta_{j}}{\beta_{k}} \leq \lambda_{j} \quad \forall(j, k) \in[n]^{2}, j \neq k
\end{align*}
$$

Although envy constraints make the solution space more complicated, we show that it retains the geometric structure of being the intersection of half-spaces. The dual of (3.7) can be derived using

Fenchel-Rockafellar's theorem:

$$
\begin{equation*}
\max _{g \in \mathbb{R}^{n}, \gamma \in \mathbb{R}_{+}^{n^{-}-n}} \mathcal{E}(g, \gamma):=\sum_{j \in[n]} \int_{\mathbb{L}_{y_{j}}(g, \gamma)} \bar{g}_{\gamma, c}\left(x, y_{j}\right) d \alpha(x)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j} \tag{3.8}
\end{equation*}
$$

where

$$
\begin{gather*}
\bar{g}_{\gamma, c}\left(x, y_{j}\right):=\left(1+\sum_{k \neq j} \gamma_{j k}\right) c\left(x, y_{j}\right)-\sum_{k \neq j} \gamma_{k j} c\left(x, y_{k}\right) \frac{\beta_{k}}{\beta_{j}}-g_{j},  \tag{3.9}\\
\mathbb{L}_{y}(g, \gamma):=\left\{x \in \mathcal{X}: y=\underset{y^{\prime} \in \mathcal{Y}}{\arg \min } \bar{g}_{\gamma, c}\left(x, y^{\prime}\right)\right\} . \tag{3.10}
\end{gather*}
$$

Theorem 3. If $\alpha$ is a continuous measure, and $\beta$ a discrete measure, then $L(\alpha, \beta, \lambda)=\max _{g, \gamma} \mathcal{E}(g, \gamma)$, and the optimal solution $\pi$ of (3.7) is given by the partition $\left\{\mathbb{L}_{y_{i}}\left(g^{*}, \gamma^{*}\right), i \in[n]\right\}: d \pi\left(x, y_{i}\right)=$ $d \alpha(x)$ if $x \in \mathbb{L}_{y_{i}}\left(g^{*}, \gamma^{*}\right), 0$ otherwise.

Note that when $c\left(x, y_{i}\right)=-x_{i}, \bar{g}_{\gamma, c}\left(x, y_{j}\right)$ is linear in $x$, which means that the new Laguerre cells $\mathbb{L}_{y}(g, y)$ given in Equation (3.10) are still intersections of half spaces (some examples are given later in Figure 3.2). Furthermore, the allocation policy can be interpreted as a greedy policy based on the adjusted utility given by (3.9), which contains additional interaction terms that take envy into account.

### 3.6 Stochastic Optimization

Theorem 3 shows that the optimal solution to our allocation problem in (3.6) is represented by the optimal dual solution from (3.8). To solve this optimization problem, first note that the objective function $\mathcal{E}(g, \gamma)$ is concave. To see this, we can rewrite the objective function as follows:

$$
\begin{equation*}
\mathcal{E}(g, \gamma)=\int_{X} \min _{i \in[n]} \bar{g}_{\gamma, c}\left(x, y_{i}\right) d \alpha(x)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j} \tag{3.11}
\end{equation*}
$$

Since $\bar{g}_{\gamma, c}\left(x, y_{j}\right)$ is linear in $g$ and $\gamma$ and taking a minimum preserves concavity, the objective function is concave. Therefore, the dual problem is a constrained convex optimization problem.

The gradient of $\mathcal{E}(g, \gamma)$ can be computed as follows:

$$
\begin{align*}
\nabla_{g} \mathcal{E}(g, \gamma)_{j} & =-\int_{\mathbb{L}_{y_{j}}(g, \gamma)} d \alpha(x)+\beta_{j}  \tag{3.12}\\
\nabla_{\gamma} \mathcal{E}(g, \gamma)_{j k} & =\int_{\mathbb{L}_{y_{j}}(g, \gamma)} c\left(x, y_{j}\right) d \alpha(x)-\int_{\mathbb{L}_{y_{k}}(g, \gamma)} c\left(x, y_{j}\right) \frac{\beta_{j}}{\beta_{k}} d \alpha(x)-\lambda_{j} \tag{3.13}
\end{align*}
$$

```
ALGORITHM 3: Projected SGD for Envy Constrained Optimal Transport
    Input: Distribution \(\alpha\), target matching distribution \(p^{*}\), timesteps \(T\).
    Initialize \(g_{0}=0, \gamma_{0}=0, \eta=\frac{1}{\sqrt{T}}\).
    for \(t \leftarrow 0,1,2, \ldots, T\) do
        Sample \(x_{t} \sim \alpha\)
        \(g_{t+1} \leftarrow g_{t}+\eta \hat{\nabla}_{g} \mathcal{E}(g, \gamma)\)
        \(\gamma_{t+1} \leftarrow\left(\gamma_{t}+\eta \hat{\nabla}_{\gamma} \mathcal{E}(g, \gamma)\right)^{+}\)
    end
    return \(\sum_{t=1}^{T} g_{t} / T, \sum_{t=1}^{T} \gamma_{t} / T\)
```

Calculating this gradient is hard, as it involves integration over an arbitrary measure $\alpha$. However, an unbiased, stochastic version of the gradient can be easily obtained from a single sample $x \sim \alpha:$

$$
\begin{align*}
\hat{\nabla}_{g} \mathcal{E}(g, \gamma)_{j} & =-\mathbb{1}\left[x \in \mathbb{L}_{y_{j}}(g, \gamma)\right]+\beta_{j}  \tag{3.14}\\
\hat{\nabla}_{\gamma} \mathcal{E}(g, \gamma)_{j k} & =c\left(x, y_{j}\right) \mathbb{1}\left[x \in \mathbb{L}_{y_{j}}(g, \gamma)\right]-c\left(x, y_{j}\right) \frac{\beta_{j}}{\beta_{k}} \mathbb{1}\left[x \in \mathbb{L}_{y_{k}}(g, \gamma)\right]-\lambda_{j} \tag{3.15}
\end{align*}
$$

The details of the algorithm is given in Algorithm 3. Standard projected SGD analysis (see for example [70]) tells us that Algorithm 3 converges at the rate $\mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}\left(\mathbb{E}\left[g_{T}\right], \mathbb{E}\left[\gamma_{T}\right]\right) \leq$ $O\left(\frac{1}{\sqrt{T}}\right)$.

### 3.7 Learning from Samples

So far we have considered the setting where the true underlying distribution is known, and assumed that we can freely draw independent samples from that distribution. In many settings,
we only have access to $\alpha$ in the form of finite number of i.i.d. samples. In this section, we focus only on the assignment cost function $c\left(x, y_{i}\right)=-x_{i}$, which models our original resource allocation problem proposed in Section 3.4. The goal of this section is to establish a sample complexity bound for solving the dual problem (3.8). Let $S=\left\{X^{1}, X^{2}, \ldots, X^{m}\right\}$ be $m$ independent samples from $\alpha$. The empirical version of the dual objective (3.11) is:

$$
\begin{equation*}
\mathcal{E}_{S}(g, \gamma)=\frac{1}{m} \sum_{t=1}^{m} \min _{i \in[n]} \bar{g}_{\gamma, c}\left(X^{t}, y_{i}\right)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j} \tag{3.16}
\end{equation*}
$$

Let $\hat{g}_{S}, \hat{\gamma}_{S}$ be the empirical maximizer given the set of samples $S:\left(\hat{g}_{S}, \hat{\gamma}_{S}\right):=\arg \max \mathcal{E}_{S}(g, \gamma)$, and $g^{*}, \gamma^{*}$ be the population maximizer $\left(g^{*}, \gamma^{*}\right)=\arg \max \mathcal{E}(g, \gamma)$. We are interested in bounding the number of samples needed so that $\mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)$ is small with high probability. Let's introduce some new notations to facilitate our later discussions. Define the following hypothesis class for each $i$ :

$$
\begin{equation*}
F_{i}=\left\{x \mapsto \bar{g}_{\gamma, c}\left(x, y_{i}\right)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j}: g \in \mathbb{R}^{n}, \gamma \in \mathbb{R}_{+}^{n(n-1)}\right\} \tag{3.17}
\end{equation*}
$$

as well as the overall hypothesis class:

$$
\begin{equation*}
F=\left\{x \mapsto \min _{i \in[n]} \bar{g}_{\gamma, c}\left(x, y_{i}\right)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j}: g \in \mathbb{R}^{n}, \gamma \in \mathbb{R}_{+}^{n(n-1)}\right\} \tag{3.18}
\end{equation*}
$$

Plugging $c\left(x, y_{i}\right)=-x_{i}$ into the definition of $\bar{g}_{\gamma, c}$, we see that for a given $g$, and $\gamma$, the corresponding hypothesis $f_{i} \in F_{i}$ can be written as $f_{i}(x)=w^{\top} x+b$, where

$$
w_{j}=\left\{\begin{array}{ll}
-\left(1+\sum_{k \neq i} \gamma_{i k}\right), & \text { if } j=i  \tag{3.19}\\
\gamma_{j i} \frac{\beta_{j}}{\beta_{i}}, & \text { if } j \neq i
\end{array},\right.
$$

and

$$
\begin{equation*}
b=-g_{i}+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j} . \tag{3.20}
\end{equation*}
$$

It also follows that

$$
F \subseteq F_{\min }:=\left\{x \mapsto \min _{i} f_{i}(x): f_{i} \in F_{i}\right\}
$$

This interpretation of the original hypothesis class as the minimum over $n$ affine hypothesis classes is the key observation to prove the sample complexity bound. We prove our main result under the following boundedness assumption:

Assumption 3. The hypothesis $f(x)=\min _{i} f_{i}(x)=\min _{i} w^{i^{\top}} x+b^{i}$ corresponding to the optimal dual solution $g^{*}, \gamma^{*}$ satisfies $\left\|w^{i}\right\|_{1} \vee\left|b^{i}\right| \leq R$ for some $R>0$. In particular, these assumptions imply that $F_{i}$ and $F$ are uniformly bounded by $R \bar{x}+R$.

From (3.19) and (3.20) we can see that this is essentially a bound on the optimal dual variables $g^{*}, \gamma^{*}$, and a bound on the ratio $\beta_{j} / \beta_{i}$, both of which are determined by the input distributions $\alpha, \beta$, and do not depend on the number of samples. In other words, $R$ is a problem dependent constant.

Theorem 4. Under Assumption 3, for a given sample size m, with probability $1-\delta, \mathcal{E}\left(g^{*}, \gamma^{*}\right)-$ $\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)<O\left(\sqrt{\frac{(\log m)^{3}+\log (1 / \delta)}{m}}\right)$.

Proof. We prove the result via uniform convergence:

$$
\begin{align*}
& \mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right) \\
= & \mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}_{S}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)+\mathcal{E}_{S}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right) \\
\leq & \mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}_{S}\left(g^{*}, \gamma^{*}\right)+\mathcal{E}_{S}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right) \\
\leq & \sup _{g, \gamma}\left(\mathcal{E}(g, \gamma)-\mathcal{E}_{S}(g, \gamma)\right)+\sup _{g, \gamma}\left(\mathcal{E}_{S}(g, \gamma)-\mathcal{E}(g, \gamma)\right) \\
\leq & 2 \sup _{g, \gamma}\left|\mathcal{E}(g, \gamma)-\mathcal{E}_{S}(g, \gamma)\right| \tag{3.21}
\end{align*}
$$

Clearly, it suffices to show that $\mathcal{E}_{S}(\cdot)$ converges uniformly to $\mathcal{E}(\cdot)$. For a given $g, \gamma$, the dual objective function can be written as an expectation over their corresponding $f \in F$,

$$
\mathcal{E}(g, \gamma)=\mathbb{E}_{\alpha}[f(X)],
$$

and similarly for the empirical objective function

$$
\mathcal{E}_{S}(g, \gamma)=\frac{1}{m} \sum_{t=1}^{m} f\left(X^{t}\right)
$$

Then we can rewrite the supremum in (3.21) as:

$$
\begin{equation*}
\sup _{g, \gamma}\left|\mathcal{E}(g, \gamma)-\mathcal{E}_{S}(g, \gamma)\right|=\sup _{f \in F}\left|\mathbb{E}_{\alpha}[f(X)]-\frac{1}{m} \sum_{X \in S} f(X)\right| \tag{3.22}
\end{equation*}
$$

Since $|f(X)| \leq(R \bar{x}+R)$ for all $f \in F, X \in \mathcal{X}$, it follows from Theorem 26.5 in [71] that with probability $1-\delta$,

$$
\begin{equation*}
\sup _{f \in F} \mathbb{E}_{\alpha}[f(X)]-\frac{1}{m} \sum_{X \in S} f(X) \leq 2 \mathbb{E}_{S}\left[\operatorname{Rad}_{m}(F \circ S)\right]+(R \bar{x}+R) \sqrt{\frac{2 \log (2 / \delta)}{m}} \tag{3.23}
\end{equation*}
$$

and the same also holds by replacing $F$ with $-F$. Here

$$
\operatorname{Rad}_{m}(F \circ S):=\mathbb{E}_{\sigma}\left[\frac{1}{m} \sup _{f} \sum_{j=1}^{m} \sigma_{j} f\left(X_{j}\right)\right]
$$

is the standard definition of Rademacher complexity of the set $F \circ S$. Since $\sigma_{i}$ are i.i.d. Rademacher random variables, it is easy to see that $\operatorname{Rad}_{m}(F \circ S)=\operatorname{Rad}_{m}(-F \circ S)$. Therefore we can use a union bound to obtain that with probability $1-\delta$,

$$
\begin{equation*}
\sup _{f \in F}\left|\mathbb{E}_{\alpha}[f(X)]-\frac{1}{m} \sum_{X \in S} f(X)\right| \leq 2 \mathbb{E}_{S}\left[\operatorname{Rad}_{m}(F \circ S)\right]+(R \bar{x}+R) \sqrt{\frac{2 \log (4 / \delta)}{m}} \tag{3.24}
\end{equation*}
$$

It remains to bound the Rademacher complexity of the $F \circ S$. To do so, we use tools from learning theory, and give the following bound on the fat-shattering dimension ([72]) of the hypothesis class $F$.

Lemma 13. Under Assumption 3, F has $\zeta$-fat-shattering dimension of at most $\frac{c_{0}(R \bar{x}+R)^{2}}{\zeta^{2}} n \log (n)$, where $c_{0}$ is some universal constant.


Figure 3.2: Allocation policy for artificial data under different envy constraints. From left to right: $\epsilon=0.2,0.1,0.0$. When the envy constraint is loose (large $\epsilon$ ), $B$ envies $A$, since both agents prefer the items on the top right, but most of them are allocated to $A$. As the envy constraint tightens, the allocation boundary tilts in the direction that makes the allocations more even between the two agents.


Figure 3.3: The trade-off curve between envy and welfare for both data-sets. The shaded region is between 25th and 75th percentile of the trials. The non-monotonicity in the plot for the simulator data is due to the stochasticity in the SGD algorithm.


Figure 3.4: Approximation gap with respect to sample size. Both $x$ and $y$ axis are in $\log$ scale. The solid line is the median and the shaded region is between the 25 th and 75 th percentile. The dashed lines show what the theoretical $1 / \sqrt{m}$ rate would look like.

Proof. Theorem 3 in [73] shows that fat $_{\zeta}\left(F_{\text {min }}\right) \leq \frac{c_{0}(R \bar{x}+R)^{2}}{\zeta^{2}} n \log n$. Since the shattering dimension is monotone in the size of the set, we are done.

The above bound on the fat-shattering dimension can be used to bound the covering number (see Definition 27.1 of [71]) of $F \circ S$. Theorem 1 from [74] states that

$$
\begin{equation*}
\mathcal{N}\left(\delta, F,\|\cdot\|_{2}\right) \leq\left(\frac{2 B}{\delta}\right)^{c_{1} \mathrm{fat}_{c_{2} \delta}(F)} \tag{3.25}
\end{equation*}
$$

where $B$ is a uniform bound on the absolute value of any $f \in F$. Let $B=(R \bar{x}+R)$, we have that

$$
\begin{aligned}
& \operatorname{Rad}_{m}(F \circ S) \\
\leq & \inf _{\delta^{\prime}>0}\left\{4 \delta^{\prime}+12 \int_{\delta^{\prime}}^{B} \sqrt{\frac{\log \mathcal{N}\left(\delta, F,\|\cdot\|_{2}\right)}{m}} d \delta\right\} \\
\leq & \inf _{\delta^{\prime}>0}\left\{4 \delta^{\prime}+12 \frac{\sqrt{c_{1} c_{0}}}{c_{2}} B \sqrt{\frac{n \log n}{m}} \int_{\delta^{\prime}}^{B} \sqrt{\log \left(\frac{2 B}{\delta}\right)} d \delta\right\} \\
= & \inf _{\delta^{\prime}>0}\left\{4 \delta^{\prime}+8 \frac{\sqrt{c_{1} c_{0}}}{c_{2}} B^{2} \sqrt{\frac{n \log n}{m}}\left(\log \left(\frac{2 B}{\delta^{\prime}}\right)^{\frac{3}{2}}+(\log 2)^{\frac{3}{2}}\right)\right\} \\
= & c^{\prime} \sqrt{\frac{n \log n(\log m)^{3}}{m}}
\end{aligned}
$$

Where in the first, second, and last step we used Dudley's chaining integral [75, 76], Lemma 13 and (3.25), and setting $\delta^{\prime}=\frac{1}{\sqrt{m}}$ respectively. Plugging the above back to (3.24) and (3.21), we see that with probability $1-\delta$,

$$
\mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right) \leq c^{\prime}\left(\sqrt{\frac{n \log n(\log m)^{3}}{m}}+\sqrt{\frac{1 \log \frac{1}{\delta}}{m}}\right)
$$

Conversely, ignoring the log terms, $m$ needs to be at most on the order of $\tilde{O}\left(\frac{n}{\epsilon^{2}}\right)$ in order for $\mathcal{E}\left(g^{*}, \gamma^{*}\right)-\mathcal{E}\left(\hat{g}_{S}, \hat{\gamma}_{S}\right)$ to be bounded by $\epsilon$ with high probability.

### 3.8 Experiments

We test our solution with both artificial data, and simulated data from a realistic simulator for blood donor matching developed by [52]. The artificial data contains two receivers, and their valuation distribution is a linearly transformed uniform distribution. This is to make visualization of the resulting allocation policy easier. The simulator data is based on geographical and population information from San Francisco, and contains 5 receivers. To set the envy budgets $\lambda \in \mathbb{R}^{n}$, we first decide on a constant $\epsilon \in \mathbb{R}_{+}$, and then multiply this by the target matching distribution $p^{*} \in \mathbb{R}^{n}$ :
$\lambda_{i j}=\epsilon p_{i}^{*} \forall i, j$. With this setup $\epsilon$ is a bound on the normalized envy for each receiver: $\frac{1}{p_{i}^{*}} \operatorname{Envy}(i) \leq$ $\epsilon \forall i$. Figure 3.2 illustrates how the allocation policy changes as we change $\epsilon$. As the envy constraint tightens, the decision boundary tilts in the direction that split the "good" (items which both agents prefer) and "bad" (items which both agents dislike) items more evenly between the receivers.

Next we investigate the tradeoff between envy and social welfare by using SGD to compute approximately optimal allocations for varying $\epsilon$. We plot the percent welfare gap (difference between the maximum welfare without envy constraints, and the welfare with envy constraints, divided by the former) with respect to realized, max normalized envy. Figure 3.3 shows the result. For the simulator data, the welfare gap is small even with a no-envy constraint, which means that aiming for envy free allocations might make sense. In the case of the artificial data however, paying 50\% of the full price of fairness reduces $65 \%$ of the envy. In such settings, one might want to sacrifice some envy for better welfare.

These experiments also highlight when envy arises. When receivers' utilities are highly correlated, but one receiver has larger variance than others, that receiver receives almost all the good items (which results in large envy for other receivers), even though others value the items almost as much. In such cases, a small reduction in welfare can reduce a large amount of envy. This seems to be the case for the simulator data. On the other hand, if utilities are correlated, but only one receiver has very strong preferences, then allowing a small amount of envy can improve the welfare significantly.

Finally, in Figure 3.4 we investigate the quality of the empirical solutions as the sample size increases. It can been seen that the approximation gap decreases faster than the theoretical rate, confirming our sample complexity bound in Theorem 4.

### 3.9 Conclusion and Future Directions

Although we believe that the model proposed here is natural, and captures the most salient aspects of some of the resource allocation problems in real life, any implementation of our proposed strategy in critical applications such as blood donation should be prefaced with more rigorous back-
testing in order to minimize the risk of unintended consequences in application specific metrics not studied in this chapter.

For future directions, one key property that we did not study in this chapter is the problem of incentive compatibility. For our motivating application of blood donation, this is not an issue because online platforms such as Meta has proprietary models that can predict the matching quality between donor and receiver. This means that the platform observes the value of matchings without having to rely on the receivers to self-report. This is also true in many other online matching problems such as sponsored ads. However, in settings where the central planner relies on the receivers to self-report their valuations for each of the items, incentive compatibility becomes a crucial issue. We are excited about the potential of using Optimal Transport in fair-division, and plan on exploring the incentive issues in future work.

# Chapter 4: Online Allocation and Learning in the Presence of Strategic Agents 

The previous Chapter focused on an offline, constrained resource allocation problem where the distribution of the items is either readily available, or can be closely approximated with existing samples. Furthermore, we assumed that the central planner can observe the agents' true valuations for the items, which means that the agents cannot misreport their valuations to game the system. In this Chapter, we relax the above assumptions by first assuming that the central planner has no a priori information about the distribution of the agents' valuations, which means that the central planner has to adjust its' allocation policy over time to achieve good performance. We further assume that the central planner relies on the agents to report their private valuations, which means that the agents can now potentially misreport their valuations in order to receive better items at the expense of social welfare. As such, the problem we study in this chapter combines all the elements from the previous chapters, which makes it especially interesting to tackle. To make the problem tractable however, we do need to make additional assumptions on the agents' valuations for the items.

### 4.1 Background and Motivation

A classic sequential resource allocation problem is to allocate $T$ sequentially arriving items to $n$ agents, where each agent must receive a predetermined fraction of the items. The goal is to maximize social welfare, i.e., the agents' total valuation of the items allocated to them. This problem is non-trivial even when the agents' valuations are stochastic and i.i.d. with a known distribution, the main difficulty being that the allocations must be performed in real-time; specifically, an item must be allocated to an agent in the current round without knowledge of their future valuations.

A more challenging (and quite useful) extension of the problem which has been the focus of recent literature considers the case where the distribution of the agents' valuations is apriori unknown to the planner. In such settings, algorithms based on online learning can be used to adaptively learn the valuation distribution from observed valuations in previous rounds, and improve the allocation policy over time (see [77, 78, 60, 61] for some examples). However, these mechanisms implicitly assume that the agents report their valuations truthfully, so that the mechanism can directly learn from the reported valuations in order to maximize the social welfare.

Many practical resource allocations settings do not conform with the truthful reporting assumption. In particular, selfish and strategic agents may have an incentive to misreport their valuations if that can lead to individual utility gain (possibly at the expense of social welfare). Hence, an allocation policy that does not take such misreporting incentives into account can incur significant loss in social welfare in presence of strategic agents. For example, consider a simple setting with two agents whose true valuations are i.i.d. and uniformly distributed between 0 and 1 . That is,

$$
X_{1}, X_{2} \stackrel{i . i . d .}{\sim} \text { Uniform }[0,1]
$$

Each agent is pre-determined to receive an equal fraction of all the items. The optimal welfare maximizing allocation policy is to allocate the item to the agent with higher valuation in (almost) every round. This policy results in $T / 3$ expected utility $\left(\mathbb{E}\left[X_{1} \mid X_{1}>X_{2}\right] / 2\right)$ for each agent, and a social welfare of $2 T / 3$. However, suppose that the first agent chooses to misreport in the following way: the agent reports a high valuation of 1 whenever her true valuation is in $[0.5,1]$ and a low valuation of 0 whenever her true valuation is in [0,0.5]. Assuming the other agent remains truthful, this will lead to the first agent receiving all the items in her top $1 / 2$ quantile, and therefore a significantly increased utility of $3 T / 8\left(\mathbb{E}\left[X_{1} \mid X_{1}>0.5\right] / 2\right)$ compared to $T / 3$ under truthful reporting. The social welfare however, goes down to $5 T / 8$ in this case. Thus under the optimal policy, each agent has an incentive to misreport her valuations in order to gain individual utility. The incentives to misreport may be further amplified under an online learning based allocation
algorithm that learns approximately optimal policies from the valuations observed in the previous rounds. In such settings, the agents can potentially mislead the online learning algorithm to learn a more favorable policy over time by repeatedly misreporting their values.

Motivated by these shortcomings, in this chapter, we consider the problem of designing an online learning and allocation mechanism in the presence of strategic agents. Specifically, we consider the problem of sequentially allocating $T$ items to $n$ strategic agents. The problem proceeds in $T$ rounds. In each round $t=1, \ldots, T$, the agents' true valuations $X_{i, t}, i=1, \ldots, n$ for the $t^{t h}$ item are generated i.i.d. from a distribution $F$ a priori unknown to the central planner. However, the central planner can only observe a value $\tilde{X}_{i, t}$ reported by each agent $i$, which may or may not be the same as her true valuation $X_{i, t}$ for the item. Using the reported valuations from the current and previous rounds, the central planner needs to make an irrevocable decision of who to allocate the current item. The allocations should be made in a way such that each agent at the end receives a fixed fraction $p_{i}^{*}$ of the $T$ items, where $p_{i}^{*}>0, \sum_{i=1}^{n} p_{i}^{*}=1$. The objective of the central planner is to maximize the total utility of the agents, where utility of each agent is defined as the sum total true valuations of items received by the agent.

Our main contribution is a mechanism that achieves both: (a) Bayesian incentive compatibility, i.e., assuming all the other agents are truthful, with high probability no single agent can gain a significant utility by deviating from the truthful reporting strategy, and; (b) near-optimal regret guarantees, namely, the utility of each individual agent under the online mechanism is "close" to that achieved under the optimal offline allocation policy.

Organization After discussing the related literature in some further detail in Section 4.2, we formally introduce the problem setting and some of the core concepts in Section 4.3. Section 4.4 describes our online learning and allocation algorithm and provides formal statements of our main results (Theorem 5 and Theorem 6). Section 4.5 and Section 4.6 provide an overview of the proofs of the above theorems. All the missing details of the proofs are provided in the appendix. Finally, in Section 4.7 we discuss some limitations and future directions.

### 4.2 Literature Review

Our work lies at the intersection of online learning and mechanism design. From an online learning perspective, our setting is closely related to the recent work on constrained online resource allocation under stochastic i.i.d. rewards/costs (e.g., see [78, 77, 59, 60, 61]). However, a crucial assumption in those settings is that the central planner can observe the true rewards/costs of an allocation, which in our setting would mean that the central planner can observe agents' true valuations of the items being allocated. Our work extends these settings to allow for selfish and strategic agents who may have incentives to misreport their valuations. As discussed in the introduction, unless the online allocation mechanism design takes these incentives into account, selfish agents may significantly misreport their valuations to cause significant loss in social welfare.

Incentives and strategic agents have been previously considered in online allocation mechanism design, however, most of that work has focused on auction design where payments are used as a key mechanism for limiting rational agents' incentives to misreport their valuations. For example, Amin et al. [79] study a posted-price mechanism in a repeated auction setting where buyers' valuations are context dependent. Golrezaei et al. [80] extend this work to multi-buyer setting, using second-price auction with dynamic personalized reserve prices. Kanoria and Nazerzadeh [81] study a similar problem in a non-contextual setting. (There is also a significant literature that studies learning in repeated auction settings from the bidder's perspective. Since this chapter focuses on the central planner's point of view, we omit references to that literature. ) All of the abovementioned works are concerned with maximizing revenue for the seller, and use money/payments as a key instrument for eliciting private information about the bidder's valuations for the items. In this chapter, we are concerned with online allocation without money, and the goal is to maximize each agent's utility.

Recently, there has been some work on studying reductions from mechanism design with money to those without money. Gorokh et al. [82] provided a black-box reduction from any oneshot, BIC mechanism with money, to an approximately BIC mechanism without money. However,
their reduction relies crucially on knowing the true value distribution of agents and therefore is not applicable to our setting. Procaccia and Tennenholtz [83] consider a specific (one-round) facility allocation problem and explicitly formulate the idea of designing mechanisms without money to achieve approximately optimal performance against mechanisms with money. Subsequently, there is a series of papers that extended the results on mechanism design without money in a single shot setting, when the bidders' value distribution is unknown [84, 85, 86, 87, 86]. These papers either use a very restricted setting with just two agents, or use very specific/simple valuation functions for the agents. Even in these basic settings, they show that the best one can hope for is a constant approximation to what one can achieve with a mechanism that uses money. It is not clear what kind of regret guarantees such reductions imply in a repeated online learning setting. Therefore, we do not consider such reductions from auction mechanisms with money to be a fruitful direction for achieving our goals of both incentive compatibility and low (sublinear) regret for our online allocation problem.

Finally, in a repeated allocation settings with known valuation distribution, there are more positive results for truthful mechanism design without money. For example, Guo et al. [88] and later Balseiro et al. [89] studied the problem of repeatedly allocating items to agents with known value distributions; both use a state-based "promised utility" framework.

To summarize, to the best of our knowledge, this is the first work to incorporate strategic agents' incentives in the well-studied online allocation problem with stochastic i.i.d. rewards and unknown distributions. Thus, it bridges the gap between the online learning and allocation literature which focuses on non-strategic inputs, and the work on learning in repeated auctions which focuses on allocation mechanisms that utilize money (payments) to achieve incentive compatibility.

### 4.3 Problem formulation

### 4.3.1 The offline problem

We first state the offline version of the problem which will serve as our benchmark for the online problem. There is a set of $n$ agents, and a distribution $\boldsymbol{F}$ over $\mathcal{X}:=[0, \bar{x}]^{n}$. Each draw
$\boldsymbol{X} \sim \boldsymbol{F}$ from this distribution represents the $n$ agents' valuations of one item: $\boldsymbol{X}=\left[X_{1}, \ldots, X_{n}\right]$. We assume that the agents' valuations are i.i.d., i.e.

$$
\boldsymbol{F}=F \otimes \ldots \otimes F
$$

A matching policy (aka allocation policy) maps, potentially with some exogenous randomness, a value vector $\boldsymbol{X}$ to one of the agents $i \in\{1, \ldots, n\}$. Specifically, given a realized value vector $\boldsymbol{X} \in[0, \bar{x}]^{n}$, a (possibly randomized) policy $\pi$ maps $\boldsymbol{X}$ to agent $\pi(\boldsymbol{X}) \in\{1, \ldots, n\}$, with the probability of agent $i$ receiving an allocation given by $\mathbb{P}(\pi(X)=i)$. The offline optimization problem is to find a social welfare-maximizing policy $\pi^{*}$ such that each agent $i$ in expectation receives a predetermined fraction $p_{i}^{*}$ of the pool of items, where $p_{i}^{*}>0, \sum_{i} p_{i}^{*}=1$. The problem of finding optimal policy can therefore be stated as the following

$$
\begin{align*}
\max _{\pi} & \mathbb{E}\left[\sum_{i=1}^{n} X_{i} \mathbb{1}(\pi(\boldsymbol{X})=i)\right]  \tag{4.1}\\
\text { s.t. } & \mathbb{P}(\pi(\boldsymbol{X})=i)=p_{i}^{*} \quad \forall i
\end{align*}
$$

where the expectations are taken both over $\boldsymbol{X} \sim \boldsymbol{F}$ and any randomness in the mapping made by policy $\pi$ given $\boldsymbol{X}$. Solving the offline problem is non-trivial, as it is an infinite dimensional optimization problem as stated in its' current form in (4.1). But it turns out to be closely related to Semi-Discrete Optimal Transport, and that the dual of (4.1) can be written as

$$
\begin{equation*}
\min _{\lambda \in \mathbb{R}^{n}} \mathcal{E}(\lambda, \boldsymbol{F}):=\sum_{i \in[n]} \int_{\mathbb{L}_{i}(\lambda)}\left(X_{i}+\lambda_{i}\right) d \boldsymbol{F}(\boldsymbol{X})-\lambda^{\top} p^{*} \tag{4.2}
\end{equation*}
$$

where $\mathbb{L}_{i}$ is the Laguerre cell that we first encountered in (3.3), restated here in a slightly different notation:

$$
\begin{equation*}
\mathbb{L}_{i}(\lambda)=\left\{\boldsymbol{X}: X_{i}+\lambda_{i}>X_{j}+\lambda_{j} \forall j \neq i,\right\} . \tag{4.3}
\end{equation*}
$$

Let $\lambda^{*}(\boldsymbol{F})$ denote an optimal solution to (4.2). When it is clear from the context, we omit the distribution $\boldsymbol{F}$. It is known that an optimal solution to (4.1) is given by the following deterministic policy defined by the Laguerre cell partition (Proposition 2.1 [54]):

$$
\begin{equation*}
\pi^{*}(\boldsymbol{X})=i \text { for all } \boldsymbol{X} \in \mathbb{L}_{i}\left(\lambda^{*}\right), i=1, \ldots, n \tag{4.4}
\end{equation*}
$$

More generally, we will refer to any policy defined by a Laguerre cell partition as a greedy policy.

Definition 3 (Greedy allocation policy). Consider any allocation policy that partitions the domain $[0, \bar{x}]^{n}$ as $\mathbb{L}_{i}(\lambda)$ (as defined in (4.4)) for some $\lambda \in \mathbb{R}^{n}$. We refer to such a policy as the greedy allocation policy with parameter $\lambda$.

Note that there are efficient algorithms for solving (4.2) (see[54]) when the distribution $\boldsymbol{F}$ is known. Therefore we will not be focusing on how to solve the offline problem, and will assume $\lambda^{*}(\hat{\boldsymbol{F}})$ can be computed efficiently for any given input distribution $\hat{\boldsymbol{F}}$.

### 4.3.2 The online problem: approximate Bayesian incentive compatibility and regret

We are interested in the case when items are sequentially allocated over $T$ rounds, and that the distribution $\boldsymbol{F}$ is initially unknown. Specifically, in each round $t=1, \ldots, T$, the agents' true valuations $\boldsymbol{X}_{t}=\left(X_{i, t}, i=1, \ldots, n\right)$ are generated i.i.d. from the distribution $\boldsymbol{F}$ a priori unknown to the central planner. However, the central planner does not observe $\boldsymbol{X}_{t}$ but only observes the reported valuations $\tilde{\boldsymbol{X}}_{t}=\left(\tilde{X}_{i, t}, i=1, \ldots, n\right)$ which may or may not be the same as the true valuations.

An online allocation mechanism consists of a sequence of allocation policies $\pi_{1}, \ldots, \pi_{t}$ where the policy $\pi_{t}$ at time $t$ may be adaptively chosen based on the observed information until before time $t$ :

$$
\begin{equation*}
\mathcal{H}_{t}=\left\{\tilde{\boldsymbol{X}}_{1}, \ldots, \tilde{\boldsymbol{X}}_{t-1}, \pi_{1}, \ldots, \pi_{t-1}\right\} . \tag{4.5}
\end{equation*}
$$

Given allocation policy $\pi_{t}$ at time $t$, the agent $i$ 's utility at time $t$ is given by

$$
u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \pi_{t}\right)=X_{i, t} \mathbb{1}\left[\pi_{t}\left(\tilde{\boldsymbol{X}}_{t}\right)=i\right]
$$

Note that since the allocation policy may be randomized, for any given value vector $\boldsymbol{X}, \pi_{t}(\boldsymbol{X})$ is a random variable. To ensure truthful reporting in presence of strategic agents, we are interested in mechanisms that are (approximately) Bayesian incentive compatible.

Definition 4 (Approximate-BIC). For an online allocation mechanism, let $\pi_{t}, t=1, \ldots, T$ be the sequence of allocations when all agents report truthfully, i.e., when $\tilde{\boldsymbol{X}}_{t}=\boldsymbol{X}_{t}, \forall t$; and let $\tilde{\pi}_{t}^{i}, t=$ $1, \ldots, T$ be the sequence when all agents except $i$ report truthfully, i.e., $X_{j, t}=\tilde{X}_{j, t}, \forall j \neq i$. Then the online allocation mechanism is called $(\alpha, \delta)$-approximate Bayesian Incentive Compatible if, for all $i=1, \ldots, n$, with probability at least $1-\delta$,

$$
\sum_{t=1}^{T} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \tilde{\pi}_{t}^{i}\right)-\sum_{t=1}^{T} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \boldsymbol{\pi}_{t}\right) \leq \alpha
$$

Here, the probability is with respect to the randomness in true valuations $\boldsymbol{X}_{t} \sim \boldsymbol{F}$ and any randomness in the online allocation policy. For the online policy to be approximate-BIC, the statement should hold for all possible misreporting of valuations $\tilde{X}_{i, t} \neq X_{i, t}$.

Therefore, if $\alpha$ is small, then an individual agent has little incentive to strategize.
Assuming that all agents are truthful, we are also interested in bounding each individual agent's regret.

Definition 5 (Individual regret). We define an individual agent $i$ 's regret under an online allocation mechanism as the difference between agent i's realized utility over $T$ rounds and the expected utility achieved in the offline expected problem. That is,

$$
\begin{equation*}
\operatorname{Regret}_{i}(T)=T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\sum_{t=1}^{T} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \pi_{t}\right) \tag{4.6}
\end{equation*}
$$

Here $\pi_{1}, \ldots, \pi_{T}$ denote the allocation policies used by the online allocation mechanism in round $t=1, \ldots, T$.

Note that since social welfare is given by the sum of all agents' utilities, a bound on individual regret implies a bound on the regret in social welfare of the mechanism.

### 4.4 Algorithm and main results

We present an online allocation mechanism that is approximately-BIC, and further achieves low regret guarantees on individual regret when all agents are truthful.

Our algorithm contains two components: a learner, and a detector. Intuitively, the detector makes sure that the mechanism is approximately BIC, and the learner adaptively learns utilitymaximizing allocation policies assuming truthful agents.

The learner runs in epochs with geometrically increasing lengths. The starting time of each epoch $k$ is given by $L_{k}=2^{k}, k=0,1, \ldots$, which is also when the allocation policy is updated. At the end of each epoch (i.e., at time $t=L_{k}-1$ for epoch $k$ ), the learner takes all the previously reported values from all the agents, and uses them to construct an empirical distribution of the agents' valuations. The learner implicitly assumes truthful agents in its computations. Therefore, since the agents' true valuations are i.i.d., it first constructs a single, one-dimensional empirical distribution $\hat{F}_{t}$, and then uses it to construct the corresponding $n$-dimensional distribution $\hat{\boldsymbol{F}}_{t}$ :

$$
\begin{gather*}
\hat{F}_{t}(x)=\frac{1}{t n} \sum_{s=1}^{t} \sum_{i=1}^{n} \mathbb{1}\left[\tilde{X}_{i, s} \leq x\right]  \tag{4.7}\\
\hat{\boldsymbol{F}}_{t}=\hat{F}_{t} \otimes \ldots \otimes \hat{F}_{t}
\end{gather*}
$$

The learning algorithm then solves the offline problem (4.2) using $\hat{\boldsymbol{F}}_{t}$, and uses the resulting greedy allocation policy characterized by $\lambda^{*}\left(\hat{\boldsymbol{F}}_{t}\right)$ to allocate the items in the following epoch.

In parallel to the learner, the detector constructs and monitors, in each time step $t$, and for each agent $i$, two empirical distributions. One using the reported valuations from agent $i$ : $\bar{F}_{t}$, and one using the reported valuations from all the other agents: $\tilde{F}_{t}$. The detection algorithm then computes the supremum between the two empirical CDFs, $\sup _{x}\left|\bar{F}_{t}(x)-\tilde{F}_{t}(x)\right|$. If this difference is greater than a predetermined threshold $\Delta_{t}$, then the detector raises a flag that there has been a violation of truthful reporting and the entire allocation game stops. Otherwise, the process continues.

The threshold $\Delta_{t}$ needs to be chosen such that if everyone is truthful, then with high probabil-
ity the detection algorithm will not pull the trigger. At the same time, if someone deviates from truthful reporting significantly, then it should detect this with high probability. The typical concentration result used in comparing empirical CDFs is the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality[90]. However, since a strategic agent can adaptively change its' misreporting strategy, we cannot directly apply the DKW inequality, which assumes i.i.d. samples. Instead, we use martingale version of the DKW inequality (Lemma 14), and use that to choose an appropriate threshold $\Delta_{t}$. The details are given in Algorithm 4 and Algorithm 5.

```
ALGORITHM 4: Epoch Based Online Allocation Algorithm
    Input: \(T, \delta\)
    Initialize: \(\lambda=[0, \ldots, 0], k=0, K=\log _{2}(T), L_{k}=2^{k}, k=0, \ldots, K\);
    for \(t \leftarrow 1,2,3, \ldots, T\) do
        Observe \(\boldsymbol{X}_{t}\)
        Run Detection Algorithm (Algorithm 5) with sample set \(S=\left\{\boldsymbol{X}_{1}, \ldots, \boldsymbol{X}_{t}\right\}\), and
            threshold \(\Delta_{t}=64 \sqrt{\frac{1}{t} \log \left(\frac{256 e t}{\delta}\right)}\)
        if Detection Algorithm Return Reject then
            Terminates.
        end
        Allocate item using greedy allocation policy \(\lambda\)
        if one of the agents \(i \in\{1, \ldots, n\}\) has reached the allocation capacity \(p_{i}^{*} T\) then
            Terminates.
        end
        if \(t=L_{k+1}-1\) then
            Compute \(\hat{F}_{t}\) from samples \(\left\{\boldsymbol{X}_{1}, \ldots, \boldsymbol{X}_{t}\right\}\) as in (4.7).
            \(\lambda \leftarrow \lambda^{*}\left(\hat{\boldsymbol{F}}_{t}\right)\)
            \(k \leftarrow k+1\)
        end
    end
```

Our main results are the following guarantees on incentive compatibility and regret of our online allocation algorithm.

Theorem 5 (Approximate-BIC). Algorithm 4 is $(O(\sqrt{n T \log (n T / \delta)}), \delta)$-approximate BIC.

Since truthful reporting constitutes an approximate equilibrium, it is reasonable to then assume that agents will act truthfully. We show the following individual regret bound assuming truthfulness.

```
ALGORITHM 5: Detection Algorithm
    Input: Sample set \(S=\left\{\boldsymbol{X}_{1}, \ldots, \boldsymbol{X}_{t}\right\}\), threshold \(\Delta_{t}\).
    for \(i \leftarrow 1, \ldots, n\) do
        Compute \(\bar{F}_{t}(x)=\frac{1}{t} \sum_{s=1}^{t} \mathbb{1}\left[\boldsymbol{X}_{i}^{s} \leq x\right]\) as the empirical CDF of the samples collected
        from agent \(i\)
        Compute \(\tilde{F}_{t}(x)=\frac{1}{t(n-1)} \sum_{s=1}^{t} \sum_{j \neq i} \mathbb{1}\left[X_{j}^{s} \leq x\right]\) be the empirical CDF of all reported
        values from the other agents.
        if \(\sup _{x}\left|\tilde{F}_{t}(x)-\bar{F}_{t}(x)\right| \geq \frac{\Delta_{t}}{2}\) then
            Return Reject
        end
    end
    Return Accept
```

Theorem 6 (Individual Regret). Assuming all agents report their valuations truthfully, then under the online allocation mechanism given by Algorithm 4, with probability $1-\delta$, every agent $i$ 's individual regret can be bounded as:

$$
\begin{aligned}
\operatorname{Regret}_{i}(T) & \leq \frac{4 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{4 n \log _{2} T+n T}{\delta}\right)} \bar{x} \\
& =O(\sqrt{n T \log (n T / \delta)})
\end{aligned}
$$

Showing approximate incentive compatibility, and then guaranteeing regret under the assumption of truthfulness, is an approach commonly seen in the online mechanism design literature (e.g. Theorem 4 in [81]). In the next section, we describe the high level proof ideas for the main results above.

### 4.5 Proof ideas

Proof ideas for Theorem 5 We establish that the mechanism is approximately BIC by showing that no single agent has incentive to significantly deviate from reporting true valuations if all the other agents are truthful. The proof consists of two parts. In Step 1, we prove that any significant deviation from the truth can be detected and will lead the mechanism to terminate. In Step 2,3, we prove that in order to achieve a significant gain in utility, an agent indeed has to report values that
significantly deviate from the truth.

Step 1 Assuming that there is only one (unidentified) strategic agent while all the other agents are truthful, we first show that if the detector does not trigger a violation by time $t$ then with high probability, the empirical distribution of valuations reported by the strategic agent is no more than $O(1 / \sqrt{t})$ away from the true distribution (Lemma 16). The key observation here is that since the agents' valuations are i.i.d., we can compare their reported values to detect if any single agent's distribution is significantly different from everyone else's. A technical challenge in making statistical comparisons here is that the strategic agent can adaptively change their reporting strategy over time based on the realized outcomes. Therefore, we derive a novel martingale version of the DKW inequality to show concentration of the empirical distribution relative to the true underlying distribution.

Step 2 In a given round, given the history, the mechanism's allocation policy is a fixed greedy allocation policy given by $\lambda$. If the distribution of strategic agent's reported values differs from the true distribution by at most $\Delta$, then the agent's expected utility gain in that round, compared to reporting truthfully, is at most $O(\Delta)$, (see Lemma 17).

Step 3 If over $t$ rounds, the distribution of the strategic agent's reported values is at most $\Delta$ away from the true distribution, then the learning algorithm will, with high probability find an allocation policy that is at most $O(\sqrt{n} \Delta)$ away (in terms of individual utility) from what it would have learned if all the agents were truthful instead (see Lemma 15).

To understand the significance and distinction between results in Step 2 vs. Step 3, note that a strategic agent has two separate ways to gain utility. The first is to report valuations in a way that the agent immediately wins more/better items under the central planner's current allocation policy. However, since the central planner is updating its' allocation policy over time, the strategic agent can also misreport in a way that benefits its' future utility, by "tricking" the central planner into learning a policy that favors him later on. Together, Step 2 and Step 3 show that the agent cannot
gain significant advantage over being truthful in either manner.
In many existing works on online auctions mechanisms design, where the central planner dynamically adjusts the reserve price over time, these two types of strategic behaviors are in conflict: the agent either sacrifices future utility to gain immediate utility; or sacrifices near-term utility for future utility. The results in those settings therefore often rely on this observation to show approximate incentive compatibility. In our case however, since there is no money involved, it is not clear if such a conflict between short and long term utility exists. Nonetheless, we are able to bound the agents' ability to strategize. Step 2 bounds the agent's short term incentive to be strategic, whereas Step 3 bounds the longer term incentive to be strategic. Combining these steps gives us a proof for Theorem 5.

Proof ideas for Theorem 6 Recall that here we assume all agents' are truthful. The proof involves two main steps.

Step 1 We show that uniformly for any $t=1, \ldots, T$, with high probability, the empirical estimate of the valuation distribution from the first $t$ rounds of samples is close (within a distance of $\tilde{O}(1 / \sqrt{n t}))$ to the true value distribution $F$. Here, the factor of $1 / \sqrt{n}$ comes from the fact that in each round we observe $n$ independent samples from the value distribution, one from each of the agents. This also implies that if all the agents are reporting truthfully, then, with high probability, the detector will not falsely trigger.

Step 2 We show that the allocation policy learned under the empirical distribution estimated from the samples is close to the the optimal allocation policy (Lemma 15). Specifically, after $t$ rounds if the empirical CDF is at most $O(1 / \sqrt{n t})$ away from the true distribution, then each agents' expected utility in one round under the learned allocation policy is at most $\sqrt{n / t}$ away (both from above and from below) from the optimal. By using an epoch based approach we can then show that each agents' individual regret is with high probability bounded by $O(\sqrt{n T})$ over the entire planning horizon $T$.

### 4.6 Proof details

We will now outline our proof in more detail. All missing proofs can be found in the Appendix. First we state the following martingale variation of the well-known Dvoretzky-Kiefer-Wolfowitz (DKW) inequality[90]. This is critical when dealing with strategic agents as they can adapt their strategy over time, resulting in non-independent (reported) values.

Lemma 14 (Martingale Version of DKW Inequality). Given a sequence of random variables $Y_{1}, \ldots, Y_{T}$, let $\mathcal{F}_{t}=\sigma\left(Y_{1}, \ldots, Y_{t}\right), t=1, \ldots, T$ be the filtration representing the information in the first $t$ variables. Let $F_{t}(y):=\operatorname{Pr}\left(Y_{t} \leq y \mid \mathcal{F}_{t-1}\right)$, and $\bar{F}_{T}(y):=\frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\left[Y_{t} \leq y\right]$. Then,

$$
\mathbb{P}\left(\sup _{y}\left|\bar{F}_{T}(y)-\frac{1}{T} \sum_{t=1}^{T} F_{t}(y)\right| \geq \alpha\right) \leq\left(\frac{128 e T}{\alpha}\right) e^{-T \alpha^{2} / 128}
$$

Next we introduce a new notation to to denote the fraction of allocation that $j$ receives under the greedy allocation policy with parameter $\lambda$ and valuation distribution $\boldsymbol{F}$ :

$$
p_{j}(\boldsymbol{F}, \lambda):=\mathbb{P}_{\boldsymbol{X} \sim \boldsymbol{F}}\left(\boldsymbol{X} \in \mathbb{L}_{j}(\lambda)\right) .
$$

We start with proving Theorem 6, as we will use this to prove Theorem 5 later.

### 4.6.1 Individual Regret Bound (Theorem 6)

In Algorithm 4, the allocation policy is trained on the empirical distribution constructed from samples. We want to show that this difference between empirical and population distribution will not impair the performance of the resulting allocation policy too much.

Lemma 15. Let $\boldsymbol{G}=G^{1} \otimes \ldots \otimes G^{n}$, and $\boldsymbol{F}=F^{1} \otimes \ldots \otimes F^{n}$ be two distributions over $[0, \bar{x}]^{n}$ where the marginals on each coordinate are independent. Suppose $\sup _{x}\left|F^{i}(x)-G^{i}(x)\right| \leq \Delta \forall i$. Let $\lambda=\lambda^{*}(\boldsymbol{G})$, and $\lambda^{*}=\lambda^{*}(\boldsymbol{F})$. Then

$$
\left|\mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}\left[\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right]\right]\right| \leq n \Delta \bar{x}
$$

Proof of Theorem 6 Now we have the main ingredients for Theorem 6. We use the DKW inequality to show that the empirical distribution constructed in (4.7) is close to the true distribution w.h.p.. Then we use Lemma 15 to show that the allocation policy selected by the learner based on the empirical distribution is almost optimal in expectation. The details can be found in the Appendix D.2.

### 4.6.2 Approximate-Bayesian Incentive Compatibility (Theorem 5)

Theorem 6 says that online utility of each agent cannot be too far below the offline optimum if everyone behaves truthfully. In order to show approximate-BIC, it suffices to show that the strategic agent cannot gain too much more than the offline optimum. To do so, we need to bound both the short term and longer term incentives for the agent to be strategic.

## Short term incentive

We start with bounding the short term strategic incentive. We first show that if agent reports from an average distribution that is very different from the truthful distribution, then with high probability Algorithm 5 can detect that. Note that given the strategic agent's strategy in a given round, his reported value is drawn from a distribution potentially different from $F$.

Lemma 16. Fix a time step $t$. Let $\Delta=64 \sqrt{\frac{\log \left(\frac{256 e t}{\delta}\right)}{t}}$. Let $F_{s}, s=1, \ldots, t$ be the strategic agent's reported value distributions in each time step given the history, i.e., $F_{s}(x):=\mathbb{P}\left(\tilde{X}_{i, s} \leq x \mid \mathcal{H}_{s}\right)$. If the average distribution $\bar{F}=\frac{1}{t} \sum_{s=1}^{t} F_{s}$ is such that $\sup _{x}|\bar{F}(x)-F(x)| \geq \Delta$, then Algorithm 4 will terminates at or before time $t$ with probability at least $1-\delta$.

Next, we show that if the agent restricts the reported distribution to not deviate more than $\Delta$ from the true distribution (so that the deviation may go undetected by the detection algorithm), then the potential gain in the agent's utility compared to truthful reporting is upper bounded by $\bar{x} \Delta$. This bounds the agent's incentive to be strategic.

Lemma 17. Fix a round $t$ and a single strategic agent $i$, so that the remaining agents are truthful, i.e., $\tilde{X}_{j, t}=X_{j, t}, \forall j \neq i$. Let $F_{r}(\cdot)$ denote the marginal distribution of values $\tilde{X}_{i, t}$ reported by the strategic agent $i$ at time $t$ conditional on the history, i.e.,

$$
F_{r}(x):=\mathbb{P}\left(\tilde{X}_{i, t} \leq x \mid \mathcal{H}_{t}\right)
$$

Suppose that $\sup _{x}\left|F(x)-F_{r}(x)\right| \leq \Delta$. Then, at any time $t$, given any greedy allocation policy $\lambda$,

$$
\mathbb{E}\left[u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \lambda\right) \mid \mathcal{H}_{t}\right]-\mathbb{E}\left[u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \lambda\right)\right] \leq \bar{x} \Delta
$$

Note that $F_{r}$ specifies only the marginal distribution of $\tilde{X}_{i, t} \mid \mathcal{H}_{t}$ and not the joint distribution of $\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}\right) \mid \mathcal{H}_{t}$. Indeed the above lemma claims that the given bound on utility gain holds for all possible joint distributions as long as the marginal $F_{r}$ of $\tilde{X}_{i, t} \mid \mathcal{H}_{t}$ is at most $\Delta$ away from $F$.

Intuitively, Lemma 16 and Lemma 17 together bound the agent's short term incentive to be strategic: if the agent deviates from the truthful distribution too much, then the mechanism will terminate early and the agent will lose out on all the future utility (Lemma 16); and given any greedy allocation strategy set by the central planner, we have that if the agents deviates within the undetectable range of Algorithm 5, then the gain in utility compared to acting truthfully is small (Lemma 17). Next, we bound an agent's incentive to lie in order to make the mechanism learn a suboptimal greedy allocation policy that is more favorable to the agent.

## Long term incentive

In order to bound the longer term incentive to be strategic, we want to show that despite agent $i$ being strategic, the central planner can still learn an allocation policy that closely approximates the offline optimal allocation policy. This means that the agent's influence over the central planner's allocation policies is limited.

Lemma 18. Fix a round $T^{\prime} \leq T$ and a strategic agent $i$. If agent $i$ is the only one being strategic, and Algorithm 5 has not been triggered by the end of time $T^{\prime}$, then with probability $1-\delta, \hat{\lambda}:=$
$\lambda^{*}\left(\hat{\boldsymbol{F}}_{T^{\prime}}\right)$ satisfies

$$
\mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \hat{\lambda})\right]-\mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \leq n \Delta_{T^{\prime}} \bar{x}
$$

where $\Delta_{T^{\prime}}=81 \sqrt{\frac{1}{n T^{\prime}} \log \left(\frac{256 e\left(T^{\prime}\right)}{\delta}\right)}$ and $\lambda^{*}=\lambda^{*}(\boldsymbol{F})$.
In particular, consider $T^{\prime}=L_{k}$. Then the lemma above shows that if an agent was not kicked out by the end of epoch $k-1$, then with high probability the greedy allocation policy in epoch $k$ will be such that the agents' expected utility by being truthful is close to what he would have received in the offline optimal solution (Lemma 18). We can now combine this result with Lemma 16 and Lemma 17 to bound the utility that any single strategic agent can gain over the entire trajectory.

Lemma 19. If agent $i$ is the only one being strategic, then with probability $1-\delta$, agent $i$ 's online utility is upper bounded by

$$
\sum_{t=1}^{T} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \tilde{\lambda}_{k_{t}}\right) \leq T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]+\frac{286 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{256 e \log _{2} T}{\delta}\right) \bar{x}}
$$

Here $k_{t}$ denotes the epoch number that time step $t$ lies in, and $\tilde{\lambda}_{k_{t}}$ denotes the allocation policy used by the central planner in that epoch.

Proof of Theorem 5 We have already proven Theorem 6 which bounds individual regret defined as the difference $T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\sum_{t=1}^{T} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \lambda_{k_{t}}\right)$, i.e., the difference between the utility of agent $i$ under the offline optimal policy and that under the allocation policy learned by the algorithm when all the agents are truthful. The proof of Theorem 5 follows from plugging in the upper bound on $T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]$ from this theorem into Lemma 19 , to obtain the desired bound on the expression $\sum_{t=1}^{T} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \tilde{\lambda}_{k_{t}}\right)-\sum_{t=1}^{T} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \lambda_{k_{t}}\right)$, i.e., on the total gain in utility achievable by misreporting under our mechanism. Further details are in Appendix D.3.

### 4.7 Conclusion and Future Directions

Although our goal is to develop mechanisms that are robust to selfish and strategic agents, real applications often involve bad faith actors that have extrinsic motivation to behave adversarially.

As such, deployment of such resource allocation mechanisms to critical applications requires significant additional validation. In future work we would like to explore the limit of relaxing the i.i.d. assumption that we place on the distribution of valuations across agents. This is a natural relaxation because if one agent thinks the item is good then it's likely that other agents would like the item as well. Furthermore it is also conceivable that agents are heterogeneous and so have different value distributions for the items. However this seems to require a completely different strategy for detecting, and disincentivize strategic behaviors, as we can no longer catch the strategic agent through comparing each agent's reported distribution with that of others.

## References

[1] R. Kleinberg and T. Leighton, "The value of knowing a demand curve: Bounds on regret for online posted-price auctions," in Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, ser. FOCS 03, USA: IEEE Computer Society, 2003, p. 594.
[2] O. Besbes and A. Zeevi, "Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms," Operations Research, vol. 57, no. 6, pp. 1407-1420, 2009.
[3] A. Badanidiyuru, R. Kleinberg, and A. Slivkins, "Bandits with knapsacks," 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, Oct. 2013.
[4] A. V. den Boer, "Dynamic pricing and learning: Historical origins, current research, and new directions," Surveys in Operations Research and Management Science, vol. 20, no. 1, pp. 1 -18, 2015.
[5] N. B. Keskin and A. Zeevi, "Chasing demand: Learning and earning in a changing environment," Mathematics of Operations Research, vol. 42, no. 2, pp. 277-307, 2017.
[6] O. Besbes, Y. Gur, and A. Zeevi, "Optimal exploration-exploitation in a multi-armed bandit problem with non-stationary rewards," Stochastic Systems, vol. 9, no. 4, pp. 319-337, 2019.
[7] F. M. Bass, "A new product growth for model consumer durables," Management Science, vol. 15, no. 5, pp. 215-227, 1969.
[8] ——,"A new product growth for model consumer durables," Management Science, vol. 50, pp. 1825-1832, Dec. 2004.
[9] B. Robinson and C. Lakhani, "Dynamic price models for new-product planning," Management Science, vol. 21, no. 10, pp. 1113-1122, 1975.
[10] S. Senecal and J. Nantel, "The influence of online product recommendations on consumers' online choices," Journal of Retailing, vol. 80, no. 2, pp. 159 -169, 2004.
[11] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," Journal of Marketing Research, vol. 43, no. 3, pp. 345-354, 2006.
[12] C. Dellarocas, X. M. Zhang, and N. F. Awad, "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," Journal of Interactive Marketing, vol. 21, no. 4, pp. 23-45, 2007.
[13] R. East, K. Hammond, and W. Lomax, "Measuring the impact of positive and negative word of mouth on brand purchase probability," International Journal of Research in Marketing, vol. 25, no. 3, pp. 215-224, 2008.
[14] R. Iyengar, C. Van den Bulte, and T. W. Valente, "Opinion leadership and social contagion in new product diffusion," Marketing Science, vol. 30, no. 2, pp. 195-212, 2011.
[15] O. Toubia, J. Goldenberg, and R. Garcia, "Improving penetration forecasts using social interactions data," Management science, vol. 60, no. 12, pp. 3049-3066, 2014.
[16] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," Management Science, vol. 49, no. 10, pp. 1407-1424, 2003.
[17] D. Lehmann, O. Netzer, and O. Toubia, "The Future of Quantitative Marketing: Results of a Survey," Customer Needs and Solutions, vol. 2, pp. 5-18, 2015.
[18] D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: A case study of rfm model-based customer segmentation using data mining," Journal of Database Marketing \& Customer Strategy Management, vol. 19, no. 3, pp. 197-208, 2012.
[19] R. J. Dolan and A. P. Jeuland, "Experience curves and dynamic demand models: Implications for optimal pricing strategies," Journal of Marketing, vol. 45, no. 1, pp. 52-62, 1981.
[20] A. V. den Boer, "Tracking the market: Dynamic pricing and learning in a changing environment," European Journal of Operational Research, vol. 247, no. 3, pp. 914 -927, 2015.
[21] O. Besbes and A. Zeevi, "On the minimax complexity of pricing in a changing environment," Operations Research, vol. 59, no. 1, pp. 66-79, 2011.
[22] Y. Chen, Z. Wen, and Y. Xie, "Dynamic pricing in an evolving and unknown marketplace," Available at SSRN, 2019.
[23] A. Garivier and E. Moulines, "On upper-confidence bound policies for switching bandit problems," in Algorithmic Learning Theory, Springer Berlin Heidelberg, 2011, pp. 174188.
[24] Y. Cao, Z. Wen, B. Kveton, and Y. Xie, "Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit," in Proceedings of Machine Learning Research, vol. 89, PMLR, Apr. 2019, pp. 418-427.
[25] Y. Russac, C. Vernade, and O. Cappé, "Weighted linear bandits for non-stationary environments," in Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019.
[26] H. Luo, C.-Y. Wei, A. Agarwal, and J. Langford, "Efficient contextual bandits in nonstationary worlds," in Proceedings of the 31st Conference On Learning Theory, vol. 75, PMLR, Jul. 2018, pp. 1739-1776.
[27] C. Tekin and M. Liu, "Online learning of rested and restless bandits," IEEE Transactions on Information Theory, vol. 58, no. 8, pp. 5588-5611, Aug. 2012.
[28] D. Bertsimas and J. Niño-Mora, "Restless bandits, linear programming relaxations, and a primal-dual index heuristic," Operations Research, vol. 48, no. 1, pp. 80-90, 2000.
[29] L. Yang and M. Wang, "Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound," in International Conference on Machine Learning, PMLR, 2020, pp. 1074610756.
[30] S. Agrawal and R. Jia, "Optimistic posterior sampling for reinforcement learning: Worstcase regret bounds," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[31] T. Jaksch, R. Ortner, and P. Auer, "Near-optimal regret bounds for reinforcement learning.," Journal of Machine Learning Research, vol. 11, no. 4, 2010.
[32] J. Grasman and M. Kornelis, "Forecasting product sales with a stochastic bass model," Journal of Mathematics in Industry, 2nd ser., vol. 9, 2019.
[33] S.-C. Niu, "A stochastic formulation of the bass model of new-product diffusion," Mathematical Problems in Engineering, vol. 8, Jan. 2002.
[34] H. Lee, S. G. Kim, H. woo Park, and P. Kang, "Pre-launch new product demand forecasting using the bass model: A statistical and machine learning-based approach," Technological Forecasting and Social Change, vol. 86, pp. 49 -64, 2014.
[35] Z.-P. Fan, Y.-J. Che, and Z.-Y. Chen, "Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis," Journal of Business Research, vol. 74, pp. $90-100,2017$.
[36] P. Yin, G. Dou, X. Lin, and L. Liu, "A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning," Kybernetes, vol. ahead-of-print, Jan. 2020.
[37] M. Zhang, H.-S. Ahn, and J. Uichanco, "Data-driven pricing for a new product," Available at SSRN 3545574, 2020.
[38] J. Baek et al., "The limits to learning a diffusion model," arXiv preprint arXiv:2006.06373, 2020.
[39] D. Jalota, M. Pavone, and Y. Ye, "Markets for efficient public good allocation," ArXiv, vol. abs/2005.10765, 2020.
[40] S. Mehrotra, H. Rahimian, M. Barah, F. Luo, and K. Schantz, "A model for supply-chain decisions for resource sharing with an application to ventilator allocation to combat covid19," medRxiv, 2020.
[41] A. Zenteno, "Models for managing surge capacity in the face of an influenza epidemic," Ph.D. dissertation, Columbia University, 2013.
[42] E. Eisenberg and D. Gale, "Consensus of subjective probabilities: The pari-mutuel method," The Annals of Mathematical Statistics, vol. 30, no. 1, pp. 165-168, 1959.
[43] X. Chen, L.-S. Huang, and S.-H. Teng, "Market equilibria with hybrid linear-leontief utilities," in Internet and Network Economics, P. Spirakis, M. Mavronicolas, and S. Kontogiannis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 274-285, ISBN: 978-3-540-68141-0.
[44] A. Ghodsi, M. Zaharia, A. Konwinski, S. Shenker, and I. Stoica, "Dominant resource fairness: Fair allocation of multiple resource types," NSDI'11: Proceedings of the 8th USENIX conference on Networked Systems Design and Implementation, pp. 323-336, 2011.
[45] D. C. Parkes, A. D. Procaccia, and N. Shah, "Beyond dominant resource fairness: Extensions, limitations, and indivisibilities," ACM Trans. Econ. Comput., vol. 3, no. 1, Mar. 2015.
[46] I. Caragiannis, D. Kurokawa, H. Moulin, A. D. Procaccia, N. Shah, and J. Wang, "The unreasonable fairness of maximum nash welfare," ACM Transactions on Economics and Computation (TEAC), vol. 7, no. 3, pp. 1-32, 2019.
[47] E. Budish, "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," Journal of Political Economy, vol. 119, no. 6, pp. 1061-1103, 2011.
[48] E. Budish, G. P. Cachon, J. B. Kessler, and A. Othman, "Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation," Operations Research, vol. 65, no. 2, pp. 314-336, 2017.
[49] P. Hall, "On representatives of subsets," Journal of the London Mathematical Society, vol. s110, no. 1, pp. 26-30, 1935. eprint: https://londmathsoc. onlinelibrary. wiley.com/doi/pdf/10.1112/jlms/s1-10.37.26.
[50] L. Gurobi Optimization, Gurobi optimizer reference manual, http://www. gurobi. com, Accessed: 2021-05-20, 2021.
[51] M. ApS, Mosek modeling cookbook, https://www.mosek. com, Accessed: 2021-0520, 2020.
[52] D. C. McElfresh et al., "Matching algorithms for blood donation," in Proceedings of the 21st ACM Conference on Economics and Computation, 2020, pp. 463-464.
[53] L. Kantorovich, "On the transfer of masses (in russian)," Doklady Akademii Nauk, vol. 37(2), pp. 227-229, 1942.
[54] G. Aude, M. Cuturi, G. Peyré, and F. Bach, "Stochastic optimization for large-scale optimal transport," arXiv preprint arXiv:1605.08527, 2016.
[55] G. Goel and A. Mehta, "Online budgeted matching in random input models with applications to adwords.," in SODA, Citeseer, vol. 8, 2008, pp. 982-991.
[56] N. R. Devanur and T. P. Hayes, "The adwords problem: Online keyword matching with budgeted bidders under random permutations," in Proceedings of the 10th ACM conference on Electronic commerce, 2009, pp. 71-78.
[57] N. R. Devanur, K. Jain, B. Sivan, and C. A. Wilkens, "Near optimal online algorithms and fast approximation algorithms for resource allocation problems," Journal of the ACM (JACM), vol. 66, no. 1, pp. 1-41, 2019.
[58] S. Agrawal, Z. Wang, and Y. Ye, "A dynamic near-optimal algorithm for online linear programming," Operations Research, vol. 62, no. 4, pp. 876-890, 2014.
[59] S. Agrawal and N. R. Devanur, "Fast algorithms for online stochastic convex programming," in Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms, SIAM, 2014, pp. 1405-1424.
[60] S. Balseiro, H. Lu, and V. Mirrokni, "The best of many worlds: Dual mirror descent for online allocation problems," arXiv preprint arXiv:2011.10124, 2020.
[61] -_, "Regularized online allocation problems: Fairness and beyond," in International Conference on Machine Learning, PMLR, 2021, pp. 630-639.
[62] S. R. Sinclair, S. Banerjee, and C. L. Yu, "Sequential fair allocation: Achieving the optimal envy-efficiency tradeoff curve," arXiv preprint arXiv:2105.05308, 2021.
[63] J. Robertson and W. Webb, Cake-cutting algorithms: Be fair if you can. CRC Press, 1998.
[64] Y. Chen, J. K. Lai, D. C. Parkes, and A. D. Procaccia, "Truth, justice, and cake cutting," Games and Economic Behavior, vol. 77, no. 1, pp. 284-297, 2013.
[65] E. Mossel and O. Tamuz, "Truthful fair division," in International Symposium on Algorithmic Game Theory, Springer, 2010, pp. 288-299.
[66] Y. J. Cohler, J. K. Lai, D. C. Parkes, and A. D. Procaccia, "Optimal envy-free cake cutting," in Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011.
[67] S. Yin, S. Wang, L. Zhang, and C. Kroer, "Dominant resource fairness with meta-types," arXiv preprint arXiv:2007.11961, 2020.
[68] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou, "The efficiency of fair division," in International Workshop on Internet and Network Economics, Springer, 2009, pp. 475-482.
[69] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," Operations research, vol. 59, no. 1, pp. 17-31, 2011.
[70] N. Harvey, "Machine learning theory lecture 17," 2018.
[71] S. Shalev-Shwartz and S. Ben-David, Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.
[72] P. L. Bartlett, P. M. Long, and R. C. Williamson, "Fat-shattering and the learnability of realvalued functions," journal of computer and system sciences, vol. 52, no. 3, pp. 434-452, 1996.
[73] A. Kontorovich and I. Attias, "Fat-shattering dimension of $k$-fold maxima," arXiv preprint arXiv:2110.04763, 2021.
[74] S. Mendelson and R. Vershynin, "Entropy and the combinatorial dimension," Inventiones mathematicae, vol. 152, no. 1, pp. 37-55, 2003.
[75] K. Sridharan, Note on refined dudley integral covering number bound, 2021.
[76] R. M. Dudley, "The sizes of compact subsets of hilbert space and continuity of gaussian processes," Journal of Functional Analysis, vol. 1, no. 3, pp. 290-330, 1967.
[77] S. Agrawal, Z. Wang, and Y. Ye, "A dynamic near-optimal algorithm for online linear programming," Operations Research, vol. 62, no. 4, 876-890, Aug. 2014.
[78] N. R. Devanur, K. Jain, B. Sivan, and C. A. Wilkens, "Near optimal online algorithms and fast approximation algorithms for resource allocation problems," J. ACM, vol. 66, no. 1, Jan. 2019.
[79] K. Amin, A. Rostamizadeh, and U. Syed, "Repeated contextual auctions with strategic buyers," Advances in Neural Information Processing Systems, vol. 27, 2014.
[80] N. Golrezaei, A. Javanmard, and V. Mirrokni, "Dynamic incentive-aware learning: Robust pricing in contextual auctions," Advances in Neural Information Processing Systems, vol. 32, 2019.
[81] Y. Kanoria and H. Nazerzadeh, "Incentive-compatible learning of reserve prices for repeated auctions," Operations Research, vol. 69, no. 2, pp. 509-524, 2021.
[82] A. Gorokh, S. Banerjee, and K. Iyer, "From monetary to nonmonetary mechanism design via artificial currencies," Mathematics of Operations Research, 2021.
[83] A. D. Procaccia and M. Tennenholtz, "Approximate mechanism design without money," ACM Transactions on Economics and Computation (TEAC), vol. 1, no. 4, pp. 1-26, 2013.
[84] M. Guo and V. Conitzer, "Strategy-proof allocation of multiple items between two agents without payments or priors.," in AAMAS, Citeseer, 2010, pp. 881-888.
[85] L. Han, C. Su, L. Tang, and H. Zhang, "On strategy-proof allocation without payments or priors," in International Workshop on Internet and Network Economics, Springer, 2011, pp. 182-193.
[86] R. Cole, V. Gkatzelis, and G. Goel, "Mechanism design for fair division: Allocating divisible items without payments," in Proceedings of the fourteenth ACM conference on Electronic commerce, 2013, pp. 251-268.
[87] ——, "Positive results for mechanism design without money," 2013.
[88] M. Guo, V. Conitzer, and D. M. Reeves, "Competitive repeated allocation without payments," in International Workshop on Internet and Network Economics, Springer, 2009, pp. 244-255.
[89] S. R. Balseiro, H. Gurkan, and P. Sun, "Multiagent mechanism design without money," Operations Research, vol. 67, no. 5, pp. 1417-1436, 2019.
[90] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," The Annals of Mathematical Statistics, pp. 642-669, 1956.
[91] P. Chareka, O. Chareka, and S. Kennendy, "Locally sub-gaussian random variables and the strong law of large numbers," in Atlantic Electronic Journal of Mathematics, vol. 1, Jan. 2006.
[92] K. Doya, S. Ishii, A. Pouget, and R. P. Rao, Bayesian brain: Probabilistic approaches to neural coding. MIT press, 2007.
[93] T. Lattimore and C. Szepesvári, Bandit algorithms. Cambridge University Press, 2020.
[94] A. Rakhlin, K. Sridharan, and A. Tewari, "Sequential complexities and uniform martingale laws of large numbers," Probability Theory and Related Fields, vol. 161, no. 1-2, pp. 111153, 2015.

## Appendix A: Appendices for Chapter1

## A. 1 A Concentration Result on Exponential Random Variables

Before moving on to the next sections, we state a concentration result that is used in many of the remaining proofs.

Lemma 20. Let $\left[I_{d}, d=1, \ldots, n\right]$ be a sequence of random variables with filtration $\mathcal{F}_{d}$ such that $I_{d} \mid \mathcal{F}_{d-1}$ is an exponential random variable with rate $\lambda_{d} \in \mathcal{F}_{d-1}$. Suppose that there exists $\underline{\lambda}>0$ such that $\lambda_{d} \geq \underline{\lambda}$ almost surely. Then for any $\epsilon \leq \frac{2 n}{\underline{\lambda}}$,

$$
\begin{aligned}
& \mathbb{P}\left(\sum_{d=1}^{n} I_{d}-\frac{1}{\lambda_{d}} \leq-\epsilon\right) \leq \exp \left(\frac{-\epsilon^{2} \underline{\lambda}^{2}}{8 n}\right) \\
& \mathbb{P}\left(\sum_{d=1}^{n} I_{d}-\frac{1}{\lambda_{d}} \geq \epsilon\right) \leq \exp \left(\frac{-\epsilon^{2} \underline{\lambda}^{2}}{8 n}\right)
\end{aligned}
$$

Proof. An exponential random variable with rate $\lambda$ satisfies (see [91])

$$
\begin{equation*}
\mathbb{E}\left[e^{s X}\right]=\frac{1}{1-s / \lambda} \leq e^{\frac{s}{\lambda}+2 \frac{s^{2}}{\lambda^{2}}} \quad \forall s \in\left[-\frac{\lambda}{2}, \frac{\lambda}{2}\right] \tag{A.1}
\end{equation*}
$$

$$
\begin{aligned}
\mathbb{P}\left(\sum_{d=1}^{n} I_{d}-\frac{1}{\lambda_{d}} \leq-\epsilon\right) & =\mathbb{P}\left(e^{s \sum_{d=1}^{n} \frac{1}{\lambda_{d}}-I_{d}} \geq e^{s \epsilon}\right) \\
\text { (Markov's Inequality) } & \leq \frac{1}{e^{s \epsilon}} \mathbb{E}\left[e^{s\left(\sum_{d=1}^{n} \frac{1}{\lambda_{d}}-I_{d}\right)}\right] \\
& \leq \frac{1}{e^{s \epsilon}} \mathbb{E}\left[e^{s\left(\sum_{d=1}^{n-1} \frac{1}{\lambda_{d}}-I_{d}\right)} \mathbb{E}\left[\left.e^{s\left(\frac{1}{\lambda_{n}}-I_{n}\right)} \right\rvert\, \mathcal{F}_{n-1}\right]\right] \\
\text { (Using (A.1) and } \left.\lambda_{n} \geq \underline{\lambda}\right) & \leq \frac{1}{e^{s \epsilon}} \mathbb{E}\left[e^{s\left(\sum_{d=1}^{n-1} \frac{1}{\lambda_{d}}-I_{d}\right)} e^{2 \frac{s^{2}}{\lambda^{2}}}\right]
\end{aligned}
$$

(Repeating the above argument) $=e^{s^{2} \frac{2 n}{\underline{1}^{2}}-s \epsilon}$

$$
\text { (solve for optimal } \left.s=\frac{\epsilon \underline{\lambda}^{2}}{4 n}\right)=\exp \left(\frac{-\epsilon^{2} \underline{\lambda}^{2}}{8 n}\right)
$$

Note that to use (A.1) in the last inequality we needed that $s=\frac{\epsilon \lambda^{2}}{4 n} \leq \frac{\lambda_{d}}{2}$ for every $d$, which is satisfied if $\epsilon \leq \frac{2 n}{\underline{\lambda}} \cdot \mathbb{P}\left(\sum_{d=1}^{n} I_{d}-\frac{1}{\lambda_{d}} \geq \epsilon\right)$ can be bounded similarly.

## A. 2 Some Properties of the Deterministic Optimal Price Curve

## A.2.1 Optimal pricing policy expression

The expressions in (1.4) (1.5) and (1.6) provided the expressions for the optimal price curve $p^{*}(x, \alpha, \beta), x \in\left[0, X_{T}^{*}\right]$ and for the total number of adoptions $X_{T}^{*}$ when optimal pricing policy is followed from an initial adoption level $X_{0}=0$ at $t=0$ to the end of the time horizon $T$. Here we derive a more general expression for the optimal pricing policy $\pi^{*}(x, \alpha, \beta, T)$ that will give the optimal price at any current adoption level $x$ and remaining planning horizon $T$ (irrespective of what pricing policy was followed for how much time to reach the adoption level $x$ ). Also, we derive an expression for $X_{T}^{*}(x)$, the adoption level at the end of time $T$ if optimal pricing policy is followed for time $T$ starting from the adoption level $X_{0}=x$ at $t=0$. These expressions will be especially useful in our lower bound derivations.

Note that under this expanded notation, $X_{T}^{*}=X_{T}^{*}(0)$. We sometimes also use the notation
$X_{T}^{*}(x, \alpha, \beta)$ to emphasis the dependence $X_{T}^{*}$ has on the market parameters. Note that by this definition, $X_{T-t}^{*}\left(X_{t}\right)=X_{T}^{*}\left(X_{0}\right)$ if $X_{t}$ is the adoption level reached at time $t$ on following the optimal price trajectory from time 0 .

The optimal price to offer at any given adoption level $x$ and remaining time $T$ can be derived using optimal control theory (see Equation (8) of [19]). and is given by the following pricing policy. Given adoption $x \in[0,1)$ and remaining time $T>0$, the optimal price is given by

$$
\begin{equation*}
\pi^{*}(x, \alpha, \beta, T):=1+\log \left(\frac{(\alpha+\beta x)(1-x)}{\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right)}\right) \tag{A.2}
\end{equation*}
$$

When the value of $\alpha, \beta$ is clear from the context, we sometimes drop the dependence on $\alpha, \beta$, and use shorter notations $p^{*}(x)$ and $\pi^{*}(x, T)$ instead of $p^{*}(x, \alpha, \beta)$ and $\pi^{*}(x, \alpha, \beta, T)$ respectively.

To see the connection (and distinction) between $p^{*}(x)$ and $\pi^{*}(x, T)$, note that $p^{*}(x)$ is the price trajectory if the optimal policy $\pi^{*}$ is followed from $t=0, X_{0}=0$ to the end of horizon $T$. That is, if $X_{t}, p_{t}$ denotes the adoption level and price at time $t \in[0, T]$ on following optimal pricing policy from $t=0, X_{0}=0$, then $p_{t}=p^{*}\left(X_{t}, \alpha, \beta\right)=\pi^{*}\left(X_{t}, \alpha, \beta, T-t\right)$.

Now consider the adoption process on starting from an initial adoption level $x$ at time $t=0$, and then following the optimal pricing policy. Again $X_{t}$ denotes the adoption level at time $t$ in this process. Plugging $p_{t}=\pi^{*}\left(X_{t}, \alpha, \beta, T-t\right)$ back into (1.2), it's easy to derive that

$$
\begin{equation*}
\frac{d X_{t}}{d t}=\frac{1}{e}\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right) \tag{A.3}
\end{equation*}
$$

This means that under the optimal policy, the rate of adoption is constant. We can integrate the above from $t=0$ to $T$, and also solve the resulting quadratic equation for $t=T$ to compute the final adoption level $X_{T}^{*}(x)$ under the optimal pricing policy when starting from an initial adoption
level $x$ at $t=0$ :

$$
\begin{array}{r}
X_{T}^{*}(x)-x=\frac{1}{e}\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right) T \\
X_{T}^{*}(x)=\frac{T(\beta-\alpha)-e+\sqrt{[T(\beta-\alpha)-e]^{2}+4 \alpha \beta T^{2}+4 e \beta x T}}{2 \beta T} \tag{A.5}
\end{array}
$$

Note that on plugging $x=0$ in (A.5), we obtain the expression for $X_{T}^{*}$ in (1.6).

## A.2.2 Price Lipschitz Bound

We start with some Lipschitz bounds on how the optimal price offered at adoption level $x$ can change with respect to $\alpha, \beta$. Note that here we are assuming $X_{0}=0$ and that the entire optimal price curve from the beginning changes if we change $\alpha, \beta$ (i.e., we use (1.4) not (A.2)). In the following lemma, $X_{T}^{*}$ denotes the adoption curve on following the optimal pricing policy for all times $t \in[0, T]$ starting from 0 adoption level under the deterministic Bass model with parameters $(\alpha, \beta)$ as given in (1.6). Using the expanded notation introduced in the previous subsection, it can also be called $X_{T}^{*}(0, \alpha, \beta)$.

Lemma 21. $0 \leq \frac{1}{1-X_{T}^{*}} \frac{\partial X_{T}^{*}}{\partial \alpha} \leq \frac{1}{\alpha}$, and $0 \leq \frac{1}{1-X_{T}^{*}} \frac{\partial X_{T}^{*}}{\partial \beta} \leq \frac{1}{\beta}$
Proof. Below, we use $X_{T}$ instead of $X_{T}^{*}$ to denote the adoption at time $T$, in the deterministic optimal trajectory starting at $X_{0}=0$. Differentiating (1.5) with respect to $\alpha$ :

$$
\begin{aligned}
\beta T X_{T}^{2}+e X_{T}-\beta T X_{T}+\alpha T X_{T}-\alpha T & =0 \\
2 \beta T \frac{\partial X_{T}}{\partial \alpha} X_{T}+e \frac{\partial X_{T}}{\partial \alpha}-\beta T \frac{\partial X_{T}}{\partial \alpha}+T X_{T}+\alpha T \frac{\partial X_{T}}{\partial \alpha}-T & =0 \\
\frac{\partial X_{T}}{\partial \alpha}\left(2 T X_{T} \beta+e-\beta T+\alpha T\right) & =T\left(1-X_{T}\right)
\end{aligned}
$$

Rearranging and substituting $X_{T}$ using (1.6):

$$
\begin{aligned}
\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \alpha} & =\frac{T}{2 \beta T X_{T}+e+(\alpha-\beta) T} \\
& =\frac{T}{T(\beta-\alpha)-e+\sqrt{4 T^{2} \alpha \beta+(T(\beta-\alpha)-e)^{2}}+(\alpha-\beta) T+e} \\
& =\frac{T}{\sqrt{4 T^{2} \alpha \beta+(T(\beta-\alpha)-e)^{2}}} \geq 0 .
\end{aligned}
$$

Note the denominator can be bounded in two ways:

$$
\begin{array}{r}
\sqrt{4 T^{2} \alpha \beta+(T(\beta-\alpha)-e)^{2}} \geq|T(\beta-\alpha)-e| \text { and }, \\
\sqrt{4 T^{2} \alpha \beta+(T(\beta-\alpha)-e)^{2}}=\sqrt{T^{2}(\alpha+\beta)^{2}+2(\alpha-\beta) T+e^{2}} \geq|T(\alpha+\beta)-e| .
\end{array}
$$

So

$$
\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \alpha} \leq \min \left(\left|\frac{1}{\alpha+\beta-\frac{e}{T}}\right|,\left|\frac{1}{\alpha-\beta+\frac{e}{T}}\right|\right)=\min \left(\left|\frac{1}{\alpha+\left(\beta-\frac{e}{T}\right)}\right|,\left|\frac{1}{\alpha-\left(\beta-\frac{e}{T}\right)}\right|\right) \leq \frac{1}{\alpha}
$$

Following a similar procedure, (differentiating (1.5) with respect to $\beta$ and using (1.6)) we can also get the following bound:

$$
\begin{aligned}
\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \beta} & =\frac{T X_{T}}{2 \beta T X_{T}+e+(\alpha-\beta) T} \\
& =\frac{T(\beta-\alpha)-e}{2 \beta \sqrt{4 T^{2} \alpha \beta+(T(-\alpha+\beta)-e)^{2}}}+\frac{1}{2 \beta} \\
0 \leq \frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \beta} & \leq \frac{1}{\beta}
\end{aligned}
$$

where in the last step we used the fact that $\frac{|T(\beta-\alpha)-e|}{\sqrt{4 T^{2} \alpha \beta+(T(-\alpha+\beta)-e)^{2}}} \leq 1$.
In our proofs we will sometimes need to compare the optimal final adoption level under two
different market parameters. To derive the results comparing two different market parameters, instead of $X_{T}^{*}(x)$, we use the notation $X_{T}^{*}(x, \alpha, \beta)$ that makes the dependence on $\alpha, \beta$ explicit.

Corollary 2. $X_{T}^{*}(0, \alpha, \beta) \geq X_{T}^{*}(0, \alpha, 0)=\frac{\alpha T}{\alpha T+e}$.
Proof. The inequality follows from the non-negativity of $\frac{\partial X^{*}}{\partial \beta}$ proved in Lemma 21. The equality can be easily derived by solving for $X_{T}^{*}$ in (1.5) after plugging in $\beta=0$.

Lemma 22. For any $0 \leq x \leq X_{T}^{*},\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \alpha}\right| \leq \frac{2+\beta / \alpha}{\alpha,},\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \beta}\right| \leq \frac{3}{\min (\alpha, \beta)}$.
Proof. Below, we use $X_{T}$ instead of $X_{T}^{*}$ to denote the optimal final adoption level, assuming that the initial adoption level is 0 . Taking the derivatives of (1.4), and by using Lemma 21:

$$
\begin{aligned}
\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \alpha}\right| & =\left|\frac{\partial \log (\alpha+\beta x)}{\partial \alpha}+\frac{\partial \log (1-x)}{\partial \alpha}-\frac{\partial \log \left(\alpha+\beta X_{T}\right)}{\partial \alpha}-\frac{\partial \log \left(1-X_{T}\right)}{\partial \alpha}\right| \\
& =\left|\frac{1}{\alpha+\beta x}-\frac{1}{\alpha+\beta X_{T}}\left(1+\beta \frac{\partial X_{T}}{\partial \alpha}\right)+\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \alpha}\right| \\
& \leq \max \left(\left|\frac{1}{\alpha+\beta x}+\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \alpha}\right|,\left|\frac{1}{\alpha+\beta X_{T}}\left(1+\beta \frac{\partial X_{T}}{\partial \alpha}\right)\right|\right) \\
& \leq \max \left(\frac{2}{\alpha}, \frac{1}{\alpha}+\frac{\beta / \alpha\left(1-X_{T}\right)}{\alpha}\right) \\
& \leq \frac{2+\beta / \alpha}{\alpha}
\end{aligned}
$$

$$
\begin{aligned}
\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \beta}\right| & =\left|\frac{\partial \log (\alpha+\beta x)}{\partial \beta}-\frac{\partial \log \left(\alpha+\beta X_{T}\right)}{\partial \beta}-\frac{\log \left(1-X_{T}\right)}{\partial \beta}\right| \\
& =\left|\frac{x}{\alpha+\beta x}-\frac{1}{\alpha+\beta X_{T}}\left(X_{T}+\beta \frac{\partial X_{T}}{\partial \beta}\right)+\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \beta}\right| \\
& \leq\left|\frac{x}{\alpha+\beta x}-\frac{1}{\alpha+\beta X_{T}}\left(X_{T}+\beta \frac{\partial X_{T}}{\partial \beta}\right)\right|+\left|\frac{1}{1-X_{T}} \frac{\partial X_{T}}{\partial \beta}\right| \\
& \leq \frac{1}{\alpha}\left(1+\beta \frac{\partial X_{T}}{\partial \beta}\right)+\frac{1}{\beta} \\
& \leq \frac{2}{\alpha}+\frac{1}{\beta} \\
& \leq \frac{3}{\min (\alpha, \beta)}
\end{aligned}
$$

## A. 3 Proof of Lemma 1 and Lemma 2

We first prove the concavity property of the deterministic optimal value stated as Lemma 1.

Lemma 1 (Concavity of deterministic optimal revenue). For any deterministic Bass model, $V^{\text {det }}(x, T)$, defined as the optimal revenue starting from adoption level $x$ and remaining time $T$, is concave in $x$, for all $T \geq 0$, and all adoption levels $x \in[0,1]$.

Proof. For simplicity of notation, in this proof we use $T$ to denote the remaining time. The optimal value function for the continuous deterministic Bass model when the remaining time is $T$ can be expressed using the following dynamic programming equation (for all $\delta \geq 0$ ),

$$
\begin{equation*}
V^{\operatorname{det}}(x, T)=\max _{p} p \lambda(p, x) \delta+V^{\operatorname{det}}(x+\lambda(p, x) \delta / m, T-\delta)+o(\delta) \tag{A.6}
\end{equation*}
$$

where $\lambda(p, x)=m e^{-p}(\alpha+\beta x)(1-x)$.
Using the Hamilton-Jacobi-Bellman equation for the deterministic Bass model (see equation (12.8) in [92]):

$$
\begin{equation*}
\frac{\partial V^{\operatorname{det}}(x, T)}{\partial T}=\max _{p} p \lambda(p, x)+\frac{\lambda(p, x)}{m} \frac{\partial V^{\operatorname{det}}(x, T)}{\partial x} \tag{A.7}
\end{equation*}
$$

And the optimal price is the price that achieves the maximum in the above expression (see equation (12.9) in [92]), i.e.,

$$
\pi^{*}(x, \alpha, \beta, T)=\underset{p}{\arg \max } p \lambda(p, x)+\frac{\lambda(p, x)}{m} \frac{\partial V^{\operatorname{det}}(x, T)}{\partial x}
$$

Solving the above maximization problem gives us an expression for the optimal price at state $x$ with $T$ time left.

$$
\begin{equation*}
\pi^{*}(x, \alpha, \beta, T)=1-\frac{1}{m} \frac{\partial V^{\operatorname{det}}(x, T)}{\partial x} \tag{A.8}
\end{equation*}
$$

From (A.2), the optimal price is also given by

$$
\pi^{*}(x, \alpha, \beta, T)=1+\log \left(\frac{(\alpha+\beta x)(1-x)}{\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right)}\right)
$$

where (refer to (A.4), (A.5))

$$
\begin{array}{r}
X_{T}^{*}(x)=x+\frac{1}{e}\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right) T \\
X_{T}^{*}(x)=\frac{T(\beta-\alpha)-e+\sqrt{[T(\beta-\alpha)-e]^{2}+4 \alpha \beta T^{2}+4 e \beta x T}}{2 \beta T}
\end{array}
$$

Therefore, substituting,

$$
\begin{align*}
\frac{1}{m} \frac{\partial V^{\operatorname{det}}(x, T)}{\partial x} & =-\log \left(\frac{(\alpha+\beta x)(1-x)}{\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right)}\right) \\
\frac{1}{m} \frac{\partial^{2} V^{\operatorname{det}}(x, T)}{\partial x^{2}} & =\frac{-\beta}{\alpha+\beta x}+\frac{1}{1-x}+\frac{\beta \frac{\partial X_{T}^{*}(x)}{\partial x}}{\alpha+\beta X_{T}^{*}(x)}-\frac{1}{1-X_{T}^{*}(x)} \frac{\partial X_{T}^{*}(x)}{\partial x} \tag{A.9}
\end{align*}
$$

Now we split (A.9) to two parts and bound them by zero individually. First note that differentiating $X_{T}^{*}(x)$ with respect to $x$ (using (A.4)) gives us

$$
\begin{equation*}
\frac{\partial X_{T}^{*}(x)}{\partial x}=\frac{e}{e+(\alpha-\beta) T+2 \beta T X_{T}^{*}(x)} \tag{A.10}
\end{equation*}
$$

Then, first we show that

$$
\frac{1}{1-x}-\frac{1}{1-X_{T}^{*}(x)} \frac{\partial X_{T}^{*}(x)}{\partial x} \leq 0
$$

which is equivalent to showing that

$$
\begin{aligned}
& 1-X_{T}^{*}(x) \leq \frac{\partial X_{T}^{*}(x)}{\partial x}(1-x) \\
& \operatorname{Using}(\mathrm{A} .4) \Longleftrightarrow 1-X_{T}^{*}(x) \leq \frac{\partial X_{T}^{*}(x)}{\partial x}\left(1-X_{T}^{*}(x)\right)+\frac{\partial X_{T}^{*}(x)}{\partial x} \frac{T}{e}\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right) \\
& \operatorname{Using}(\mathrm{A} .10) \Longleftrightarrow 1-X_{T}^{*}(x) \leq\left(1-X_{T}^{*}(x)\right)\left(\frac{e}{e+(\alpha-\beta) T+2 \beta T X_{T}^{*}(x)}\right)\left(\frac{e+T\left(\alpha+\beta X_{T}^{*}(x)\right)}{e}\right) \\
& \Longleftrightarrow 1-X_{T}^{*}(x) \leq\left(1-X_{T}^{*}(x)\right) \cdot \frac{e+\alpha T+\beta T X_{T}^{*}(x)}{e+(\alpha-\beta) T+2 \beta T X_{T}^{*}(x)}
\end{aligned}
$$

Using expression for $X_{T}^{*}(x)$ from (A.5), we have

$$
e+(\alpha-\beta) T+2 \beta T X_{T}^{*}(x)=\sqrt{[T(\beta-\alpha)-e]^{2}+4 \alpha \beta T^{2}+4 e \beta x T} \geq 0
$$

Then, since $\beta \geq 0, T \geq 0,0 \leq X_{T}^{*}(x) \leq 1$, we have that the fraction in the last inequality is at least 1 , and therefore the last inequality holds.

Now we bound the remaining terms in the RHS of (A.9) by 0 . This requires showing that

$$
\begin{aligned}
& \begin{aligned}
& \frac{-\beta}{\alpha+\beta x}+\frac{\beta \frac{\partial X_{T}^{*}(x)}{\partial x}}{\alpha+\beta X_{T}^{*}(x)} \leq 0 \\
& \Longleftrightarrow \quad(\alpha+\beta x) \frac{\partial X_{T}^{*}(x)}{\partial x} \leq \alpha+\beta X_{T}^{*}(x) \\
& \text { Using (A.4) } \Longleftrightarrow \frac{\partial X_{T}^{*}(x)}{\partial x}\left(\alpha+\beta X_{T}^{*}(x)-\beta \frac{T}{e}\left(\alpha+\beta X_{T}^{*}(x)\right)\left(1-X_{T}^{*}(x)\right)\right) \leq \alpha+\beta X_{T}^{*}(x) \\
& \Longleftrightarrow \\
&\left(\alpha+\beta X_{T}^{*}(x)\right) \frac{\partial X_{T}^{*}(x)}{\partial x}\left(1-\frac{\beta T}{e}\left(1-X_{T}^{*}(x)\right)\right) \leq \alpha+\beta X_{T}^{*}(x) \\
& \text { Using (A.10) } \Longleftrightarrow \\
& e+(\alpha-\beta) T+2 \beta X_{T}^{*}(x) T
\end{aligned} \leq 1
\end{aligned}
$$

Since $\alpha, \beta \geq 0, T \geq 0, X_{T}^{*}(x) \geq 0$, the last inequality holds.
Therefore the sum of all the terms in the right hand side of (A.9) is bounded by 0 . This proves the lemma statement.

Now we use the above concavity property to show that for any starting point and remaining time, the optimal revenue in the deterministic model is at least the optimal expected revenue in
the stochastic model. Let $V^{\text {stoch }}(d, T)$ be the optimal expected revenue one can achieve in the stochastic Bass model with $d$ current adopters and $T$ time remaining.

Lemma 23. For any $d \in\{0, \ldots, m\}, x=\frac{d}{m}$, and any $T \geq 0$ :

$$
V^{d e t}(x, T) \geq V^{s t o c h}(d, T)
$$

Proof. Given any $\delta \geq 0$, let $\Delta(p, d)$ be the random number of adoptions that take place in the next $\delta$ time in the discrete stochastic Bass model when the current price is $p$ and current number of adopters is $d$. Let $x=\frac{d}{m}$, then $\mathbb{E}[\Delta(p, x)]=\lambda(p, x) \delta+o(\delta)$. Using dynamic programming, we have:

$$
\begin{aligned}
V^{\text {stoch }}(d, T) & =\max _{p} \mathbb{E}\left[p \Delta(p, d)+V^{\text {stoch }}(d+\Delta(p, d), T-\delta)\right]+o(\delta) \\
& =\max _{p} p \mathbb{E}[\Delta(p, d)]+\mathbb{E}\left[V^{\text {stoch }}(d+\Delta(p, d), T-\delta)\right]+o(\delta) .
\end{aligned}
$$

Also,

$$
V^{\operatorname{det}}(x, T)=\max _{p} p \mathbb{E}[\Delta(p, d)]+V^{\operatorname{det}}(x+\mathbb{E}[\Delta(p, d)] / m, T-\delta)+o(\delta)
$$

We use induction to prove the lemma by working from the end of the planning horizon (no time remaining). We know $V^{\operatorname{det}}(x, 0)=V^{\text {stoch }}(d, 0)=0$ for all $d, x=\frac{d}{m}$. Suppose the inequality holds for $T-\delta$, and any $d, x=\frac{d}{m}$. Then,

$$
V^{\operatorname{det}}(x, T)=\max _{p} p \mathbb{E}[\Delta(p, d)]+V^{\operatorname{det}}(x+\mathbb{E}[\Delta(p, d)] / m, T-\delta)+o(\delta)
$$

(using concavity from Lemma 1 ) $\geq \max _{p} p \mathbb{E}[\Delta(p, d)] m+\mathbb{E}\left[V^{\mathrm{det}}(x+\Delta(p, d) / m, T-\delta)\right]+o(\delta)$ (inductive assumption on $T-\delta) \geq \max _{p} p \mathbb{E}[\Delta(p, d)] m+\mathbb{E}\left[V^{\text {stoch }}(d+\Delta(p, d), T-\delta)\right]+o(\delta)$ $=V^{\text {stoch }}(d, T)+o(\delta)$

Then, taking $\delta \rightarrow 0$, we obtain the lemma statement.

Finally, we prove the following lemma that will allow us to establish an upper bound on the deterministic optimal revenue compared to the stochastic optimal revenue.

Lemma 24. Fix any $\alpha, \beta, T$ such that

$$
m X_{T}^{*} \geq 8 \log ^{2}(4 m \log (e+(\alpha+\beta) T))+32
$$

then

$$
V^{\text {det }}(0, T)-V^{\text {stoch }}(0, T) \leq O\left(\log ((\alpha+\beta) T) \sqrt{m X_{T}^{*} \log (m)}\right)
$$

Proof. The proof constructs a fixed price sequence such that the expected revenue in the stochastic Bass model under these prices is within $O(\sqrt{m})$ of the optimal deterministic revenue. Then, since the optimal expected stochastic revenue is at least as much as that obtained under the given price sequence, we will obtain the lemma statement.

Consider the following pricing scheme: for all time instances after arrival of $(d-1)^{\text {th }}$ customer, and until arrival of $d^{t h}$ customer, post price $p_{d}^{*}$ given by: $p_{d}^{*}:=p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right)$ for $d=d_{0}+$ $1, \ldots,\left\lfloor m X_{T}^{*}\right\rfloor$. Here $p^{*}(\cdot, \alpha, \beta)$ and $X_{T}^{*}$ are given by the optimal price curve and adoption levels for the deterministic Bass model, as defined in (1.4) and (1.6).

Now, since in our price sequence, the prices are fixed between two arrivals, we know that (in the stochastic Bass model) the inter-arrival time between customer $d-1$ and $d$ is an exponential random variable $I_{d} \sim \operatorname{Exp}\left(\lambda\left(p_{d}^{*}, \frac{d-1}{m}\right)\right)$, where $\lambda(p, x)=e^{-p}(\alpha+\beta x)(1-x) m$. Note that from (A.3) and (A.4) we have that $\lambda\left(p_{d}^{*}, \frac{d-1}{m}\right)=\frac{m X_{T}^{*}}{T}$, and that $\sum_{d=1}^{n} \frac{1}{\lambda\left(p_{d}^{*}, \frac{d-1}{m}\right)}=\frac{n T}{m X_{T}^{*}}$ for any $n \leq\left\lfloor m X_{T}^{*}\right\rfloor$.

Let $\tau_{n}$ denotes the time of arrival of the $n^{\text {th }}$ customer in the stochastic model. Set $n=\left\lfloor m X_{T}^{*}-\right.$ $\left.\sqrt{8 m X_{T}^{*} \log \left(\frac{2}{\delta}\right)}\right\rfloor, \underline{\lambda}=\frac{m X_{T}^{*}}{T}, \epsilon=\sqrt{\frac{8 n}{\lambda^{2}} \log \left(\frac{2}{\delta}\right)}$, for some $\delta \geq 0$ to be specified later. Assume $n \geq$ $2 \log (2 / \delta)$ for now, then by Lemma 20 , we have with probability $1-\delta$ :

$$
\left|\tau_{n}-\left(1-\sqrt{\frac{8 \log \left(\frac{2}{\delta}\right)}{m X_{T}^{*}}}\right) T\right| \leq T \sqrt{\frac{8}{X_{T}^{*} m} \log \left(\frac{2}{\delta}\right)}
$$

Therefore, with probability at least $1-\delta, \tau_{n} \leq T$, which means that the total number of adoptions $d_{T}$ observed in the stochastic Bass model in time $T$ is at least $n$. Let $p_{x}^{*}=p^{*}(x, \alpha, \beta), p_{\text {max }}^{*}=$ $\max _{x} p^{*}(x, \alpha, \beta)$. Then,

$$
\begin{aligned}
V^{\mathrm{det}}(0, T)-V^{\text {stoch }}(0, T) & \leq m \int_{0}^{X_{T}^{*}} p_{x}^{*} d x-\sum_{d=1}^{d_{T}} p_{\frac{d-1}{m}}^{*} \\
(\text { Corollary 3) } & \leq \sum_{1}^{\left\lfloor m X_{T}^{*}\right\rfloor} p_{\frac{d-1}{m}}^{*}-\sum_{d=1}^{d_{T}} p_{\frac{d-1}{m}}^{*}+\frac{\beta}{2 \alpha}+\frac{1}{2} \log (m)+3 p_{\max }^{*} \\
& \leq \sum_{n+1}^{\left\lfloor m X_{T}^{*}\right\rfloor} p_{\frac{d-1}{m}}^{*}+\frac{\beta}{2 \alpha}+\frac{1}{2} \log (m)+3 p_{\max }^{*} \quad \text { w.p. } 1-\delta \\
(\text { Lemma 7) } & =O\left(\log ((\alpha+\beta) T) \sqrt{m X_{T}^{*} \log \left(\frac{1}{\delta}\right)}\right) \quad \text { w.p. } 1-\delta
\end{aligned}
$$

where we borrowed Corollary 3 from Appendix A.4.2, and used the price upper bound $p_{\max }^{*} \leq$ $O(\log ((\alpha+\beta) T))$ from Lemma 7 .

Note that the third step holds with probability $1-\delta$. When it does not hold (with probability at most $\delta$ ), we can bound the gap between deterministic and stochastic revenue by a trivial upper bound of $p_{\max }^{*} m X_{T}^{*}$ on the deterministic revenue. We set $\delta=\frac{1}{\sqrt{m X_{T}^{*} p_{\max }^{*}}}$ to get that

$$
\begin{aligned}
& V^{\mathrm{det}}(0, T)-V^{\mathrm{stoch}}(0, T) \\
\leq & (1-\delta) O\left(\log ((\alpha+\beta) T) \sqrt{m X_{T}^{*} \log \left(\frac{1}{\delta}\right)}\right)+\delta p_{\max }^{*} m X_{T}^{*} \\
\leq & O\left(\log ((\alpha+\beta) T) \sqrt{m X_{T}^{*} \log (m)}\right)
\end{aligned}
$$

Finally, one can verify that the condition on $m X_{T}^{*}$ implies that $n \geq 2 \log (2 / \delta)$. Let $M=m X_{T}^{*}$, and
assume that $\log \left(2 \sqrt{M p_{\text {max }}^{*}}\right) \geq 1$ :

$$
\begin{aligned}
n \geq 2 \log (2 / \delta) & \Longleftrightarrow M \geq 2 \sqrt{2 M \log \left(2 \sqrt{M p_{\max }^{*}}\right)}+2 \log \left(2 \sqrt{M p_{\max }^{*}}\right) \\
& \Longleftarrow M \geq(2 \sqrt{2 M}+2) \log \left(2 \sqrt{M p_{\max }^{*}}\right) \\
& \Longleftarrow \sqrt{M} \geq 2 \sqrt{2} \log \left(4 M p_{\max }^{*}\right) \\
(\text { Lemma } 7) & \Longleftarrow \sqrt{M} \geq 2 \sqrt{2} \log (4 M \log (e+(\alpha+\beta) T))
\end{aligned}
$$

If $\log \left(2 \sqrt{M p_{\max }^{*}}\right)<1$, then the first line is implied by $M \geq 2 \sqrt{2 M}+2$, which is satisfied by $M \geq 32$.

The proof of Lemma 2 can now be completed using the upper and lower bounds on deterministic optimal revenue compared to the stochastic optimal revenue proven above.

Proof of Lemma 2. The first part of Lemma 2 follows directly from Lemma 23 because

$$
\text { Pseudo-Regret }(T) \geq \operatorname{Regret}(T) \Longleftrightarrow V^{\operatorname{det}}(0, T) \geq V^{\text {stoch }}(0, T)
$$

Similarly, the second part follows from Lemma 24. Note that if the condition on $m X_{T}^{*}$ does not hold, then the gap $V^{\text {det }}(0, T)-V^{\text {stoch }}(0, T)$ can be trivially bounded by $O\left(\log T \log ^{2}(m \log T)\right)$ :

$$
\begin{aligned}
V^{\operatorname{det}}(0, T)-V^{\text {stoch }}(0, T) & \leq m X_{T}^{*} p_{\text {max }}^{*}-0 \\
(\operatorname{Lemma} 7) & \leq\left[8 \log ^{2}(4 m \log (e+(\alpha+\beta) T))+32\right] O(\log ((\alpha+\beta) T)) \\
& =O\left(\log ((\alpha+\beta) T) \log ^{2}(m \log ((\alpha+\beta) T))\right)
\end{aligned}
$$

## A. 4 Upper Bound Proofs

## A.4.1 Step 1: Bounding the estimation errors (Proof of Lemma 3, Lemma 4)

We prove the estimation bound on $\hat{\alpha}$ and $\hat{\beta}_{i}$ separately in Lemma 25 and Lemma 26 respectively. Lemma 3 follows directly from these two results.

Lemma 25. Assuming that $m^{1 / 3} \geq \frac{16(\alpha+\beta)}{\alpha} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, then with probability $1-\delta,|\alpha-\hat{\alpha}| \leq A$.
Proof. From (1.8) we have the following expression for the estimator error of $\hat{\alpha}$ :

$$
|\alpha-\hat{\alpha}|=\gamma e^{p_{0}}\left|\frac{1}{\gamma m \mathbb{E}\left[\tau_{1}\right]}-\frac{1}{\tau_{\gamma m}}\right|
$$

Recall that $\tau_{d}$ denotes the (stochastic) time of arrival of $d^{t h}$ customer in the stochastic Bass model under the pricing decisions made by the algorithm. Note that in our algorithm prices do not change between customer arrivals. Therefore, the interarrival time $I_{d}=\tau_{d}-\tau_{d-1}$ between $d-1$ and $d$ customer follows an exponential distribution. Specifically, since the prices were fixed as $p_{0}$ for the first $\gamma m$ customers, we have $I_{d} \sim \operatorname{Exp}\left(\lambda\left(p_{0}, \frac{d-1}{m}\right)\right)$ for $d \in\{1, \ldots, \gamma m\}$ where $\lambda(p, x)=$ $m e^{-p}(\alpha+\beta x)(1-x)$ and $\lambda\left(p_{0}, \frac{d-1}{m}\right) \geq \underline{\lambda}:=e^{-p_{0}} \alpha(1-\gamma) m$ for $d \in\{1, \ldots, \gamma m\}$.

Therefore, $\mathbb{E}\left[\tau_{1}\right]=\mathbb{E}\left[I_{1}\right]=\frac{1}{e^{-p} \alpha m}$, and $\mathbb{E}\left[\tau_{\gamma m}\right]=\mathbb{E}\left[\sum_{d=1}^{\gamma m} I_{d}\right]$. Using Lemma 20 we have:

$$
\begin{aligned}
& \quad \mathbb{P}\left(\left|\tau_{\gamma m}-\sum_{d=1}^{\gamma m} \mathbb{E}\left[I_{d}\right]\right| \geq \epsilon\right) \leq 2 \exp \left(-\frac{\epsilon^{2} \underline{\underline{\lambda}}^{2}}{8 \gamma m}\right), \\
& \text { so that }\left|\tau_{\gamma m}-\sum_{d=1}^{\gamma m} \frac{1}{\lambda\left(p_{0}, d\right)}\right| \leq \frac{e^{p_{0}}}{\alpha(1-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}} \quad \text { with probability } 1-\delta,
\end{aligned}
$$

where we set $\epsilon=\sqrt{\frac{8 \log \left(\frac{2}{\delta}\right) \gamma m}{\underline{\lambda}^{2}}}$. Since $m^{1 / 3} \geq \sqrt{2 \log \left(\frac{2}{\delta}\right)} \Longrightarrow \epsilon \leq \frac{2 \gamma m}{\underline{\lambda}}$, the condition for

Lemma 20 is satisfied. To bound the estimation error of $\alpha$ :

$$
\begin{aligned}
\left|\tau_{\gamma m}-\gamma m \mathbb{E}\left[\tau_{1}\right]\right| & =\left|\tau_{\gamma m}-\sum_{d=1}^{\gamma m} \frac{1}{\lambda\left(p_{0}, d\right)}\right|+\left|\sum_{d=1}^{\gamma m} \frac{1}{\lambda\left(p_{0} \frac{d-1}{m}\right)}-\gamma m \mathbb{E}\left[\tau_{1}\right]\right| \\
\text { (with probability } 1-\delta) & \leq \frac{e^{p_{0}}}{\alpha(1-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\left|\sum_{d=1}^{\gamma m}\left(\frac{1}{\lambda\left(p_{0} \frac{d-1}{m}\right)}-\frac{1}{e^{-p_{0}} \alpha m}\right)\right| \\
& \leq \frac{e^{p_{0}}}{\alpha(1-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\left|\sum_{d=1}^{\gamma m} \frac{\left|e^{-p_{0}} \alpha m-e^{-p_{0}}\left(\alpha+\beta \frac{d-1}{m}\right)(m-d+1)\right|}{\left(e^{\left.-p_{0} \alpha(1-\gamma) m\right)^{2}}\right.}\right| \\
& \leq \frac{e^{p_{0}}}{\alpha(1-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\left|\sum_{d=1}^{\gamma m} \frac{\max (\alpha, \beta) d}{e^{-p_{0} \alpha^{2}(1-\gamma)^{2} m^{2}}}\right| \\
& \leq \frac{e^{p_{0}}}{\alpha(1-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\frac{e^{p_{0}} \max (\alpha, \beta) \gamma^{2}}{\alpha^{2}(1-\gamma)^{2}} \\
& \leq \frac{2 e^{p_{0}}(\alpha+\beta)}{\alpha^{2}(1-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right) m^{-2 / 3}}
\end{aligned}
$$

Denote the above bound by $\mathcal{B}_{\alpha}$. Plug this and $p_{0}=0$ into the $|\hat{\alpha}-\alpha|$ expression above we have with probability $1-\delta$ :

$$
\begin{aligned}
|\hat{\alpha}-\alpha| & \leq \gamma \frac{\mathcal{B}_{\alpha}}{\left(\gamma m \mathbb{E}\left[\tau_{1}\right]-\mathcal{B}_{\alpha}\right)^{2}} \\
& \leq \gamma \frac{4 \mathcal{B}_{\alpha}}{\gamma^{2} m^{2} \mathbb{E}\left[\tau_{1}\right]^{2}} \\
& =\frac{8(\alpha+\beta)}{(1-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right) m^{-1 / 3}}
\end{aligned}
$$

where in the second inequality we used the fact that $m^{1 / 3} \geq \frac{16(\alpha+\beta)}{\alpha} \sqrt{8 \log \left(\frac{2}{\delta}\right)} \Longrightarrow \mathcal{B}_{\alpha} \leq$ $\frac{1}{2} \gamma m \mathbb{E}\left[\tau_{1}\right]$.

Lemma 26. Assuming that $m^{1 / 3} \geq 64 \frac{(\alpha+\beta)^{2}}{\alpha^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, then for any $i=1, \ldots, K$, with probability $1-\delta,\left|\beta-\hat{\beta}_{i}\right| \leq B_{i}$.

Proof. Note that $\mathbb{E}\left[I_{\gamma_{i} m+1}\right]=\mathbb{E}\left[\tau_{\gamma_{i} m+1}-\tau_{\gamma_{i} m}\right]=\frac{1}{\lambda\left(p_{0}, \gamma_{i}\right)}=\frac{1}{e^{-p_{0}}\left(\alpha+\beta \gamma_{i}\right)\left(1-\gamma_{i}\right) m}$. From (1.9) we can
bound the estimation error of $\hat{\beta}$ as follows. We have

$$
\begin{aligned}
\hat{\alpha}+\hat{\beta}_{i} \gamma_{i} & =\frac{\gamma m}{e^{-p_{0}}\left(1-\gamma_{i}\right) m\left(\tau_{\left(\gamma_{i}+\gamma\right) m}-\tau_{\gamma_{i} m}\right)}, \\
\alpha+\beta \gamma_{i} & =\frac{\gamma m}{e^{-p_{0}}\left(1-\gamma_{i}\right) m \mathbb{E}\left[\tau_{\gamma_{i} m+1}-\tau_{\gamma_{i} m}\right] \gamma m} .
\end{aligned}
$$

Therefore,

$$
\left|\beta-\hat{\beta}_{i}\right| \leq \frac{|\hat{\alpha}-\alpha|}{\gamma_{i}}+\frac{\gamma m}{\gamma_{i} e^{-p_{0}}\left(1-\gamma_{i}\right) m}\left|\frac{1}{\tau_{\left(\gamma_{i}+\gamma\right) m}-\tau_{\gamma_{i} m}}-\frac{1}{\gamma m \mathbb{E}\left[\tau_{\gamma_{i} m+1}-\tau_{\gamma_{i}}\right]}\right| .
$$

Similar to the estimation bound of $\hat{\alpha}$, the main step is to bound the arrival times. Note that $\lambda\left(p_{0}, \frac{d-1}{m}\right) \geq e^{-p_{0}}\left(\alpha+\beta \gamma_{i}\right)\left(1-\gamma_{i}-\gamma\right) m \geq \underline{\lambda}:=e^{-p_{0}}\left(\alpha+\beta \gamma_{i}\right)(1 / 3-\gamma) m$ for $d \in\left\{\gamma_{i} m+\right.$ $\left.1, \ldots,\left(\gamma_{i}+\gamma\right) m\right\}$, where we used the fact that by the construction of Algorithm $1, \gamma_{i} \leq 2 / 3$ for all $i=1, \ldots, K$. Using Lemma 20 we have:

$$
\begin{aligned}
& \mathbb{P}\left(\left|\tau_{\gamma m}-\sum_{d=1}^{\gamma m} \mathbb{E}\left[I_{d}\right]\right| \geq \epsilon\right) \leq 2 \exp \left(-\frac{\epsilon^{2} \underline{\lambda}^{2}}{8 \gamma m}\right) \\
& \text { so that }\left|\tau_{\gamma m}-\sum_{d=1}^{\gamma m} \frac{1}{\lambda\left(p_{0}, d\right)}\right| \leq \frac{e^{p_{0}}}{\left(\alpha+\beta \gamma_{i}\right)(1 / 3-\gamma)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}} \quad \text { with probability } 1-\delta
\end{aligned}
$$

where we set $\epsilon=\sqrt{\frac{8 \log \left(\frac{2}{\delta}\right) \gamma m}{\underline{\lambda}^{2}}}$. Since $m^{1 / 3} \geq \sqrt{2 \log \left(\frac{2}{\delta}\right)} \Longrightarrow \epsilon \leq \frac{2 \gamma m}{\underline{\lambda}}$, the condition for Lemma 20 is satisfied. To bound the estimation error of $\beta$ :

$$
\begin{aligned}
& \left|\left(\tau_{\left(\gamma_{i}+\gamma\right) m}-\tau_{\gamma_{i} m}\right)-\gamma m \mathbb{E}\left[\tau_{\gamma_{i} m+1}-\tau_{\gamma_{i} m}\right]\right|=\left|\sum_{d=\gamma_{i} m+1}^{\left(\gamma_{i}+\gamma\right) m} I_{d}-\frac{\gamma m}{\lambda\left(p_{0}, \gamma_{i}\right)}\right| \\
\leq & \left|\sum_{d=\gamma_{i} m+1}^{\left(\gamma_{i}+\gamma\right) m}\left(I_{d}-\frac{1}{\lambda\left(p_{0}, \frac{d-1}{m}\right)}\right)\right|+\left|\sum_{d=\gamma_{i} m+1}^{\left(\gamma_{i}+\gamma\right) m} \frac{1}{\lambda\left(p_{0}, \frac{d-1}{m}\right)}-\frac{\gamma m}{\lambda\left(p_{0}, \gamma_{i}\right)}\right| \\
\text { (w.p. } 1-\delta) \leq & \frac{e^{p_{0}}}{(1 / 3-\gamma)\left(\alpha+\beta \gamma_{i}\right)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\left|\sum_{d=\gamma_{i} m+1}^{\gamma_{i} m+\gamma m}\left(\frac{1}{\lambda\left(p_{0}, d\right)}-\frac{1}{\lambda\left(p_{0}, \gamma_{i}\right)}\right)\right| \\
\leq & \frac{e^{p_{0}}}{(1 / 3-\gamma)\left(\alpha+\beta \gamma_{i}\right)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}} \\
& +\left|\sum_{d=\gamma_{i} m+1}^{\gamma_{i} m+\gamma m} \frac{\left\lvert\, e^{-p_{0}}\left(\alpha+\beta \frac{d-1}{m}\right)(m-d\right.}{\left(e^{-p_{0}}\left(\alpha+\beta \gamma_{i}\right)\left(1-\gamma_{i}-\gamma\right) m\right)^{2}}\right| \\
\leq & \frac{e^{p_{0}}}{(1 / 3-\gamma)\left(\alpha+\beta \gamma_{i}\right)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\left|\sum_{d=\gamma_{i} m+1}^{\gamma_{i} m+\gamma m}\left(\frac{p_{0}}{e^{-p_{0}}\left(\alpha+\beta \gamma_{i}\right)^{2}\left(1-\gamma_{i}-\gamma\right)^{2} m}\right)\right| \\
\leq & \frac{e^{p_{0}}}{(1 / 3-\gamma)\left(\alpha+\beta \gamma_{i}\right)} \sqrt{8 \log \left(\frac{2}{\delta}\right) \frac{\gamma}{m}}+\frac{e^{p_{0}}(\alpha+\beta) \gamma^{2}}{\left(\alpha+\beta \gamma_{i}\right)^{2}(1 / 3-\gamma)^{2}} \\
\leq & \frac{2 e^{p_{0}}(\alpha+\beta)}{\left(\alpha+\beta \gamma_{i}\right)^{2}(1 / 3-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right) m^{-2 / 3}}
\end{aligned}
$$

Let $\mathcal{B}_{\beta}$ denote this bound. Plug this result back into the bound on $\left|\beta-\hat{\beta}_{i}\right|$ :

$$
\begin{aligned}
\left|\beta-\hat{\beta}_{i}\right| & \leq \frac{|\hat{\alpha}-\alpha|}{\gamma_{i}}+\frac{\gamma m}{\gamma_{i} e^{-p_{0}}\left(1-\gamma_{i}\right) m}\left(\frac{1}{\tau_{\left(\gamma_{i}+\gamma\right) m}-\tau_{\gamma_{i} m}}-\frac{1}{\gamma m \mathbb{E}\left[\tau_{\gamma_{i} m+1}-\tau_{\gamma_{i}}\right]}\right) \\
& \leq \frac{|\hat{\alpha}-\alpha|}{\gamma_{i}}+\frac{e^{p_{0}} \gamma}{\gamma_{i}\left(1-\gamma_{i}\right)}\left(\frac{\mathcal{B}_{\beta}}{\left(\gamma m \mathbb{E}\left[I_{\gamma_{i} m+1}\right]-\mathcal{B}_{\beta}\right)^{2}}\right) \\
\left(^{*}\right) & \leq \frac{|\hat{\alpha}-\alpha|}{\gamma_{i}}+\frac{e^{p_{0} \gamma}}{\gamma_{i}\left(1-\gamma_{i}\right)}\left(\frac{4 \mathcal{B}_{\beta}}{\gamma^{2} m^{2} \mathbb{E}\left[I_{\gamma_{i} m+1}\right]^{2}}\right) \\
& \leq \frac{|\hat{\alpha}-\alpha|}{\gamma_{i}}+\frac{8(\alpha+\beta)\left(1-\gamma_{i}\right)}{\gamma_{i}(1 / 3-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right) m^{-1 / 3}} \\
& \leq \frac{16(\alpha+\beta)\left(1-\gamma_{i}\right)}{\gamma_{i}(1 / 3-\gamma)^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right) m^{-1 / 3}}
\end{aligned}
$$

where for the $\left(^{*}\right)$ step we used the fact that $m^{1 / 3} \geq 64 \frac{(\alpha+\beta)^{2}}{\alpha^{2}} \sqrt{8 \log \left(\frac{2}{\delta}\right)} \Longrightarrow \mathcal{B}_{\beta} \leq \frac{1}{2} \gamma m \mathbb{E}\left[I_{\gamma_{i} m+1}\right]$.

Lemma 4 (Error bounds for estimated prices). Given any market parameters $\alpha, \beta$ and their estimators $\hat{\alpha}, \hat{\beta}_{i}$ that satisfy $|\alpha-\hat{\alpha}| \leq A,\left|\beta-\hat{\beta}_{i}\right| \leq B_{i}, \hat{\alpha}-A>0, \hat{\beta}_{i}-B_{i}>0$, then for every $d=0, \ldots, m-1$,

$$
\left|p^{*}\left(\frac{d}{m}, \hat{\alpha}, \hat{\beta}_{i}\right)-p^{*}\left(\frac{d}{m}, \alpha, \beta\right)\right| \leq L_{\alpha} A+L_{\beta i} B_{i},
$$

where $L_{\alpha}, L_{\beta i}$ are as defined in (1.11).
Proof. Clearly, $\hat{\alpha}-A \leq \alpha, \hat{\beta}_{k}-B_{k} \leq \beta$, and $\hat{\beta}_{k}+B_{k} \geq \beta$. Then using Lemma 22 we have

$$
\begin{aligned}
\left|p^{*}\left(x, \hat{\alpha}, \hat{\beta}_{k}\right)-p^{*}(x, \alpha, \beta)\right| \leq & \max _{\alpha^{\prime} \in[\alpha, \hat{\alpha}] \text { or } \alpha^{\prime} \in[\hat{\alpha}, \alpha]}\left\{\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \alpha}\right|_{\alpha=\alpha^{\prime}}\right\}|\alpha-\hat{\alpha}| \\
& +\max _{\beta^{\prime} \in[\beta, \hat{\beta}] \text { or } \beta^{\prime} \in[\hat{\beta}, \beta]}\left\{\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial \beta}\right|_{\beta=\beta^{\prime}}\right\}\left|\beta-\hat{\beta}_{k}\right| \\
\leq & L_{\alpha} A+L_{\beta i} B_{k}
\end{aligned}
$$

## A.4.2 Step 2: Proof of Lemma 5

Since the prices that we offer in the stochastic Bass model is based on a discretized version of the continuous price curve in the deterministic Bass model, we first need to prove a result that says that this discretization does not introduce a lot of error. Lemma 27 below shows that it only introduces a logarithmic (in $m$ ) amount of error, for any fixed $\alpha, \beta, T$.

Lemma 27. For any fixed $T, \alpha, \beta$ and $n=1, \ldots, m X_{T}^{*}$,

$$
\left|\sum_{d=1}^{n} p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right)-m \int_{0}^{n / m} p^{*}(x, \alpha, \beta) d x\right| \leq \frac{n}{2 m} \frac{\beta}{\alpha}+\frac{1}{2} \log (m)+2 p_{\max }^{*}
$$

where $p_{\text {max }}^{*}$ denotes an upper bound on the optimal prices $p^{*}(x, \alpha, \beta)$ for all $x$.

Proof. Using (1.4),

$$
\left|\frac{\partial p^{*}(x, \alpha, \beta)}{\partial x}\right| \leq \frac{\beta}{\alpha}+\frac{1}{1-x}
$$

In below we abbreviate $p^{*}(x, \alpha, \beta)$ as $p_{x}^{*}$.

$$
\begin{aligned}
& \left|\sum_{d=1}^{n} p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right)-m \int_{0}^{n / m} p^{*}(x, \alpha, \beta) d x\right| \\
\leq & \sum_{d=1}^{n-2}\left|p_{\frac{d-1}{m}}^{*}-m \int_{\frac{d-1}{m}}^{d / m} p_{x}^{*} d x\right|+2 p_{\text {max }}^{*} \\
\leq & \sum_{d=1}^{n-2}\left[m \int_{\frac{d-1}{m}}^{d / m}\left(\frac{\beta}{\alpha}+\frac{1}{1-d / m}\right)\left(x-\frac{d-1}{m}\right) d x\right]+2 p_{\text {max }}^{*} \\
= & \frac{n}{2 m} \frac{\beta}{\alpha}+\sum_{d=1}^{n-2} \frac{1}{2(1-d / m) m}+2 p_{\text {max }}^{*} \\
\leq & \frac{n}{2 m} \frac{\beta}{\alpha}+\frac{1}{2 m} \int_{1}^{n-1} \frac{1}{(1-g / m)} d g+2 p_{\text {max }}^{*} \\
\leq & \frac{n}{2 m} \frac{\beta}{\alpha}+\frac{1}{2} \log \left(\frac{1}{1-\frac{n-1}{m}}\right)+2 p_{\text {max }}^{*} \\
\leq & \frac{n}{2 m} \frac{\beta}{\alpha}+\frac{1}{2} \log (m)+2 p_{\text {max }}^{*}
\end{aligned}
$$

The first inequality follows because we know $p^{*}(x, \alpha, \beta)$ is bounded below and above by 0 and $p_{\max }^{*}$. Therefore the difference between $p^{*}$ evaluated at two different $x$ values is at most $p_{\text {max }}^{*}$.

Corollary 3. For any fixed $T, \alpha, \beta$,

$$
\left|\sum_{d=1}^{\left\lfloor m X_{T}^{*}\right\rfloor} p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right)-m \int_{0}^{X_{T}^{*}} p^{*}(x, \alpha, \beta) d x\right| \leq \frac{\beta}{2 \alpha}+\frac{1}{2} \log (m)+3 p_{\max }^{*}
$$

Proof. Let $n=\left\lfloor m X_{T}^{*}\left(x_{0}\right)\right\rfloor-d_{0}$. Since rounding $m X_{T}^{*}$ introduces at most at additional $p_{\max }^{*}$ difference in revenue, the result then immediately follows from Lemma 27.

For the lemma below, we define

$$
V_{i}^{\operatorname{det}}(T):=m \int_{\gamma_{i}}^{2 \gamma_{i} \wedge X_{T}^{*}} p^{*}(x, \alpha, \beta) d x
$$

$$
\operatorname{Rev}_{i}:=\sum_{d=\left\lfloor\gamma_{i} m\right\rfloor+1}^{\left\lfloor 2 \gamma_{i} m\right\rfloor \wedge\left\lfloor X_{T}^{*} m\right\rfloor} p_{d}
$$

Lemma 5. For any epoch $i$ in the algorithm, with probability $1-\delta$,

$$
V_{i}^{d e t}-\operatorname{Rev}_{i} \leq O\left(m^{2 / 3} \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right)=\tilde{O}\left(m^{2 / 3}\right) .
$$

Proof. Let $A, B_{i}$ be the bound on the estimation error stated in Lemma 25, Lemma 26. Recall that first $\gamma m$ customers in every epoch are offered an exploration price $p_{0}=0$. Let $p_{d}$ be the price paid by customer $d$ as specified in Algorithm 1 and (1.11). Recall also that $p^{*}(x, \alpha, \beta)$ is the optimal price curve as specified in (1.4). We use the short hand notations $p_{x}^{*}:=p^{*}(x, \alpha, \beta), p_{d}^{*}:=$ $p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right), p_{\text {max }}^{*}:=\max _{x \in[0,1)} p_{x}^{*}$ in the following calculations.

$$
\begin{aligned}
V_{i}^{\operatorname{det}}-\operatorname{Rev}_{i} & =m \int_{\gamma_{i}}^{2 \gamma_{i} \wedge X_{T}^{*}} p_{x}^{*} d x-\sum_{d=\gamma_{i} m+1}^{2 \gamma_{i} m \wedge\left\lfloor X_{T}^{*} m\right\rfloor} p_{d} \\
& =\sum_{d=\gamma_{i} m+1}^{2 \gamma_{i} m \wedge\left\lfloor X_{T}^{*} m\right\rfloor}\left[p_{d}^{*}-p_{d}\right]+\sum_{d=\gamma_{i} m+1}^{2 \gamma_{i} m \wedge\left\lfloor X_{T}^{*} m\right\rfloor} p_{d}^{*}-m \int_{\gamma_{i}}^{2 \gamma_{i} \wedge X_{T}^{*}} p_{x}^{*} d x \mid \\
(\text { Corollary 3)} & \leq \sum_{d=\gamma_{i} m+1}^{2 \gamma_{i} m \wedge\left\lfloor X_{T}^{*} m\right\rfloor}\left[p_{d}^{*}-p_{d}\right]+\frac{\beta}{2 \alpha}+\frac{1}{2} \log (m)+3 p_{\max }^{*} \\
& \leq(\gamma m+3) p_{\text {max }}^{*}+\sum_{d=\left(\gamma_{i}+\gamma\right) m+1}^{2 \gamma_{i} m \wedge\left\lfloor X_{T}^{*} m\right\rfloor}\left[p_{d}^{*}-p_{d}\right]+\frac{\beta}{2 \alpha}+\frac{1}{2} \log (m)
\end{aligned}
$$

$\left(\right.$ Lemma 4, Lemma 25, Lemma 26) $\leq(\gamma m+3) p_{\max }^{*}+2 \gamma_{i} m\left(L_{\alpha} A+L_{\beta i} B_{i}\right)+\frac{\beta}{2 \alpha}+\frac{1}{2} \log (m) \quad$ w.p.1-2 $\delta$

$$
\left(\text { Lemma 7) } \leq O\left(m^{2 / 3} \log ((\alpha+\beta) T)\right)+2 \gamma_{i} m\left(L_{\alpha} A+L_{\beta i} B_{i}\right)\right.
$$

where $L_{\alpha}, L_{\beta i}$ are defined in (1.11).
The second to last step follows because we know that $\left|\underline{p_{i}^{*}}\left(\frac{d-1}{m}\right)-p_{d}^{*}\right| \leq 2\left(L_{\alpha} A+L_{\beta i} B_{i}\right)$ using the definition of $\underline{p_{i}^{*}}$ in (1.11) and Lemma 4, and that from Lemma 25 and Lemma 26 we know that the error bounds $A, B_{i}$ hold with probability $1-2 \delta$.

Assuming that $m^{1 / 3} \geq \frac{64 \phi}{\alpha} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, and $\gamma_{i} m^{1 / 3} \geq \frac{640 \phi}{\beta} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, one can check this implies
that $A \leq \frac{\alpha}{4}$ and $B_{i} \leq \frac{\beta}{4}$, which in turn implies that $\hat{\alpha}-A \geq \frac{\alpha}{2}$ and $\hat{\beta}-B_{i} \geq \frac{\beta}{2}$. Applying these bounds to (1.11) we have:

$$
L_{\alpha} \leq \frac{4}{\alpha}+\frac{6 \beta}{\alpha^{2}} \quad L_{\beta i} \leq 6\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)
$$

Plug in the expressions of $A$ and $B_{i}$, as well as the above bounds on $L_{\alpha}, L_{\beta i}$, we have with probability $1-\delta$,

$$
V_{i}^{\operatorname{det}}(T)-\operatorname{Rev}_{i} \leq O\left(m^{2 / 3} \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right)
$$

When $m^{1 / 3} \leq \frac{64 \phi}{\alpha} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, or $\gamma_{i} m^{1 / 3} \leq \frac{640 \phi}{\beta} \sqrt{8 \log \left(\frac{2}{\delta}\right)}$, the regret can be trivially bounded by $p_{\text {max }}^{*} \gamma_{i} m \leq O\left(\log ((\alpha+\beta) T)\left(\frac{\phi}{\alpha} \sqrt{\log \left(\frac{1}{\delta}\right)}\right)^{3}+\log ((\alpha+\beta) T) \frac{\phi}{\beta} \sqrt{\log \left(\frac{1}{\delta}\right)} m^{2 / 3}\right)$, where we used Lemma 7 to bound $p_{\text {max }}^{*}$. Since the first component is a constant with respect to $m$ and the second is no larger than the expression from before, we are done.

## A.4.3 Step 3: Proof of Lemma 6 and Lemma 7

Lemma 6. If the seller follows Algorithm 1, then with probability at least $1-\delta \log (m)$, the number of adoptions at the end of time horizon $T$ is lower bounded as:

$$
d_{T} \geq m X_{T}^{*}-\sqrt{8 m X_{T}^{*} \log \left(\frac{4}{\delta}\right)} .
$$

Proof. Let $p_{d}$ be the prices that the algorithm offers for customer $d$. Since Algorithm 1 offers either $p_{0}=0$ or the lower confidence bound price defined in (1.11), we know from Corollary 1 , as well as Lemma 25, Lemma 26, that with probability $1-\delta K, p_{d} \leq p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right) \forall d \leq d_{T} \wedge m X_{T}^{*}$, where $K$ is the total number of epochs. This means that $\lambda\left(p_{d}, \frac{d-1}{m}\right) \geq \lambda\left(p^{*}\left(\frac{d-1}{m}, \alpha, \beta\right), \frac{d-1}{m}\right)=$ $\frac{1}{e}\left(\alpha+\beta X_{T}^{*}\right)\left(1-X_{T}^{*}\right) m$. Let $\underline{\lambda}:=\frac{1}{e}\left(\alpha+\beta X_{T}^{*}\right)\left(1-X_{T}^{*}\right) m$, which by (1.5) is equal to $\frac{X_{T}^{*} m}{T}$.

Set $n=\left\lfloor m X_{T}^{*}-\sqrt{8 m X_{T}^{*} \log \left(\frac{2}{\delta}\right)}\right\rfloor, \underline{\lambda}=\frac{m X_{T}^{*}}{T}, \epsilon=\sqrt{\frac{8 n}{\underline{\lambda}^{2}} \log \left(\frac{2}{\delta}\right)}$, then by Lemma 20, we have with
probability $1-\delta$ :

$$
\left|\tau_{n}-\left(1-\sqrt{\frac{8 \log \left(\frac{2}{\delta}\right)}{m X_{T}^{*}}}\right) T\right| \leq T \sqrt{\frac{8}{X_{T}^{*} m} \log \left(\frac{2}{\delta}\right)}
$$

This means that with probability $1-\delta(K+1), \tau_{n} \leq T$, which means that the total number of adoptions observed in the stochastic Bass model is at least $n$. The result follows by observing that there can be at most $\log (m)$ epochs.

Lemma 7 (Upper bound on optimal prices). All prices in the optimal price curve for deterministic Bass model are upper bounded as:

$$
p^{*}(x, \alpha, \beta) \leq \log (e+(\alpha+\beta) T)
$$

Proof. Here we use the expanded notation of $X_{T}^{*}(x, \alpha, \beta)$ introduced in Appendix A.2. Using (1.4), we have for any $x \leq X_{T}^{*}(0, \alpha, \beta)$ :

$$
\begin{aligned}
p^{*}(x, \alpha, \beta) & \leq 1+\log \left(\frac{1-x}{1-X_{T}^{*}(0, \alpha, \beta)}\right) \\
& \leq 1+\log \left(\frac{1}{1-X_{T}^{*}(0, \alpha, \beta)}\right) \\
& \leq \log (e+(\alpha+\beta) T)
\end{aligned}
$$

The last step follows from the following upper bound on $X_{T}^{*}$ :

$$
X_{T}^{*}(0, \alpha, \beta) \leq X_{T}^{*}(0, \alpha+\beta, 0)=\frac{(\alpha+\beta) T}{(\alpha+\beta) T+e}
$$

The first inequality is easy to see from (1.5): by replacing $\alpha+\beta X_{T}^{*}$ with $\alpha+\beta$, we can see that $X_{T}^{*} /\left(1-X_{T}^{*}\right)$ increases, and since this quantity is strictly monotone in $X_{T}^{*}, X_{T}^{*}$ must be larger. And the last equality follows from solving (1.5) after replacing $\alpha+\beta X_{T}^{*}$ with $\alpha+\beta$.
A.4.4 Step 4: Putting it all together for proof of Theorem 1

## Proof of Theorem 1.

$$
\begin{aligned}
\text { Pseudo-Regret }= & \sum_{i=1}^{K}\left[V_{i}^{\mathrm{det}}-\operatorname{Rev}_{i}\right]+\sum_{d=d_{T}+1}^{m X_{T}^{*}} p_{d} \\
\leq & \log (m) O\left(m^{2 / 3} \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right) \quad \text { w.p. } 1-(K+1) \delta \\
& +O(\log ((\alpha+\beta) T) \sqrt{m}) \\
= & O\left(m^{2 / 3} \log (m) \log \left(\frac{T}{\delta}\right)\left(\frac{1}{\alpha}+\frac{1}{\beta}+\frac{\beta}{\alpha^{2}}\right) \phi\right)
\end{aligned}
$$

where the per epoch regrets are bounded using Lemma 5 and the "uncaptured revenue" term is bounded using Lemma 6 and 7.

The inequality holds with probability $1-(K+1) \delta$, and $K$ is the total number of epochs, which is bounded by $\log (m)$ since the epoch length is defined with respect to the number of customers and increases geometrically.

## A. 5 Lower Bound Proofs

## A.5.1 Step 1: missing lemmas and proofs

Lemma 8. At any adoption level $x$ and remaining time $T^{\prime}$, the disadvantage of offering a suboptimal price $p$ in the deterministic Bass model is lower bounded as:

$$
A^{d e t}\left(x, T^{\prime}, p\right) \geq \lambda(p, x) \min \left(\frac{1}{4}\left(\pi^{*}\left(x, T^{\prime}\right)-p\right)^{2}, \frac{1}{10}\right)
$$

where $\pi^{*}\left(x, T^{\prime}\right)=\arg \min _{p} A^{\text {det }}\left(x, T^{\prime}, p\right)$ denotes the optimal price at $x, T^{\prime}$.

Proof. Let $\pi^{*}$ denote $\pi^{*}\left(x, T^{\prime}\right)$ for the remainder of this proof. Using (A.6), we can rewrite the left
hand side of the lemma:

$$
\begin{align*}
& A^{\operatorname{det}}\left(x, T^{\prime}, p\right) \\
= & \lim _{\delta \rightarrow 0} \frac{V^{\operatorname{det}}\left(x, T^{\prime}\right)-p \lambda(p, x) \delta-V^{\operatorname{det}}\left(x+\lambda(p, x) \delta / m, T^{\prime}-\delta\right)}{\delta} \\
= & \lim _{\delta \rightarrow 0} \frac{\pi^{*} \lambda\left(\pi^{*}, x\right) \delta+V^{\operatorname{det}}\left(x+\lambda\left(\pi^{*}, x\right) \delta / m, T^{\prime}-\delta\right)-p \lambda(p, x) \delta-V^{\operatorname{det}}\left(x+\lambda(p, x) \delta / m, T^{\prime}-\delta\right)}{\delta} \\
= & G^{\operatorname{det}}\left(x, T^{\prime}, \pi^{*}\right)-G^{\operatorname{det}}\left(x, T^{\prime}, p\right) \tag{A.11}
\end{align*}
$$

where we define

$$
\begin{align*}
G^{\operatorname{det}}\left(x, T^{\prime}, p\right) & =\lim _{\delta \rightarrow 0} p \lambda(p, x)+\frac{V^{\operatorname{det}}\left(x+\lambda(p, x) \delta / m, T^{\prime}-\delta\right)-V^{\operatorname{det}}\left(x, T^{\prime}-\delta\right)}{\delta} \\
& =p \lambda(p, x)+\frac{\lambda(p, x)}{m} \frac{\partial V^{\operatorname{det}}\left(x, T^{\prime}\right)}{\partial x} \tag{A.12}
\end{align*}
$$

The distribution of limits is valid since both limits exist. The new quantity $G^{\operatorname{det}}\left(x, T^{\prime}, p\right)$ will help us quantify the instantaneous impact, or the (dis)Advantage, of offering a suboptimal price.

From above, note that $\pi^{*}\left(x, T^{\prime}\right):=\arg \min _{p} A^{\text {det }}\left(x, T^{\prime}, p\right)=\arg \max _{p} G^{d e t}\left(x, T^{\prime}, p\right)$. We derived the expression for $\pi^{*}\left(x, T^{\prime}\right)$ earlier in the proof of Lemma 1 (equation (A.8)) as

$$
\pi^{*}\left(x, T^{\prime}\right)=1-\frac{1}{m} \frac{\partial V^{\operatorname{det}}\left(x, T^{\prime}\right)}{\partial x}
$$

We can now bound (A.11) using the derivative of $G^{\text {det. }}$

$$
\begin{aligned}
\frac{\partial G^{\operatorname{det}}\left(x, T^{\prime}, p\right)}{\partial p} & =\lambda(p, x)-p \lambda(p, x)-\frac{\lambda(p, x)}{m} \frac{\partial V^{\operatorname{det}}\left(x, T^{\prime}\right)}{\partial x} \\
& =\left(\pi^{*}-p\right) \lambda(p, x) \\
G^{\operatorname{det}}\left(x, T^{\prime}, \pi^{*}\right)-G^{\operatorname{det}}\left(x, T^{\prime}, p\right) & =\int_{p}^{\pi^{*}}\left(\pi^{*}-\rho\right) \lambda(\rho, x) d \rho \\
& =\lambda(p, x)\left(\pi^{*}-p\right)+\lambda\left(\pi^{*}, x\right)-\lambda(p, x) \\
& =\lambda(p, x)\left(e^{p-\pi^{*}}-1-\left(p-\pi^{*}\right)\right) \\
& \geq \lambda(p, x) \min \left(\frac{1}{10}, \frac{1}{4}\left(p-\pi^{*}\right)^{2}\right)
\end{aligned}
$$

To see the last step, consider $f(x)=e^{x}-1-x$ which is a convex function minimized at $x=0$. For $x \geq-\frac{1}{2}$, it's easy to show that $f^{\prime \prime}(x) \geq e^{-1 / 2} \geq \frac{1}{2}$. So by standard strong convexity argument $f(x) \geq \frac{1}{4} x^{2}$. For $x \leq-\frac{1}{2}$ we have $f(x) \geq f\left(-\frac{1}{2}\right) \geq \frac{1}{10}$.

Before we translate the above result into a regret bound in terms of cumulative pricing error, we explain the proof idea with some more details.

Given any arbitrary pricing algorithm, let

$$
\left[\left(p_{1}, I_{1}\right), \ldots,\left(p_{n}, I_{n}\right)\right]
$$

be the first $n$ observations (tuples of price $p_{d}$ and inter-arrival times $I_{d}$ between customer $d-1$ and $d$ ) in the stochastic Bass model, on following the algorithm's pricing policy. Here we assume that the algorithm is allowed to continue running even after the planning horizon $T$ has passed. If the algorithm is undefined after time $T$, we assume that it offers 0 price. We use these observations to define a continuous price trajectory $p_{x}, 0 \leq x \leq n / m$ as follows: set $p_{x}=p_{d}, \forall x \in\left[\frac{d-1}{m}, \frac{d}{m}\right), d=1, \ldots, n$. We call $p_{x}$ the induced price trajectory of $p_{d}$. Let $t_{n / m}^{d e t}$ denote the time when the adoption level hits $x$ in the deterministic Bass model following this induced pricing trajectory $p_{x}$. In other words, $t_{x}^{d e t}=\int_{0}^{x} \frac{m}{\lambda\left(x^{\prime}, p_{x^{\prime}}\right)} d x^{\prime}$. Note that $t_{n / m}^{d e t}$ is a stochastic quantity because it depends on stochastic trajectory $p_{x}$, which in turn depends on the prices $p_{d}$ offered to the first $n$ customers in the stochastic model.

Recall that $\tau_{n}=\sum_{d=1}^{n} I_{d}$ denotes the arrival time of customer $n$ in the stochastic model. First we show that the total time for $n$ customer arrivals in the deterministic vs. stochastic model (i.e., $t_{n / m}^{d e t}$ vs. $\tau_{n}$ ) under the two price trajectories ( $p_{x}$ vs. $p_{d}$ ) is roughly the same.

Lemma 28. Given $n \leq m, \delta \in(0,1)$ such that $n \geq 2 \log \left(\frac{2}{\delta}\right)$. Then for any algorithm satisfying Assumption 1, with probability at least $1-\delta$,

$$
\left|t_{n / m}^{d e t}-\tau_{n}\right| \leq \frac{e^{p_{\max }}}{\alpha(m-n)} \sqrt{8 n \log \left(\frac{2}{\delta}\right)}+\frac{e^{p_{\max }}(\alpha+\beta) n}{2 \alpha^{2}(m-n-1)^{2} m}
$$

where $p_{\text {max }}$ is an upper bound on the prices $p_{d}, 1 \leq d \leq d_{T}$ offered by algorithm.
Proof. Since algorithm's prices are fixed between arrival of two customers (by Assumption 1), we have that inter-arrival times follow an exponential distribution. That is, for any $d$, given $p_{d+1}$, the price paid by $(d+1)^{t h}$ customer, $\tau_{d+1}-\tau_{d}$ follows the exponential distribution $\operatorname{Exp}\left(\lambda\left(p_{d+1}, \frac{d}{m}\right)\right)$, where $\lambda(p, x)=e^{-p}(\alpha+\beta x)(1-x) m$. Note that $\lambda\left(p_{d+1}, \frac{d}{m}\right) \geq \underline{\lambda}:=e^{-p_{\max }} \alpha(m-n)$ for all $d \leq n$. Set $\epsilon=\sqrt{\frac{8 n}{\underline{\lambda}^{2}} \log \left(\frac{2}{\delta}\right)}$, then Lemma 20 in Appendix A. 1 provides that with probability $1-\delta$ :

$$
\left|\tau_{n}-\sum_{d=1}^{n} \mathbb{E}\left[\tau_{d}-\tau_{d-1} \mid \mathcal{F}_{d-1}\right]\right| \leq \sqrt{\frac{8 n}{\underline{\lambda}^{2}} \log \left(\frac{2}{\delta}\right)}
$$

Note that the condition on $\epsilon$ in Lemma 20 is satisfied since $\sqrt{\frac{8 n}{\lambda^{2}} \log \left(\frac{2}{\delta}\right)} \leq \frac{2 n}{\underline{\lambda}} \Longleftrightarrow n \geq 2 \log \left(\frac{2}{\delta}\right)$.
On the other hand, for any $d \leq n, t_{\frac{d+1}{m}}^{d e t}-t_{\frac{d}{m}}^{d e t}=m \int_{d / m}^{\frac{d+1}{m}} \frac{1}{\lambda\left(p_{d+1}, x\right)} d x$. It's easy to check that $\left|\frac{\partial}{\partial x} \frac{1}{\lambda(p, x)}\right|=\left|\frac{1}{m} \frac{e^{p}[\beta(1-x)-(\alpha+\beta x)]}{(\alpha+\beta x)^{2}(1-x)^{2}}\right| \leq \frac{e^{p}(\alpha+\beta)}{\alpha^{2}(1-x)^{2} m}$. So

$$
\begin{aligned}
\left|t_{\frac{d+1}{m}}^{d e t}-t_{d / m}^{d e t}-\mathbb{E}\left[\tau_{d+1}-\tau_{d} \mid \mathcal{F}_{d}\right]\right| & =\left|m \int_{d / m}^{\frac{d+1}{m}} \frac{1}{\lambda\left(p_{d+1}, x\right)} d x-\frac{1}{\lambda\left(p_{d+1}, d / m\right)}\right| \\
& \leq\left|m \int_{d / m}^{\frac{d+1}{m}} \frac{1}{\lambda\left(p_{d+1}, \frac{d}{m}\right)}+\frac{e^{p_{d+1}}(\alpha+\beta)}{\alpha^{2}\left(1-\frac{d+1}{m}\right)^{2} m}\left(x-\frac{d}{m}\right) d x-\frac{1}{\lambda\left(p_{d+1}, d / m\right)}\right| \\
& \leq\left|\frac{e^{p_{d+1}}(\alpha+\beta)}{\alpha^{2}\left(1-\frac{d+1}{m}\right)^{2}} \int_{d / m}^{\frac{d+1}{m}}\left(x-\frac{d}{m}\right) d x\right| \\
& \leq \frac{e^{p_{d+1}(\alpha+\beta)}}{2 \alpha^{2}\left(1-\frac{d+1}{m}\right)^{2} m^{2}}
\end{aligned}
$$

In the second to last step we canceled the first and third term from the previous step. Combining this with above, we have with probability $1-\delta$

$$
\left|t_{n / m}^{d e t}-\tau_{n}\right| \leq \frac{e^{p_{\max }}}{\alpha(m-n)} \sqrt{8 n \log \left(\frac{2}{\delta}\right)}+\frac{e^{p_{\max }}(\alpha+\beta) n}{2 \alpha^{2}(m-n-1)^{2} m}
$$

Using $t_{x}^{d e t}$ and the induced pricing trajectory $p_{x}$ as defined right before Lemma 28, we can now obtain the following result.

Lemma 29. Fix any $\alpha$, $\beta$. Then for any $m, T$ such that $m X_{T}^{*} \geq n:=\left\lfloor\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right\rfloor$, and any algorithm satisfying Assumption 1 and 2,
$\mathbb{E}[\operatorname{Pseudo-Regret}(T)] \geq \mathbb{E}\left[m \int_{0}^{n / m} \min \left(\frac{1}{4}\left(\pi_{x}^{*}-p_{x}\right)^{2}, \frac{1}{10}\right) d x\right]-O\left(\frac{(\alpha+\beta)^{1 / 3} T e^{C}}{\alpha^{1 / 3}} m^{1 / 3} \sqrt{\log (m)}\right)$
where $\pi_{x}^{*}:=\pi^{*}\left(x, \alpha, \beta, T-t_{x}^{\text {det }}\right)$ denotes the deterministic optimal price at adoption level $x, p_{x}:=$ $p_{d}$ denotes the price offered by the algorithm for customer $d=\lfloor m x\rfloor+1$, and $t_{x}^{\text {det }}$ is the time at which adoption level reaches $x$ on following the price trajectory $p_{x^{\prime}}, \forall x^{\prime} \leq x$ in the deterministic Bass model.

Proof. Let $X_{t}$ be the trajectory of adoption levels in the deterministic Bass model on following price curve $p_{x}$. Recall that by (1.2), $m \frac{d X_{t}}{d t}=\lambda\left(p_{t}, X_{t}\right)$. Let $p_{t}:=p_{X_{t}}$ be the same price trajectory as $p_{x}$ but indexed by time.

First, note that for $m$ large enough we have $t_{n / m}^{\operatorname{det}} \leq \frac{n}{e^{-p \operatorname{pax} \alpha(m-n)}} \leq \frac{T e^{C} n}{\alpha(m-n)} \leq T$, This is because the last inequality holds for $m \geq n+\frac{n}{\alpha} e^{C}$, i.e., for any $m$ satisfying $m^{1 / 3} \geq\left(\frac{(\alpha+1) \alpha^{1 / 3}}{(\alpha+\beta)^{4 / 3}}\right) e^{C}$. Here we used the upper bound $p_{\max }:=\log (T)+C$ from Assumption 2. Therefore for the rest of the proof
we can assume that $t_{n / m}^{d e t} \leq T$.

$$
\begin{align*}
& \mathbb{E}[\text { Pseudo-Regret }(T)] \\
&= \mathbb{E}\left[V^{\operatorname{det}}(0, T)-\sum_{d=1}^{n} p_{d}-\sum_{d=n+1}^{d_{T}} p_{d}\right] \\
&= \mathbb{E}\left[V^{\operatorname{det}}(0, T)-\sum_{d=1}^{n} p_{d}\right]-\mathbb{E}\left[\mathbb{E}\left[\sum_{d=n+1}^{d_{T}} p_{d} \mid \mathcal{F}_{n}\right]\right] \\
& \geq \mathbb{E}\left[V^{\operatorname{det}}(0, T)-\sum_{d=1}^{n} p_{d}\right]-\mathbb{E}\left[V^{\text {stoch }}\left(\frac{n}{m}, T-\tau_{n}\right)\right] \\
&= \mathbb{E}\left[V^{\operatorname{det}}(0, T)-\sum_{d=1}^{n} p_{d}-V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{d e t}\right)\right]+\mathbb{E}\left[V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{d e t}\right)-V^{\mathrm{stoch}}\left(n, T-\tau_{n}\right)\right] \tag{A.13}
\end{align*}
$$

Given a particular sequence of $p_{d}$ for $d=1, \ldots, n$, and its' induced continuous version $p_{x}$ as defined in the lemma statement, the quantity inside the first expectation brackets is the cumulative "disadvantage" that the $p_{x}$ incurs on the deterministic Bass model, where disadvantage is defined in Lemma 8. Therefore we can bound it as follows:

$$
\begin{aligned}
& \mathbb{E}\left[V^{\operatorname{det}}(0, T)-\sum_{d=1}^{n} p_{d}-V^{\operatorname{det}}\left(n / m, T-t_{n / m}^{d e t}\right)\right] \\
= & \mathbb{E}\left[\int_{0}^{t_{n / m}^{d e t}} V^{\operatorname{det}}\left(X_{t}, T-t\right)-Q^{\operatorname{det}}\left(X_{t}, T-t, p_{t}, d t\right)\right] \\
= & \mathbb{E}\left[\int_{0}^{t_{n / m}^{d e t}} \frac{V^{\operatorname{det}}\left(X_{t}, T-t\right)-p \lambda\left(p_{t}, X_{t}\right) d t-V^{\operatorname{det}}\left(X_{t}+\lambda\left(p_{t}, X_{t}\right) d t / m, T-t-d t\right)}{d t} d t\right] \\
\geq & \mathbb{E}\left[\int_{0}^{t_{n / m}^{d e t}} \lambda\left(p_{t}, X_{t}\right) \min \left(\left(\frac{1}{4} \pi_{X_{t}}^{*}-p_{t}\right)^{2}, \frac{1}{10}\right) d t\right] \\
= & m \mathbb{E}\left[\int_{0}^{n / m} \min \left(\frac{1}{4}\left(\pi_{x}^{*}-p_{x}\right)^{2}, \frac{1}{10}\right) d x\right]
\end{aligned}
$$

where in the last step we applied change of variables $\lambda\left(p_{t}, X_{t}\right) d t=m d X_{t}$
The second part of (A.13) can be bounded by using Lemma 23 and bounding the difference
between $\tau_{n}$ and $t_{n / m}^{d e t}$ :

$$
\begin{aligned}
& \mathbb{E}\left[V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{d e t}\right)-V^{\mathrm{stoch}}\left(n, T-\tau_{n}\right)\right] \\
= & \mathbb{E}\left[V^{\operatorname{det}}\left(\frac{n}{m}, T-\tau_{n}\right)-V^{\text {stoch }}\left(n, T-\tau_{n}\right)\right]+\mathbb{E}\left[V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{\text {det }}\right)-V^{\operatorname{det}}\left(n, T-\tau_{n}\right)\right] \\
\geq & \mathbb{E}\left[V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{d e t}\right)-V^{\operatorname{det}}\left(n, T-\tau_{n}\right)\right]
\end{aligned}
$$

Now recall from (A.7) and (A.8) that $\pi^{*}\left(x, \alpha, \beta, T^{\prime}\right)=1-m \frac{\partial V^{\text {det }}\left(x, T^{\prime}\right)}{\partial x}$ and

$$
\begin{equation*}
\frac{\partial V^{\operatorname{det}}\left(x, T^{\prime}\right)}{\partial T^{\prime}}=\pi^{*} \lambda\left(\pi^{*}, x\right)+\lambda\left(\pi^{*}, x\right)\left(1-\pi^{*}\right)=\lambda\left(\pi^{*}, x\right) \leq(\alpha+\beta) m \tag{A.14}
\end{equation*}
$$

Using Lemma 28, which bounds the difference between $T-t_{n / m}^{d e t}$ and $T-\tau_{n}$, we have with probability $1-\delta$ :

$$
\left|V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{\operatorname{det}}\right)-V^{\operatorname{det}}\left(\frac{n}{m}, T-\tau_{n}\right)\right| \leq O\left(\frac{e^{p_{\max }}(\alpha+\beta)^{1 / 3}}{\alpha^{1 / 3}} m^{1 / 3} \sqrt{\log \left(\frac{2}{\delta}\right)}\right)
$$

Since $V^{\operatorname{det}}\left(\frac{n}{m}, T-\tau_{n}\right) \leq V^{\operatorname{det}}\left(\frac{n}{m}, T\right) \leq m p_{\text {max }}^{*}=O(m \log ((\alpha+\beta) T))$, we can set $\delta=\frac{1}{m}$ and use Assumption 2 to conclude that

$$
\begin{equation*}
\mathbb{E}\left[\left|V^{\operatorname{det}}\left(\frac{n}{m}, T-t_{n / m}^{\operatorname{det}}\right)-V^{\operatorname{det}}\left(\frac{n}{m}, T-\tau_{n}\right)\right|\right] \leq O\left(\frac{(\alpha+\beta)^{1 / 3} T e^{C}}{\alpha^{1 / 3}} m^{1 / 3} \sqrt{\log (m)}\right) \tag{A.15}
\end{equation*}
$$

Here we also assumed that $m \geq\left(2 \log \left(\frac{2}{\delta}\right)\right)^{3 / 2}\left(\frac{\alpha+\beta}{\alpha}\right)^{2}$, which implies that $n \geq 2 \log \left(\frac{2}{\delta}\right)$. This satisfies the condition for Lemma 28. If this assumption on $m$ does not hold, then since the expected pseudo-regret is lower bounded by zero, and the first term in the lemma statement is at most $\frac{n}{10} \leq \frac{1}{5} \log \left(\frac{2}{\delta}\right)$, we have that the lemma statement still holds for $\delta=\Theta\left(\frac{1}{m}\right)$.

## A.5.2 Step 2: missing lemmas and proofs

Lemma 30. For any $i$, let $\left[\left(p_{1}, I_{1}\right), \ldots\left(p_{i}, I_{i}\right)\right]$ be sequence of prices offered and inter-arrival times $\left(I_{d}:=\tau_{d}-\tau_{d-1}\right)$ for the first $i$ customers, and let $\mathcal{F}_{i}$ be the filtration generated by these
random variables. Here, prices could have been chosen adaptively using an arbitrary strategy as long as $p_{i} \in \mathcal{F}_{i-1}$. Let $\pi_{1}^{*} \neq \pi_{2}^{*}$ be any two fixed prices. Fix any deterministic algorithm that takes in the past $n$ observations as input and outputs a single price $\pi \in \mathcal{F}_{n}$. Then, for any $\epsilon>0$ and $n$ such that $n \leq\left(\frac{\alpha m}{\epsilon}\right)^{2 / 3}$, at least one of the following holds:

$$
\begin{aligned}
\mathbb{P}_{\alpha, \beta}\left(\mid \pi-\pi_{2}^{*}\right)\left|\leq\left|\pi-\pi_{1}^{*}\right|\right) & \geq \frac{1}{4}, \text { or }, \\
\mathbb{P}_{\alpha, \beta+\epsilon}\left(\left|\pi-\pi_{1}^{*}\right| \leq\left|\pi-\pi_{2}^{*}\right|\right) & \geq \frac{1}{4},
\end{aligned}
$$

where $\mathbb{P}_{\alpha, \beta}$ denotes the probability distribution under the stochastic Bass model with parameters $\alpha, \beta$. Note that the only random quantity is $\pi$, which depends on the first $n$ observations.

Proof. Since $p_{i} \in \mathcal{F}_{i-1}$, the probability of observing the sequence $\left[\left(p_{1}, I_{1}\right), \ldots\left(p_{n}, I_{n}\right)\right]$ is the product of the probabilities of observing $I_{i}$ given $\mathcal{F}_{i-1}$.

$$
\begin{aligned}
\mathbb{P}_{\alpha, \beta}\left(\left[\left(p_{1}, I_{1}\right), \ldots\left(p_{n}, I_{n}\right)\right]\right) & =\prod_{i=1}^{n} \mathbb{P}_{\alpha, \beta}\left(I_{i} \mid \mathscr{F}_{i-1}\right) \\
\mathbb{P}_{\alpha, \beta+\epsilon}\left(\left[\left(p_{1}, I_{1}\right), \ldots\left(p_{n}, I_{n}\right)\right]\right) & =\prod_{i=1}^{n} \mathbb{P}_{\alpha, \beta+\epsilon}\left(I_{i} \mid \mathscr{F}_{i-1}\right)
\end{aligned}
$$

The $(\alpha, \beta)$ subscript denotes the fact that $I_{i}$ 's are generated according to the stochastic Bass model with parameters $(\alpha, \beta)$. Since customer arrivals are Poisson, we know that given price $p_{i}$, the arrival time $I_{i}$ between customer $i-1$ and $i$ is exponentially distributed. Specifically, $I_{i} \sim \operatorname{Exp}\left(\lambda_{\alpha, \beta}\left(p_{i}, \frac{i-1}{m}\right)\right)$ in the $(\alpha, \beta)$ market, and $I_{i} \sim \operatorname{Exp}\left(\lambda_{\alpha, \beta+\epsilon}\left(p_{i}, \frac{i-1}{m}\right)\right)$ in the $(\alpha, \beta+\epsilon)$ market, where $\lambda_{\alpha, \beta}(p, x)=e^{-p}(\alpha+\beta x)(1-x) m$. We can now bound the KL divergence between the
joint distributions of the first $n$ observations between the two markets:

$$
\begin{align*}
\operatorname{KL}\left(\mathbb{P}_{\alpha, \beta+\epsilon}, \mathbb{P}_{\alpha, \beta}\right) & =\operatorname{KL}\left(\prod_{i=1}^{n} \mathbb{P}_{\alpha, \beta+\epsilon}\left(I_{i} \mid \mathscr{F}_{i-1}\right), \prod_{i=1}^{n} \mathbb{P}_{\alpha, \beta}\left(I_{i} \mid \mathscr{F}_{i-1}\right)\right) \\
& =\sum_{d=1}^{n} \operatorname{KL}\left(\operatorname{Exp}\left(\lambda_{\alpha, \beta+\epsilon}\left(p_{d}, \frac{d-1}{m}\right)\right), \left.\operatorname{Exp}\left(\lambda_{\alpha, \beta}\left(p_{d}, \frac{d-1}{m}\right)\right) \right\rvert\,\right) \\
& \leq \sum_{d=1}^{n} \frac{\left(\epsilon \frac{d-1}{m}\right)^{2}}{2\left(\alpha+\beta \frac{d-1}{m}\right)^{2}} \\
& \leq n \frac{\left(\epsilon \frac{n}{m}\right)^{2}}{2 \alpha^{2}} \tag{A.16}
\end{align*}
$$

where the second equality follows from the standard decomposition of KL divergence (see for example Lemma 15.1 of [93]) and the inequality follows from the following bound on the KL divergence of the two exponential distributions:

The KL divergence for a general pair of exponentials is $\operatorname{KL}\left(\operatorname{Exp}(\lambda), \operatorname{Exp}\left(\lambda_{0}\right)\right)=\log \left(\frac{\lambda_{0}}{\lambda}\right)+$ $\frac{\lambda}{\lambda_{0}}-1$

$$
\begin{aligned}
& \operatorname{KL}\left(\operatorname{Exp}\left(\lambda_{\alpha, \beta+\epsilon}\left(p_{d}, \frac{d-1}{m}\right)\right), \operatorname{Exp}\left(\lambda_{\alpha, \beta}\left(p_{d}, \frac{d-1}{m}\right)\right)\right) \\
= & \frac{\alpha+\beta \frac{d-1}{m}+\epsilon \frac{d-1}{m}}{\alpha+\beta \frac{d-1}{m}}-1+\ln \left(\frac{\alpha+\beta \frac{d-1}{m}}{\alpha+\beta \frac{d-1}{m}+\epsilon \frac{d-1}{m}}\right) \\
= & \frac{\epsilon \frac{d-1}{m}}{\alpha+\beta \frac{d-1}{m}}-\frac{\epsilon \frac{d-1}{m}}{\alpha+\beta \frac{d-1}{m}}+\frac{\left(\epsilon \frac{d-1}{m}\right)^{2}}{2\left(\alpha+\beta \frac{d-1}{m}\right)^{2}}-\frac{2\left(\epsilon \frac{d-1}{m}\right)^{3}}{3!\left(\alpha+(\beta+\tilde{\epsilon}) \frac{d-1}{m}\right)^{3}} \\
\leq & \frac{\left(\epsilon \frac{d-1}{m}\right)^{2}}{2\left(\alpha+\beta \frac{d-1}{m}\right)^{2}}
\end{aligned}
$$

where in the last equality $\tilde{\epsilon}$ is some value in between 0 and $\epsilon$. Now let $A_{n}$ be a sequence of $n$ observations such that the policy outputs a price that is closer to $\pi_{1}^{*}$. Using Pinsker's inequality and
(A.16), we can bound the difference in probability of observing this sequence in the two markets:

$$
\begin{align*}
& 2\left(\mathbb{P}_{\alpha, \beta}\left(A_{n}\right)-\mathbb{P}_{\alpha, \beta+\epsilon}\left(A_{n}\right)\right)^{2} \leq \operatorname{KL}\left(\mathbb{P}_{\alpha, \beta+\epsilon}, \mathbb{P}_{\alpha, \beta}\right) \\
& \quad\left|\mathbb{P}_{\alpha, \beta}\left(A_{n}\right)-\mathbb{P}_{\alpha, \beta+\epsilon}\left(A_{n}\right)\right| \leq \sqrt{n \frac{\left(\epsilon \frac{n}{m}\right)^{2}}{4 \alpha^{2}}}=\frac{\epsilon n^{3 / 2}}{2 \alpha m} \tag{A.17}
\end{align*}
$$

Plugging in the upper bound on on $n$ from the lemma statement to (A.17) gives us $\mathbb{P}_{\alpha, \beta}\left(A_{n}\right)$ $\mathbb{P}_{\alpha, \beta+\epsilon}\left(A_{n}\right) \left\lvert\,<\frac{1}{2}\right.$.

However, suppose neither inequality in the lemma statement holds, then by the definition of $A_{n}$, we have that $\left|\mathbb{P}_{\alpha, \beta}\left(A_{n}\right)-\mathbb{P}_{\alpha, \beta+\epsilon}\left(A_{n}\right)\right| \geq\left|\frac{3}{4}-\frac{1}{4}\right|=\frac{1}{2}$ for $n^{3 / 2} \leq \frac{\alpha m}{\epsilon}$, which is a contradiction.

Above lemma holds for any two prices $\pi_{1}^{*} \neq \pi_{2}^{*}$. Readers should think of $\pi_{1}^{*}, \pi_{2}^{*}$ as the optimal prices for customer $n+1$ in the $(\alpha, \beta)$ and $(\alpha, \beta+\epsilon)$ market respectively. To reduce clutter in the following Corollary statement, let $\pi_{1}^{*}=\pi^{*}\left(x, \alpha, \beta, T^{\prime}\right), \pi_{2}^{*}=\pi^{*}\left(x, \alpha, \beta+\varepsilon, T^{\prime}\right)$, and let $\pi_{M}^{*}(x)$ be the optimal price for market $M$ i.e., $\pi_{M}^{*}=\pi_{1}^{*}$ if $M=(\alpha, \beta)$ and $\pi_{M}^{*}=\pi_{2}^{*}$ otherwise.

Corollary 4. Consider market $M$ sampled uniformly at random from $\{(\alpha, \beta),(\alpha, \beta+\epsilon)\}$, let $M^{\prime}$ be the other market. Suppose $n$ is such that $n^{3 / 2} \leq \frac{\alpha m}{\epsilon}$. Let $\left[\left(p_{1}, I_{1}\right), \ldots,\left(p_{n}, I_{n}\right)\right]$ be a sequence of $n$ observations generated from the market $M$, where $p_{i} \in \mathcal{F}_{i-1}$. Fix any pricing algorithm that outputs $\pi$ based on the first $n$ observations. Then for any $x \in[0,1)$, any $T^{\prime}$ such that $\pi_{1}^{*} \neq \pi_{2}^{*}$ :

$$
\mathbb{P}\left(\left|\pi-\pi_{M^{\prime}}^{*}(x)\right| \leq\left|\pi-\pi_{M}^{*}(x)\right|\right) \geq \frac{1}{8}
$$

where the probability is taken both with respect to the randomness from the stochastic arrival times, as well as the uniform sampling of the market parameters.

Proof. This directly follows from Lemma 30. The extra factor of $1 / 2$ in the probability comes from the fact that we randomly picked one market out of two.

## A.5.3 Step 3: Lipschitz bound on the optimal pricing policy

Lemma 31. For any remaining time $T \geq \frac{(1+\sqrt{2}) e}{\alpha+\beta}$ and $x \leq \frac{\alpha^{2} e}{4(\alpha+\beta)^{3} T}$

$$
\frac{\partial \pi^{*}(x, \alpha, \beta, T)}{\partial \beta} \leq \frac{-\alpha e}{4(\alpha+\beta)^{3} T}
$$

Proof. Differentiating both sides of (A.4) with respect to $\beta$ gives us

$$
\frac{1}{1-X_{T}^{*}(x)} \frac{\partial X_{T}^{*}(x)}{\partial \beta}=\frac{T X_{T}^{*}(x)}{2 \beta T X_{T}^{*}(x)+(\alpha-\beta) T+e}
$$

We omit the initial state argument $x$ and denote $X_{T}^{*}=X_{T}^{*}(x)$ in the following proof.

$$
\begin{aligned}
\frac{\partial \pi^{*}(x, \alpha, \beta, T)}{\partial \beta} & =\frac{1}{1-X_{T}^{*}} \frac{\partial X_{T}^{*}}{\partial \beta}-\frac{X_{T}^{*}}{\alpha+\beta X_{T}^{*}}-\frac{\beta \frac{\partial X_{T}^{*}}{\partial \beta}}{\alpha+\beta X_{T}^{*}}+\frac{x}{\alpha+\beta x} \\
& =\frac{T X_{T}^{*}}{2 \beta T X_{T}^{*}+(\alpha-\beta) T+e}-\frac{X_{T}^{*}}{\alpha+\beta X_{T}^{*}}-\frac{\beta}{\alpha+\beta X_{T}^{*}} \frac{\left(1-X_{T}^{*}\right) T X_{T}^{*}}{2 \beta T X_{T}^{*}+(\alpha-\beta) T+e}+\frac{x}{\alpha+\beta x} \\
& =\frac{T X_{T}^{*}\left(\alpha+\beta X_{T}^{*}\right)-X_{T}^{*}\left(2 \beta T X_{T}^{*}+(\alpha-\beta) T+e\right)-\left(1-X_{T}^{*}\right) \beta T X_{T}^{*}}{\left(\alpha+\beta X_{T}^{*}\right)\left(2 \beta T X_{T}^{*}+(\alpha-\beta) T+e\right)}+\frac{x}{\alpha+\beta x} \\
& =\frac{-e X_{T}^{*}}{\left(\alpha+\beta X_{T}^{*}\right)\left(2 \beta T X_{T}^{*}+(\alpha-\beta) T+e\right)}+\frac{x}{\alpha+\beta x} \\
& =\frac{-e X_{T}^{*}}{\left(\alpha+\beta X_{T}^{*}\right) \sqrt{(\alpha+\beta)^{2} T^{2}+2 e(\alpha-\beta) T+e^{2}+4 e \beta x T}}+\frac{x}{\alpha+\beta x} \\
& \leq \frac{-e X_{T}^{*}}{\left(\alpha+\beta X_{T}^{*}\right)((\alpha+\beta) T+e)}+\frac{x}{\alpha+\beta x}
\end{aligned}
$$

We replaced $X_{T}^{*}$ with (A.5) in the last equality. The last inequality follows from the fact that $\sqrt{(\alpha+\beta)^{2} T^{2}+2 e(\alpha-\beta) T+e^{2}+4 e \beta x T} \leq(\alpha+\beta) T+e$. In particular, if $T \geq \frac{(1+\sqrt{2}) e}{\alpha+\beta}$, then $\sqrt{(\alpha+\beta)^{2} T^{2}+2 e(\alpha-\beta) T+e^{2}+4 e \beta x T} \leq \sqrt{2}(\alpha+\beta) T$. Also

$$
X_{T}^{*}(x, \alpha, \beta) \geq X_{T}^{*}(0, \alpha, \beta) \geq X_{T}^{*}(0, \alpha, 0)=\frac{\alpha T}{\alpha T+e} \geq \frac{1+\sqrt{2}}{2+\sqrt{2}} \frac{\alpha}{\alpha+\beta}
$$

The first inequality is easy to see from (A.5), the second follows from Corollary 2, and the last
inequality follows from the assumption on $T$. So the above can be simplified to

$$
\frac{\partial \pi^{*}(x, \alpha, \beta, T)}{\partial \beta} \leq \frac{-e \frac{1+\sqrt{2}}{2+\sqrt{2}} \frac{\alpha}{\alpha+\beta}}{(\alpha+\beta)^{2} T \sqrt{2}}+\frac{x}{\alpha+\beta x}=\frac{-\alpha e}{2(\alpha+\beta)^{3} T}+\frac{x}{\alpha+\beta x}
$$

Then, it easy to verify that for $x \leq \frac{\alpha^{2} e}{4(\alpha+\beta)^{3} T}, \frac{\partial \pi^{*}(x, \alpha, \beta, T)}{\partial \beta} \leq \frac{-\alpha e}{4(\alpha+\beta)^{3} T}$

## A.5.4 Step 4: Putting it all together for proof of Theorem 2

We are now ready to prove our main lower bound result Theorem 2. In the following proof, as defined earlier in Step 1, $p_{x}$ denotes the induced price trajectory from the algorithm's offered prices, and $t_{x}^{\text {det }}$ is the time when adoption level hits $x$ in the deterministic Bass model on following $p_{x}$ (both were defined in detail in the paragraphs before Lemma 28).
Proof of Theorem 2. Let $T=\frac{2(1+\sqrt{2})}{\alpha+\beta}, \epsilon=\frac{(\alpha+\beta)^{2}}{\alpha}, n=\left\lfloor\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right\rfloor$. Randomly draw the Bass model parameters from $\{(\alpha, \beta),(\alpha, \beta+\epsilon)\}$ with equal probabilities. We denote the chosen market as $M$, and the other market $M^{\prime}$. In the calculations that follow, we use $\mathbb{E}_{M}$ to indicate that the expectation is taken with respect to both the random choice of Bass model parameters as well as the randomness in the stochastic Bass model.

To reduce clutter, let $\pi_{1}^{*}(x)=\pi^{*}\left(x, \alpha, \beta, T-t_{x}^{d e t}\right), \pi_{2}^{*}(x)=\pi^{*}\left(x, \alpha, \beta+\varepsilon, T-t_{x}^{d e t}\right)$, and let $\pi_{M}^{*}(x)$ be the optimal price i.e., $\pi_{M}^{*}(x)=\pi_{1}^{*}(x)$ if $M=(\alpha, \beta)$ and $\pi_{M}^{*}(x)=\pi_{2}^{*}(x)$ otherwise.

## $\mathbb{E}_{M}[$ Pseudo-Regret $(T)]$

(Lemma 29) $\geq \mathbb{E}_{M}\left[m \int_{0}^{n / m} \min \left(\frac{1}{4}\left(\pi_{M}^{*}(x)-p_{x}\right)^{2}, \frac{1}{10}\right) d x\right]-O\left(\frac{(\alpha+\beta)^{1 / 3} T e^{C}}{\alpha^{1 / 3}} m^{1 / 3} \sqrt{\log (m)}\right)$

$$
\begin{aligned}
& \geq m \int_{0}^{n / m} \mathbb{E}_{M}\left[\min \left(\frac{1}{4}\left(\frac{\pi_{1}^{*}(x)-\pi_{2}^{*}(x)}{2}\right)^{2}, \frac{1}{10}\right) \mathbb{1}\left(\left|p_{x}-\pi_{M^{\prime}}^{*}(x)\right| \leq\left|p_{x}-\pi_{M}^{*}(x)\right|\right)\right] d x \\
& \quad-O\left(\frac{m^{1 / 3} e^{C}}{(\alpha+\beta)^{2 / 3} \alpha^{1 / 3}} \sqrt{\log (m)}\right)
\end{aligned}
$$

(Lemma 31) $\geq m \int_{0}^{n / m} \mathbb{E}_{M}\left[\min \left(\left(\frac{\alpha e \varepsilon}{16(\alpha+\beta)^{3}\left(T-t_{x}^{d e t}\right)}\right)^{2}, \frac{1}{10}\right) \mathbb{1}\left(\left|p_{x}-\pi_{M^{\prime}}^{*}(x)\right| \leq\left|p_{x}-\pi_{M}^{*}(x)\right|\right)\right] d x$
$-O\left(\frac{m^{1 / 3} e^{C}}{(\alpha+\beta)^{2 / 3} \alpha^{1 / 3}} \sqrt{\log (m)}\right)$
$\left(T \geq T-t_{x}^{d e t}\right) \geq m \int_{0}^{n / m} \min \left(\left(\frac{e}{32(1+\sqrt{2})}\right)^{2}, \frac{1}{10}\right) \mathbb{E}_{M}\left[\mathbb{1}\left(\left|p_{x}-\pi_{M^{\prime}}^{*}(x)\right| \leq\left|p_{x}-\pi_{M}^{*}(x)\right|\right)\right] d x$
$-O\left(\frac{m^{1 / 3} e^{C}}{(\alpha+\beta)^{2 / 3} \alpha^{1 / 3}} \sqrt{\log (m)}\right)$
(Corollary 4$) \geq \frac{n}{8} \min \left(\left(\frac{e}{32(1+\sqrt{2})}\right)^{2}, \frac{1}{10}\right)-O\left(\frac{m^{1 / 3} e^{C}}{(\alpha+\beta)^{2 / 3} \alpha^{1 / 3}} \sqrt{\log (m)}\right)$

$$
\begin{aligned}
& =\Omega\left(\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right)-O\left(\frac{m^{1 / 3} e^{C}}{(\alpha+\beta)^{2 / 3} \alpha^{1 / 3}} \sqrt{\log (m)}\right) \\
& =\Omega\left(\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right)
\end{aligned}
$$

Using the fact that the maximum expected regret between the two markets must be at least the average, we have that for at least one of $(\alpha, \beta)$ and $(\alpha, \beta+\epsilon)$,

$$
\mathbb{E}[\text { Pseudo-Regret }(T)] \geq \Omega\left(\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{2 / 3}\right)
$$

Finally, we still need to check the assumptions of Lemma 29, Lemma 31 and Corollary 4 are satisfied for large enough $m$.

To apply Lemma 29, we need $X_{T}^{*} \geq\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{-1 / 3}=\frac{n}{m}$. Using the expanded notation
$X^{*}(x, \alpha, \beta)$ introduced in Appendix A.2, we know from Corollary 2 that $X_{T}^{*}(0, \alpha, \beta) \geq X_{T}^{*}(0, \alpha, 0)=$ $\frac{\alpha T}{\alpha T+e}$. It is easy to verify that for large enough $m$, specifically when $m^{1 / 3} \geq\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3}\left(2+\frac{e(\alpha+\beta)}{\alpha(1+\sqrt{2})}\right)$ and $T=2(1+\sqrt{2}) e /(\alpha+\beta), \frac{n}{m} \leq \frac{\alpha T}{\alpha T+e} \leq X_{T}^{*}(0, \alpha, \beta)=X_{T}^{*}$.

To apply Lemma 31 we need the conditions $T_{x}^{\prime}:=T-t_{x}^{d e t} \geq \frac{(1+\sqrt{2}) e}{\alpha+\beta}$ and $x \leq \frac{\alpha^{2} e}{4(\alpha+\beta)^{3} T_{x}^{\prime}}$ for all $x \leq \frac{n}{m}$. For any fixed $\alpha$, $\beta$, for large enough $m$, specifically for for $m^{1 / 3} \geq \frac{\alpha^{1 / 3}}{(\alpha+\beta)^{4 / 3}}\left(\alpha+e^{C}\right)$, where $C$ is a function of $\alpha, \beta$ as defined in Assumption 2, we have that $t_{n / m}^{d e t} \leq \frac{n}{e^{-p \max \alpha(m-n)}}=\frac{e^{p \max }}{\alpha(m-n)} \leq$ $\frac{(1+\sqrt{2}) e}{\alpha+\beta}$. The last inequality followed from simple algebraic calculations using the assumption on $m$ and Assumption 2. Therefore, $T-t_{x}^{d e t} \geq \frac{(1+\sqrt{2}) e}{\alpha+\beta}$ for all $x \leq \frac{n}{m}$. This satisfies the first condition. Furthermore, for $m^{1 / 3} \geq 8(1+\sqrt{2})\left(\frac{\alpha+\beta}{\alpha}\right)^{2 / 3}, x \leq n / m=\left(\frac{\alpha}{\alpha+\beta}\right)^{4 / 3} m^{-1 / 3} \leq \frac{\alpha^{2}}{8(1+\sqrt{2})(\alpha+\beta)^{2}}=$ $\frac{\alpha^{2} e}{4(\alpha+\beta)^{3} T} \leq \frac{\alpha^{2} e}{4(\alpha+\beta)^{3} T_{x}^{\prime}}$. This satisfies the second condition.

Finally, to apply Corollary 4 we needed the condition $n^{3 / 2} \leq \frac{\alpha m}{\epsilon}$. By plugging in our choice of $m, \epsilon$, it is easy to verify that this condition is satisfied.

## A. 6 Auxiliary Results

Lemma 32. Let $\bar{p}$ be a constant upper bound on the prices that the seller can offer: $p_{t} \leq \bar{p}$. Then there exists constants $C_{0}, C_{1}$ independent of $m$ such that for $T \geq C_{0} \log (m)+C_{1}$, there is a trivial policy that achieves $O\left(\log \left(\frac{1}{\delta}\right)\right)$ regret.

Proof. Consider the trivial pricing policy where the seller sets the price to the highest possible value $\bar{p}$ for the entire planning horizon. The rate of arrival of customer $d$ is given by $\lambda\left(\bar{p}, \frac{d-1}{m}\right)=$ $e^{-\bar{p}}\left(\alpha+\beta \frac{d-1}{m}\right)(m-d+1) \geq e^{-\bar{p}} \alpha(m-d+1)$

We now divide the market of $m$ customers into a sequence of segments with geometrically decreasing lengths. Let $m_{0}=0, m_{1}=\lfloor m / 2\rfloor, m_{2}=m_{1}+\lfloor m / 4\rfloor, \ldots, m_{i}=m_{i-1}+\left\lfloor m / 2^{i}\right\rfloor$. Let $K=\left\lfloor\log _{2}(m)-\log _{2} \log \left(\frac{1}{\delta}\right)-2\right\rfloor$ such that $m_{K}=m-O\left(\log \left(\frac{1}{\delta}\right)\right)$. We call customers $d$ for $m_{i-1}<d \leq m_{i}$ the customers in segment $i$.

Since the price is constant, we use the following short hand notation $\lambda_{d}:=\lambda\left(\bar{p}, \frac{d-1}{m}\right)$. Note that by construction, $m-m_{i} \geq m / 2^{i}$. So for $m_{i-1}<d \leq m_{i}, \lambda_{d} \geq \underline{\lambda}_{i}:=e^{-\bar{p}} \alpha \frac{m}{2^{i}}$.

Let $I_{d}$ be the stochastic inter-arrival time between customer $d-1$ and $d$, then using Lemma 20, we can obtain the following bound on the time it takes for all customers in segment $i$ to arrive:

$$
\mathbb{P}\left(\sum_{d=m_{i-1}+1}^{m_{i}} I_{d}-\frac{1}{\lambda_{d}} \geq \epsilon_{i}\right) \leq \exp \left(\frac{-\epsilon_{i}^{2} \underline{\lambda}_{i}^{2}}{8\left(m_{i}-m_{i-1}\right)}\right)
$$

Setting this to $\delta$ and solving for $\epsilon_{i}$, one can verify that with probability $1-\delta$,

$$
\sum_{d=m_{i-1}+1}^{m_{i}} I_{d}-\frac{1}{\lambda_{d}} \geq \epsilon_{i}:=\frac{4 e^{\bar{p}}}{\alpha} \sqrt{\log \left(\frac{1}{\delta}\right) \frac{2^{i}}{m}}
$$

Note that we needed $\epsilon_{i} \leq \frac{2\left(m_{i}-m_{i-1}\right)}{\underline{\lambda}_{i}}$ to apply Lemma 20. This condition is satisfied for $i=1, \ldots, K$ because:
$\epsilon_{i} \leq \frac{2\left(m_{i}-m_{i-1}\right)}{\underline{\lambda}_{i}} \Longleftrightarrow \sqrt{\log \left(\frac{1}{\delta}\right) \frac{2^{i}}{m}} \leq \frac{1}{2} \Longleftrightarrow i \leq 2 \log _{2}\left(\frac{1}{2} \sqrt{\frac{m}{\log (1 / \delta)}}\right)=\log (m)-\log _{2} \log \left(\frac{1}{\delta}\right)-2$
Applying a union bound to this result for all segments, we have that with probability at least $1-\delta \log _{2}(m)$, we have that

$$
\begin{aligned}
\sum_{d=1}^{m_{K}} I_{d} & \leq \sum_{d=1}^{m_{K}} \frac{1}{\lambda_{d}}+\sum_{i=1}^{K} \epsilon_{i} \\
& \leq \sum_{i=1}^{K} \frac{m_{i}-m_{i-1}}{\underline{\lambda}_{i}}+\sum_{i=1}^{K} \frac{4 e^{\bar{p}}}{\alpha} \sqrt{\log \left(\frac{1}{\delta}\right) \frac{2^{i}}{m}} \\
& =2 K \frac{e^{\bar{p}}}{\alpha}+\frac{4 e^{\bar{p}}}{\alpha} \sqrt{\log \left(\frac{1}{\delta}\right) \frac{1}{m} \frac{\sqrt{2}\left(2^{K / 2}-1\right)}{\sqrt{2}-1}} \\
& \leq \frac{e^{\bar{p}}}{\alpha}\left(2 \log _{2}(m)+\frac{4 \sqrt{2}}{\sqrt{2}-1} \sqrt{\log \left(\frac{1}{\delta}\right)}\right) \\
& =C_{0} \log (m)+C_{1}
\end{aligned}
$$

This means that if $T \geq C_{0} \log (m)+C_{1}$, then with probability $1-\delta \log _{2}(m)$, the realized revenue is at least $m_{K} \bar{p}$. Since the maximum possible revenue is $m \bar{p}$, the regret is at most $\bar{p}\left(m-m_{K}\right)=$ $O\left(\log \left(\frac{1}{\delta}\right)\right)$.

## Appendix B: Appendices for Chapter2

## B. 1 Missing Proofs of Results

## B.1.1 Proof of Equivalence in Definition 2

Proof. Suppose the second definition holds but the first one does not. Then by the definition of eliminated resource, there exists an optimal solution such that for every $g_{l} \in G_{i}$, there exists $r \in g_{l}$ such that $\sum_{i \in N} x_{i r}<S_{r}$. Then for every $g_{l} \in G_{i}$ we can assign $i$ a little more of the resource type above, and have $y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}}<\sum_{r \in g_{l}} x_{i r}$. This contradicts the second definition.

Now suppose the first definition holds but the second definition does not. This means that there exists an optimal solution such that $y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}}<\sum_{r \in g_{l}} x_{i r}$ for every $g_{l} \in G_{i}$. Consider the $g_{l}$ such that $g_{l} \cap R_{t+1}=\emptyset$ by the first definition (every $r \in g_{l}$ is eliminated by the end of round $t$ ). We can reduce the allocation of resources in that demand group to $i$ by a little bit without sacrificing optimality because the allocation constraints were satisfied strictly. But this means we have an optimal solution that does not use up the supply of $r \in g_{l}$ : this contradicts the elimination of these resources in the first definition.

## B.1.2 Proof of Claim 2

Proof. The first part is straightforward. If $q_{i g}>0$ is the dual variable for the allocation constraint for some agent $i \in N_{t}, g \in G_{i}$, then by complementary slackness every optimal solution needs to satisfy $y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}}=\sum_{r \in g_{l}} x_{i r}$, which means agent $i$ needs to be eliminated by Definition 2 .

Let's now rewrite the linear program solved at time $t$ :

$$
\begin{array}{clr}
\max y_{t} & & \\
\text { s.t.- } \sum_{r \in g_{l}} x_{i r}+ & \frac{d_{i l}}{d_{i *}} w_{i} y_{t} & \leq 0 \\
-\sum_{r \in g_{l}} x_{i r} & & \forall-\frac{d_{i l}}{d_{i *}} \gamma_{i} w_{i} \\
\sum_{i \in N} x_{i r} & & \forall i \notin N_{t}, g_{l} \in C  \tag{B.4}\\
x_{i r} & \geq S_{r} & g_{l} \in C \\
& \geq 0 & \forall r \in R \\
& & \forall i \in N, r \in R
\end{array}
$$

This LP is in canonical form, where the objective coefficient vector is $c^{T}=[0, \ldots, 0,1]$. Let $q_{i g}$ be the dual variables that correspond to the allocation constraints (constraint (B.2) and (B.3)), and $q_{r}$ the dual variables corresponding to the supply constraints (constraint (B.4)). Let $y_{t}^{*}$ be the value of $y_{t}$ in an optimal solution to the linear program and let $\bar{q}$ be the optimal solution to the corresponding dual program. By complementary slackness we know that $\bar{q}^{\top} A_{y}=c_{y}=1$, where $A_{y}$ is the last column of the primal constraint matrix. Note that the entries in $A_{y}$ are either positive or zero. Therefore, $\bar{q}_{i g}$ must be positive for some $i \in N_{t}, g \in G_{i}$. This finishes the proof of the second part.

## B.1.3 Proof of Lemma 9 and Fact 1

Proof. Suppose $x$ is the output of Algorithm 2 and there exists allocation $x^{\prime}$ such that agent $i$ is strictly better off while other agents have just as much utility. Let $y^{\prime} \times w_{i *}$ be the fraction of the dominant resource meta-type that $i$ receives with allocation $x^{\prime}$.

Let $t$ be the round in which $i$ was eliminated in Algorithm 2. Since $i$ is strictly better off with allocation $x^{\prime}, y^{\prime}>y_{t}^{*}$. Now we construct a new allocation by scaling down agent $i$ 's allocation from $x_{i}^{\prime}$ to $x_{i}^{\prime} \frac{y_{t}^{*}}{y^{\prime}}$. Since we know other agents have at least as much utility as with allocation $x$, this new solution has an LP objective value at least as high as before, satisfies all the allocation/supply
constraints, and does not use up all the resources that $i$ cares about. This contradicts $i$ being eliminated in round $t$. This concludes the proof for Lemma 9

Note that by the same argument as above we know that the allocation constraint in Equation 2.2 for the eliminated agents has to be satisfied with equality (otherwise we can scale this allocation down to make the constraint tight, and that agent would not have been eliminated in an earlier round).

## B.1.4 Proof of Lemma 10

Proof. For any pair of agents $i, j \in N$, we will show that $i$ does not envy $j$. Let $x$ be the allocation returned by Algorithm 2. Starting from the LHS of the definition of weighted envy-freeness:

$$
\begin{aligned}
& u_{i}\left(x_{j r} \frac{w_{i l}}{w_{j l}} \forall r \in g_{l}^{i}, l \in[L]\right) \\
& =\min _{g_{l} \in G_{i}} \frac{\sum_{r \in g_{l}^{i}} x_{j r} \frac{w_{i l}}{w_{j l}}}{d_{i l}} \\
& =\min _{g_{l} \in G_{i}} \frac{\sum_{r \in g_{l}^{j} \cap g_{l}^{i}} x_{j r} \frac{w_{i l}}{w_{j l}}}{d_{i l}} \\
& \leq \min _{g_{l} \in G_{i}} \frac{\sum_{r \in g_{l}^{j}} x_{j r} \frac{w_{i l}}{w_{j l}}}{d_{i l}} \\
& =\min _{g_{l} \in G_{i}} \frac{\frac{w_{i l}}{w_{j l}} \sum_{r \in g_{l}^{j}} x_{j r}}{d_{i l}} .
\end{aligned}
$$

The first equality is the definition of Leontief utility in (2.1). The second equality holds because $x_{j r}=0$ for $r \in \Omega_{l}$ but $r \notin g_{l}^{j}$ (If the output allocation does contain inaccessible resources then we can simply remove them without affecting the utilities of agents). The inequality follows from non-negativity of $x_{j r}$.

Now let $t_{i}, t_{j}$ be the rounds in which agent $i$ and $j$ are eliminated respectively. Note that from
the LP in Equation 2.2, we know $\sum_{r \in g_{l}^{j}} x_{j r}=y_{t_{j}}^{*} w_{j *} d_{j l} / d_{j *}$. So

$$
\begin{aligned}
\min _{g_{l} \in G_{i}} \frac{\frac{w_{i l}}{w_{j l}} \sum_{r \in g_{l}^{j}} x_{j r}}{d_{i l}} & =\min _{g_{l} \in G_{i}} \frac{\frac{w_{i l}}{w_{j l}} y_{t_{j}}^{*} w_{j *} d_{j l} / d_{j *}}{d_{i l}} \\
& =\min _{g_{l} \in G_{i}} \frac{w_{j *}}{d_{j *}} \frac{d_{j l}}{w_{j l}} y_{t_{j}}^{*} \frac{w_{i l}}{d_{i l}} \\
& \leq \min _{g_{l} \in G_{i}} y_{t_{j}}^{*} \frac{w_{i l}}{d_{i l}}=y_{t_{j}}^{*} \frac{w_{i *}}{d_{i *}}
\end{aligned}
$$

where the inequality follows from the definition of dominant resource meta-type $\left(\frac{w_{j *}}{d_{*}^{j}}=\min _{l} \frac{w_{j l}}{d_{j l}}\right)$. If $y_{t_{j}}^{*} \leq y_{t_{i}}^{*}$ (which means $t_{j} \leq t_{i}$, by Fact 2 and Fact 1 ), we have

$$
y_{t_{j}}^{*} \frac{w_{i *}}{d_{i *}} \leq y_{t_{i}}^{*} \frac{w_{i *}}{d_{i *}}=u_{i}\left(x_{i}\right) .
$$

Now suppose $y_{t_{j}}^{*}>y_{t_{i}}^{*}$ (which means $t_{j}>t_{i}$ ), and that $i$ envies $j$. Note that this implies that for every group $g_{l}^{i} \in G_{i}$, there exists at least one $r \in g_{l}^{i}$ such that $x_{j r}>0$.

Consider an alternative allocation $x^{\prime}$ that scales the allocation to agent $j$ to $\frac{y_{t_{i}^{*}}^{*}}{y_{t_{j}}^{*}} x_{j}$ while keeping the allocations to other agents the same as in $x$, namely, $x_{j}^{\prime}=x_{j} \frac{y_{t_{i}^{*}}^{*}}{y_{t_{j}}}$ and $x_{k}^{\prime}=x_{k} \forall k \neq j$. This alternative allocation gives every agent as much utility as they had before in round $t_{i}$ while maintaining slack in at least one resource from each demand group of $G_{i}$. This contradicts the definition of $t_{i}$ because agent $i$ was eliminated in round $t_{i}$ (see Definition 2).

## B.1.5 Proof of Lemma 11

Our proof approach is adapted from [45] with important modifications. We first introduce some new notations and prove two helpful results. Let $i$ be the only agent who reports her demands untruthfully. Let $d$ be the true demand vector for all agents and $d^{\prime}$ be an alternative demand where only the elements belonging to agent $i$ might be different. Let $t^{*}$ be the first round in which agent $i$ is eliminated in Algorithm 2, either with truthful or untruthful reporting (minimum of the two). Let $N_{t}, N_{t}^{\prime}$, and $y_{t}^{*}, y_{t}^{* \prime}$ represent the remaining active agents at the beginning of round $t$, and the
optimal LP objective in round $t$, under $d$ and $d^{\prime}$ respectively,

Claim 3. If agent $i$ is not eliminated in round $t$, then if we remove the allocation constraint for agent $i$ and omit the variables related to agent $i$ from the supply constraints in Equation 2.2, the optimal value as well as agents eliminated in that round do not change.

Proof. First we show that $x_{i r}=0$ if $r$ is one of the eliminated resources in that round. Suppose $x_{i r}>0$. Since $i$ is not eliminated, there must exist another resource $r^{\prime}$ in the same demand group of $r$ for agent $i$ that is not eliminated. This means that we could replace some of the allocation of $r$ with a little more allocation of $r^{\prime}$. But this would then contradict $r$ being an eliminated resource. Note that by the same logic $x_{i r}=0$ holds in all future rounds too.

This allows us to remove $x_{i r}$ from the supply constraints. Now the allocation constraint can be written as

$$
y_{t} \times w_{i *} \times \frac{d_{i l}}{d_{i *}} \leq \sum_{r \in g_{l} \cap R_{t+1}} x_{i r} \quad \forall g_{l} \in G_{i}
$$

Since the remaining resources are not constrained by supply, this inequality can always hold without posing limits on other variables. So we can remove this constraint completely.

Claim 4. For all $t \leq t^{*}, N_{t}=N_{t}^{\prime}$. For all $t<t^{*}, y_{t}^{*}=y_{t}^{* \prime}$.

Proof. We use proof by induction. $t=0$ holds trivially.
We assume the claim holds for $t$. Suppose $t+1<t^{*}$. Then by Claim 3, we can remove the constraints related to agent $i$ from the optimization problem. But the only differences between these two optimization problems are those related to agent $i$, so they have the same solutions and we are eliminating the same agents.

## Now we prove Lemma 11.

Proof. Let $x$ and $x^{\prime}$ be the allocations returned by Algorithm 2 given demand $d$ (truthful reporting) and $d^{\prime}$ (agent $i$ misreports) respectively. We consider the following four cases separately.

- $y_{t^{*}}^{*} \leq y_{t^{*}}^{* \prime}$ and agent $i$ is eliminated in $t^{*}$ reporting $d$. By Claim 4, we know $N_{t^{*}}=N_{t^{*}}^{\prime}$. Although we do not know the exact round in which agents in $N_{t^{*}}^{\prime}$ are eventually eliminated under $d^{\prime}$, we know that their dominant resource shares are all at least $y_{t^{*}}^{* \prime} \geq y_{t^{*}}^{*}$, because the optimal objective value of the optimization problem can only increase over time by Fact 2. Suppose $u_{i}\left(x_{i}^{\prime}\right)>u_{i}\left(x_{i}\right)$. Now consider $x^{\prime}$ as a candidate solution for the optimization problem in round $t^{*}$ of truthful reporting. Every agent $j$ in $N_{t^{*}}$ receives at least $y_{t^{*}}^{* \prime} w_{j *} \geq$ $y_{t^{*}}^{*} w_{j *}$ fraction of their dominant resource meta-type, while agent $i$ receives strictly more. This contradicts either the optimality of $y_{t}^{*}$ or the fact that agent $i$ was eliminated in round $t^{*}$ reporting $d$ (see Definition 2).
- $y_{t^{*}}^{*} \geq y_{t^{*}}^{* \prime}$ and the agent is eliminated in $t^{*}$ reporting $d^{\prime}$. Suppose the dominant resource meta-type is the same under $d^{\prime}$. Since $y_{t^{*}}^{* \prime} \times w_{i *}$ is the fraction of the total supply of dominant resource that $i$ receives, agent $i$ must be receiving less of that under $d^{\prime}$.

Now suppose the reported dominant resource meta-type is different under $d^{\prime}$. Let $\bullet$ denote the new dominant resource meta-type. Let $w_{i \bullet}, d_{i \bullet}^{\prime}$, be the new dominant resource weight and demand. Let $d_{i *}^{\prime}$ be the new demand for the original dominant resource meta-type. The amount of original dominant resource meta-type $i$ receives is

$$
y_{t^{*}}^{* *} \frac{w_{i \bullet}}{d_{i \bullet}^{\prime}} d_{i *}^{\prime} \leq y_{t^{*}}^{* \prime} w_{i *} \leq y_{t^{*}}^{*} w_{i *}
$$

The first inequality follows from the definition dominant resource meta-type: $\frac{w_{i}^{i}}{d_{i}^{i}}:=\min _{l \in[L]} \frac{w_{i l}}{d_{l}^{i}}$. The final expression is the amount of original dominant resource that agent $i$ receives under truthful reporting.

- $y_{t^{*}}^{* \prime}>y_{t^{*}}^{*}$ and the agent is not eliminated reporting $d$ but is eliminated reporting $d^{\prime}$. We argue that this case cannot happen. By Claim 4, we can remove the allocation constraints related to $i$ in round $t^{*}$ under truthful reporting. But then we are left with two optimization problems with the same constraints, except that with untruthful reporting the optimization
problem has extra allocation constraint (for agent $i$ ), and an extra non-negative term in the supply constraints. Extra constraints and extra terms in the supply constraints can only make the optimization problem harder.
- $y_{t^{*}}^{* \prime}<y_{t^{*}}^{*}$ and the agent is eliminated reporting $d$ but not eliminated reporting $d^{\prime}$. This is the symmetric case as the previous one and so cannot happen either.

Finally, a closer inspection of the above shows we did not need the group structure of agent $i$ to stay the same, so the result holds for misreporting group structures as well.

## B.1.6 Proof of Lemma 12

Proof. Recall that we use $s_{i l}$ denote the proportion that is both accessible to and contributed by agent $i$. Each agent might have access to other people's contributions as well. We set $w_{i l}=s_{i l}$. Since an agent might not have access to all of the supplies that she brings, $\sum_{i \in N} w_{i l}$ might be strictly less than one. In that case we can pretend that there is a phantom agent with weight $1-\sum_{i \in N} w_{i l}$ for each meta-type $l$, and that his demand vector is zero. Note that we do not need to implement this phantom agent when running the algorithm, because DRF-MT is invariant to the scale of weights. We are only adding this weight to make our definition of sharing incentive consistent with the assumption $\sum_{i} w_{i l}=1$

Now note that $\sum_{r \in U_{i \in N^{\prime}} g_{l}^{i}} S_{r} \geq \sum_{i \in N^{\prime}} s_{i l}$ because each agent has access to her own accessible supply. By the definition of dominant resource $\frac{w_{i *}}{d_{i *}} d_{i l} \leq w_{i l}$. So for any $N^{\prime} \subset N, l \in[L]$

$$
\left(\sum_{r \in \mathrm{U}_{i \in N^{\prime}} g_{l}^{i}} S_{r}\right) /\left(\sum_{i \in N^{\prime}} w_{i *} \frac{d_{i l}}{d_{i *}}\right) \geq \frac{\sum_{i \in N^{\prime}} s_{i l}}{\sum_{i \in N^{\prime}} w_{i l}}=1 .
$$

After rearranging the terms, we have

$$
\begin{equation*}
\sum_{r \in \cup_{i \in N^{\prime}} g_{l}^{i}} S_{r} \geq\left(\sum_{i \in N^{\prime}} w_{i *} \frac{d_{i l}}{d_{i *}}\right) \forall N^{\prime} \subseteq N, l \in\{1 \ldots L\} . \tag{B.5}
\end{equation*}
$$

For every agent $i$ and every meta-type $l$, consider $w_{i *} \frac{d_{i l}}{d_{i *}}$ as the "total demand" for resource meta-type $l$ from agent $i$.

Then, we construct a bipartite graph as follows: for the left-hand nodes, we create a node for every $\epsilon$ unit of total demand from each agent for each resource meta-type. Thus, each node is associated with some specific agent $i$ and resource meta-type $l$. For the right-hand nodes, we create a node for every $\epsilon$ unit of supply of each resource type $(r \in R)$. Note that since there is a finite number of agents and resource types, there exists an $\epsilon$ small enough that it can perfectly divide up all the demands and supplies, assuming that all the weights are rational.

Next, we create an edge between each pair of left and right-hand side nodes if and only if the supply side node belongs to the demand group of that agent for that meta-type: $r \in g_{l}^{i}$.

Eq.(B.5) now tells us that for every subset of the demand side nodes, the number of neighbors of that subset is greater than or equal to the size of the subset. This is precisely the condition in Hall's Theorem, which states that if this condition holds, then there exists a matching in the bipartite graph such that the demand side nodes are covered.

Consider such a matching obtained via Hall's Theorem. We construct a solution $x$ by setting $x_{i r}$ equal to $\epsilon$ times the number of matched edges corresponding to $i r$. This yields an assignment that gives each agent $w_{i *} \frac{d_{i l}}{d_{i *}}$ of each meta-type. By the construction of the matching this is a legal allocation. Then, we can set $y_{t}=1$ to obtain a feasible solution to the optimization problem in (2.2).

This means that after the first round of DRF-MT, agent $i$ 's utility is at least

$$
u_{i}\left(x_{i}\right)=\min _{g_{l} \in G_{i}} w_{i *} \frac{d_{i l}}{d_{i *}} \frac{1}{d_{i l}}=\frac{w_{i *}}{d_{i *}}=\min _{l: d_{i l} \neq 0} \frac{s_{i l}}{d_{i l}}
$$

The final expression is exactly the utility agent $i$ gets from her own supplies.


Figure B.1: Running times comparison with meta-types: $\Omega_{1}=\{0\}, \Omega_{2}=\{1,2\}, \Omega_{3}=$ $\{3,4,5\}, \Omega_{4}=\{6,7,8,9\}, \Omega_{5}=\{10,11,12,13,14\}$.

## B. 2 Experimental Setup and Additional Experiments

Solving for MNW is an Exponential Cone program, and solving for Discrete MNW is a Mixed Integer Exponential Cone program. Both of which did not have a reliable commercial solver until recently. This changed with the introduction of MOSEK version 9, which added support for such cones [51]. We implemented the MNW and Discrete MNW using the MOSEK solver and our DRF-MT approach using GUROBI. We made little effort to optimize either approach beyond the off-the-shelf implementations, and all experiments are run on a 2019 16-inch Macbook Pro with a 6 core Intel i7 processor.

First we describe in more details the random instance generating procedure used in Section 2.5. Recall that we fixed a meta-type structure. From there, for each agent, the group structure is generated by first uniformly sample a size between 0 and $\left|\Omega_{l}\right|$, and randomly pick a subset of that size from $\Omega_{l}$ as the demand group. The demands and weights (before normalization) are sampled uniformly from $[1,10]$, and the number of agents range from $n=5$ to $n=1000$. The supply for each resource is uniformly sampled from $[n \times 500, n \times 1000]$.


Figure B.2: Running time comparison with respect to number of resources

Using the same procedure, we also compare the running times on a larger instance (more metatypes, and more types within a meta-type). As shown in Figure B.1, the relative performances remain the same.

Next, we scale up the number of meta-types instead of number of agents, even though we think it is more natural to have problems with large number of agents. Here we assume that each metatype has five types and fix the number of agents to 50 . We see in Figure B. 2 again that the running time for Discrete-MNW quickly blows up while DRF-MT remains the fastest method out of the three.

Since Discrete-MNW runs much slower than DRF-MT and MNW, we next focus on the running time comparison of just MNW and DRF-MT in our next set of experiments. The setups are the same as the ones used for Figure 2.2 (left) and Figure B.2, except this time we scale the problem instances to much larger ones. We see from Figure B. 3 and Figure B. 4 that DRF-MT is 3-4 times faster than MNW in terms of running time.

Finally, as mentioned in Section 2.5, normalized difference in social welfare is calculated from subtracting the social welfare of Discrete MNW from that of DRF-MT and then divide the differ-


Figure B.3: Running time comparison between MNW and DRF-MT. The meta types are $\Omega_{1}=$ $\{0\}, \Omega_{2}=\{1,2\}, \Omega_{3}=\{3,4,5\}, \Omega_{4}=\{6,7,8,9\}$


Figure B.4: Running time comparison between MNW and DRF-MT with respect to number of resources.
ence by the social welfare of Discrete MNW. Here, instead of looking at the aggregated normalized difference in social welfare as in Figure 2.2 (right), we group the results by the number of agents. The box plot in Figure B. 5 shows that in most of the trials, DRF-MT generates social welfare that


Figure B.5: Normalized difference in social welfare between Discrete MNW and DRF-MT, grouped by number of agents.
is comparable to that of Discrete MNW, and sometimes even higher, with no significant variation in performance with respect to the number of agents. Readers might wonder how does Discrete DMW compare to (rounded) MNW in terms of social welfare. In Figure B. 6 we see that the two have virtually the same social welfare on almost all instances, with Discrete MNW having a slight edge over MNW. This means that compared to rounded MNW, our rounded DRF-MT algorithm also achieves at least $90 \%$ of the social welfare on most instances.

## B. 3 Beyond Sharing Incentive: Proportionality

Proportionality is defined as follows:


Figure B.6: Normalized difference in social welfare between Discrete MNW and MNW over all trials. Normalized difference is calculated by subtracting the social welfare of Discrete MNW from that of DRF-MT and then dividing by the social welfare of Discrete MNW.

Proportionality An allocation $x$ satisfies proportionality if $u_{i}\left(x_{i}\right) \geq u_{i}\left(x_{i}^{\prime}\right)$ for all $i$, where $x_{i r}^{\prime}=$ $w_{i l} S_{r}$ for each $r \in g_{l}^{i}$ and $l \in[L] . u_{i}\left(x^{\prime}\right)$ can be explicitly written out as

$$
u_{i}\left(x^{\prime}\right):=\min _{g_{l} \in G_{i}}\left\{\frac{w_{i l}}{d_{i l}} \sum_{r \in g_{l}} S_{r}\right\} .
$$

In the existing resource allocation literature, sharing incentive and proportionality are often used interchangeably. Indeed, when the priority weights are set according to the agents' accessible contributions to the resource pool ( $w_{i l}=s_{i l}$ ), the two notions are equivalent in settings where there are no accessibility constraints. With accessibility constraints, however, they are not the same. Under our definition of proportionality, the amount of accessible resource meta-type $l$ that agent $i$ receives is $s_{i l} \times \sum_{r \in g_{l}} S_{r}$. Since $\sum_{r \in g_{l}^{i}} S_{r}<1$ if agent $i$ cannot access the entire supply of meta-type $l, u_{i}\left(x^{\prime}\right)$ can be smaller than $u_{i}^{o}$. Therefore when priority weights are set according to agents' contributions, proportionality is a weaker notion than sharing incentive. However, since proportionality can be defined for arbitrary weights, regardless of whether or not we are in a setting
where agents bring their own supplies, it is a more flexible concept. Unfortunately proportionality does not hold generally. We prove proportionality under the following assumption:

## Assumption 4.

$$
\begin{aligned}
& \min _{N^{\prime} \subseteq N, l \in\{1 \ldots L\}}\left(\sum_{r \in \cup_{i \in N^{\prime}} g_{l}^{i}} S_{r}\right) /\left(\sum_{i \in N^{\prime}} w_{i *} d_{i l} / d_{i *}\right) \\
& \geq \max _{i \in N}\left\{\min _{l: d_{i l} \neq 0} \frac{w_{i l} \sum_{r \in g_{l}^{i}} S_{r}}{w_{i *} d_{i l} / d_{i *}}\right\}
\end{aligned}
$$

Lemma 33. Assume that demands, weights and supplies are all rational numbers. Then under Assumption 4, DRF-MT satisfies proportionality.

The proof is very similar to that of Lemma 12, but we first give some intuition for Assumption 4.

Since $\sum_{r \in g_{l}^{i}} S_{r} \leq 1$ for every $i, l$, and $\min _{l} \frac{w_{i l}}{d_{i l}}=\frac{w_{i *}}{d_{i *}}$ for every $i$, the right hand side of Assumption 4 is upper bounded by 1. $\sum_{r \in \cup_{i \in N^{\prime}} g_{l}^{i}} S_{r}$ is the union of the acceptable supply of resource meta-type $l$ from every agent in $N^{\prime} . \sum_{i \in N^{\prime}} w_{i} d_{i l} / d_{i *}$ is the total weighted demand from agents in set $N^{\prime}$. Note that $d_{i l} \leq d_{i *}$.

So, what the condition says intuitively is that whenever there is a group of agents who have a large combined weighted demand on meta-type $l$, they need to also collectively have access to/accept a large fraction of the total supply of $l$.

To provide more intuition for this assumption, we look at two examples. First we check that Example 1 satisfies Assumption 4. The RHS of the assumption evaluates to 1 (with hospital 1 and resource meta-type 1). One can check that the minimum on the LHS is achieved by picking $N^{\prime}=N$ and $l=1$ which gives us $\frac{16}{15}>1$. Thus Assumption 4 is satisfied. The resulting allocation and utilities using DRF-MT is given in Table B.1. Clearly our allocation is better for everyone than the proportional allocation.

| DRF-MT <br> Allocations | Hospital 1 <br> $\left(w_{1}=1 / 4\right)$ | Hospital 2 <br> $\left(w_{2}=1 / 4\right)$ | Hospital 3 <br> $\left(w_{3}=1 / 2\right)$ |
| :---: | :---: | :---: | :---: |
| Doctor A |  | 100 | 400 |
| Doctor B | 400 |  | 100 |
| Nurse C | 100 | 400 |  |
| Nurse D |  |  | 500 |
| Utilities |  |  |  |
| DRF-MT | 100 | 100 | 500 |
| Proportional | 62.5 | 31.25 | 250 |

Table B.1: Allocations from DRF-MT in Example 1 and the comparison of the resulting utilities with utilities of proportional allocation.

However, by adjusting the weights of the hospitals we can also construct an example that does not satisfy Assumption 4. Take the same parameters of Example 1 with the following modification on weights: $w_{1}=0.49, w_{2}=0.49, w_{3}=0.02$. The RHS value of Assumption 4 does not change. However, because the weights of hospitals 1,2 now dominate the market, the minimum of LHS is achieved with $N^{\prime}=\{1,2\}$ and $l=2$, which gives us $\frac{1 / 2}{0.49 \times 1+0.49 \times 1 / 4}<1$. So the assumption is violated. Intuitively, the problem with this setup is that even though hospitals 1 and 2 account for vast majority of the weighted demand for the nurse meta-type, they are both severely constrained to the same half of the total supply of nurses.

Under this setup, the DRF-MT assignments/utilities do not change. With proportional allocation however, the utilities for the three agents are [122.5, 61.25, 10.0]. So Agent 1 received more utility under proportional allocation than the allocation given by DRF-MT, at the expense of significantly hurting the social welfare: the sum of the utilities is less than 200, compared to 700 generated by the DRF-MT allocation. Now we prove Lemma 33

Proof. Let $\hat{y}$ denote the RHS of Assumption 4. After rearranging we have for all $N^{\prime} \subseteq N$ and $l \in\{1 \ldots L\}:$

$$
\sum_{r \in \mathrm{U}_{i \in N^{\prime}} g_{l}^{i}} S_{r} \geq \hat{y}\left(\sum_{i \in N^{\prime}} w_{*}^{i} d_{i l} / d_{i *}\right)
$$

For every agent $i$ and every meta-type $l$, consider $\hat{y} w_{i *} d_{i l} / d_{i *}$ as the "total demand" for resource meta-type $l$ from agent $i$.

Then we construct a bipartite graph and apply Hall's theorem the same way as in the proof of Lemma 12. This yields an assignment that gives each agent at least $\hat{y} w_{i *} d_{i l} / d_{i *}$ of each meta-type. By the definition of $\hat{y}$, it follows that the utility of each agent after the first round is at least:

$$
\begin{aligned}
\frac{d_{i l}}{d_{i *}} \hat{y} w_{i *} \frac{1}{d_{i l}}=\frac{w_{i *}}{d_{i *}} \hat{y} & \geq \frac{w_{i *}}{d_{i *}} \min _{l: d_{i l} \neq 0} \frac{w_{i l} \sum_{r \in g_{l}^{i}} S_{r}}{w_{i l} d_{i l} / d_{i *}} \\
& =\min _{l: d_{i l} \neq 0} \frac{\sum_{r \in g_{l}^{i}} S_{r}}{d_{i l}}
\end{aligned}
$$

Note that the right-most quantity is the utility of the proportional allocation. This means that after the first round, every agent already achieves at least as much utility as the proportional allocation. Fact 2 finishes the proof.

## B. 4 Alternative Design of DRF-MT

Instead of the algorithm described in Section 2.4, an alternative is to make each remaining agent receive a $y_{t} \times w_{i *} \times d_{i l} / d_{i *}$ fraction of the total supply from each of its resource group $g_{l_{i}^{*}}$ (instead of the supply from the entire meta-type). This idea might seem more intuitive since agent $i$ can only derive utilities from resources in $g_{l_{i}^{*}} \subseteq \Omega_{l_{i}^{*}}$. To do so, we can multiply the left hand side of the allocation constraints in Equation 2.2 by $\sum_{r \in g_{l}^{i}} S_{r}$. This alternative setup, however, does not lead to a mechanism with envy-freeness and strategy-proofness.

As a simple example, assume that there are five agents $1,2,3,4,5$ of equal weights, one metatype, and two resource types $A, B$ that fall under this meta-type, with equal supply. Agent 1,2 accept only type $A$; agent $3,4,5$ accept only type $B$. With simple calculation, we have that the largest $y_{1}$ we can get is $1 / 3$ : everyone receives $1 / 3$ of their accepted supply. The only possible allocation to achieve that is by assigning $1 / 3$ of $A$ each to agents 1,2 , and $1 / 3$ of $B$ each to agents $3,4,5$. However, if agent 2 strategically stated that he could take both $A$ and $B$, the resulting
allocation would be assigning $1 / 3$ of A to agent $1,2 / 3$ of $A$ to agent 2 , and $1 / 3$ of $B$ each to agents $3,4,5$. In this new allocation, the largest $y_{1}$ is still $1 / 3$, but since the total accepted supply for agent 2 is larger, he receives more. Furthermore, agent 1 would now envy agent 2.

## Appendix C: Appendices for Chapter3

## C. 1 Proof of Theorem 3

Proof. Using Fenchel-Rockafellar's duality theorem, the dual of (3.7) can be written as

$$
\begin{array}{ll} 
& \max _{f, g, \gamma \geq 0} \int_{\mathcal{X}} f(x) d \alpha(x)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j}  \tag{C.1}\\
\text { s.t. } & \left(1+\sum_{k \neq j} \gamma_{j k}\right) c(x, j)-\sum_{k \neq j} \gamma_{k y} c(x, k) \frac{\beta_{k}}{\beta_{j}} \\
& -f(x)-g_{j} \geq 0 \quad \forall x \in \mathcal{X}, y_{j} \in \mathcal{Y}
\end{array}
$$

Fixing $g \in \mathbb{R}^{n}$ and $\gamma \in \mathbb{R}^{n(n-1)}$, we can check using first order conditions that the optimal $f(x)$ has the closed form expression:

$$
\min _{j \in[n]} \bar{g}_{\gamma, c}\left(x, y_{j}\right):=\left(1+\sum_{k \neq j} \gamma_{j k}\right) c(x, j)-\sum_{k \neq j} \gamma_{k j} c(x, k) \frac{\beta_{k}}{\beta_{j}}-g_{j}
$$

Using this, the infinite dimensional optimization problem in (C.1) can be transformed to a finite dimensional optimization problem:

$$
\begin{equation*}
\max _{g, \gamma \geq 0} \mathcal{E}(g, \gamma):=\int_{X} \min _{j \in[n]} \bar{g}_{\gamma, c}\left(x, y_{j}\right) d \alpha(x)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j} \tag{C.2}
\end{equation*}
$$

Alternatively, we can adapt the Laguerre cell notation in (3.2) to (C.2):

$$
\mathcal{E}(g, \gamma)=\sum_{i \in[n]} \int_{\mathbb{L}_{y_{i}}(g, \gamma)} \bar{g}_{\gamma, c}\left(x, y_{i}\right) d \alpha(x)+g^{\top} \beta-\sum_{j, k, j \neq k} \gamma_{j k} \lambda_{j}
$$

where $\mathbb{L}_{y_{i}}(g, \gamma)=\left\{x \in \mathcal{X}: y_{i}=\underset{y_{j}}{\arg \min } \bar{g}_{\gamma, c}\left(x, y_{j}\right)\right\}$.
sectionExperimental Setup For the artificial data, the value utility vectors are generated from $X=[1,0.7]-Z\left[\begin{array}{cc}0.2, & 0 \\ 0.8, & 0.4\end{array}\right]$ where $Z \sim \operatorname{Unif}(0,1) \times \operatorname{Unif}(0,1)$. For finding the optimal allocation policy on the artificial data, we used Algorithm 3 with $T=2 \cdot 10^{5}$. For simulator data, we used $T=2.5 \cdot 10^{6}$. To generate Figure 3.2 we sampled 6000 points from the distribution and plotted them, colored by the allocation. For Figure 3.4, for each $m$ we ran 16 trials, sampling a different set of $m$ data points as our training data per trial. All experiments are run on a 2019, 6-core Macbook Pro laptop. The simulator code is open sourced by [52] at https://github.com/ duncanmcelfresh/blood-matching-simulations, and also included in the supplementary material.

## Appendix D: Appendices for Chapter4

## D. 1 Concentration Results

Lemma 34 (DKW Inequality [90]). Given i.i.d. samples $X_{1}, \ldots, X_{T}$ from a distribution $F$ (cdf), let $\hat{F}_{T}(x)=\frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\left[X_{t} \leq x\right]$. Then,

$$
\mathbb{P}\left(\sup _{x}\left|\hat{F}_{T}(x)-F(x)\right| \geq \alpha\right) \leq 2 e^{-2 T \alpha^{2}}
$$

## D.1.1 Proof of Lemma 14

Lemma 14 (Martingale Version of DKW Inequality). Given a sequence of random variables $Y_{1}, \ldots, Y_{T}$, let $\mathcal{F}_{t}=\sigma\left(Y_{1}, \ldots, Y_{t}\right), t=1, \ldots, T$ be the filtration representing the information in the first $t$ variables. Let $F_{t}(y):=\operatorname{Pr}\left(Y_{t} \leq y \mid \mathcal{F}_{t-1}\right)$, and $\bar{F}_{T}(y):=\frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\left[Y_{t} \leq y\right]$. Then,

$$
\mathbb{P}\left(\sup _{y}\left|\bar{F}_{T}(y)-\frac{1}{T} \sum_{t=1}^{T} F_{t}(y)\right| \geq \alpha\right) \leq\left(\frac{128 e T}{\alpha}\right) e^{-T \alpha^{2} / 128}
$$

Proof. This follows from sequential uniform convergence, see Lemma 10,11 in [94], and the fact that indicator functions have fat-shattering dimension 1.

A more convenient way to use Lemma 14 is the following corollary:

Corollary 5. Given a sequence of random variables $Y_{1}, \ldots, Y_{T}$, let $\mathcal{F}_{t}=\sigma\left(Y_{1}, \ldots, Y_{t}\right), t=1, \ldots, T$ be the filtration representing the information in the first t variables. Suppose $F_{t}(y)=\operatorname{Pr}\left(Y_{t} \leq y \mid \mathscr{F}_{t-1}\right)$, and let $\bar{F}_{T}(y):=\frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\left[Y_{t} \leq y\right]$. If $\alpha \geq 16 \sqrt{\frac{\log \left(\frac{128 e t}{\delta}\right)}{T}}$, then with probability $1-\delta$

$$
\sup _{x}\left|\bar{F}_{T}(x)-\frac{1}{T} \sum_{t=1}^{T} F_{t}(x)\right| \leq \alpha
$$

Proof.

$$
\begin{aligned}
& \left(\frac{128 e T}{\alpha}\right) e^{-T \alpha^{2} / 128} \leq \delta \\
\Longleftrightarrow & \alpha^{2} \geq \frac{128 \log \left(\frac{128 e T}{\delta}\right)}{T}+\frac{128}{T} \log \left(\frac{1}{\alpha^{2}}\right) \\
\Longleftarrow & \alpha^{2} \geq \frac{256 \log \left(\frac{128 e T}{\delta}\right)}{T} \\
\Longleftrightarrow & \alpha \geq 16 \sqrt{\frac{\log \left(\frac{128 e T}{\delta}\right)}{T}}
\end{aligned}
$$

## D. 2 Proof of Theorem 6

## D.2.1 Proof of Lemma 15

We first state two helper claims. The first one states that for any fixed greedy allocation policy $\lambda$, if the two distributions of valuations are similar, then the final allocation sizes for each receiver will also be close.

Claim 5. Fix a greedy allocation policy $\lambda$. Let $\boldsymbol{G}=G^{1} \otimes \ldots \otimes G^{n}$, and $\boldsymbol{F}=F^{1} \otimes \ldots \otimes F^{n}$ be two distributions over $[0, \bar{x}]^{n}$ where the marginals in each coordinate are independent. Suppose $\sup _{x}\left|F^{i}(x)-G^{i}(x)\right| \leq \Delta \forall i$. Then

$$
\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \leq n \Delta .
$$

Next we show that if the allocation sizes are similar for two different greedy allocation policies, then the corresponding allocation decisions (domain partitions) are also similar.

Claim 6. Let $\lambda^{\prime}, \lambda$ be any two fixed greedy allocation policies, and $\boldsymbol{F}$ a distribution over $[0, \bar{x}]^{n}$. For all $j$, let $\Omega_{j}^{\prime}=\mathbb{L}_{j}\left(\lambda^{\prime}\right)$, and $\Omega_{j}=\mathbb{L}_{j}(\lambda)$. Suppose $\sum_{j}\left(p_{j}\left(\boldsymbol{F}, \lambda^{\prime}\right)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-\right.$
$\left.p_{j}\left(\boldsymbol{F}, \lambda^{\prime}\right)\right)^{+} \leq \Delta$. Then

$$
\mathbb{P}\left(\Omega_{j}^{\prime} \backslash \Omega_{j}\right) \leq \Delta \quad \forall j
$$

Using these two Claims, we can now prove Lemma 15. The proofs for these two helper Claims follow after the proof of Lemma 15.

Proof of Lemma 15. Claim 5 shows that

$$
\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \leq n \Delta
$$

Note that by definition of $\lambda^{*}$ (due to the constraint $\operatorname{Pr}\left(\Omega_{j}\right)=p_{j}^{*}$ in problem (4.1)), we have $p_{j}(\boldsymbol{G}, \lambda)=p_{j}^{*}=p_{j}\left(\boldsymbol{F}, \lambda^{*}\right)$ for all $j$. This means that

$$
\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}\left(\boldsymbol{F}, \lambda^{*}\right)\right)^{+}=\sum_{j}\left(p_{j}\left(\boldsymbol{F}, \lambda^{*}\right)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \leq n \Delta
$$

Now we can apply Claim 6 and conclude that

$$
\begin{equation*}
\mathbb{P}\left(\Omega_{j}^{*} \backslash \Omega_{j}\right) \leq n \Delta \quad \forall j, \quad \text { and } \quad \mathbb{P}\left(\Omega_{j} \backslash \Omega_{j}^{*}\right) \leq n \Delta \quad \forall j, \tag{D.1}
\end{equation*}
$$

where $\Omega_{j}=\mathbb{L}_{j}(\lambda)$ and $\Omega_{j}^{*}=\mathbb{L}_{j}\left(\lambda^{*}\right)$. Therefore,

$$
\begin{aligned}
& \mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}\left[\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right]\right] \\
= & \int_{\boldsymbol{X} \in \Omega_{i}} X_{i} d \boldsymbol{F}(\boldsymbol{X})-\int_{\boldsymbol{X} \in \Omega_{i}^{*}} X_{i} d \boldsymbol{F}(\boldsymbol{X}) \\
\leq & \int_{\boldsymbol{X} \in \Omega_{i} \backslash \Omega_{i}^{*}} X_{i} d \boldsymbol{F}(\boldsymbol{X}) \\
\leq & n \Delta \bar{x}
\end{aligned}
$$

Using the same steps as above we can also show that

$$
\mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\mathbb{E}_{\boldsymbol{X} \sim \boldsymbol{F}}\left[u_{i}[\boldsymbol{X}, \boldsymbol{X}, \lambda]\right] \leq n \Delta \bar{x}
$$

## Proof of Claim 5

We first show variant of Claim 5 where the two distributions only differ in one coordinate:

Claim 7. Fix a greedy allocation policy $\lambda$. Let $\boldsymbol{G}=G^{1} \otimes G^{2} \ldots \otimes G^{n}$, and $\boldsymbol{F}=F^{1} \otimes F^{2} \ldots \otimes F^{n}$ be two distributions over $[0, \bar{x}]^{n}$ where the marginals in each coordinate are independent. Assume that $\boldsymbol{G}$ and $\boldsymbol{F}$ differ only in one coordinate, w.l.o.g. say coordinate i. Then, if $\sup _{x}\left|F^{i}(x)-G^{i}(x)\right| \leq \Delta$,

$$
\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \leq \Delta .
$$

Proof. We start with the distribution $\boldsymbol{G}=F^{1} \cdots F^{i-1} \otimes G^{i} \otimes F^{i+1} \cdots F^{n}$ and replace $G^{i}$ with a distribution $F^{i}$ to construct $\boldsymbol{F}=F^{1} \otimes F^{2} \cdots \otimes F^{n}$. We will construct $F^{i}$ in such a way that it is at most $\Delta$ away from $G^{i}$ and the changes in the allocation proportions are maximized. Note that since $\sum_{i} p_{i}(\boldsymbol{F}, \lambda)=1$ and $\sum_{i} p_{i}(\boldsymbol{G}, \lambda)=1$, we know that

$$
L H S(\boldsymbol{F}):=\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+}:=R H S(\boldsymbol{F})
$$

is always true for any $\boldsymbol{F}, \boldsymbol{G}, \lambda$. This means that we can focus on either maximizing either the LHS or the RHS of the above equation. There are two types of $F^{i}$ that we can use. One is such that $p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda) \geq 0$ and the other is $p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda)<0$. We can therefore bound the above quantity under these two scenarios separately:

$$
\begin{align*}
& F^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda) \geq 0  \tag{D.2}\\
& \text { max } \Longleftrightarrow \max _{F^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda) \geq 0} \sum_{j \neq i}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+}  \tag{D.3}\\
& F^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda)<0
\end{align*} L H S(\boldsymbol{F}) \Longleftrightarrow \max _{F^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda)<0} \sum_{j \neq i}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}+
$$

Therefore for the rest of the proof we can focus on bounding the right hand side of (D.2) and (D.3).

Bounding the RHS of (D.2) Let $\tilde{F}(x):=\left(G^{i}(x)-\Delta\right)^{+} \forall x<\bar{x}, \tilde{F}(\bar{x}):=1$. We claim that the $F^{i}$ that maximizes $\sum_{j \neq i}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+}$, while being at most $\Delta$ away, is $\tilde{F}$. To see this, consider a different distribution $F^{\prime}$ on the support $[0, \bar{x}]$ such that $\sup _{x}\left|F^{\prime}(x)-G^{i}(x)\right| \leq \Delta$. We know that $F^{\prime}(x) \geq \tilde{F}(x)$.

Later in Claim 8, we show that for any two distributions $G$ and $F$, we can sample $X \sim F$ using $Y$ sampled from $G$ by performing the following transformation:

$$
F^{-1}\left(G^{u}(Y)\right)
$$

where $G^{u}$ is the random variable defined for distribution $G$ in (D.12) and $F^{-1}:=\inf \{x \in \mathbb{R}$ : $F(x) \geq p\}$ denotes the generalized inverse, sometimes also referred to as the quantile function. This is essentially the inverse CDF method applied to a general distribution (instead of a uniformly sampled variable). In particular, let $G^{i u}$ be the following random function:

$$
G^{i u}(y)= \begin{cases}G^{i}(y) & \text { if } G^{i}(y)=G^{i}\left(y_{-}\right) \\ \text {Uniform }\left[G^{i}\left(y_{-}\right), G^{i}(y)\right] & \text { if } G^{i}(y)>G^{i}\left(y_{-}\right)\end{cases}
$$

Now, denote by $\tilde{\boldsymbol{F}}, \boldsymbol{F}^{\prime}$, the joint distribution that we get from $\boldsymbol{G}$ on replacing $G^{i}$ with $\tilde{F}$ and $F^{\prime}$, respectively. Then, the winning probabilities for the agents in these two cases are:

$$
\begin{align*}
p_{i}(\tilde{\boldsymbol{F}}, \lambda) & =\int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(\tilde{x}+\lambda_{i}-\lambda_{j}\right) d \tilde{F}(\tilde{x}) \\
& =\int_{0}^{\bar{x}} \mathbb{E}_{G^{i u}(x)}\left[\prod_{j \neq i} F^{j}\left(\tilde{F}^{-1}\left(G^{i u}(x)\right)+\lambda_{i}-\lambda_{j}\right)\right] d G^{i}(x),  \tag{D.4}\\
p_{j}(\tilde{\boldsymbol{F}}, \lambda) & =\int_{x} \tilde{F}\left(x+\lambda_{j}-\lambda_{i}\right) \prod_{k \notin\{i, j\}} F^{k}\left(x+\lambda_{j}-\lambda_{k}\right) d F^{j}(x), \quad \forall j \neq i \tag{D.5}
\end{align*}
$$

$$
\begin{align*}
& p_{i}\left(\boldsymbol{F}^{\prime}, \lambda\right)=\int_{x} \mathbb{E}_{G^{i u}(x)}\left[\prod_{j \neq i} F^{j}\left(F^{\prime-1}\left(G^{i u}(x)\right)+\lambda_{i}-\lambda_{j}\right)\right] d G^{i}(x),  \tag{D.6}\\
& p_{j}\left(\boldsymbol{F}^{\prime}, \lambda\right)=\int_{x} F^{\prime}\left(x+\lambda_{j}-\lambda_{i}\right) \prod_{k \notin\{i, j\}} F^{k}\left(x+\lambda_{j}-\lambda_{k}\right) d F^{j}(x), \quad \forall j \neq i \tag{D.7}
\end{align*}
$$

Since $\tilde{F}(x) \leq F^{\prime}(x) \forall x \in[0, \bar{x}]$ by construction, $\tilde{F}^{-1}(p) \geq F^{\prime-1}(p) \forall p \in[0,1]$. It's easy to see that $($ D.4) $\geq$ (D.6), and (D.5) $\leq$ (D.7). Using this we have

$$
\sum_{j \neq i}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\tilde{\boldsymbol{F}}, \lambda)\right)^{+} \geq \sum_{j \neq i}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}\left(\boldsymbol{F}^{\prime}, \lambda\right)\right)^{+} .
$$

and substituting $F^{\prime}$ by $G^{i}$ and again using (D.4) $\geq$ (D.6), and (D.5) $\leq$ (D.7), we get

$$
\begin{aligned}
p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\tilde{\boldsymbol{F}}, \lambda) & \geq 0 \quad \forall j \neq i, \\
p_{i}(\boldsymbol{G}, \lambda)-p_{i}(\tilde{\boldsymbol{F}}, \lambda) & \leq 0
\end{aligned}
$$

This shows that $\tilde{F}$ is the maximizer of the RHS of (D.2) among all distributions that are at most $\Delta$ away from $G^{i}$. Using this we have

$$
\begin{aligned}
& \operatorname{Fix}^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda) \geq 0 \\
= & \sum_{j \neq i}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \\
= & \left.p_{i}(\tilde{\boldsymbol{F}}, \lambda)-p_{i}(\boldsymbol{G}, \lambda)-p_{j}(\tilde{\boldsymbol{F}}, \lambda)\right)^{+} \\
= & \int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d \tilde{F}(x)-\int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x) \\
= & \int_{x_{\Delta}}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x)+\prod_{j \neq i} F^{j}\left(\bar{x}+\lambda_{i}-\lambda_{j}\right) \Delta-\int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x) \\
\leq & \Delta
\end{aligned}
$$

Bounding the RHS of (D.3) Let $\hat{F}(x):=\min \left(G^{i}(x)+\Delta, 1\right)$. Then we can use the same steps as above for LHS to show that $\hat{F}(x)$ maximizes $\sum_{j \neq i}\left(p_{j}(\hat{\boldsymbol{F}}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}$, and that

$$
\begin{gathered}
p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\hat{\boldsymbol{F}}, \lambda) \leq 0 \forall j \neq i \\
p_{i}(\boldsymbol{G}, \lambda)-p_{i}(\hat{\boldsymbol{F}}, \lambda) \geq 0
\end{gathered}
$$

This shows that $\hat{F}$ is the maximizer of the RHS of (D.3).From there, we have

$$
\begin{aligned}
& \max ^{i}: p_{i}(\boldsymbol{F}, \lambda)-p_{i}(\boldsymbol{G}, \lambda)<0 \\
= & \sum_{j \neq i}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+} \\
= & \left.p_{i}(\boldsymbol{G}, \lambda)-p_{i}(\hat{\boldsymbol{F}}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+} \\
= & \int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x)-\int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d \hat{F}(x) \\
= & \int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x)-\Delta \prod_{j \neq i} F^{j}\left(\lambda_{i}-\lambda_{j}\right)-\int_{0}^{x^{\Delta}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d F(x) \\
= & \int_{0}^{\bar{x}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d G^{i}(x)-\int_{0}^{x^{\Delta}} \prod_{j \neq i} F^{j}\left(x+\lambda_{i}-\lambda_{j}\right) d F(x) \\
\leq & \Delta
\end{aligned}
$$

where $x^{\Delta}=F^{-1}(1-\Delta)$.

Using Claim 7 we can now easily prove the original Claim 5.

Proof of Claim 5. First we construct the following sequence of distributions where for any two
adjacent distributions they only differ on one coordinate.

$$
\begin{aligned}
& \boldsymbol{G}_{0}=\boldsymbol{G}=G^{1} \otimes \ldots \otimes G^{n}, \\
& \boldsymbol{G}_{1}=F^{1} \otimes G^{2} \otimes \ldots \otimes G^{n}, \\
& \boldsymbol{G}_{2}=F^{1} \otimes F^{2} \otimes G^{3} \otimes \ldots \otimes G^{n}, \\
& \ldots \\
& \boldsymbol{G}_{n}=\boldsymbol{F}
\end{aligned}
$$

Then we can decompose the difference between $p(\boldsymbol{F}, \lambda)$ and $p^{*}$ into a sum of differences:

$$
\begin{aligned}
\|p(\boldsymbol{F}, \lambda)-p(\boldsymbol{G}, \lambda)\|_{1} & =\left\|p\left(\boldsymbol{G}_{n}, \lambda\right)-p\left(\boldsymbol{G}_{0}, \lambda\right)\right\|_{1} \\
& \leq \sum_{i=1}^{n}\left\|p\left(\boldsymbol{G}_{i}, \lambda\right)-p\left(\boldsymbol{G}_{i-1}, \lambda\right)\right\|_{1} \\
& \leq 2 n \Delta
\end{aligned}
$$

where the last step follows from Claim 7. Since $\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-\right.$ $\left.p_{j}(\boldsymbol{F}, \lambda)\right)^{+}=\frac{1}{2}\|p(\boldsymbol{F}, \lambda)-p(\boldsymbol{G}, \lambda)\|_{1}$, we have

$$
\sum_{j}\left(p_{j}(\boldsymbol{F}, \lambda)-p_{j}(\boldsymbol{G}, \lambda)\right)^{+}=\sum_{j}\left(p_{j}(\boldsymbol{G}, \lambda)-p_{j}(\boldsymbol{F}, \lambda)\right)^{+} \leq n \Delta
$$

## Proof of Claim 6

Proof. WLOG, assume that $\lambda_{1}-\lambda_{1}^{\prime} \geq \lambda_{2}-\lambda_{2}^{\prime} \ldots \geq \lambda_{n}-\lambda_{n}^{\prime}$. Let $Q_{j k}=\Omega_{j} \cap \Omega_{k}^{\prime}$ be the "error flow" of items from $j$ to $k$. It is easy to see that for $k<j, \boldsymbol{X}_{j}+\lambda_{j}>\boldsymbol{X}_{k}+\lambda_{k} \Longrightarrow \boldsymbol{X}_{j}+\lambda_{j}^{\prime}>\boldsymbol{X}_{k}+\lambda_{k}^{\prime}$. Therefore $Q_{j k}=\emptyset$ for $k<j$. Then it follows that

$$
\Omega_{j}^{\prime} \backslash \Omega_{j} \subseteq \bigcup_{i: i<j} Q_{i j} \subseteq \bigcup_{i: i<j} \bigcup_{k: k \geq j} Q_{i k} .
$$

The right hand side above is the net outflow from the set $\{i: i<j\}$. However, we know that each individual agents' net in flow is $p_{j}\left(\boldsymbol{F}, \lambda^{\prime}\right)-p_{j}(\boldsymbol{F}, \lambda)$, so we can bound the RHS by

$$
\mathbb{P}\left(\bigcup_{i: i<j} \bigcup_{k: k \geq j} Q_{i k}\right) \leq \sum_{i: i<j} p_{j}\left(\boldsymbol{F}, \lambda^{\prime}\right)-p_{j}(\boldsymbol{F}, \lambda) \leq \Delta
$$

## D.2.2 Proof of Theorem 6

Proof. Fix an epoch $k$, let $\Delta=\sqrt{\frac{1}{2 n\left(L_{k}-1\right)} \log \left(\frac{2}{\delta}\right)}, \hat{\lambda}=\lambda^{*}\left(\hat{\boldsymbol{F}}_{L_{k}-1}\right)$.

$$
\begin{align*}
& \text { (Lemma 34) } \\
& \sup _{x}\left|\hat{F}_{L_{k}-1}(x)-F(x)\right| \leq \Delta \\
& \text { w.p. } 1-\delta / 2 \\
& \text { (Lemma 15) } \Longrightarrow \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \hat{\lambda})\right] \leq n \Delta \bar{x} \quad \forall i  \tag{D.8}\\
& \text { w.p. } 1-\delta / 2 \\
& \text { (Chernoff bound) } \left.\Longrightarrow \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]\left(L_{k+1}-L_{k}\right)-\sum_{t=L_{k}}^{L_{k+1}-1} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \hat{\lambda}\right)\right] \\
& \leq n \Delta \bar{x}\left(L_{k+1}-L_{k}\right)+\bar{x} \sqrt{\frac{\left(L_{k+1}-L_{k}\right)}{2} \log \left(\frac{2}{\delta}\right)} \\
& \text { w.p. } 1-\delta \\
& =\sqrt{2^{k} n \log \left(\frac{2}{\delta}\right)} \bar{x}+\sqrt{2^{k-1} \log \left(\frac{2}{\delta}\right)} \bar{x} \tag{D.9}
\end{align*}
$$

The above bounds the regret in one epoch if the algorithm does not terminate before the epoch ends. It remains to show that the algorithm with high probability does not terminate too early. This involves showing that with high probability, no agent hits their capacity constraint $p_{j}^{*} T$ significantly earlier than $T$, and that the detection algorithm does not falsely trigger.

Continuing from (D.9), for any time step $T^{\prime} \leq T$, we have

$$
\begin{array}{rlrl} 
& T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\sum_{t=1}^{T^{\prime}} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \lambda_{k_{t}}\right) & \\
\leq & \left.\sum_{k=0}^{\log _{2} T^{\prime}}\left[\left(L_{k+1}-L_{k}\right) \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\sum_{t=L_{k}}^{L_{k+1}-1} u_{i}\left(\boldsymbol{X}_{t}, \boldsymbol{X}_{t}, \lambda_{k}\right)\right]\right] & \\
& +\left(T-T^{\prime}\right) \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] & & \\
= & \sum_{k=1}^{\log _{2} T} \sqrt{n 2^{k} \log \left(\frac{2}{\delta}\right) \bar{x}+\sqrt{\frac{2^{k}}{2} \log \left(\frac{2}{\delta}\right) \bar{x}}+\left(T-T^{\prime}\right) \bar{x}} & \text { w.p } 1-\delta \log _{2} T \\
\leq & 2 \sqrt{n \log \left(\frac{2}{\delta}\right) \bar{x} \sum_{k=1}^{\log _{2} T} \sqrt{2^{k}}+\left(T-T^{\prime}\right) \bar{x}} & \text { w.p } 1-\delta \log _{2} T \\
\leq & \frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right) \bar{x}+\left(T-T^{\prime}\right) \bar{x}} & \text { w.p } 1-\delta \log _{2} T \tag{D.10}
\end{array}
$$

where the second inequality follows from(D.9) and union bound. Now, since there are at most $\log _{2}(T)$ epochs for any $T^{\prime} \leq T$, above holds for all epochs and therefore for all $T^{\prime}$ with probability $1-\delta \log _{2}(T)$. Now we show that with high probability, for all $T^{\prime} \leq T-\frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right)}$ and for any fixed agent $i$, the constraint of total allocation to agent $i$ to be less than $p_{i}^{*} T$ will be satisfied. Note that a byproduct of applying Lemma 15 in (D.8) is that $\left|p_{i}\left(\boldsymbol{F}, \lambda_{k}\right)-p_{i}^{*}\right| \leq n \Delta_{L_{k}-1}$ (See (D.1)). Fix a time step $\tau$,

$$
\begin{array}{rll} 
& \sum_{i=1}^{\tau} \mathbb{1}\left[\underset{j}{\arg \max } \boldsymbol{X}_{j}+\lambda_{k_{t} j}=i\right] \\
\text { (Chernoff) } & \leq \sum_{k=1}^{\log _{2} \tau}\left(L_{k+1}-L_{k}\right) p_{i}\left(\boldsymbol{F}, \lambda_{k}\right)+\sqrt{\frac{\left(L_{k+1}-L_{k}\right)}{2} \log \left(\frac{2}{\delta}\right)} & \text { w.p. } 1-\delta \log _{2} \tau \\
& \leq \sum_{k=1}^{\log _{2} \tau}\left(L_{k+1}-L_{k}\right)\left(p_{i}^{*}+n \Delta_{L_{k}-1}\right)+\sqrt{\frac{\left(L_{k+1}-L_{k}\right)}{2} \log \left(\frac{2}{\delta}\right)} & \text { w.p. } 1-\delta \log _{2} \tau \\
\left(L_{k}=2^{k}\right) & \leq p_{i}^{*} \tau+\sum_{k=1}^{\log _{2} \tau}\left(\sqrt{n 2^{k} \log \left(\frac{2}{\delta}\right)}+\sqrt{2^{k-1} \log \left(\frac{2}{\delta}\right)}\right) & \text { w.p. } 1-\delta \log _{2} \tau \\
& \leq p_{i}^{*} \tau+\frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n \tau \log \left(\frac{2}{\delta}\right)} & \text { w.p. } 1-\delta \log _{2} \tau
\end{array}
$$

This means that for all $\tau \leq T-\frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right)}$, with probability $1-\delta \log _{2} T$,

$$
\sum_{t=1}^{\tau} \mathbb{1}\left[\underset{j}{\arg \max } \boldsymbol{X}_{j}+\lambda_{k_{t} j}=i\right] \leq p_{i}^{*} T
$$

Combining above with (D.10), we have that with probability $1-2 \delta \log _{2} T$, for any fixed $i$, if the algorithm terminates at $T^{\prime}$ due to allocation limit reached for agent $i$, then $T^{\prime} \geq \frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right)}$, so that

$$
T \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]-\sum_{t=1}^{T^{\prime}} u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda_{k_{t}}\right) \leq \frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right)} \bar{x}+\frac{2 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T \log \left(\frac{2}{\delta}\right) \bar{x}}
$$

Finally, we also have to bound the probability that the detection algorithm falsely triggers. For a given time $t$ and for each $i$, let

$$
\begin{aligned}
F_{t}^{i}(x) & =\frac{1}{t} \sum_{t=1}^{t} \mathbb{1}\left[X_{i, t} \leq x\right] \\
\tilde{F}_{t}(x) & =\frac{1}{t(n-1)} \sum_{t=1}^{t} \sum_{j \neq i} \mathbb{1}\left[\boldsymbol{X}_{j, t} \leq x\right]
\end{aligned}
$$

be the empirical CDF for agent $i$ and the rest of the agents. Since all agents are truthful, using Lemma 34 we have that with probability $1-\delta$,

$$
\begin{aligned}
& \sup _{x}\left|F_{t}^{i}(x)-F(x)\right| \leq \sqrt{\frac{1}{2 t} \log \left(\frac{2}{\delta}\right)} \\
& \sup _{x}\left|\tilde{F}_{t}(x)-F(x)\right| \leq \sqrt{\frac{1}{2 t(n-1)} \log \left(\frac{2}{\delta}\right)}
\end{aligned}
$$

This means that $\sup _{x}\left|F_{t}^{i}(x)-\tilde{F}_{t}(x)\right| \leq \sqrt{\frac{1}{t} \log \left(\frac{2}{\delta}\right)} \leq 32 \sqrt{\frac{1}{t} \log \left(\frac{256 e t}{\delta}\right)}=\Delta_{t} / 2$, which means that Algorithm 5 is not triggered by agent $i$. Using union bound, we know that with probability $1-\delta n T$, the algorithm will not end early because of a false trigger (by any agent).

The result follows by replacing $\delta$ with $\frac{\delta}{n\left(2 \log _{2} T+T\right)}$ and take the union bound over all agents.

## D. 3 Proof of Theorem 5

## D.3.1 Proof of Lemma 16

Proof. Let $\alpha=\frac{\Delta}{4}$. We first check that the given condition on $\Delta$ satisfies $\left(\frac{128 e t}{\alpha}\right) e^{-t \alpha^{2} / 128} \leq \frac{\delta}{2}$ and that $2 e^{-2 t(n-1) \alpha^{2}} \leq \frac{\delta}{2}$

$$
\begin{aligned}
& \left(\frac{128 e t}{\alpha}\right) e^{-t \alpha^{2} / 128} \leq \frac{\delta}{2} \\
\Longleftrightarrow & \alpha^{2} \geq \frac{128 \log \left(\frac{256 e t}{\delta}\right)}{t}+\frac{64}{t} \log \left(\frac{1}{\alpha^{2}}\right) \\
\Longleftarrow & \alpha^{2} \geq \frac{256 \log \left(\frac{256 e t}{\delta}\right)}{t} \\
\Longleftrightarrow & \Delta \geq 64 \sqrt{\frac{\log \left(\frac{256 e t}{\delta}\right)}{t}}
\end{aligned}
$$

$$
\begin{aligned}
& 2 e^{-2 t(n-1) \alpha^{2}} \leq \frac{\delta}{2} \\
\Longleftrightarrow & \alpha \geq \sqrt{\frac{1}{2 t(n-1)} \log \left(\frac{4}{\delta}\right)} \\
\Longleftarrow & \Delta \geq 64 \sqrt{\frac{\log \left(\frac{256 e t}{\delta}\right)}{t}}
\end{aligned}
$$

Let $\bar{F}_{t}(x)=\frac{1}{t} \sum_{s=1}^{t} \mathbb{1}\left[\tilde{X}_{i, s} \leq x\right]$ be the empirical CDF of the samples collected from agent $i$. Let $\tilde{F}_{t}(x)=\frac{1}{(n-1) t} \sum_{s=1}^{t} \sum_{j \neq i} \mathbb{1}\left[\tilde{X}_{j, s} \leq x\right]$ be the empirical CDF of all reported values from the other agents. Let $\bar{F}(x)=\frac{1}{t} \sum_{s=1}^{t} F_{s}(x)$, where $F_{s}(x)=\mathbb{P}\left(\tilde{X}_{i, s} \leq x \mid \mathcal{H}_{s}\right)$. Lemma 14 tells us that with probability $1-\delta / 2$,

$$
\begin{equation*}
\sup _{x}\left|\bar{F}_{t}(x)-\bar{F}(x)\right| \leq \frac{\Delta}{4} \tag{D.11}
\end{equation*}
$$

Since other agents are truthful, their reported values are independent, and we can use the regular DKW inequality to bound the empirical distribution constructed from their values. Using Lemma 34 we can show that with probability $1-\delta / 2$,

$$
\sup _{x}\left|\tilde{F}_{t}(x)-F(x)\right| \leq \frac{\Delta}{4}
$$

Using union bound, we can conclude that if $\sup _{x}|\bar{F}(x)-F(x)| \geq \Delta$, then with probability $1-\delta$ :

$$
\sup _{x}\left|\tilde{F}_{t}(x)-\bar{F}_{t}(x)\right|>\frac{\Delta}{2}
$$

which means that Algorithm 5 would have returned Reject.

## D.3.2 Proof of Lemma 17

First we state a technical result on monotone mapping between two distributions. Given a cumulative distribution function $F$, we define the following random function:

$$
F^{u}(y)= \begin{cases}F(y) & \text { if } F(y)=F\left(y_{-}\right)  \tag{D.12}\\ \operatorname{Uniform}\left[F\left(y_{-}\right), F(y)\right] & \text { if } F(y)>F\left(y_{-}\right)\end{cases}
$$

If $F$ is a continuous distribution then $F^{u}$ is deterministic and is the same as $F$. However if $F$ contains point masses, then at points where $F$ jumps, $F^{u}$ is uniformly sampled from the interval of that jump. It is easy to see that $F^{u}$ has the nice property that if $Y \sim F$, then $F^{u}(Y) \sim$ Uniform $[0,1]$.

Claim 8. Let $G$ be any distribution ( $c d f$ ) over $\mathcal{X} \subseteq \mathbb{R}$, and $F$ over $\mathcal{Y} \subseteq \mathbb{R}$. Then there exists a unique joint distribution $r$ over $\mathcal{X} \times \mathcal{Y}$ with marginals $G, F$ such that the conditional distribution $r(\cdot \mid Y)$ has the following monotonicity property: define $\bar{x}_{r}(\cdot), \underline{x}_{r}(\cdot)$ so that $X \in\left[\underline{x}_{r}(Y), \bar{x}_{r}(Y)\right]$
almost surely, i.e.,

$$
\begin{gathered}
\bar{x}_{r}(y)=\inf \{x: \mathbb{P}(X>x \mid Y=y)=0\} \\
\underline{x}_{r}(y)=\sup \{x: \mathbb{P}(X<x \mid Y=y)=0\},
\end{gathered}
$$

then

$$
\bar{x}_{r}\left(y_{1}\right) \leq \underline{x}_{r}\left(y_{2}\right) \quad \forall y_{1}<y_{2} .
$$

In particular, the random variable $X \mid Y \sim r(\cdot \mid Y)$ can be sampled as $G^{-1}\left(F^{u}(Y)\right)$, where $F^{u}$ is the random function defined in (D.12) and $G^{-1}:=\inf \{x \in \mathbb{R}: G(x) \geq p\}$ denotes the generalized inverse, sometimes also referred to as the quantile function.

The proof of this Claim is in Appendix D.4.1. Using the above result, we derive the following key result that will provide insight into a strategic agent's best response to a greedy allocation strategy. Note that given a particular marginal distribution $G$ for the agent $i$ 's reported values and the true value distribution $F$, there are many potential joint distributions between the true and reported valuations. In the following lemma, we show that the "best" joint distribution among these, in terms of agent $i$ 's utility maximization, is the one characterized in Claim 8.

Claim 9. Fix a greedy allocation policy $\lambda$. Let $\boldsymbol{X} \in[0, \bar{x}]^{n}$ be drawn from $F \otimes \ldots \otimes F$. Fix another distribution $G$ over $[0, \bar{x}]$. Given $\boldsymbol{X}$, define $\tilde{X}^{*}$ as follows: let $\tilde{X}_{i}^{*}=G^{-1}\left(F^{u}\left(X_{i}\right)\right)$, and $\tilde{X}_{j}^{*}=X_{j} \forall j \neq i$. Let $\mathcal{R}$ be the set of all joint distributions over $[0, \bar{x}]^{2}$ such that the marginals are $F$ and $G$; and for any $r \in \mathcal{R}$, given $\boldsymbol{X}$ define $\tilde{\boldsymbol{X}}^{r}$ as follows: $\tilde{X}_{i}^{r} \sim r\left(\cdot \mid X_{i}\right)$, and $\tilde{X}_{j}^{r}=X_{j} \forall j \neq i$. Then

$$
\mathbb{E}\left[u_{i}\left(\tilde{\boldsymbol{X}}^{*}, \boldsymbol{X}, \lambda\right)\right] \geq \max _{r \in \mathcal{R}} \mathbb{E}\left[u_{i}\left(\tilde{\boldsymbol{X}}^{r}, \boldsymbol{X}, \lambda\right)\right] .
$$

Proof. First we show that for any joint distribution that is not monotone (i.e., does not have the monotonicity property defined in Claim 8), there is a monotone one that obtains at least as much utility. Suppose $r$ is one such joint distribution that is not monotone, i.e., $\exists x_{1}<x_{2}$, s.t. $\bar{x}_{r}\left(x_{1}\right)>$
$\underline{x}_{r}\left(x_{2}\right)$ (as defined in Claim 8). First recall that since $X_{j} \sim F, \forall j$ are independent, the expected utility can be written as the following:

$$
\mathbb{E}\left[u_{i}\left(\tilde{\boldsymbol{X}}^{r}, \boldsymbol{X}, \lambda\right)\right]=\int_{0}^{\bar{x}} \int_{\underline{x}_{r}(x)}^{\bar{x}_{r}(x)} x \prod_{j \neq i} F\left(\tilde{x}+\lambda_{i}-\lambda_{j}\right) d r(\tilde{x} \mid x) d F(x)
$$

Now consider a pair of values $\tilde{x}_{1}>\tilde{x}_{2}$ such that $\left(\tilde{x}_{1}, x_{1}\right)$ and ( $\tilde{x}_{2}, x_{2}$ ) has a non-zero probability density under distribution $r$. This pair exists because $\bar{x}_{r}\left(x_{1}\right)>\underline{x}_{r}\left(x_{2}\right)$. Then using the fact that for $a, b, c, d>0, a<b, c<d: a c+b d>a d+b c$, we can see that:

$$
x_{1} \prod F\left(\tilde{x}_{1}+\lambda_{i}-\lambda_{j}\right)+x_{2} \prod F\left(\tilde{x}_{2}+\lambda_{i}-\lambda_{j}\right)<x_{1} \prod F\left(\tilde{x}_{2}+\lambda_{i}-\lambda_{j}\right)+x_{2} \prod F\left(\tilde{x}_{1}+\lambda_{i}-\lambda_{j}\right)
$$

This means that if we exchanged the probability mass between the two conditionals of $x_{1}, x_{2}$, the utility would be at least as much as before, if not higher. This means that at least one monotone joint distribution belongs in the set of utility maximizing joint distributions. Since Claim 8 showed that the distribution of $\left(G^{-1}\left(F^{u}(X)\right), X\right)$ is the unique joint distribution that is monotone, we conclude that $\tilde{\boldsymbol{X}}^{*}$ as defined in the lemma statement is indeed utility maximizing.

## Proof of Lemma 17

Proof. Let $G(x):=(F(x)-\Delta)^{+} \forall x<\bar{x}, G(\bar{x}):=1$ be the distribution whose CDF is shifted down from $F$ by $\Delta$. Let $\tilde{r}$ be the utility maximizing joint distribution from Claim 9 . Let $\hat{r}, \hat{F}$ be a different pair of joint and marginal distribution such that $\sup _{x}|F(x)-\hat{F}(x)| \leq \Delta$. We know that $\hat{F}(x) \geq G(x)$ for all $x$. Agent $i$ 's utilities for using $\hat{r}$ and $\tilde{r}$ respectively, are:

$$
\begin{align*}
\mathbb{E}_{\hat{r}}\left[u_{i}(\hat{\boldsymbol{X}}, \boldsymbol{X}, \lambda)\right] & =\int_{0}^{\bar{x}} x \int_{\underline{x}_{\hat{r}}(x)}^{\bar{x}_{\hat{f}}(x)} \prod_{j \neq i} F\left(\hat{x}+\lambda_{i}-\lambda_{j}\right) d \hat{r}(\hat{x} \mid x) d F(x) \\
& =\int_{0}^{\bar{x}} x \mathbb{E}_{F^{u}(x)}\left[\prod_{j \neq i} F\left(\hat{F}^{-1}\left(F^{u}(x)\right)+\lambda_{i}-\lambda_{j}\right)\right] d F(x) \tag{D.13}
\end{align*}
$$

and

$$
\begin{equation*}
\mathbb{E}_{\tilde{r}}\left[u_{i}(\tilde{\boldsymbol{X}}, \boldsymbol{X}, \lambda)\right]=\int x \mathbb{E}_{F^{u}(x)}\left[\prod_{j \neq i} F\left(G^{-1}\left(F^{u}(x)\right)+\lambda_{i}-\lambda_{j}\right)\right] d F(x) \tag{D.14}
\end{equation*}
$$

respectively. Since $\hat{F}(x) \geq G(x)$, we know $\hat{F}^{-1}(p) \leq G^{-1}(p)$. Clearly (D.13) $\leq$ (D.14). We conclude that given a greedy allocation policy $\lambda$, true valuation $X_{i, t}$ and truthful agents $j \neq i$ (with $\left.\tilde{X}_{j, t}=X_{j, t}\right)$, reporting $\tilde{X}_{i, t} \sim \tilde{r}\left(\cdot \mid X_{i, t}\right)$ is a strategy for agent $i$ that maximizes $\mathbb{E}\left[u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \lambda\right)\right]$ subject to the marginal distribution constraint $\sup _{x}\left|F(x)-F_{r}(x)\right| \leq \Delta$. That is,

$$
\mathbb{E}_{r}\left[u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \lambda\right)\right] \leq \mathbb{E}_{\tilde{r}}\left[u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \lambda\right)\right] \quad \forall r \text { s.t. } \sup _{x}\left|F_{r}(x)-F(x)\right| \leq \Delta
$$

It remains to bound the difference $\mathbb{E}_{\tilde{r}}\left[u_{i}(\tilde{\boldsymbol{X}}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]$. First note that $G^{-1}(p)=$ $F^{-1}(p+\Delta)$. Then we have that

$$
\begin{align*}
& \mathbb{E}_{r}\left[u_{i}(\tilde{\boldsymbol{X}}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]  \tag{D.15}\\
= & \int_{0}^{\bar{x}} x\left(\mathbb{E}_{F^{u}(x)}\left[\prod_{j \neq i} F\left(F^{-1}\left(F^{u}(x)+\Delta\right)+\lambda_{i}-\lambda_{j}\right)\right]-\prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right)\right) d F(x) \\
\leq & \bar{x} \int_{0}^{\bar{x}}\left(\mathbb{E}_{F^{u}(x)}\left[\prod_{j \neq i} F\left(F^{-1}\left(F^{u}(x)+\Delta\right)+\lambda_{i}-\lambda_{j}\right)\right]-\prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right)\right) d F(x) \tag{D.16}
\end{align*}
$$

where the inequality follows from the fact that $F^{-1}\left(F^{u}(x)+\Delta\right) \geq x w . p .1$ for all $x$. To bound the remaining expression in the integral, we can use the fact that since the marginal distribution of $\tilde{x}$
under the joint distribution $\tilde{r}(\tilde{x}, x)$ is $G$, we have

$$
\begin{align*}
& \int_{0}^{\bar{x}} \int_{0}^{\bar{x}} \prod_{j \neq i} F\left(\tilde{x}+\lambda_{i}-\lambda_{j}\right) d \tilde{r}(\tilde{x} \mid x) d F(x) \\
= & \int_{0}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d G(x) \\
= & \int_{x_{\Delta}}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x)+\prod_{j \neq i} F\left(\bar{x}+\lambda_{i}-\lambda_{j}\right) \Delta \\
\leq & \int_{x_{\Delta}}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x)+\Delta \tag{D.17}
\end{align*}
$$

where $x_{\Delta}:=F^{-1}(\Delta)$. Similarly,

$$
\begin{align*}
& \int_{0}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x) \\
= & \int_{0}^{x_{\Delta}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x)+\int_{x_{\Delta}}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x) \\
\geq & \int_{x_{\Delta}}^{\bar{x}} \prod_{j \neq i} F\left(x+\lambda_{i}-\lambda_{j}\right) d F(x) \tag{D.18}
\end{align*}
$$

Plugging (D.17) and (D.18) back to (D.16), we can now bound the expression in (D.15), and thereby the profit from strategizing, by $\bar{x} \Delta$.

## D.3.3 Proof of Lemma 18

Proof. Let $\bar{F}$ be the average distribution that agent $i$ reported from up to round $T^{\prime}: \bar{F}=\frac{1}{T^{\prime}} \sum_{t=1}^{T^{\prime}} F_{t}$, where $F_{t}$ is the reported value distribution of agent $i$ in time $t: F_{t}(x):=\mathbb{P}\left(\tilde{X}_{i, t} \leq x \mid \mathcal{H}_{t}\right)$. Since the the detection algorithm has not been triggered, we can conclude using Lemma 16 that with
probability $1-\delta$,

$$
\begin{array}{r}
\sup _{x}|\bar{F}(x)-F(x)|<\Delta:=64 \sqrt{\frac{\log \left(\frac{256 e T^{\prime}}{\delta}\right)}{T^{\prime}}}, \\
\text { and } \sup _{x}\left|\bar{F}_{T^{\prime}}(x)-\bar{F}(x)\right|<\frac{\Delta}{4}=16 \sqrt{\frac{\log \left(\frac{256 e T^{\prime}}{\delta}\right)}{T^{\prime}}} .
\end{array}
$$

The second inequality holds because the proof of Lemma 16 uses the second inequality to show the first (see Equation D.11). Combining the above two steps, we have

$$
\begin{equation*}
\sup _{x}\left|\bar{F}_{T^{\prime}}(x)-F(x)\right|<\frac{\Delta}{4}=80 \sqrt{\frac{\log \left(\frac{256 e T^{\prime}}{\delta}\right)}{T^{\prime}}} \quad \text { w.p. } 1-\delta \text {. } \tag{D.19}
\end{equation*}
$$

This shows that if the detection algorithm has not been triggered, the empirical CDF of strategic agent's reported values are close to the true CDF. Let $\tilde{F}_{T^{\prime}}(x)=\frac{1}{(n-1) T^{\prime}} \sum_{t=1}^{T^{\prime}} \sum_{j \neq i} \mathbb{1}\left[X_{j, t} \leq x\right]$ be the emipircal distriution from all agents other than $i$. We know from Lemma 34 that

$$
\begin{equation*}
\sup _{x}\left|\tilde{F}_{T^{\prime}}(x)-F(x)\right| \leq \sqrt{\frac{1}{2(n-1) T^{\prime}} \log \left(\frac{2}{\delta}\right)} \quad \text { w.p. } 1-\delta . \tag{D.20}
\end{equation*}
$$

Combining (D.19) and (D.20), we can now bound the error in the combined estimation, $\hat{F}_{T^{\prime}}=$ $\frac{1}{n T^{\prime}} \sum_{t=1}^{T^{\prime}} \sum_{j=1}^{n} \mathbb{1}\left[\boldsymbol{X}_{j}^{t} \leq x\right]:$

$$
\begin{array}{rlr} 
& \sup _{x}\left|\hat{F}_{T^{\prime}}(x)-F(x)\right| & \\
= & \sup _{x}\left|\frac{1}{n} \bar{F}_{T^{\prime}}(x)+\frac{n-1}{n} \tilde{F}_{T^{\prime}}(x)-F(x)\right| & \\
= & \sup _{x}\left|\frac{1}{n} \bar{F}_{T^{\prime}}(x)-\frac{1}{n} F(x)+\frac{n-1}{n} \tilde{F}_{T^{\prime}}(x)-\frac{n-1}{n} F(x)\right| & \\
\leq & \sup _{x}\left|\frac{1}{n} \bar{F}_{T^{\prime}}(x)-\frac{1}{n} F(x)\right|+\sup _{x}\left|\frac{n-1}{n} \tilde{F}_{T^{\prime}}(x)-\frac{n-1}{n} F(x)\right| & \\
\leq & 80 \sqrt{\frac{\log \left(\frac{256 e T^{\prime}}{\delta}\right)}{n T^{\prime}}}+\sqrt{\frac{1}{2 n T^{\prime}} \log \left(\frac{2}{\delta}\right)} & \text { w.p. } 1-2 \delta \\
\leq & 81 \sqrt{\frac{\log \left(\frac{256 e T^{\prime}}{\delta}\right)}{n T^{\prime}}} & \text { w.p. } 1-2 \delta \tag{D.21}
\end{array}
$$

Let $\hat{\boldsymbol{F}}_{T^{\prime}}=\hat{F}_{T^{\prime}} \otimes \ldots \otimes \hat{F}_{T^{\prime}}$, and $\lambda=\lambda^{*}\left(\hat{\boldsymbol{F}}_{T^{\prime}}\right)$, and $\Delta_{T^{\prime}}=81 \sqrt{\frac{\log \left(\frac{2566 T^{\prime}}{\delta}\right)}{n T^{\prime}}}$. Applying Lemma 15 to (D.21) we have

$$
\begin{gathered}
\sup _{x}\left|\hat{F}_{T^{\prime}}(x)-F(x)\right| \leq \Delta_{T^{\prime}} \quad \text { w.p. } 1-2 \delta \\
\left(\text { Lemma 15) } \Longrightarrow \mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \leq n \Delta_{T^{\prime}} \bar{x}\right.
\end{gathered}
$$

## D.3.4 Proof of Lemma 19

Proof. Let $F_{t}, t=1, \ldots, T$ be the distributions that agent $i$ reports from in each round given the history, i.e. $\tilde{X}_{i, t} \mid \mathcal{H}_{t} \sim F_{t}$. First we try to bound the utility that the strategic agent can get from a single epoch. Fix an epoch $k$. Suppose $T^{\prime}$ is the time when either detection algorithm is triggered, or the first time some receiver hits his allocation budget of $p_{j}^{*} T$. Let $\tau=\min \left(T^{\prime}, L_{k+1}-1\right)$. We
now define three distributions:

$$
\begin{aligned}
& \bar{F}^{1}=\frac{1}{L_{k}-1} \sum_{t=1}^{L_{k}-1} F_{t} \\
& \bar{F}^{2}=\frac{1}{\tau-1} \sum_{t=1}^{\tau-1} F_{t} \\
& \bar{F}^{3}=\frac{1}{\tau-L_{k}} \sum_{t=L_{k}}^{\tau-1} F_{t}
\end{aligned}
$$

These are the average distributions that agent $i$ reported from, averaged across three time periods: $\left[1, L_{k}\right),[1, \tau)$ and $\left[L_{k}, \tau\right)$. In particular, $\bar{F}^{3}$ is the average distribution that the strategic agent reports from in epoch $k$. From Lemma 16 we know that with probability $1-2 \delta$ :

$$
\begin{aligned}
& \sup _{x}\left|\bar{F}^{1}(x)-F(x)\right| \leq 64 \sqrt{\frac{\log \left(\frac{256 e\left(L_{k}-1\right)}{\delta}\right)}{n\left(L_{k}-1\right)}} \\
& \sup _{x}\left|\bar{F}^{2}(x)-F(x)\right| \leq 64 \sqrt{\frac{\log \left(\frac{256 e(\tau-1)}{\delta}\right)}{n(\tau-1)}}
\end{aligned}
$$

which together means that

$$
\begin{aligned}
& \sup _{x}\left|\bar{F}^{2}(x)-F(x)\right|=\sup _{x}\left|\frac{L_{k}}{\tau}\left(\bar{F}^{1}(x)-F(x)\right)+\frac{\tau-L_{k}}{\tau}\left(\bar{F}^{3}(x)-F(x)\right)\right| \\
\Longrightarrow & \sup _{x}\left|\bar{F}^{2}(x)-F(x)\right| \geq \sup _{x}\left|\frac{\tau-L_{k}}{\tau}\left(\bar{F}^{3}(x)-F(x)\right)\right|-\sup _{x}\left|\frac{L_{k}}{\tau}\left(\bar{F}^{1}(x)-F(x)\right)\right| \\
\Longrightarrow & \sup _{x}\left|\bar{F}^{3}(x)-F(x)\right| \leq \bar{\Delta}_{k}:=\min \left(\frac{128 \tau}{\tau-L_{k}} \sqrt{\frac{\log \left(\frac{256 e(\tau-1)}{\delta}\right)}{n(\tau-1)}}, 1\right)
\end{aligned}
$$

Note that the last step also uses the fact that the difference between two CDFs cannot be bigger than 1 . Let $r$ be any joint distribution for agent $i$ 's reported and true valuation $(\tilde{x}, x)$ such that the marginal for the reported valuation is equal to $\bar{F}^{3}$, i.e.,

$$
\bar{X}_{i, t} \sim r\left(\cdot \mid X_{i, t}\right), X_{i, t} \sim F \Longrightarrow F_{r}(x):=\mathbb{P}\left(\bar{X}_{i, t} \leq x\right)=\bar{F}^{3}
$$

Let $\overline{\boldsymbol{X}}$ denote the reported value vector when $i$ is the only strategic agent and uses $r\left(\cdot \mid X_{i}\right)$ to pick his reported value: $\bar{X}_{j}=X_{j} \forall j \neq i, \bar{X}_{i} \sim r\left(\cdot \mid X_{i}\right)$. Let $\Delta_{L_{k}-1}=81 \sqrt{\frac{\log \left(\frac{256 e\left(L_{k}-1\right)}{\delta\left(L_{k}-1\right)}\right.}{n}}$. Using this, we have

$$
\begin{align*}
&\left(\text { Lemma 18) } \Longrightarrow \mathbb{E}\left[u_{i}(\boldsymbol{X}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \leq n \Delta_{L_{k}-1} \bar{x}\right. \\
&(\text { Lemma 17) } \Longrightarrow \mathbb{E}\left[u_{i}(\overline{\boldsymbol{X}}, \boldsymbol{X}, \lambda)\right]-\mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \leq n \Delta_{L_{k}-1} \bar{x}+\bar{\Delta}_{k} \bar{x} \\
& \text { (Corollary 5) } \Longrightarrow \sum_{t=L_{k}}^{\tau-1} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \tilde{\lambda}_{k_{t}}\right)-\left(\tau-L_{k}\right) \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \\
& \leq\left(n \Delta_{L_{k}-1}+\bar{\Delta}_{k}\right) \bar{x}\left(\tau-L_{k}\right)+16 \sqrt{\left(\tau-L_{k}\right) \log \left(\frac{128 e\left(\tau-L_{k}\right)}{\delta}\right)} \bar{x} \\
& \text { w.p. } 1-\delta  \tag{D.22}\\
& \leq 81 \sqrt{\frac{n\left(\tau-L_{k}\right)^{2}}{2\left(L_{k}-1\right)} \log \left(\frac{256 e L_{k}}{\delta}\right)} \bar{x}+144 \sqrt{2 \tau \log \left(\frac{256 e \tau}{\delta}\right) \bar{x}} \quad \text { w.p. } 1-\delta
\end{align*}
$$

The above is a high probability bound on how much an agent can get in one epoch. We can now bound the strategic agent's utility over the full horizon.

$$
\begin{aligned}
& \sum_{t=1}^{T^{\prime}} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \tilde{\lambda}_{k_{t}}\right)-T^{\prime} \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right] \\
\leq & \sum_{k=0}^{\log _{2} T^{\prime}-1}\left[\sum_{t=L_{k}}^{L_{k+1}-1} u_{i}\left(\tilde{\boldsymbol{X}}_{t}, \boldsymbol{X}_{t}, \lambda\right)-\left(L_{k+1}-L_{k}\right) \mathbb{E}\left[u_{i}\left(\boldsymbol{X}, \boldsymbol{X}, \lambda^{*}\right)\right]\right]
\end{aligned}
$$

$($ Using $(\mathrm{D} .22)) \leq \bar{x}\left(L_{1}-1\right)+\sum_{k=0}^{\log _{2} T^{\prime}-1}\left(81 \sqrt{\frac{n\left(L_{k+1}-L_{k}\right)^{2}}{2\left(L_{k}-1\right)} \log \left(\frac{256 e\left(L_{k}-1\right)}{\delta}\right)} \bar{x}\right.$
$\left.+144 \sqrt{2 L_{k+1} \log \left(\frac{256 e L_{k+1}}{\delta}\right) \bar{x}}\right)$
w.p $1-\delta \log _{2} T$
$\left(L_{k}=2^{k}\right) \leq \bar{x}+\sum_{k=0}^{\log _{2} T^{\prime}-1} 285 \sqrt{n 2^{k} \log \left(\frac{256 e T^{\prime}}{\delta}\right)} \bar{x}$
$\leq\left(\frac{285 \sqrt{2}}{\sqrt{2}-1} \sqrt{n T^{\prime} \log \left(\frac{256 e}{\delta}\right)}+1\right) \bar{x}$ w.p $1-\delta \log _{2} T$ w.p $1-\delta \log _{2} T$

The result follows by replacing the original $\delta$ with $\frac{\delta}{\log _{2} T}$.

## D. 4 Auxiliary Proofs

## D.4.1 Proof of Claim 8

Claim 8. Let $G$ be any distribution (cdf) over $\mathcal{X} \subseteq \mathbb{R}$, and $F$ over $\mathcal{Y} \subseteq \mathbb{R}$. Then there exists a unique joint distribution $r$ over $\mathcal{X} \times \mathcal{Y}$ with marginals $G, F$ such that the conditional distribution $r(\cdot \mid Y)$ has the following monotonicity property: define $\bar{x}_{r}(\cdot), \underline{x}_{r}(\cdot)$ so that $X \in\left[\underline{x}_{r}(Y), \bar{x}_{r}(Y)\right]$ almost surely, i.e.,

$$
\begin{gathered}
\bar{x}_{r}(y)=\inf \{x: \mathbb{P}(X>x \mid Y=y)=0\} \\
\underline{x}_{r}(y)=\sup \{x: \mathbb{P}(X<x \mid Y=y)=0\},
\end{gathered}
$$

then

$$
\bar{x}_{r}\left(y_{1}\right) \leq \underline{x}_{r}\left(y_{2}\right) \quad \forall y_{1}<y_{2} .
$$

In particular, the random variable $X \mid Y \sim r(\cdot \mid Y)$ can be sampled as $G^{-1}\left(F^{u}(Y)\right)$, where $F^{u}$ is the random function defined in (D.12) and $G^{-1}:=\inf \{x \in \mathbb{R}: G(x) \geq p\}$ denotes the generalized inverse, sometimes also referred to as the quantile function.

Proof. We first prove existence by constructing a joint distribution with the desired marginals and monotonicity, then we show uniqueness.

Existence. We will construct the joint distribution by defining the conditional distribution of $X$ given $Y=y$ for every $y$. Note that if $F$ is a continuous distribution, then we can easily construct $r(\cdot \mid Y=y)$ using the inverse-CDF method:

$$
X \mid y=G^{-1}(F(y))
$$

where $G^{-1}:=\inf \{x \in \mathbb{R}: G(x) \geq p\}$ is the generalized inverse. This works because $F(Y) \sim$ Uniform $[0,1]$. If $F$ contains point masses, then $F(Y)$ is no longer uniformly distributed, and the inverse-CDF method does not work. To resolve this, we construct a different random variable $F^{u}(y)$ for each value $y$. For a given sample $y$, If $F(y) \neq F\left(y_{-}\right)$, let $F^{u}(y) \sim \operatorname{Uniform}\left[F\left(y_{-}\right), F(y)\right]$. Otherwise, let $F^{u}(y)=F(y)$. Now we let

$$
X \mid y=G^{-1}\left(F^{u}(y)\right)
$$

To see that $X$ sampled using this process has the marginal distribution $G$, we just need to show that $F^{u}(Y)$ is uniformly distributed. For a given $p$, if $\exists$ y s.t. $F(y)=p$, then $\mathbb{P}\left(F^{u}(Y) \leq p\right)=\mathbb{P}(F(Y) \leq$ $p)=\mathbb{P}(Y \leq y)=p$. Otherwise that means $\exists y$ s.t. $p_{1}:=F\left(y_{-}\right) \leq p$ and $p_{2}:=F(y)>p$.

$$
\begin{aligned}
& \mathbb{P}\left(F^{u}(Y) \leq p\right) \\
= & \mathbb{P}(Y<y)+\mathbb{P}\left(F^{u}(y) \leq p \mid Y=y\right) \mathbb{P}(Y=y) \\
= & p_{1}+\frac{p-p_{1}}{p_{2}-p_{1}}\left(p_{2}-p_{1}\right) \\
= & p
\end{aligned}
$$

This construction also satisfies monotonicity, since if $y_{1}<y_{2}$, then $F^{u}\left(y_{1}\right) \leq F\left(y_{1}\right)$ w.p.1. and $F^{u}\left(y_{2}\right) \geq F\left(y_{1}\right)$ w.p.1.

Uniqueness Now we show uniqueness. For a given $(x, y)$ pair, suppose $x<\bar{x}_{r}(y)$. Then from monotonicity we know $\underline{x}_{r}\left(y^{\prime}\right) \geq \bar{x}_{r}(y)>x$ for all $y^{\prime}>y$, which implies that

$$
\mathbb{P}_{r}(X \leq x, Y \leq y)=G(x) .
$$

If $x \geq \bar{x}_{r}(y)$, then from monotonicity we know $\bar{x}_{r}\left(y^{\prime}\right) \leq \bar{x}_{r}(y) \leq x$ for all $y^{\prime}<y$, which implies that

$$
\mathbb{P}_{r}(X \leq x, Y \leq y)=F(y)
$$

Since $G$ and $F$ are fixed, we have shown that all joint distributions $r$ with monotonicity and the required marginals are the same.


[^0]:    ${ }^{1}$ Technically our lower bound only holds for algorithms that satisfy certain conditions (Assumption 1 and 2) stated in Section 1.7. However, we believe it is still applicable to the algorithm in [37]. Their algorithm with its constant upper bound on prices seems to satisfy both of our assumptions.

[^1]:    ${ }^{2}$ A counting process $\left\{d_{t}, t \geq 0\right\}$ is called a non-homogeneous Poisson process with rate $\lambda_{t}$ if all the following conditions hold: (a) $d_{0}=0$; (b) $d_{t}$ has independent increments; (c) for any $t \geq 0$ we have $\operatorname{Pr}\left(d_{t+\delta}-d_{t}=0\right)=$ $1-\lambda_{t} \delta+o(\delta), \operatorname{Pr}\left(d_{t+\delta}-d_{t}=1\right)=\lambda_{t} \delta+o(\delta), \operatorname{Pr}\left(d_{t+\delta}-d_{t} \geq 2\right)=o(\delta)$.

[^2]:    ${ }^{3}$ A subtle but important difference between this deterministic model vs. the stochastic Bass model introduced earlier is that in the deterministic model, the increment $m d X_{t}$ in adoptions is fractional. On the other hand, in our stochastic model, the number of customer arrivals (or adoptions) $d_{t}$ is a counting process with discrete increments. This difference will be taken into account later when we compare our pseudo-regret to the original definition of regret.

[^3]:    ${ }^{4}$ In Lemma 32 in Appendix A.6, we show that if a constant upper bound is required on the price, then the problems becomes trivial for any $T \geq \Omega(\log (m)))$.

[^4]:    ${ }^{1}$ Manhattan, the Bronx, and Brooklyn are three boroughs of New York City.

[^5]:    ${ }^{2}$ A closely related concept called proportionality is also often seen in the literature. We focus on Sharing Incentive in the main paper and include a discussion of proportionality in the Appendix.

[^6]:    ${ }^{3}$ We also considered an alternative design that did not yield a strategy-proof and envy-free algorithm. See Appendix D.

