Learning Algorithms in Strategic Environments

Jon Schneider

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Computer Science
Adviser: Professor Mark Braverman

September 2018

Abstract

Learning algorithms are often analyzed under the assumption their inputs are drawn from stochastic or adversarial sources. Increasingly, these algorithms are being applied in strategic settings, where we can hope for stronger guarantees. This thesis aims to understand the performance of existing learning algorithms in these settings, and to design new algorithms that perform well in these settings.

This thesis is divided into three parts. In Part I, we address the question of how agents should learn to bid in repeated non-truthful auctions – and conversely, how should we design auctions whose participants are learning agents.

In Part II, we study the dynamic pricing problem: the question of how should a large retailer learn how to set prices for a sequence of disparate goods over time, based on observing demands for goods at various prices. Previous work has demonstrated how to obtain $O(\log T)$ regret for this problem. We show how to achieve regret $O(\log \log T)$, which is tight. Our algorithm uses ideas from integral geometry (most notably the concept of intrinsic volumes).

Finally, in Part III, we study how to learn the ranking of a set of N items from pairwise comparisons that may be strategic or noisy. In particular, we design mechanisms for a variety of settings (choosing the winner of a round-robin tournament, aggregating the top-K items under the strong stochastic transitivity noise model) which outperform the naive rule of ranking items according to the total number of pairwise comparisons won.

Acknowledgements

First and foremost, I would like to thank my advisor, Mark Braverman. Mark has been an incredible advisor over the last couple years, always guiding me as a researcher, encouraging me to work on important problems, and always being willing to discuss anything I find interesting. Mark's ability to generate insights and new ideas in many different fields is truly inspirational – without fail, every time I talk with Mark I walk away understanding something better, regardless of whether we're talking about complexity theory, learning algorithms, or the state of healthcare in the United States.

I would also like to thank Matt Weinberg, who introduced me to the field of algorithmic mechanism design and acted in many ways as an informal co-advisor to me. Matt's insight and intuition guided a lot of the research I've done over the last couple years, and if it were not for Matt's presence at Princeton, this dissertation would look very different.

In the Fall of 2017, I did an internship at Google Research in NYC. This ended up being one of the most enjoyable and most productive research experiences of my PhD. I'd like to thank Vahab Mirrokni for inviting me to intern in his group, Mohammad Mahdian for mentoring me over the course of the internship, and Santiago Balseiro, Negin Golrezaie, and Renato Paes Leme for working with me on interesting research problems over the course of my internship.

The Computer Science department at Princeton has been an excellent environment for research over the last five years. I'd like to thank the broader research community at Princeton, including all professors whose classes I took and all my officemates over the year. I'd especially like to thank Bernard Chazelle, Elad Hazan, Matt Weinberg, and Omri Weinstein for taking time out of their schedules to serve on my dissertation committee, and all my coauthors that I've had the pleasure of working with over the course of my PhD: Santiago Balseiro, Mark Braverman, Xi

Chen, Sumegha Garg, Negin Golrezaie, Sivakanth Gopi, Mohammad Mahdian, Jieming Mao, Vahab Mirrokni, Renato Paes Leme, Cristobal Rojas, Ariel Schvartzman, and Matt Weinberg.

I'm thankful to all my friends and professors from my undergraduate studies at MIT, who taught me a lot and shaped me into who I am today. I'd especially like to thank Richard Stanley who mentored me as an undergraduate researcher and really inspired me to pursue research.

Finally, I am incredibly grateful to my family – Mom, Dad, and Julia – for their support over the last five years of my PhD and the last twenty six years of my life. None of this would be possible without their constant encouragement. It is to them that I dedicate this thesis.

Contents

Chapter 1

Introduction

Increasingly, the fields of algorithmic mechanism design and machine learning are starting to overlap. Machine learning algorithms are being applied in strategic settings, where rational agents have control over the inputs to these algorithms and have a vested stake in the outputs of these algorithms. This is evident in applications of machine learning methods to problems like fraud detection, spam filtering, and university admissions – all problems where the providers of the data may have incentives to misreport their true values. Similarly, machine learning algorithms are increasingly being used to learn how to participate in specific games and mechanisms. The best learning algorithms can now easily beat master-level human players in games like Poker and Go [133, 32], and automated machine-learning algorithms are increasingly being applied to tasks like trading in stock markets [145].

The goal of this thesis is to understand this overlap and make progress towards answering the following questions:

- **How should we design mechanisms whose participants are learning agents?**
- **How should we design learning agents to participate in specific mechanisms?**

For example, how should we design a repeated auction if we know that all the participants are learning over time how to optimally bid? Or, alternatively, how should we design a learning algorithm to learn how to optimally bid over time in this auction?

A priori, it may not seem obvious that these questions are any different from the central questions studied in mechanism design or machine learning. For example, if we design an "optimal" (in some sense) mechanism for some problem under the assumption that the participants are rational, self-interested agents, then one might surmise that such a mechanism would also be optimal if the participants were agents learning to play over time. Likewise, we might assume that if we design learning agents with theoretically guaranteed good behavior even under adversarial inputs, these agents will also perform well in strategic settings. One of the recurring morals of this thesis is that this is *not* necessarily the case.

Partly this occurs when the problems we consider have more structure to exploit than the most general adversarial problems. For instance, in Chapter 4 we consider a pricing problem where a standard low-regret bandit algorithm obtains regret $\tilde{O}(\sqrt{T})$ over T rounds, but where we propose a learning algorithm that obtains regret $O(\log \log T)$. This is a significant improvement, but not very surprising, no different than an improvement we should expect when specializing any general learning algorithm to a specific task.

More surprising is the following phenomenon: there are situations where learning algorithms with strong adversarial guarantees (e.g. algorithms for the multi-armed bandit problem that achieve sublinear regret) perform *as badly as possible* in strategic settings. We show in Chapter 2 that if bidders use common low-regret algorithms to learn how to bid over time (e.g. EXP3), then the seller can design a mechanism which extracts maximum revenue leaving the bidders with *zero* net utility. In Chapter 3, which studies a variant of the multi-armed bandit problem where the arms are

strategic agents that can withhold rewards, we show that any adversarial low-regret bandit algorithm leads to bad equilibria where the algorithm receives asymptotically zero total reward (yet there exist other algorithms which receive positive total reward).

In the remainder of this introduction, we describe the results of this thesis in further detail. We begin with a quick introduction to the multi-armed bandit problem (Section 1.1) and then briefly describe the main results of each chapter (Section 1.2).

1.1 Online learning, bandit algorithms, and regret

With the exception of Part III of this thesis where we look at rank-aggregation mechanisms, all of the learning agents we consider learn over time and can be thought of as instances of algorithms for the *multi-armed bandits problem* (or one of its variants). In this section, we provide a brief introduction to the multi-armed bandits problem. Additional details will be presented in the individual chapters as necessary. For a more detailed overview, we recommend the reader consult the survey by Bubeck and Cesa-Bianchi [34].

The classic multi-armed bandit problem is a problem where a learner must choose one of K actions ('arms') per round, over T rounds. On round t, the learner receives some reward $r_{i,t} \in [0, 1]$ for choosing ('pulling') arm i, where the values $r_{i,t}$ are possibly drawn stochastically from some arm-specific distribution or chosen adversarially, depending on the specific problem setup. The learner's goal is to maximize their total reward.

We measure the quality of an algorithm for the multi-armed bandits problem in terms of its *regret*; the difference between the total reward it obtains, and the total reward obtained by the best individual arm (or in some cases, some other appropriately chosen benchmark). More formally, let I_t denote the arm pulled by the principal at round t. The *regret* of an algorithm $\mathcal{A}$ for the learner is then given by the random

variable $\mathsf{Reg}(\mathcal{A}) = \max_i \sum_{t=1}^{T} r_{i,t} - \sum_{t=1}^{T} r_{I_t,t}$. We say an algorithm $\mathcal{A}$ is *no-regret* (alternatively "low-regret") if $\mathbb{E}[\mathsf{Reg}(\mathcal{A})] \leq o(T)$ (that is, its expected regret is sublinear in T).

A seminal result in online learning is that there exist simple no-regret algorithms for the multi-armed bandits problem. The best known of these algorithms, UCB1 and EXP3 both achieve $\tilde{O}(\sqrt{KT})$ regret (when rewards are drawn stochastically and adversarially, respectively).

In some settings, a learner additionally receives some information per round, and can use this information to influence their choice of action. This is captured by the *contextual bandit problem*. In the contextual bandit problem, the learner now begins by receiving some context c_t (belonging to some finite set $\mathcal{C}$ of cardinality C) at the beginning of round t. Based on this context, the learner must choose one of K actions. If the learner chooses action i on round t in context c, they receive reward $r_{i,t}(c)$ (where again, these rewards might be chosen adversarially or generated stochastically). The learner once again wants to maximize their total reward.

As before, it is possible to define a notion of regret for contextual bandits, where the regret of a contextual bandit algorithm is the difference between the reward it obtains and the best reward obtained by an algorithm in some class of policies. Throughout this thesis, we will only be concerned with the case where this class of policies is the set of all stationary policies; that is, functions mapping the set of contexts to the set of actions. In this case there again exist algorithms with regret sublinear in T; one simple construction is to just run a separate instance of EXP3 for every different context. This obtains expected regret $\tilde{O}(\sqrt{CKT})$.

1.2 Organization of this thesis

The remainder of this thesis is organized into chapters, each chapter focusing on a specific problem somewhere in the intersection of algorithmic mechanism design and learning. Each chapter is self-contained and can be read out of order.

Very broadly, the chapters can be divided into three themes: *learning how to bid* (Part I), *learning how to price* (Part II), and *learning how to rank* (Part III). We summarize each part in further detail below.

1.2.1 Learning how to bid

Imagine participating in an auction for an item. How should you decide how much to bid? If the auction is truthful (e.g., a second price auction), then bidding is easy: in such auctions, simply bidding your own value is a dominant strategy. Unfortunately, many auctions commonly used in practice – first price auctions, generalized second price auctions – are not truthful. How should you bid in these more complex auctions?

Traditional economic wisdom states that you should establish priors for other participants' value for the item, compute the Bayes-Nash equilibrium of the resulting game, and play according to this equilibrium. There are many difficulties with this approach. Estimating priors can be difficult, computation of equilibria can be intractable [49], and other participants in the auction may not even be playing rationally.

On the other hand, if this auction is repeated, a much simpler option is to learn how to bid over time by employing online learning algorithm – indeed, this can naturally be thought of as a contextual bandits problem, where the bidder receives a context (their value) every round, must choose an action (a bid), and receives a corresponding reward (their net utility). This leads to a couple natural questions. What sort of learning algorithm should a learner use to learn how to bid in a complicated

auction? Conversely, as an auctioneer, how should you design an auction if you expect that bidders will learn how to bid over time (in say, a low-regret manner)?

In Chapter 2, we study this problem in the setting of a seller selling a single item every round to a learning buyer. Interestingly, even in this simplified setting, a wealth of interesting phenomena emerge. For example, if the buyer is using a low-regret algorithm in a class of algorithms we call "mean-based" (which includes a variety common learning algorithms, such as EXP3, Multiplicative Weights, and Follow the Perturbed Leader), we show that there is a mechanism for the seller which can extract the full expected utility of the buyer. Even when we restrict the seller to more restricted classes of "reasonable mechanisms" (e.g. auctions where overbidding is a dominated strategy), we show it is possible for the seller to achieve average revenue significantly larger than Myerson revenue (the best they can hope to achieve in the one-shot strategic variant of this problem).

The issue with existing learning algorithms in this setting is that they are almost all "mean-based": with high probability, they always pick one of the best historically performing actions. While intuitively this may seem like a useful property to have, our results show it can be exploited in strategic settings. Indeed, we also demonstrate a (non-mean-based) low-regret learning algorithm for the buyer which guarantees that the seller receives no more than the Myerson revenue per round.

One can also flip this setting around and examine learning from the perspective of the *auctioneer*. Faced with a collection of strategic bidders with unknown value distributions and structure, how should the auctioneer adapt their mechanism over time?

In Chapter 3 we consider a very simplified model of this problem which we call *bandits with strategic arms*. In this model, the auctioneer does not solicit bids, but instead simply selects one of the bidders every round to give the item to. The bidder then receives the item and pays the auctioneer some amount of the bidder's choosing.

A bidder is allowed to pay zero for the item, but the auctioneer may react by picking the bidder less in future rounds. What sort of mechanism should the auctioneer run to maximize their revenue?

Since every round the auctioneer chooses one of K choices (which bidder to select), this is a type of multi-armed bandit problem for the auctioneer but where the arms are strategic – instead of passing on their full reward (their value for the item), they instead get to choose what fraction of reward to pass on. One reasonable choice of mechanism for the auctioneer is to run some no-regret algorithm for the multi-armed bandits problem, like EXP3.

In this chapter, we show that adversarial no-regret algorithms (like EXP3) are *far from strategyproof*; if the auctioneer uses such an algorithm, then there is an ε-approximate Nash equilibrium for the arms where the auctioneer receives 0 total revenue. Unlike many other collusive equilibria, this equilibrium requires no explicit communication between bidders, and can arise from bidders simply trying to maintain an equal market share (i.e. being picked equally often). In contrast, there are simple mechanisms for the auctioneer (similar to auctioning off the right to be played for all T rounds) that are not no-regret which guarantee the auctioneer positive revenue (approximately the second highest average value per round over all the bidders).

1.2.2 Learning how to price

How should large retailers set prices for the items they sell? One approach is to learn prices over time – set a price for an item, observe the demand for the item at that price, and use that information when setting prices in the next time period or for similar items. This is the core of the problem of *dynamic pricing*.

Consider the following very simple model of dynamic pricing, where a retailer tries to repeatedly sell an item to a single consumer. Every round t (for T rounds), the retailer must set a price for a new item. This item has some collection of relevant

features, which are observed by the retailer can be described as a vector $u_t \in \mathbb{R}^d$. Based on these features, the retailer must set a price p_t for this item. Now, the consumer has a fixed valuation vector $\theta \in \mathbb{R}^d$, where θ_i represents the value they assign to the ith feature. The consumer is willing to pay up to $\langle \theta, u_t \rangle$ for this item. If p_t is less than this inner product, the consumer will purchase the item at price p_t, and if p_t is greater, the consumer will refuse to purchase the item. The goal of the retailer is to maximize their revenue.

Note that again, this can be thought of as a special case of the contextual bandits problem for the retailer. Every round the retailer receives a context (the features u_t of the item), must take an action (setting a price p_t) and receives a reward (p_t if the consumer buys the item, 0 otherwise). Indeed, it is possible to apply standard bandits algorithms to this problem and obtain $\tilde{O}(\sqrt{T})$ regret.

In [44] and [99], the authors showed how to obtain $O(\log T)$ regret by using geometric approaches to efficiently narrow down the set of possible values of the valuation vector θ. In Chapter 4, we demonstrate an algorithm for this problem which achieves $O(\log \log T)$ regret, which is tight due to a lower bound of Kleinberg and Leighton [90].

Our algorithm relies heavily on ideas from integral geometry, notably the concept of *intrinsic volumes*. Many people are familiar with two common "measures" of a d-dimensional convex body that are invariant under rigid motions – its volume and surface area (or more accurately, the higher-dimensional analogues of these concepts). One of the fundamental results in integral geometry is there are in fact d "distinct" invariant measures of a d-dimensional convex body S, one for each dimension from 1 to d (of which surface area and volume are just 2). These invariant measures are the intrinsic volumes of S and have many nice mathematic properties (for example, they occur as the coefficients of Steiner's formula).

1.2.3 Learning how to rank

Consider the following scenario. You have N items, and you are given some number of pairwise comparisons between each pair of items, with the caveat that these comparisons might be noisy, or non-transitive, or strategically chosen in some way. How should you, from the information you have, decide which item is the "best" item (or which K items are the "best" K items)?

Such problems arise naturally in fields like crowdsourcing, recommendation systems, and tournament design. A commonly used strategy in practice (known as the Copeland strategy in social choice theory) is to simply choose the item that wins the largest proportion of pairwise comparisons. In the simplest models, this algorithm can be shown to perform well (and even optimally) [132].

In Chapter 5, we consider the problem of deciding the winner of a round-robin tournament. This is a strategic variant of this problem, where the items themselves are strategic agents and they have some incentive to be chosen as the "best item".

In tournament design, if one player beats all other players, that player should definitely be chosen as the winner. In addition, you want to design a tournament rule that is manipulation-proof; in particular, it should be hard for two players to increase the chance that one of them wins the tournament by fixing the outcome of the match between them.

The Copeland mechanism is often used in practice to decide the winner of such tournaments. But the Copeland mechanism is manipulable; there are situations where collusion can increase the probability of a pair of players winning from near 0 to almost 1. In this chapter we examine tournament structures that minimize the incentive for pairs of players to collude, and show that random single-elimination tournaments are optimal. In future work, we hope to understand which tournament structures minimize the incentive for larger coalitions of players to collude (in this case, we know that random single-elimination tournaments are *not* optimal).

In Chapter 6, we study the top-K problem under a model for noisy pairwise comparisons known as the Strong Stochastic Transitivity (SST) model, which assumes there is an underlying ordering of the items, and asserts that the probability p_{ik} of i beating k in a pairwise comparision is larger than p_{jk} if i occurs above j in the true ordering of items. This SST model subsumes many parametric models commonly used in practice. In this chapter we develop new $O(\sqrt{N})$-competitive algorithms for identifying the top K items in this model (i.e. our algorithm requires at most $O(\sqrt{N})$ times as many samples to correctly identify the top K as any algorithm, even one especially tailored to the instance). Previous algorithms for this problem, including the Copeland mechanism, were all $\Omega(N)$-competitive at best.

Part I

Learning how to bid

Chapter 2

Selling to a No-Regret Buyer

This chapter is joint work with Mark Braverman, Jieming Mao, and Matthew Weinberg [30].

2.1 Introduction

Consider a bidder trying to decide how much to bid in an auction (for example, a sponsored search auction). If the auction happens to be the truthful Vickrey-Clarke-Groves auction [142, 42, 70], then the bidder's decision is easy: simply bid your value. If instead, the bidder is participating in a Generalized First-Price (GFP) or Generalized Second-Price (GSP) auction, the optimal strategy is less clear. Bidders can certainly attempt to compute a Bayes-Nash equilibrium of the associated game and play accordingly, but this is unrealistic due to the need for accurate priors and extensive computation.

Alternatively, the bidders may try to learn a best-response over time (possibly offloading the learning to commercial bid optimizers). We specifically consider bidders who *no-regret learn*, as empirical work of [115] shows that bidder behavior on Bing is largely consistent with no-regret learning (i.e. for most bidders, there exists a per-click value such that their behavior guarantees no-regret for this value). From the

perspective of a revenue-maximizing auction designer, this motivates the following question: **If a seller knows that buyers are no-regret learning over time, how should they maximize revenue?**

This question is already quite interesting even when there is just a single item for sale to a single buyer. We consider a model where in every round t, the seller solicits a bid $b_t \in [0, 1]$ from the buyer, then allocates the item according to some allocation rule $x_t(\cdot)$ and charges the bidder according to some pricing rule $p_t(\cdot)$ (satisfying $p_t(b) \leq b \cdot x_t(b)$ for all t, b).[1] Note that the allocation and pricing rules (henceforth, auction) can differ from round to round, and that the auction need not be truthful. Each round, the bidder has a value v_t drawn independently from $\mathcal{D}$, and uses some no-regret learning algorithm to decide which bid to place in round t, based on the outcomes in rounds $1, \ldots, t-1$ (we will make clear exactly what it means for a buyer with changing valuation to play no-regret in Section 2.2, but one can think of v_t as providing a "context" for the bidder during round t). The same mathematical model can also represent a population $\mathcal{D}$ of many indistinguishable buyers with fixed values who each separately no-regret learn - see Section 2.2.3 for further details.

One default strategy for the seller is to simply to set Myerson's revenue-optimal reserve price for $\mathcal{D}$, $r(\mathcal{D})$, in every round (that is, $x_t(b_t) = I(b_t \geq r(\mathcal{D}))$, $p_t(b_t) = r(\mathcal{D}) \cdot I(b_t \geq r(\mathcal{D}))$ for all t, where $I(\cdot)$ is the indicator function). It's not hard to see that *any* no-regret learning algorithm will eventually learn to submit a winning bid during all rounds where $v_t > r(\mathcal{D})$, and a losing bid whenever $v_t < r(\mathcal{D})$. Note that this observation appeals only to the fact that the buyer guarantees no-regret, and makes no reference to any specific algorithm the buyer might use. So if $\mathsf{Rev}(\mathcal{D})$ denotes the expected revenue of the optimal reserve price when a single buyer is drawn from $\mathcal{D}$, the default strategy guarantees the seller revenue $T \cdot \mathsf{Rev}(\mathcal{D}) - o(T)$

[1] Of course, the pricing rule can be implemented by charging $p_t(b)/x_t(b)$ whenever the item is awarded if ex-post individual rationality is desired.

over T rounds. The question then becomes whether or not the seller can beat this benchmark, and if so by how much.

The answer to this question isn't a clear-cut yes or no, so let's start with the following instantiation: how much revenue can the seller extract if the buyer runs EXP3 [16]? In Theorem 2.3.1, we show that the seller can actually do *much* better than the default strategy: it's possible to extract revenue per round equal to (almost) the full expected welfare! That is, if $\mathsf{Val}(\mathcal{D}) = \mathbb{E}_{v \leftarrow \mathcal{D}}[v]$, there exists an auction that extracts revenue $T \cdot \mathsf{Val}(\mathcal{D}) - o(T)$ for all $\mathcal{D}$.[2] It turns out this result holds not only for EXP3, but for any learning algorithm with the following (roughly stated) property: if at time t, the mean reward of action a is significantly larger than the mean reward of action b, the learning algorithm will choose action b with negligible probability. We call a learning algorithm with this property a "mean-based" learning algorithm and note that many commonly used learning algorithms - EXP3, Multiplicative Weights Update [12], and Follow-the-Perturbed-Leader [73, 83, 84] - are 'mean-based' (see Section 2.2 for a formal definition).

We postpone all intuition until Section 2.3.1 with a worked-through example, but just note here that the auction format is quite unnatural: it "lures" the bidder into submitting high bids early on by giving away the item for free, and then charging very high prices (but still bounded in $[0, 1]$) near the end. The transition from "free" to "high-price" is carefully coordinated across different bids to achieve the revenue guarantee.

This result motivates two further directions. First, do there exist other no-regret algorithms for which full surplus extraction is impossible for the seller? In Theorem 2.3.2, we show that the answer is yes. In fact, there is a simple no-regret algorithm $\mathcal{A}$, such that when the bidder uses algorithm $\mathcal{A}$ to bid, the default strategy (set the Myerson reserve every round) is optimal for the seller. We again postpone a

[2]The order of quantifiers in this sentence is correct: it is actually the same auction format that works for all $\mathcal{D}$.

formal statement and intuition to Section 2.3.2, but just note here that the algorithm is a natural adaptation of EXP3 (or in fact, any existing no-regret algorithm) to our setting.

Finally, it is reasonable to expect that bidders might use off-the-shelf no-regret learning algorithms like EXP3, so it is still important to understand what the seller can hope to achieve if the buyer is specifically using such a "mean-based" algorithm (formal definition in Section 2.2). Theorem 2.3.1 is perhaps unsatisfying in this regard because the proposed auction is so unnatural. It turns out that the key property separating natural untruthful auctions (e.g. GSP/GFP) from the unnatural auction above is whether overbidding is a dominated strategy. That is, in our unnatural auction, if the bidder truly hopes to guarantee low regret they must seriously consider overbidding (and this is how the auction lures them into bidding way above their value). In both GSP and GFP, overbidding is dominated, so the bidder can guarantee no regret while overbidding with probability 0 in every round.

The final question we ask is the following: if the buyer is using EXP3 (or any "mean-based" algorithm), never overbids (we call such a bidder *conservative*), how much revenue can the seller extract using an auction where overbidding is dominated in every round? It turns out that the auctioneer can still outperform the default strategy, but not extract full welfare. Instead, we identify a linear program (as a function of $\mathcal{D}$) that tightly characterizes the optimal revenue the seller can achieve in this setting when the buyer's values are drawn from $\mathcal{D}$. Moreover, we show that the auction that achieves this guarantee is natural, and can be thought of as a pay-your-bid auction with decreasing reserves over time. Finally, we show that this "mean-based revenue" benchmark, $\mathsf{MBRev}(\mathcal{D})$ lies truly in between the Myerson revenue and the expected welfare: for all c, there exists a distribution $\mathcal{D}$ over values such that $c \cdot T \cdot \mathsf{Rev}(\mathcal{D}) < \mathsf{MBRev}(\mathcal{D}) < \frac{1}{c} \cdot T \cdot \mathsf{Val}(\mathcal{D})$. In other words, the seller's mean-based revenue may be unboundedly better than the default strategy, yet simultaneously

unboundedly far from the expected welfare. We provide formal statements and a detailed proof overview of these results in Section 2.3.3. To briefly recap, our main results are the following:

1. If the buyer uses a "mean-based" learning algorithm like EXP3, the seller can extract revenue $(1-\varepsilon)T \cdot \mathsf{Val}(\mathcal{D}) - o(T)$ for any constant $\varepsilon > 0$ (Theorem 2.3.1).

2. There exists a natural no-regret algorithm $\mathcal{A}$ such that when the buyer bids according to $\mathcal{A}$, the seller's default strategy (charging the Myerson reserve every round) is optimal (Theorem 2.3.2).

3. If the buyer uses a "mean-based" algorithm only over undominated strategies, the seller can extract revenue $\mathsf{MBRev}(\mathcal{D})$ using an auction where overbidding is dominated in every round. Moreover, we characterize $\mathsf{MBRev}(\mathcal{D})$ as the value of a linear program, and show it can be simultaneously unboundedly better than $T \cdot \mathsf{Rev}(\mathcal{D})$ and unboundedly worse than $T \cdot \mathsf{Val}(\mathcal{D})$ (Theorems 2.3.6, 2.3.4 and 2.3.8).

Our plan for the remaining sections is as follows. Below, we overview our connection to related work. Section 2.2 formally defines our model. Section 2.3 works through a concrete example, providing intuition for all three results. Section 2.4 discusses conclusions and open problems.

2.1.1 Related Work

There are two lines of work that are most related to ours. The first is that of *dynamic auctions*, such as [116, 14, 106, 107, 98]. Like our model, there are T rounds where the seller has a single item for sale to a single buyer, whose value is drawn from some distribution every round. However, the buyer is fully strategic and processes fully how their choices today affect the seller's decisions tomorrow (e.g. they engage with deals of the form "pay today to get the item tomorrow"). Additional closely related

work is that of Devanur et al. studying the Fishmonger problem [55, 75]. Here, there is again a single buyer and seller, and T rounds of sale. Unlike our model, the buyer draws a value from $\mathcal{D}$ once during round 0 and that value is fixed through all T rounds (so the seller could try to learn the buyer's value over time). Also unlike our model, they study perfect Bayesian equilibria (where again the buyer is fully strategic, and reasons about how their actions today affect the seller's behavior tomorrow).

In contrast to these works, while buyers in our model do care about the future (e.g. they value learning), they don't reason about how their actions today might affect the seller's decisions tomorrow. Our model better captures settings where full information about the auction is not public (and fully strategic reasoning is simply impossible without the necessary information).

Other related work considers the *Price of Anarchy* of simple combinatorial auctions when bidders no-regret learn [123, 139, 115, 50]. One key difference between this line of work and ours is that these all study welfare maximization for combinatorial auctions with rich valuation functions. In contrast, our work studies revenue maximization while selling a single item. Additionally, in these works the seller commits to a publicly known auction format, and the only reason for learning is due to the strategic behavior of other buyers. In contrast, buyers in our model have to learn *even when they are the only buyer*, due to the strategic nature of the seller.

Recent work has also considered learning from the perspective of the seller. In these works, the buyer's (or buyers') valuations are drawn from an unknown distribution, and the seller's goal is to learn an approximately optimal auction with as few samples as possible [45, 53, 108, 109, 69, 35, 56]. These works consider numerous different models and achieve a wide range of guarantees, but all study the learning problem from the perspective of the *seller*, whereas the buyer is simply myopic and participates in only one round. In contrast, it is the buyer in our model who does the

learning (and there is no information for the seller to learn: the buyer's values are drawn fresh in every round).

Finally, no-regret learning in online decision problems is an extremely well-studied problem. When feedback is revealed for every possible action, one well-known solution is the multiplicative weight update rule which has been rediscovered and applied in many fields (see survey [12] for more details). Another algorithmic scheme for the online decision problem is known as Follow the Perturbed Leader [73, 83, 84]. When only feedback for the selected action is revealed, the problem is referred to as the multi-armed bandit problem. Here, similar ideas to the MWU rule are used in developing the EXP3 algorithm [16] for adversarial bandit model, and also for the contextual bandit problem [95]. Our algorithm in Theorem 2.3.2 bears some similarities to the low swap regret algorithm introduced in [24]. See the survey [34] for more details about the multi-armed bandit problem. Our results hold in both models (i.e. whether the buyer receives feedback for every bid they could have made, or only the bid they actually make), so we will make use of both classes of algorithms.

In summary, while there is already extensive work related to repeated sales in auctions, and even no-regret learning with respect to auctions (from both the buyer and seller perspective), our work is the first to address how a seller might adapt their selling strategy when faced with a no-regret buyer.

2.2 Model and Preliminaries

We consider a setting with 1 buyer and 1 seller. There are T rounds, and in each round the seller has one item for sale. At the start of each round t, the buyer's value $v(t)$ (known only to the buyer) for the item is drawn independently from some distribution $\mathcal{D}$ (known to both the seller and the buyer). For simplicity, we assume

$\mathcal{D}$ has a finite support[3] of size m, supported on values $0 \leq v_1 < v_2 < \cdots < v_m \leq 1$. For each $i \in [m]$, v_i has probability q_i of being drawn under $\mathcal{D}$.

The seller then presents K options for the buyer, which can be thought of as "possible bids" (we will interchangeably refer to these as *options*, *bids*, or *arms* throughout this chapter, depending on context). Each arm i is labelled with a bid value $b_i \in [0, 1]$, with $b_1 < \ldots, < b_K$. Upon pulling this arm at round t, the buyer receives the item with some allocation probability $a_{i,t}$, and must pay a price $p_{i,t} \in [0, a_{i,t} \cdot b_i]$. These values $a_{i,t}$ and $p_{i,t}$ are chosen by the seller during time t, but remain unknown to the buyer until he plays an arm, upon which he learns the values for that arm. All of our positive results (i.e. strategies for the seller) are *non-adaptive* (in some places called *oblivious*), in the sense that that $a_{i,t}, p_{i,t}$ are set before the first round starts. All of our negative results (i.e. upper bounds on how much a seller can possibly attain) hold even against *fully adaptive* sellers, where $a_{i,t}$ and $p_{i,t}$ can be set *even after learning the distribution of arms the buyer intends to pull in round t.*

In order for the selling strategies to possibly represent natural auctions, we require the allocation/price rules to be monotone. That is, if $i > j$, then for all t, $a_{i,t} \geq a_{j,t}$ and $p_{i,t} \geq p_{j,t}$. In other words, bidding higher should result in a (weakly) higher probability of receiving the item and (weakly) higher expected payment. We'll also insist on the existence of an arm 0 with bid $b_0 = 0$ and $a_{0,t} = 0$ for all t; i.e., an arm which charges nothing but does not give the item. Playing this arm can be thought of as not participating in the auction.

2.2.1 Bandits and experts

Our goal is to understand the behavior of such mechanisms when the buyer plays according to some no-regret strategy for the multi-armed bandit problem. In the

[3]If $\mathcal{D}$ instead has infinite support, all our results hold approximately after discretization to multiples of ε. If $\mathcal{D}$ is bounded in $[0, H]$, then all our results hold after normalizing $\mathcal{D}$ by dividing by H.

classic multi-armed bandit problem a learner (in our case, the buyer) chooses one of K arms per round, over T rounds. On round t, the learner receives a reward $r_{i,t} \in [0,1]$ for pulling arm i (where the values $r_{i,t}$ are possibly chosen adversarially). The learner's goal is to maximize his total reward.

Let I_t denote the arm pulled by the principal at round t. The *regret* of an algorithm $\mathcal{A}$ for the learner is the random variable $\mathsf{Reg}(\mathcal{A}) = \max_i \sum_{t=1}^{T} r_{i,t} - \sum_{t=1}^{T} r_{I_t,t}$. We say an algorithm $\mathcal{A}$ for the multi-armed bandit problem is *δ-no-regret* if $\mathbb{E}[\mathsf{Reg}(\mathcal{A})] \leq \delta$ (where the expectation is taken over the randomness of $\mathcal{A}$). We say an algorithm $\mathcal{A}$ is *no-regret* if it is δ-no-regret for some $\delta = o(T)$.

In the multi-armed bandits setting, the learner only learns the value $r_{i,t}$ for the arm i which he pulls on round t. In our setting, the learner will learn $a_{i,t}$ and $p_{i,t}$ explicitly (from which they can compute $r_{i,t}$). Our results (both positive and negative) also hold when the learner learns the value $r_{i,t}$ for *all* arms i (we refer this full-information setting as the *experts setting*, in contrast to the partial-information *bandits setting*). Simple no-regret algorithms exist in both the experts setting and the bandits setting. Of special interest in this chapter will be a class of learning algorithms for the bandits problem and experts problem which we term 'mean-based'.

Definition 2.2.1 (Mean-Based Learning Algorithm)**.** *Let $\sigma_{i,t} = \sum_{s=1}^{t} r_{i,s}$. An algorithm for the experts problem or multi-armed bandits problem is γ-mean-based if it is the case that whenever $\sigma_{i,t} < \sigma_{j,t} - \gamma T$, then the probability that the algorithm pulls arm i on round t is at most γ. We say an algorithm is mean-based if it is γ-mean-based for some $\gamma = o(1)$.*

Intuitively, 'mean-based' algorithms will rarely pick an arm whose current mean is significantly worse than the current best mean. Many no-regret algorithms, including commonly used variants of EXP3 (for the bandits setting), the Multiplicative Weights algorithm (for the experts setting) and the Follow-the-Perturbed-Leader algorithm (experts setting), are mean-based (Appendix A.4).

Contextual bandits

In our setting, the buyer has the additional information of their current value for the item, and hence is actually facing a *contextual bandits* problem. In (our variant of) the contextual bandits problem, each round t the learner is additionally provided with a *context* c_t drawn from some distribution $\mathcal{D}$ supported on a finite set C (in our setting, $c_t = v(t)$, the buyer's valuation for the item at time t). The adversary now specifies rewards $r_{i,t}(c)$, the reward the learner receives if he pulls arm i on round t while having context c. If we are in the full-information (experts) setting, the learner learns the values of $r_{i,t}(c_t)$ for all arms i after round t, where as if we are in the partial-information (bandits) setting, the learner only learns the value of $r_{i,t}(c_t)$ for the arm i that he pulled.

In the contextual bandits setting, we now define the regret of an algorithm $\mathcal{A}$ in terms of regret against the best "context-specific" policy π; that is, $\mathsf{Reg}(\mathcal{A}) = \max_{\pi:C\to[K]} \sum_{t=1}^{T} r_{\pi(c_t),t}(c_t) - \sum_{t=1}^{T} r_{I_t,t}(c_t)$, where again I_t is the arm pulled by M on round t. As before, we say an algorithm is δ-low regret if $\mathbb{E}[\mathsf{Reg}(M)] \leq \delta$, and say an algorithm is no-regret if it is δ-no-regret for some $\delta = o(T)$.

If the size of the context set C is constant with respect to T, then there is a simple way to construct a no-regret algorithm M' for the contextual bandits problem from a no-regret algorithm M for the classic bandits problem: simply maintain a separate instance of M for every different context $v \in C$ (in the contextual bandits literature, this is sometimes referred to as the S-EXP3 algorithm [34]). We call the algorithm we obtain this way its *contextualization*, and denote it as $\mathsf{cont}(M)$.

If we start with a mean-based learning algorithm, then we can show that its contextualization satisfies an analogue of the mean-based property for the contextual-bandits problem (proof in Appendix A.4).

Definition 2.2.2 (Mean-Based Contextual Learning Algorithm)**.** *Let* $\sigma_{i,t}(c) = \sum_{s=1}^{t} r_{i,s}(c)$. *An algorithm for the contextual bandits problem is* γ*-mean-based if it*

is the case that whenever $\sigma_{i,t}(c) < \sigma_{j,t}(c) - \gamma T$*, then the probability* $p_{i,t}(c)$ *that the algorithm pulls arm* i *on round* t *if it has context* c *satisfying* $p_{i,t}(c) < \gamma$*. We say an algorithm is mean-based if it is* γ*-mean-based for some* $\gamma = o(1)$*.*

Theorem 2.2.3. *If an algorithm for the experts problem or multi-armed bandits problem is mean-based, then its contextualization is also a mean-based algorithm for the contextual bandits problem.*

Finally, we will refer to learning algorithms that never overbid as *conservative*. We will sometimes abuse notation and instead refer to a buyer employing a conservative algorithm as conservative.

2.2.2 Welfare and monopoly revenue

In order to evaluate the performance of our mechanisms for the seller, we will compare the revenue the seller obtains to two benchmarks from the single-round setting of a seller selling a single item to a buyer with value drawn from distribution $\mathcal{D}$.

The first benchmark we consider is the *welfare* of the buyer, the expected value the buyer assigns to the item. This quantity clearly upper bounds the expected revenue that the seller can hope to extract per round.

Definition 2.2.4. *The* welfare*,* $\mathsf{Val}(\mathcal{D})$ *is equal to* $\mathbb{E}_{v\sim\mathcal{D}}[v]$*.*

The second benchmark we consider is the *monopoly revenue*, the maximum possible revenue attainable by the seller in one round against a rational buyer. Seminal work of Myerson [112] shows that this revenue is attainable by setting a fixed price ("monopoly/Myerson reserve") for the item, and hence can be characterized as follows.

Definition 2.2.5. *The* monopoly revenue *(alternatively,* Myerson revenue*)* $\mathsf{Mye}(\mathcal{D})$ *is equal to* $\max_p p \cdot \Pr_{v\sim\mathcal{D}}[v \geq p]$*.*

2.2.3 A final note on the model

For concreteness, we chose to phrase our problem as one where a single bidder whose value is repeatedly drawn independently from $\mathcal{D}$ each round engages in no-regret learning with their value as context. Alternatively, we could imagine a population of m different buyers, each with a *fixed* value v_i. Each round, exactly one buyer arrives at the auction, and it is buyer i with probability q_i. The buyers are indistinguishable to the seller, and each buyer no-regret learns (without context, because their value is always v_i). This model is mathematically equivalent to ours, so all of our results hold in this model as well if the reader prefers this interpretation instead.

2.3 An Illustrative Example

In this section, we overview an illustrative example to show the difference between mean-based and non-mean-based learning algorithms, and between conservative and non-conservative learners. We will not prove all claims in this section (nor carry out all calculations) as it is only meant to illustrate and provide intuition. Throughout this section, the running example will be when $\mathcal{D}$ samples $1/4$ with probability $1/2$, $1/2$ with probability $1/4$, and 1 with probability $1/4$. Note that $\mathsf{Val}(\mathcal{D}) = 1/2$ and $\mathsf{Rev}(\mathcal{D}) = 1/4$.

2.3.1 Mean-Based Learning

Let's first consider what the seller can do with an auction when the buyer is running a mean-based (non-conservative) learning algorithm like EXP3. The seller will let the buyer bid 0 or 1. If the buyer bids 0, they pay nothing but do not receive the item (recall that an arm of this form is required). If the buyer bids 1 in round t, they receive the item and pay some price p_t as follows: for the first half of the game

$(1 \leq t \leq T/2)$, the seller sets $p_t = 0$. For the second half of the game $(T/2 < t \leq T)$, the seller sets $p_t = 1$.

Let's examine the behaviour of the buyer, recalling that they run a mean-based learning algorithm, and therefore (almost) always pull the arm with highest cumulative utility. The buyer with value 1 will happily bid 1 all the way through, since he is always offered the item for less than or equal to his value for the item. The buyer with value $1/2$ will bid 1 for the first $T/2$ rounds, accumulating a surplus (i.e., negative regret) of $1/2$ per round. For the next $T/2$ rounds, this surplus slowly disappears at the rate of $1/2$ per round until it disappears at time T, so the bidder with value $1/2$ will bid 1 all the way through. Finally, the bidder with value $1/4$ will bid 1 for the first $T/2$ rounds, accumulating surplus at a rate of $1/4$ per round. After round $T/2$, this surplus decreases at a rate of $3/4$ per round, until at round $2T/3$ his cumulative utility from bidding 1 reaches 0 and he switches to bidding 0.

Now let's compute the revenue. From round $T/2$ through $2T/3$, the buyer always buys the item at a price of 1, so the seller obtains $T/6$ revenue. Finally, from round $2T/3$ through T, the buyer purchases the item with probability $1/2$ and pays 1. The total revenue is $0 + T/6 + T/6 = T/3$. Note that if the seller used the default strategy, they would extract revenue only $T/4$.

Where did our extra revenue come from? First, note that the welfare of the buyer in this example is quite high: the bidder gets the item the whole way through when $v \geq 1/2$, and two-thirds of the way through when $v = 1/4$. One reason why the welfare is so high is because we give the item away for free in the early rounds. But notice also that the utility of the buyer is quite low: the buyer actually has zero utility when $v \leq 1/2$, and utility $1/2$ when $v = 1$. The reason we're able to keep the utility low, despite giving the item away for free in the early rounds is because we overcharge the bidders in later rounds (and they choose to overpay, exactly because their learning is mean-based).

In fact, by offering additional options to the buyer, we show that *it is possible for the seller to extract up to the full welfare from the buyer* (e.g. a net revenue of $T/2 - o(T)$ for this example). As in the above example, our mechanism makes use of arms which are initially very good for the buyer (giving the item away for free, accumulating negative regret), followed by a period where they are very bad for the buyer (where they pay more than their value). The trick in the construction is making sure that the good/bad intervals line up so that: a) the buyer purchases the item in every round, no matter their value (this is necessary in order to possibly extract full welfare) and b) by round T, the buyer has zero (arbitrarily small) utility, no matter their value.

Getting the intervals to line up properly so that any mean-based learner will pick the desired arms still requires some work. But interestingly, our constructed mechanism is non-adaptive and prior-independent (i.e. the same mechanism extracts full welfare *for all* $\mathcal{D}$). Theorem 2.3.1 below formally states the guarantees. The construction itself and the proof appear in Appendix A.2.

Theorem 2.3.1. *If the buyer is running a mean-based algorithm, for any constant* $\varepsilon > 0$*, there exists a strategy for the seller which obtains revenue at least* $(1-\varepsilon)\mathsf{Val}(\mathcal{D})T - o(T)$*.*

Two properties should jump out as key in enabling the result above. The first is that the buyer *only* has no regret towards fixed arms and *not* towards the policy they would have used with a lower value (this is what leads the buyer to continue bidding 1 with value 1/2 even though they have already learned to bid 0 with value 1/4). This suggests an avenue towards an improved learning algorithm: have the bidder attempt to have no regret not only towards each fixed arm, but also towards the policy of play produced when having different values. This turns out to be exactly the right idea, and is discussed in the following subsection below.

The second key property is that we were able to "lure" the bidders into playing an arm with a free item, then overcharge them later to make up for lost revenue. This requires that the bidder consider pulling an arm with maximum bid exceeding their value, which will never happen for a conservative bidder. It turns out it is still possible to do better than the default strategy against conservative bidders, but not as well as against non-conservative mean-based bidders. Section 2.3.3 explores conservative mean-based bidders for this example.

2.3.2 Better Learning

In our bad example above, the buyer with value 1/2 for the item slowly spends the second half of the game losing utility. While his behaviour is still no-regret (he ends up with zero net utility, which indeed is at least as good as only bidding 0), he would have been much happier to follow the actions of the buyer with value 1/4, who started bidding 0 at $2T/3$.

Using this idea, we show how to construct a no-regret algorithm for the buyer (Algorithm 1) such that the seller receives at most the Myerson revenue every round. We accomplish this by extending an arbitrary no-regret algorithm (e.g. EXP3) by introducing "virtual arms" for each value, so that each buyer with value v has low regret not just with respect to every fixed bid, but also no-regret with respect to the policy of play as if they had a different value v' for the item (for all $v' < v$). In some ways, our construction is very similar to the construction of low internal-regret (or swap-regret) algorithms from low external-regret algorithms. The main difference is that instead of having low regret with respect to swapping actions, we have low regret with respect to swapping *contexts* (i.e. values). Theorem 2.3.2 below states that the seller cannot outperform the default strategy against buyers who use such algorithms to learn.

Theorem 2.3.2. *There exists a no-regret algorithm (Algorithm 1) for the buyer against which every seller strategy extracts no more than* $\mathsf{Mye}(\mathcal{D})T + O(m\sqrt{\delta T})$ *revenue.*

Algorithm 1 No-regret algorithm for buyer where the seller achieves no more than $\mathsf{Mye}(\mathcal{D})T + o(T)$ revenue.

Let M be a δ-no-regret algorithm for the classic multi-armed bandit problem, with $\delta = o(T)$. Initialize m copies of M, M_1 through M_m.
Instance M_i of M will learn over $K + i - 1$ arms.
The first K arms of M_i ("bid arms") correspond to the K possible menu options $b_1, b_2, \ldots, b_K$.
The last $i - 1$ arms of M_i ("value arms") correspond to the $i - 1$ possible values (contexts) $v_1, \ldots, v_{i-1}$.
for $t = 1$ to T **do**
  if buyer has value v_i **then**
    Use M_i to pick one arm from the $K + i - 1$ arms.
    if the arm is a bid arm b_j **then**
      Pick the menu option j (i.e. bid b_j).
    else if the arm is a value arm v_j **then**
      Sample an arm from M_j (but don't update its state). If it is a bid arm, pick the corresponding menu option. If it is a value arm, recurse.
    end if
    Update the state of algorithm M_i with the utility of this round.
  end if
end for

A more further discussion of the algorithm along with a proof of Theorem 2.3.2 appear in Appendix A.1. The key observation in the proof is that "not regretting playing as if my value were v'" sounds a lot like "not preferring to report value v' instead of v." This suggests that the aggregate allocation probabilities and prices paid by any buyer using our algorithm should satisfy the same constraints as a truthful auction, proving that the resulting revenue cannot exceed the default strategy (and indeed the proof follows this approach).

Finally, observe that the following corollary immediately follows. Because the seller cannot hope to get more than $\mathsf{Mye}(\mathcal{D})T + o(T)$ per round when the buyer is using Algorithm 1, and the buyer cannot hope to do better than telling the truth

against a truthful auction, it is in fact a Nash for the buyer to use Algorithm 1 and the seller to set price equal to the Myerson reserve every round.

Corollary 2.3.3. *It is an $o(T)$-Nash equilibrium for the seller to set the Myerson reserve $p(\mathcal{D})$ in every round (any bid $\geq p(\mathcal{D})$ reserve wins the item and pays $p(\mathcal{D})$), and the buyer to use Algorithm 1.*

2.3.3 Mean-Based Learning and Conservative Bidders

Recall in our example that to extract revenue $T/3$, bidders with values $1/4$ and $1/2$ had to consider bidding 1. If bidders are conservative, they will simply never do this.

Although the auction in Section 2.3.1 is no longer viable, consider the following auction instead: in addition to the zero arm, the bidder can bid $1/4$ or $1/2$. If they bid $1/2$ in any round, they will get the item with probability 1 and pay $1/2$. If they bid $1/4$ in round $t \leq T/3$, they get nothing. If they bid $1/4$ in round $t \in (T/3, T]$, they get the item and pay $1/4$. Let's again see what the bidder will choose to do, remembering that they will always pull the arm that has provided highest cumulative utility (due to being mean-based).

Clearly, the bidder with value $1/4$ will bid $1/4$ every round (since they are conservative, they won't even consider bidding $1/2$), making a total payment of $2T/3 \cdot 1/4 \cdot 1/2 = T/12$. The bidder with value $1/2$ will bid $1/2$ for the first $T/3$ rounds, and then immediately switch to bidding $1/4$, making a total payment of $T/3 \cdot 1/2 \cdot 1/4 + 2T/3 \cdot 1/4 \cdot 1/4 = T/12$.

The bidder with value 1 will actually bid $1/2$ for the entire T rounds. To see this, observe that their cumulative surplus through round t from bidding $1/2$ is $t \cdot 1/2 \cdot 1/4 = t/8$ (t rounds by utility $1/2$ per round by probability $1/4$ of having value 1). Their cumulative surplus through round t from bidding $1/4$ is instead $(t - T/3) \cdot 3/4 \cdot 1/4 = 3t/16 - T/16 \leq t/8$ (for $t \leq T$). Because they are mean-based, they will indeed bid $1/2$ for the entire duration due to its strictly higher utility. So their total payment will

be $T \cdot 1/2 \cdot 1/4 = T/8$. The total revenue is then $7T/24 > T/4$, again surpassing the default strategy (but not reaching the $T/3$ achieved against non-conservative buyers).

Let's again see where our extra revenue comes from in comparison to a truthful auction. Notice that the bidder receives the item with probability 1 conditioned on having value $1/2$, and also conditioned on having value 1. Yet somehow the bidder pays an average of $1/3$ conditioned on having value $1/2$, but an average of $1/2$ conditioned on having value 1. *This could never happen in a truthful auction*, as the bidder would strictly prefer to pretend their value was $1/2$ rather than 1. But it is entirely possible when the buyer does mean-based learning, as evidenced by this example.

In Appendix A.3, we define $\mathsf{MBRev}(\mathcal{D})$ as the value of the LP in Figure 2.1. In Theorems 2.3.6 and 2.3.4, we show that $\mathsf{MBRev}(\mathcal{D})T$ tightly characterizes (up to $\pm o(T)$) the optimal revenue a seller can extract against a conservative buyer. The proofs can be found in Appendix A.3.1.

$$\begin{aligned} \textbf{maximize} \quad & \sum_{i=1}^{m} q_i(v_i x_i - u_i) \\ \textbf{subject to} \quad & u_i \geq (v_i - v_j) \cdot x_j, \quad \forall\, i, j \in [m] : i > j \\ & u_i \geq 0, 1 \geq x_i \geq 0, \quad \forall\, i \in [m] \end{aligned}$$

Figure 2.1: The mean-based revenue LP.

Before stating our theorems, let us parse this LP. q_i is a constant representing the probability that the buyer has value v_i (also a constant). x_i is a variable representing the average probability that the bidder gets the item with value v_i, and u_i is a variable representing the average utility of the bidder when having value v_i. Therefore, this bidder's average value is $v_i x_i$, the average price they pay is $v_i x_i - u_i$, and the objective function is simply the average revenue. The second constraints are just normalization, ensuring that everything lies in $[0, 1]$. The first line of constraints are

the interesting ones. These look a lot like IC constraints that a truthful auction must satisfy, but something's missing: the LHS is clearly the utility of the buyer with value v_i for "telling the truth," but the utility of the buyer for "reporting v_j instead" is $(v_i - v_j) \cdot x_j + u_j$ (so the u_j term is missing on the RHS).

Here is a brief proof outline for why no seller can extract more revenue than $\mathsf{MBRev}(\mathcal{D})$:

1. Since the buyer has no regret conditioned on having value v_i, their utility is at least as high as playing arm j every round, for all $j \leq i$.

2. Since the auction never charges arm j more than v_j (conditioned on awarding the item), the buyer's utility for playing arm j every round is at least $y_j \cdot (v_i - v_j)$, where y_j is the average probability that arm j awards the item.

3. Since the auction is monotone, and the buyer never considers overbidding, if the buyer gets the item with probability x_j conditioned on having value v_j, we must have $y_j \geq x_j$.

These three facts together show that no seller can extract more than $\mathsf{MBRev}(\mathcal{D})$ against a no-regret buyer who doesn't overbid. Observe also that step 3 is *exactly* the step that doesn't hold for buyers who consider overbidding (and is exactly what's violated in our example in Section 2.3.1): if the buyer ever overbids, then they might receive the item with higher probability than had they just played their own arm every round.

Theorem 2.3.4. *Any strategy for the seller achieves revenue at most* $\mathsf{MBRev}(\mathcal{D})T + o(T)$ *against a conservative buyer.*

The full proof of Theorem 2.3.4 appears in the appendix - all of the key ideas have been overviewed above.

It turns out that the previous theorem is tight; there exists an auction (taking the form of a first-price auction with descending reserve) which achieves revenue

$\mathsf{MBRev}(\mathcal{D})T$ against a conservative mean-based buyer. More specifically, this auction is defined by a threshold r_t that decreases over time. If at time t you bid $b_t \geq r_t$, then you receive the item and must pay b_t; otherwise, you receive nothing and pay nothing. Moreover, the threshold function r_t which achieves optimal revenue is determined from the optimal solution to the mean-based LP: the threshold r_t drops from v_i to v_{i+1} at round x_i (where the x_i belong to some optimal solution).

To show that this is a valid strategy for the seller, we need to show that the values x_i are monotone increasing. Luckily, this follows simply from the structure of the mean-based revenue LP.

Lemma 2.3.5. *Let $x_1, x_2, \ldots, x_m, u_1, u_2, \ldots, u_m$ be an optimal solution to the mean-based revenue LP. Then for all $i < j$, $x_i < x_j$.*

Proof. We proceed by contradiction. Suppose that the sequence of x_i are not monotone; then there exists an $1 \leq i \leq m-1$ such that $x_i > x_{i+1}$. Now consider another solution of the LP, where we increase x_{i+1} to x_i, keeping the value of all other variables the same. This new solution does not violate any constraints in the LP since for all $j > i+1$, $u_j \geq (v_j - v_i) \cdot x_i \geq (v_j - v_{i+1}) \cdot x_i$. However this change increases the value of the objective by $v_{i+1}q_{i+1}(x_i - x_{i+1}) > 0$, thus contradicting the fact that $x_1, \ldots, x_m, u_1, ..., u_m$ was an optimal solution of the mean-based revenue LP. □

We now show that this strategy indeed achieves $\mathsf{MBRev}(\mathcal{D})T$ against a conservative buyer.

Theorem 2.3.6. *For any constant $\varepsilon > 0$, there exists a strategy for the seller gets revenue at least $(\mathsf{MBRev}(\mathcal{D}) - \varepsilon)T - o(T)$ against a buyer running a mean-based algorithm who overbids with probability 0. The strategy sets a decreasing cutoff r_t and for all t awards the item with probability 1 to any bid $b_t \geq r_t$ for price b_t, and with probability 0 to any bid $b_t < r_t$.*

Proof. We will show that: i) the buyer with value v_i receives the item for at least $x_iT - o(T)$ turns (receiving $v_ix_iT - o(T)$ total utility from the items), and ii) this buyer's net utility is at most $(u_i + \varepsilon)T + o(T)$. This implies that this buyer pays the seller at least $x_iv_iT - (u_i + \varepsilon)T - o(T)$ over the course of the T rounds; taking expectation over all v_i completes the proof.

Assume the buyer is running a γ-mean-based learning algorithm. Consider the buyer when they have value v_i. Note that

$$\sigma_{j,t}(v_i) = (v_i - v_j + \varepsilon) \cdot \max(0, t - (1 - x_j)T).$$

We first claim that after round $(1-x_i)T+\gamma T/\varepsilon$, the buyer will buy the item (i.e., choose an option that results in him getting the item) each round with probability at least $1 - m\gamma$. To see this, first note that $\sigma_{i,t}(v_i) \geq \gamma T$ when $t \geq (1 - x_i)T + \gamma T/\varepsilon$. Then, since the cumulative utility of any arm is 0 until it starts offering the item, it follows from the mean-based condition that the buyer will pick a specific arm that is not offering the item with probability at most γ, and therefore choose some good arm with probability at least $1 - m\gamma$. It follows that, in expectation, the buyer with value v_i receives the item for at least $(1 - m\gamma)(x_iT - \gamma T/\varepsilon) = x_iT - o(T)$ turns.

We now proceed to upper bound the overall expected utility of the buyer. For each index $j \leq i$, let S_j be the set of t where $\sigma_{j,t}(v_i) > \sigma_{j',t}(v_i)$ for all other j'. Note that since each $\sigma_{j,t}(v_i)$ is a linear function in t (when positive), each S_j is either the empty set or an interval (y_jT, z_jT). Since all the v_i are distinct, note that these intervals partition the interval $((1-x_i)T, T)$ (with the exception of up to m endpoints of these intervals); in particular, $\sum_{j\geq i}(z_j - y_j) = x_i$.

Let $\varepsilon' = \min_j(v_{j+1} - v_j)$. Note that, if $t \in (y_jT + \gamma T/\varepsilon', z_jT - \gamma T/\varepsilon')$, then for all $j' \neq j$, $\sigma_{j,t}(v_i) > \sigma_{j',t}(v_i) + \gamma T$. This follows since $\sigma_{j,t}(v_i) - \sigma_{j',t}(v_i)$ is linear in t with slope $v_j - v_{j'}$, and $|v_j - v_{j'}| > \varepsilon'$. It follows that if t is in this interval, then the

buyer will choose option j with probability at least $1 - m\gamma$ (by a similar argument as before).

Define $j(t) = \arg\max_j \sigma_{j,t}(v_i)$ to be the index of the arm with the current largest cumulative reward, and let $\sigma_{max,t}(v_i) = \sum_{s=1}^{t} r_{j(s),s}(v_i)$ be the cumulative utility of always playing the arm with the current highest cumulative reward for the first t rounds. The following lemma shows that $\sigma_{max,T}(v_i)$ is close to $\max_j \sigma_{j,T}(v_i)$. (In other words, playing the best arm every round and playing the best-at-the-end arm every round have similar payoffs if the historically best arm does not change often).

Lemma 2.3.7. $|\sigma_{max,T}(v_i) - \max_j \sigma_{j,T}(v_i)| \leq m$.

Proof. Let $W = |\{t|j(t) \neq j(t+1)\}|$ equal the number of times the best arm switches values; note that since each $\sigma_{j,t}(v_i)$ is linear, W is at most m. Let $t_1 < t_2 < \cdots < t_W$ be the values of t such that $j(t) \neq j(t+1)$. Additionally define $t_0 = 1$ and $t_{W+1} = T$. Then, dividing the cumulative reward $\sigma_{max,t}$ into intervals by these t_i, we get that

$$\begin{aligned}
\sigma_{max,t}(v_i) &= \sum_{s=1}^{t} r_{j(s),s}(v_i) \\
&= \sum_{i=1}^{W+1} (\sigma_{j(t_i),t_i}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) \\
&= \sigma_{j(T),T}(v_i) + \sum_{i=1}^{W+1} (\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) \\
&= \max_j \sigma_{j,t}(v_i) + \sum_{i=1}^{W+1} (\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i))
\end{aligned}$$

It therefore suffices to show that $|\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)| \leq 1$ for all i. To see this, note that (by the definition of $j(t)$), $\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i) > 0$, and that $\sigma_{j(t_{i-1}),t_{i-1}+1}(v_i) - \sigma_{j(t_i),t_{i-1}+1}(v_i) < 0$. However,

$$(\sigma_{j(t_{i-1}),t_{i-1}+1}(v_i) - \sigma_{j(t_i),t_{i-1}+1}(v_i)) =$$

$$(\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) + (r_{j(t_{i-1}),t_{i-1}+1}(v_i) - r_{j(t_i),t_{i-1}+1}(v_i)) \quad (2.1)$$

Since $0 \leq r_{j,t}(u) \leq 1$, it follows that $|\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)| \leq 1$. This completes the proof. □

Let $\sigma_T(v_i) = \sum_{t=1}^{T} \mathbb{E}[r_{I_t,t}(v_i)]$ denote the expected cumulative utility of this buyer at time T. We claim that $\sigma_T \leq \max_j \sigma_{j,T}(v_i) + o(T)$. To see this, recall that, for $t \in (y_j T + \gamma T/\varepsilon', z_j T - \gamma T/\varepsilon')$, $\Pr[I_t \neq j] \leq m\gamma$, and therefore $\mathbb{E}[r_{I_t,t}] \leq r_{j,t} + m\gamma$. Furthermore, note that for $t \in S_j$, $j(t) = j$, so $r_{j,t} = r_{j(t),t}$ and $\mathbb{E}[r_{I_t,t}] \leq r_{j(t),t} + m\gamma$. It follows that

$$
\begin{aligned}
\sigma_T(v_i) &= \sum_{t=1}^{T} \mathbb{E}[r_{I_t,t}(v_i)] \\
&\leq \sum_{t=(1-x_i)T}^{T} \mathbb{E}[r_{I_t,t}(v_i)] \\
&= \sum_{j=1}^{i} \sum_{t=y_jT}^{z_jT} \mathbb{E}[r_{I_t,t}(v_i)] \\
&\leq \sum_{j=1}^{i} \left(\frac{2\gamma T}{\varepsilon'} + \sum_{t=y_jT+\gamma T/\varepsilon'}^{z_jT-\gamma T/\varepsilon'} \mathbb{E}[r_{I_t,t}(v_i)] \right) \\
&\leq \sum_{j=1}^{i} \left(\frac{2\gamma T}{\varepsilon'} + \sum_{t=y_jT+\gamma T/\varepsilon'}^{z_jT-\gamma T/\varepsilon'} (r_{j(t),t}(v_i) + m\gamma) \right) \\
&\leq \frac{2m\gamma T}{\varepsilon'} + m\gamma T + \sum_{t=1}^{T} r_{j(t),t}(v_i) \\
&= \frac{2m\gamma T}{\varepsilon'} + m\gamma T + \sigma_{max,T}(v_i) \\
&\leq \frac{2m\gamma T}{\varepsilon'} + m\gamma T + m + \max_j \sigma_{j,T}(v_i) \\
&= \max_j \sigma_{j,T}(v_i) + o(T).
\end{aligned}
$$

Finally, note that

$$
\begin{aligned}
\max_j \sigma_{j,T}(v_i) &= \max_{j<i}(v_i - v_j + \varepsilon)x_j T \\
&\leq (\max_{j<i}(v_i - v_j)x_j + \varepsilon)T \\
&= (u_i + \varepsilon)T
\end{aligned}
$$

It follows that $\sigma_T(v_i) \leq (u_i + \varepsilon)T + o(T)$, as desired.

□

Finally, we show that this quantity $\mathsf{MBRev}(\mathcal{D})$ is in fact significantly different from both $\mathsf{Val}(\mathcal{D})$ and $\mathsf{Rev}(\mathcal{D})$; in particular, it is a constant-factor approximation to neither. In particular, the multiplicative gap between $\mathsf{MBRev}(\mathcal{D})$ and $\mathsf{Rev}(\mathcal{D})$ can grow as large as $\log \log H$ for distributions $\mathcal{D}$ supported on $[1, H]$. In comparison, the gap between $\mathsf{Val}(\mathcal{D})$ and $\mathsf{Rev}(\mathcal{D})$ can grow as large as $\log H$ on this same interval, and in fact both gaps are maximized for the same distribution: the equal-revenue curve $\mathcal{D}_{ERC}$ truncated at H.

Theorem 2.3.8. *For distributions $\mathcal{D}$ supported on $[1, H]$, $\mathsf{MBRev}(\mathcal{D}) = O(\log \log H)$, and there exist $\mathcal{D}$ supported on $[1, H]$ such that $\mathsf{MBRev}(\mathcal{D}) = \Theta(\log \log H)$. For this same $\mathcal{D}$, $\mathsf{Val}(\mathcal{D}) = \Theta(\log H)$.*

The proof of Theorem 2.3.8 is included in Appendix A.3. The proof is divided into two parts (after extending the definition of $\mathsf{MBRev}(\mathcal{D})$ to hold for continuous distributions $\mathcal{D}$): 1. showing that $\mathsf{MBRev}(\mathcal{D}_{ERC}) \leq O(\log \log H)$, and 2. showing that $\mathsf{MBRev}(\mathcal{D}_{ERC}) \geq O(\log \log H)$.

To show the first part, it suffices to simply demonstrate a solution to the mean-based LP with value at least $O(\log \log H)$. We see in Theorem A.3.11 that it suffices to choose $x(v) = \frac{\log v}{\log H}$ (equivalently, the reserve for the associated second-price auction should exponentially decay over time).

To show the second part, we examine the dual of the LP. Effectively, this involves rewriting $\mathsf{MBRev}(\mathcal{D})$ in the form

$$\mathsf{MBRev}(\mathcal{D}) = \max_{x} \mathbb{E}_{v_i \sim \mathcal{D}} \left[v_i x_i - \max_{j} (v_i - v_j) x_j \right]$$

(in particular, note that for a fixed choice of x, $u_j = \max_j (v_i - v_j) x_j$), and finding an appropriate function $j(i)$ (which corresponds to an assignment to the dual).

2.3.4 A Final Note on the Example

While reading through our examples, the reader may think that the mean-based learner's behavior is clearly irrational: why would you continue paying above your value? Why would you continue paying more than necessary, when you can safely get the item for less?

But this is exactly the point: a more thoughtful learner can indeed do better (for instance, by using the algorithm of Section 2.3.2). It is also perhaps misleading to believe that the bidder should "obviously" stop overpaying: we only know this because we know the structure of the example. But in principle, how is the bidder supposed to know that the overcharged rounds are the new norm and not an anomaly? Given that most standard no-regret algorithms are mean-based, it's important to nail down the seller's options for exploiting this behavior.

2.4 Conclusion and Future Directions

We consider a revenue-maximizing seller with a single item (each round) to sell to a single buyer. We show that when the buyer uses mean-based algorithms like EXP3, the seller can extract revenue equal to the expected welfare with an unnatural auction. We then provide a modified no-regret algorithm $\mathcal{A}$ such that the seller cannot extract revenue exceeding the monopoly revenue when the buyer bids according to $\mathcal{A}$. Finally, we consider a mean-based buyer who never overbids. We tightly characterize the seller's optimal revenue with a linear program, and show that a pay-your-bid auction with decreasing reserves over time achieves this guarantee. Moreover, we show that the mean-based revenue can be unboundedly better than the monopoly revenue while simultaneously worse than the expected welfare. In particular, for the equal revenue curve truncated at H, the monopoly revenue is 1, the expected welfare is $\ln(H)$, and the mean-based revenue is $\Theta(\ln(\ln(H)))$.

While our work has already shown the single-buyer problem is quite interesting, the most natural direction for future work is understanding revenue maximization with multiple learning buyers. Of our three main results, only Theorem 2.3.2 extends easily (that if every buyer uses our modified learning, the default strategy, which now runs Myerson's optimal auction every round, is optimal; see Theorem A.1.5 for details). Our work certainly provides good insight into the multi-bidder problem, but there are still clear barriers. For example, in order to obtain revenue equal to the expected welfare, the auction must necessarily also maximize welfare. In our single-bidder model, this means that we can give away the item for free for $\Omega(T)$ rounds, but with multiple bidders, such careless behaviour would immediately make it impossible to achieve the optimal welfare. Regarding the mean-based revenue, while there is a natural generalization of our LP to multiple bidders, it's no longer clear how to achieve this revenue against conservative bidders, as all the relevant variables now implicitly depend on the actions of the other bidders. These are just examples of concrete barriers, and there are likely interesting conceptual barriers for this extension as well.

Another interesting direction is understanding the consequences of our work from the perspective of the buyer. Aside from certain corner configurations (e.g. the seller extracting the buyer's full welfare), it's not obvious how the buyer's utility changes. For instance, is it possible that the buyer's utility actually *increases* as the seller switches from the default strategy to the optimal mean-based revenue? Does the buyer ever benefit from using an "exploitable" learning strategy, so that the seller can exploit it and make them both happier?

Chapter 3

Multi-armed bandits with strategic arms

This chapter is joint work with Mark Braverman, Jieming Mao, and Matthew Weinberg [29].

3.1 Introduction

In this chapter, we consider a strategic model for the multi-armed bandit problem where each arm is an individual strategic agent and each round one arm is pulled by an agent we refer to as the *principal*. Each round, the pulled arm receives a private reward $v \in [0, 1]$ and then decides what amount x of this reward gets passed on to the principal (upon which the principal receives utility x and the arm receives utility $v - x$). Each arm therefore has a natural tradeoff between keeping most of its reward for itself and passing on the reward so as to be chosen more frequently. Our goal is to design mechanisms for the principal which simultaneously learns which arms are valuable while also incentivizing these arms to pass on most of their rewards.

This model captures a variety of dynamic agency problems, where at each time step the principal must choose to employ one of K agents to perform actions on the

principal's behalf, where the agent's cost of performing that action is unknown to the principal (for example, hiring one of K contractors to perform some work, or hiring one of K investors with external information to manage some money - the important feature being that the principal doesn't know exactly how much they will pay/receive/etc. until the job is done, and the agent has a lot of freedom to set this ex-post). In this sense, this model can be thought of as a multi-agent generalization of the principal-agent problem in contract theory when agents are allowed private savings (see Section 3.1.2 for references). The model also captures, for instance, the interaction between consumers (as the principal) and many sellers deciding how steep a discount to offer the consumers - higher prices now lead to immediate revenue, but offering better discounts than your competitors will lead to future sales. In all domains, our model aims to capture settings where the principal has little domain-specific or market-specific knowledge, and can really only process the reward they get for pulling an arm and not any external factors that contributed to that reward.

There are two "obvious" approaches to try and solve these problems: Option one is to treat it like a procurement auction and run a reverse second-price auction. This doesn't quite work, however, in the case where the agents don't initially know how much reward they'll generate, so some amount of learning needs to enter the picture for a solution to be viable. Using the contractor as a *toy* running example: the contractor will not initially know how much it costs her to work on your home, but after working on your home several times *they* will start to learn how much the next one will cost (you will only learn how much they charge you). In any case, one cannot simply treat it like an auctions problem and ignore learning completely.

The second "obvious" approach is just to treat it as a learning problem, and ignore incentives completely. In fact, one oft-cited motivation for considering adversarial rewards in bandit settings is that arms might be strategic. Indeed, this is because even if the arms' rewards are stochastic, the utility they strategically pass on

to the principal is unlikely to follow any distribution. Algorithms like EXP3 which guarantee low-regret in adversarial settings then seem like the natural "pure learning" approach. Interestingly, our main "negative result" shows that *any* adversarial learning algorithm admits a really bad approximate Nash equilibrium (more details below).

It follows that auctions alone cannot solve the problem, nor can learning alone. To complement our main negative result, we show that the right combination of auctions and learning yields a positive result: an algorithm such that all approximate Nash result in good utility for the principal. We now overview our results in more detail.

3.1.1 Our results

Low-regret algorithms are far from strategyproof

Many algorithms for the multi-armed bandit problem are designed to work in worst-case settings, where an adversary can adaptively decide the value of each arm pull. Here, algorithms such as EXP3 ([16]) guarantee that the principal receives almost as much as if he had only pulled the best arm. Formally, such algorithms guarantee that the principal experiences at most $O(\sqrt{T})$ regret over T rounds compared to any algorithm that only plays a single arm (when the adversary is oblivious).

Given these worst-case guarantees, one might naively expect low-regret algorithms such as EXP3 to also perform well in our strategic variant. It is important to note, however, that single arm strategies perform dismally in this strategic setting; if the principal only ever selects one arm, the arm has no incentive to pass along any surplus to the principal. In fact, we show that the objectives of minimizing adversarial regret and performing well in this strategic variant are fundamentally at odds.

Theorem 3.1.1 (informal restatement of Theorem B.1.3)**.** *Let M be a low-regret algorithm for the classic multi-armed bandit problem with adversarially chosen values.*

Then there exists an instance of the strategic multi-armed bandit problem and an $o(T)$-Nash equilibrium for the arms where a principal running M receives at most $o(T)$ revenue.

While not immediately apparent from the statement of Theorem 3.1.1, these instances where low-regret algorithms fail are far from pathological; in particular, there is a problematic equilibrium for any instance where arm i receives a fixed reward v_i each round it is pulled, as long as the the gap between the largest and second-largest v_i is not too large (roughly 1/#arms).

Here we assume the game is played under a *tacit* observational model, meaning that arms can only observe which arms get pulled by the principal, but not how much value they give to the principal. In particular, this means that arms can achieve this equilibrium despite not communicating directly with each other and not observing the actions of the other arms. This rules out various sorts of "grim trigger" collusion strategies (similar to collusion that occurs in the setting of repeated auctions, see [135]), where arms agree on a protocol ahead of time and immediately defect as soon as one arm deviates from this protocol. (Indeed, in an *explicit* observational model, where arms can see both which arms get pulled and how much value they pass on, it is easy to show even stronger results via such strategies; see Appendix B.1.2 for details).

Instead, the strategies in the equilibrium of Theorem 3.1.1 take the form of *market-sharing strategies*, where arms calibrate their actions so that they each get played some proportion (e.g. $1/K$) of the time while passing on little utility to the principal. For example, consider a simple instance of this problem with two strategic arms, where the principal is using the low-regret EXP3 algorithm, and where arm 1 always gets private reward 1 if pulled and arm 2 always gets private reward 0.8. By always reporting some value slightly larger than 0.8, arm 1 can incentivize the principal to almost always pull it in the long run. This gains arm 1 roughly 0.2 utility per round

(and arm 2 nothing). On the other hand, if arm 1 and arm 2 never pass along any surplus to the principal, they will likely be played equally often, gaining arm 1 roughly 0.5 utility per round and arm 2 0.4 utility per round.

To show such a market-sharing strategy works for general low-regret algorithms, much more work needs to be done. The arms must be able to enforce an even split of the principal's pulls (as soon as the principal starts lopsidedly pulling one arm more often than the others, the remaining arms can defect and start reporting their full value whenever pulled). As long as the principal guarantees good performance in the non-strategic adversarial case (achieving $o(T)$ regret), we show that the arms can (at $o(T)$ cost to themselves, and without explicitly communicating) cooperate so that they are all played equally often.

Mechanisms for strategic arms with stochastic values

We next show that, in contrast to Theorem 3.1.1, it is in fact possible for the principal to extract positive values from the arms per round, if we do not restrict the principal to use an adversarial low-regret algorithm (and hence there is a price to being adversarial low-regret).

We consider a setting where each arm i's reward when pulled is drawn independently from some distribution D_i with mean μ_i (unknown to the principal). In this case the principal can extract the value of the second-best arm (which is the best possible, as we show in Lemma 3.4.3). In the below statement, we are using the term "truthful mechanism" quite loosely as shorthand for "strategy that induces a game among the arms where each arm has a dominant strategy."

Theorem 3.1.2 (restatement of Corollary 3.4.5)**.** *Let μ' be the second largest mean amongst the set of μ_is. Then there exists a truthful mechanism for the principal that guarantees revenue at least $\mu' T - o(T)$ when the arms are playing according to any $o(T)$-Nash equlibrium.*

The mechanism in Theorem 3.1.2 can be thought of as a combination of a second-price auction with the explore-then-exploit strategy from multi-armed bandits. The principal divides the time horizon into three "phases". In the first phase (of size $o(T)$), the principal begins by asking each arm i to simply report their value each round, thus allowing the principal to learn which arm is the most valuable. In the second phase (which comprises the vast majority of the rounds), the principal asks the most valuable arm (the arm with the highest mean in the first phase) to give him the second-largest mean worth of value per round. If this arm fails to comply in any round, the principal avoids picking this arm for the remainder of the rounds. Finally, in the third phase, the principal uses a proper scoring rule to recompensate all arms for reporting truthfully in the first phase. (A more detailed description of the mechanism can be seen in Mechanisms 2 and 3 in Section 3.4).

As an added bonus, we show that this mechanism has similar guarantees in the setting where some arms are strategic and some arms are non-strategic (and our mechanism does not know which arms are which).

Theorem 3.1.3 (restatement of Theorem 3.4.7)**.** *Let μ_s be the second largest mean amongst the means of the strategic arms, and let μ_n be the largest mean amongst the means of the non-strategic arms. Then there exists a truthful mechanism for the principal that guarantees (with probability $1-o(1/T)$) revenue at least $\max(\mu_s, \mu_n)T - o(T)$ when arms play according to any $o(T)$-Nash equilibrium.*

In particular, this implies that Mechanism 3 has low-regret in the classical *stochastic* multi-armed bandits setting, and so the adversarial aspect of the low-regret guarantees is actually essential for the proof of Theorems 3.1.1.

A fair critique of this mechanism is that most of the work of learning the distributions of the arms is offloaded to the beginning of the game. This is appealing because it makes it much feasible to "slide in" some auction design and scoring rules to handle incentives. It is an interesting problem whether learning can still be done

adaptively over time in this model, as such a procedure would necessitate a much more sophisticated treatment of incentives; see Section 3.5 for further discussion.

3.1.2 Related work

The study of classical multi-armed bandit problems was initiated by [122], and has since grown into an active area of study. The most relevant results for this chapter concern the existence of low-regret bandit algorithms in the adversarial setting, such as the EXP 3 algorithm ([16]), which achieves regret $\tilde{O}(\sqrt{KT})$. Other important results in the classical setting include the upper confidence bound (UCB) algorithm for stochastic bandits ([94]) and the work of [68] for Markovian bandits. For further details about multi-armed bandit problems, see the survey [34].

One question that arises in the strategic setting (and other adaptive settings for multi-armed bandits) is what the correct notion of regret is; standard notions of regret guarantee little, since the best overall arm may still have a small total reward. [11] considered the multi-armed bandit problem with an adaptive adversary and introduced the quantity of "policy regret", which takes the adversary's adaptiveness into account. They showed that any multi-armed bandit algorithm will get $\Omega(T)$ policy regret. This indicates that it is not enough to treat strategic behaviors as an instance of adaptively adversarial behavior; good mechanisms for the strategic multi-armed bandits problem must explicitly take advantage of the rational self-interest of the arms.

Our model bears some similarities to the principal-agent problem of contract theory, where a principal employs an more informed agent to make decisions on behalf of the principal, but where the agent may have incentives misaligned from the principal's interests when it gets private savings (for example [37]). For more details on principal-agent problem, see the book [93]. Our model can be thought of as a sort of multi-armed version of the principal-agent problem, where the principal has many

agents to select from (the arms) and can try to use competition between the agents to align their interests with the principal.

Our negative results are closely related to results on collusions in repeated auctions. Existing theoretical work [102, 15, 81, 9, 10, 135] has shown that collusive schemes exist in repeated auctions in many different settings, e.g., with/without side payments, with/without communication, with finite/infinite typespace. In some settings, efficient collusion can be achieved, i.e., bidders can collude to allocate the good to the bidders who values it the most and leave 0 asymptotically to the seller. Even without side payments and communication, [135] showed that tacit collusion exists and can achieve asymptotic efficiency with a large cartel.

Our truthful mechanism uses a proper scoring rule [31, 103] implicitly. In general, scoring rules are used to assessing the accuracy of a probabilistic prediction. In our mechanisms, we use a logarithmic scoring rule to incentivize arms to truthfully report their average rewards.

Our setting is similar to settings considered in a variety of work on dynamic mechanism design, often inspired by online advertising. [23] considers the problem where a buyer wants to buy a stream of goods with an unknown value from two sellers, and examines Markov perfect equilibria in this model. [18, 54, 17] study truthful pay-per-click auctions where the auctioneer wishes to design a truthful mechanism that maximizes the social welfare. [92, 64] consider the scenario where the principal cannot directly choose which arm to pull, and instead must incentivize a stream of strategic players to prevent them from acting myopically. [6, 7] consider a setting where a seller repeatedly sells to a buyer with unknown value distribution, but the buyer is more heavily discounted than the seller. [82] develops a general method for finding optimal mechanisms in settings with dynamic private information. [113] develops an ex ante efficient mechanism for the Cost-Per-Action charging scheme in online advertising.

3.2 Our Model

3.2.1 Classic Multi-Armed Bandits

We begin by reviewing the definition of the classic multi-armed bandits problem and associated quantities.

In the classic multi-armed bandit problem a learner (the *principal*) chooses one of K choices (arms) per round, over T rounds. On round t, the principal receives some reward $v_{i,t} \in [0, 1]$ for pulling arm i. The values $v_{i,t}$ are either drawn independently from some distribution corresponding to arm i (in the case of *stochastic bandits*) or adaptively chosen by an adversary (in the case of *adversarial bandits*). Unless otherwise specified, we will assume we are in the adversarial setting.

Let I_t denote the arm pulled by the principal at round t. The *revenue* of an algorithm M is the random variable

$$\mathsf{Rev}(M) = \sum_{t=1}^{T} v_{I_t,t}$$

and the the *regret* of M is the random variable

$$\mathsf{Reg}(M) = \max_i \sum_{t=1}^{T} v_{i,t} - \mathsf{Rev}(M)$$

Definition 3.2.1 (δ-Low Regret Algorithm)**.** *Mechanism M is a δ-low regret algorithm for the multi-armed bandit problem if*

$$\mathbb{E}[\mathsf{Reg}(M)] \leq \delta.$$

Here the expectation is taken over the randomness of M and the adversary.

Definition 3.2.2 ((ρ, δ)-Low Regret Algorithm)**.** *Mechanism M is a (ρ, δ)-low regret algorithm for the multi-armed bandit problem if with probability $1 - \rho$,*

$$\mathsf{Reg}(M) \leq \delta.$$

There exist $O(\sqrt{KT \log K})$-low regret algorithms and $(\rho, O(\sqrt{KT \log(K/\rho)}))$-low regret algorithms for the multi-armed bandit problem; see Section 3.2 of [34] for details.

3.2.2 Strategic Multi-Armed Bandits

The strategic multi-armed bandits problem builds upon the classic multi-armed bandits problem with the notable difference that now arms are strategic agents with the ability to withhold some payment from the principal. Instead of the principal directly receiving a reward $v_{i,t}$ when choosing arm i, now arm i receives this reward and passes along some amount $w_{i,t}$ to the principal, gaining the remainder $v_{i,t} - w_{i,t}$ as utility.

For simplicity, in the strategic setting, we will assume the rewards $v_{i,t}$ are generated stochastically; that is, each round, $v_{i,t}$ is drawn independently from a distribution D_i (where the distributions D_i are known to all arms but not to the principal). While it is possible to pose this problem in the adversarial setting (or other more general settings), this comes at the cost of there being no clear notion of strategic equilibrium for the arms.

This strategic variant comes with two additional modeling assumptions. The first is the informational model of this game; what information does an arm observe when some other arm is pulled. We define two possible observational models:

1. **Explicit:** After each round t, every arm sees the arm played I_t along with the quantity $w_{I_t,t}$ reported to the principal.

2. **Tacit:** After each round t, every arm only sees the arm played I_t.

In both cases, only arm i knows the size of the original reward $v_{i,t}$; in particular, the principal also only sees the value $w_{i,t}$ and learns nothing about the amount withheld by the arm. Collusion between arms is generally significantly easier in the explicit observational model than in the tacit observational model, and for this reason we will assume we are in the tacit observational model unless otherwise stated.

The second modeling assumption is whether to allow arms to go into debt while paying the principal. In the *restricted payment* model, we impose that $w_{i,t} \leq v_{i,t}$; an arm cannot pass along more than it receives in a given round. In the *unrestricted payment* model, we let $w_{i,t}$ be any value in $[0, 1]$. We prove our negative results in the restricted payment model and our positive results in the unrestricted payment model, but our proofs for our negative results work in both models (in particular, it is easier to collude and prove negative results in the unrestricted payment model) and Mechanism 3 can be adapted to work in the restricted payment model (see discussion in Section 3.4.2).

Finally, we proceed to define the set of strategic equilibria for the arms. We assume the mechanism M of the principal is fixed ahead of time and known to the K arms. If each arm i is using a (possibly adaptive) strategy S_i, then the expected utility of arm i is defined as

$$u_i(M, S_1, \ldots, S_K) = \mathbb{E}\left[\sum_{t=1}^{T}(v_{i,t} - w_{i,t}) \cdot \mathbb{1}_{I_t=i}\right].$$

An ε-Nash equilibrium for the arms is then defined as follows.

Definition 3.2.3 (ε-Nash Equilibrium for the arms)**.** *Strategies $(S_1, ..., S_K)$ form an ε-Nash equilibrium for the strategic multi-armed bandit problem if for all $i \in [n]$ and any deviating strategy S_i',*

$$u_i(S_1, \ldots, S_i, \ldots, S_K) \geq u_i(S_1, \ldots, S_i', \ldots, S_K) - \varepsilon.$$

Similarly as before, the revenue of the principal in this case is the random variable

$$\mathsf{Rev}(M, S_1, \ldots, S_K) = \sum_{t=1}^{T} w_{I_t,t}.$$

The goal of the principal is to choose a mechanism M which guarantees large revenue in any ε-Nash Equilibrium for the arms.

In Section 3.4, we will construct mechanisms for the strategic multi-armed bandit problem which are truthful for the arms. We define the related terminology below.

Definition 3.2.4 (Dominant Strategy)**.** *When the principal uses mechanism M, we say S_i is a dominant strategy for arm i if for any deviating strategy S_i' and any strategies for other arms $S_1, .., S_{i-1}, S_{i+1}, ..., S_K$,*

$$u_i(M, S_1, \ldots, S_i, \ldots, S_K) \geq u_i(M, S_1, \ldots, S_i', \ldots, S_K).$$

Definition 3.2.5 (Truthfulness)**.** *We say that a mechanism M for the principal is truthful, if all arms have some dominant strategies.*

3.3 Negative Results Overview

In this section we give a sketch of the proof of our main theorem, Theorem B.1.3. The full list of our negative results and proofs can be found in Appendix B.1.

Theorem 3.3.1. *[(restatement of Theorem B.1.3)] Let mechanism M be a (ρ, δ)-low regret algorithm for the multi-armed bandit problem with K arms, where $K \leq T^{1/3}/\log(T)$, $\rho \leq T^{-2}$, and $\delta \geq \sqrt{T \log T}$. Then in the strategic multi-armed bandit problem under the tacit observational model, there exist distributions D_i and an $O(\sqrt{KT\delta})$-Nash Equilibrium for the arms where the principal gets at most $O(\sqrt{KT\delta})$ revenue.*

Proof Sketch. The underlying idea here is that the arms work to try to maintain an equal market share, where each of the K arms are each played approximately $1/K$ of the time. To ensure this happens, arms collude so that arms that aren't as likely to be pulled pass along more than arms that have been pulled a lot or are more likely to be pulled; this ends up forcing any low-regret algorithm for the principal to choose all the arms equally often. Interestingly, this collusion strategy is *mechanism dependent*, as arms need to estimate the probability they will be pulled in the next round.

More formally, let μ_i denote the mean of the ith arm's distribution D_i. Without loss of generality, further assume that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$. We will show that as long as $\mu_1 - \mu_2 \leq \frac{\mu_1}{K}$, there exists some $O(\sqrt{KT\delta})$-Nash equilibrium for the arms where the principal gets at most $O(\sqrt{KT\delta})$ revenue.

We begin by describing the equilibrium strategy S^* for the arms. Let $c_{i,t}$ denote the number of times arm i has been pulled up to time t. Set $B = 7\sqrt{KT\delta}$ and set $\theta = \sqrt{\frac{K\delta}{T}}$. The equilibrium strategy for arm i at time t is as follows:

1. If at any time $s \leq t$ in the past, there exists an arm j with $c_{j,s} - c_{i,s} \geq B$, defect and offer your full value $w_{i,t} = \mu_i$.

2. Compute the probability $p_{i,t}$, the probability that the principal will pull arm i conditioned on the history so far.

3. Offer $w_{i,t} = \theta(1 - p_{i,t})$.

The main technical challenge in proving that this strategy is an equilibrium involves showing that, if all arms are following this strategy and the principal is using a low-regret mechanism, then with high probability the arms will not defect. Here the low-regret property of the mechanism M is essential (indeed, as our positive results imply, the theorem is not true without this assumption). In particular, by the construction of $w_{i,t}$ in terms of $p_{i,t}$, the principal's expected total regret (here defined to be the sum of the principal's regrets with respect to each arm) will increase each

round by some amount proportional to the variance of the $p_{i,t}$. Intuitively, this implies that the values $p_{i,t}$ cannot be too far from uniform for too many rounds, and therefore that each arm should be picked approximately the same proportion of the time. This is formalized in the following lemma:

Lemma 3.3.2. *If all arms are using strategy S^*, then with probability $\left(1 - \frac{3}{T}\right)$, $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$.*

Proof. As before, assume that all arms are playing the strategy S^* with the modification that they never defect. This does not change the probability that $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$.

Define $R_{i,t} = \sum_{s=1}^{t} w_{i,s} - \sum_{s=1}^{t} w_{I_s,s}$ be the regret the principal experiences for not playing only arm i up until time t. We begin by showing that with probability at least $1 - \frac{2}{T}$, $R_{i,t}$ lies in $[-K\theta\sqrt{T \log T} - (K-1)\delta, \delta]$ for all $t \in [T]$ and $i \in [K]$.

To do this, first note that since the principal is using a (T^{-2}, δ)-low-regret algorithm, with probability at least $1 - T^{-2}$ the regrets $R_{i,t}$ are all upper bounded by δ at any fixed time t. Via the union bound, it follows that $R_{i,t} \leq \delta$ for all i and t with probability at least $1 - \frac{1}{T}$.

To lower bound $R_{i,t}$, we will first show that $\sum_{i=1}^{K} R_{i,t}$ is a submartingale in t. Note that, with probability $p_{j,t}$, $R_{i,t+1}$ will equal $R_{i,t} + \theta((1 - p_{j,t}) - (1 - p_{i,t}))$. We then have

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{K} R_{i,t+1} \middle| \sum_{i=1}^{K} R_{i,t}\right] &= \sum_{i=1}^{K} R_{i,t} + \sum_{i=1}^{K} p_{i,t} \sum_{j=1}^{K} \theta((1-p_{j,t}) - (1-p_{i,t})) \\
&= \sum_{i=1}^{K} R_{i,t} + \sum_{i=1}^{K} p_{i,t} \sum_{j=1}^{K} \theta(p_{i,t} - p_{j,t}) \\
&= \sum_{i=1}^{K} R_{i,t} + \theta \sum_{i=1}^{K} p_{i,t}(Kp_{i,t} - 1) \\
&= \sum_{i=1}^{K} R_{i,t} + \theta \left(K \sum_{i=1}^{K} p_{i,t}^2 - \sum_{i=1}^{K} p_{i,t} \right) \\
&\geq \sum_{i=1}^{K} R_{i,t}
\end{aligned}$$

where the last inequality follows by Cauchy-Schwartz. It follows that $\sum_{i=1}^{K} R_{i,t}$ forms a submartingale.

Moreover, note that (since $|p_i - p_j| \leq 1$) $|R_{i,t+1} - R_{i,t}| \leq \theta$. It follows that $\left|\sum_{i=1}^{K} R_{i,t+1} - \sum_{i=1}^{K} R_{i,t}\right| \leq K\theta$ and therefore by Azuma's inequality that, for any fixed $t \in [T]$,

$$\Pr\left[\sum_{i=1}^{K} R_{i,t} \leq -2K\theta\sqrt{T \log T}\right] \leq \frac{1}{T^2}.$$

With probability $1 - \frac{1}{T}$, this holds for all $t \in [T]$. Since (with probability $1 - \frac{1}{T}$) $R_{i,t} \leq \delta$, this implies that with probability $1 - \frac{2}{T}$, $R_{i,t} \in \left[-2K\theta\sqrt{T \log T} - (K-1)\delta, \delta\right]$.

We next proceed to bound the probability that $c_{i,t} - c_{j,t} > B$ for a i, j, and t. Define

$$S_t^{(i,j)} = \left(c_{i,t} - c_{j,t} + \frac{1}{\theta}(R_{i,t} - R_{j,t})\right).$$

We claim that $S_t^{(i,j)}$ is a martingale. To see this, we first claim that $R_{i,t+1} - R_{j,t+1} = R_{i,t} - R_{j,t} - \theta(p_{i,t} - p_{j,t})$. Note that, if arm k is pulled, then $R_{i,t+1} = R_{i,t} + \theta((1 - p_{i,t}) - (1 - p_{k,t})) = R_{i,t} + \theta(p_{k,t} - p_{i,t})$ and similarly, $R_{j,t+1} = R_{j,t} + \theta(p_{k,t} - p_{j,t})$. It follows that $R_{i,t+1} - R_{j,t+1} = R_{i,t} - R_{j,t} - \theta(p_{i,t} - p_{j,t})$.

Secondly, note that (for any arm k) $\mathbb{E}[c_{k,t+1} - c_{k,t}|p_t] = p_{k,t}$, and thus $\mathbb{E}[c_{i,t+1} - c_{j,t+1} - (c_{i,t} - c_{j,t})|p_t] = p_{i,t} - p_{j,t}$. It follows that

$$\begin{aligned}
\mathbb{E}[S_{t+1}^{(i,j)} - S_t^{(i,j)}|p_t] &= \mathbb{E}[(c_{i,t+1} - c_{j,t+1}) - (c_{i,t} - c_{j,t})|p_t] \\
&\quad + \frac{1}{\theta}\mathbb{E}[(R_{i,t+1} - R_{j,t+1}) - (R_{i,t} - R_{j,t})|p_t] \\
&= (p_{i,t} - p_{j,t}) - (p_{i,t} - p_{j,t}) \\
&= 0
\end{aligned}$$

and thus that $\mathbb{E}[S_{t+1}^{(i,j)}|S_t^{(i,j)}] = S_t^{(i,j)}$, and thus that $S_t^{(i,j)}$ is a martingale. Finally, note that $|S_{t+1}^{(i,j)} - S_t^{(i,j)}| \leq 2$, so by Azuma's inequality

$$\Pr\left[S_t^{(i,j)} \geq 4\sqrt{T\log(TK)}\right] \leq (TK)^{-2}$$

Taking the union bound, we find that with probability at least $1 - \frac{1}{T}$, $S^{(i,j)} \leq 4\sqrt{T\log(TK)}$ for all i, j, and t. Finally, since with probability at least $1 - \frac{2}{T}$ each $R_{i,t}$ lies in $\left[-2K\theta\sqrt{T\log T} - (K-1)\delta, \delta\right]$, with probability at least $1 - \frac{3}{T}$ we have that (for all i, j, and t)

$$
\begin{aligned}
c_{i,t} - c_{j,t} &= S_t^{(i,j)} - \frac{1}{\theta}(R_{i,t} - R_{j,t}) \\
&\leq 4\sqrt{T\log(TK)} + \frac{1}{\theta}\left|R_{i,t} - R_{j,t}\right| \\
&\leq 4\sqrt{T\log(TK)} + 2K\sqrt{T\log T} + \frac{K\delta}{\theta} \\
&\leq \frac{7K\delta}{\theta} \\
&= 7K\sqrt{T\delta} \\
&= B
\end{aligned}
$$

□

Lemma 3.3.2 implies that if each arm plays strategy S^*, then each arm i will receive on average μ_i/K per round. To finish the proof, it suffices to note that by deviating and playing a different strategy S from S^*, one of two things can occur. If playing this different strategy S does not trigger the defect condition in (1), then still each arm will be played roughly $1/K$ of the time (and your total utility is unchanged up to $o(T)$ additive factors). On the other hand, once the defect condition is triggered, you can receive at most $\mu_1 - \mu_2$ utility per round (and only if you are arm 1). This implies that as long as $\mu_1/K > \mu_1 - \mu_2$, there is no incentive to deviate.

Additional details are provided in Appendix B.1. □

While the theorem above merely claims that a bad set of distributions for the arms exists, the proof shows it is possible to collude in a wide range of instances - in particular, any collection of distributions which satisfies $\mu_1 - \mu_2 \leq \mu_1/K$. A natural question is whether we can extend the above results to show that it is possible to collude in any set of distributions.

One issue with the collusion strategy in the above proof is that if $\mu_1 - \mu_2 > \mu_1/K$, then arm 1 will have an incentive to defect in any collusive strategy that plays all

the arms evenly (arm 1 can report a bit over μ_2 per round, and make $\mu_1 - \mu_2$ every round instead of μ_1 every K rounds). One solution to this is to design a collusive strategy that plays some arms more than others in equilibrium (for example, playing arm 1 90% of the time). We show how to modify our result for two arms to achieve an arbitrary market partition and thus work over a broad set of distributions.

Theorem 3.3.3. *Let mechanism M be a (ρ, δ)-low regret algorithm for the multi-armed bandit problem with two arms, where $\rho \leq T^{-2}$ and $\delta \geq \sqrt{T \log T}$. Then, in the strategic multi-armed bandit problem under the tacit observational model, for* any *distributions D_1, D_2 of values for the arms (supported on $[\sqrt{\delta/T}, 1]$), there exists an $O(\sqrt{T\delta})$-Nash Equilibrium for the arms where a principal using mechanism M gets at most $O(\sqrt{T\delta})$ revenue.*

Proof. See Appendix B.1. □

Unfortunately, it as not as easy to modify the proof of Theorem B.1.3 to prove the same result for K arms. It is an interesting open question whether there exist collusive strategies for K arms that can achieve an arbitrary partition of the market.

3.4 Positive Results

In this section we will show that, in contrast to the previous results on collusion, there exists a mechanism for the principal that can obtain $\Theta(T)$ revenue from the arms when they play according to an $o(T)$-Nash equilibrium.

We begin by demonstrating a simpler version of our mechanism (Mechanism 2) that guarantees the principal $\Theta(T)$ revenue whenever the arms play according to their dominant strategies. In Section 3.4.2, we then show how to make this mechanism more robust (Mechanism 3) so that the principal is guaranteed $\Theta(T)$ revenue whenever the arms play according to any $o(T)$-approximate Nash equilibrium (thus showing a separation between the power of adversarial low-regret algorithms and general learning

algorithms in this model). As an added bonus, we show that this mechanism also works for a combination of strategic and non-strategic arms (and therefore achieves low regret in the classical stochastic multi-armed bandits setting).

Throughout this section we will assume we are working in the tacit observational model and the unrestricted payment model (unless otherwise specified). We postpone all the proofs of this section to Appendix B.2.

3.4.1 Good dominant strategy equilibria

This mechanism essentially incentivizes each arm to report the mean of its distribution and then runs a second-price auction, asking the arm with the highest mean for the second-highest mean each round.

Define μ_i as the mean of distribution D_i for $i = 1, \ldots, K$, let $\mu_{min} = \min_{i:\mu_i \neq 0}(\mu_i)$, and $u = -\log \mu_{min} + 1$. We assume throughout that $u = o(T/K)$.

Algorithm 2 Truthful mechanism for strategic arms with known stochastic values in the tacit model

Play each arm once (i.e. play arm 1 in the first round, arm 2 in the second round, etc.). Let w_i be the value arm i reports in round i.
Let $i^* = \arg\max w_i$ (breaking ties lexicographically), and let $w' = \max_{i \neq i^*} w_i$.
Tell arm i^* the value of w'. Play arm i^* for $R = T - (u+2)K - 1$ rounds. If arm i^* ever reports a value different from w', stop playing it immediately. If arm i^* always gives w', play it for one bonus round (ignoring the value it reports).
For each arm i such that $i \neq i^*$, play it for one round.
For each arm i satisfying $u + \log(w_i) \geq 0$, play it $\lfloor u + \log(w_i) \rfloor$ times. Then, with probability $u + \log(w_i) - \lfloor u + \log(w_i) \rfloor$, play arm i for one more round.

We will first show that the dominant strategy of each arm in this mechanism includes truthfully reporting their mean at the beginning, and then then compute the principal's revenue under this dominant strategy.

Lemma 3.4.1. *The following strategy is the dominant strategy for arm i in Mechanism 2:*

1. *(line 1 of Mechanism 2) Report the mean value μ_i of D_i the first time when arm i is played.*

2. *(lines 3,4 of Mechanism 2) If $i = i^*$, for the R rounds that the principal expects to see reported value w', report the value w'. For the bonus round, report 0. If $i \neq i^*$, report 0.*

3. *(line 5 of Mechanism 2) For all other rounds, report* 0.

Corollary 3.4.2. *Under Mechanism 2, the principal will receive revenue at least $\mu' T - o(T)$ when arms use their dominant strategies, where μ' is the second largest mean in the set of means μ_i.*

We additionally show that the performance of Mechanism 2 is as good as possible; no mechanism can do better than the second-best arm in the worst case.

Lemma 3.4.3. *Let μ and μ' be the largest and second largest values respectively among the μ_i. Then for any constant $\alpha > 0$, no truthful mechanism can guarantee $(\alpha\mu + (1-\alpha)\mu')T$ revenue in the worst case.*

3.4.2 Good approximate Nash equilbria

One issue with Mechanism 2 is that, while the principal achieves $\Theta(T)$ revenue when the arms play according to their dominant strategies, there can exist ϵ-Nash equilibria for the arms which still leave the principal with negligible revenue. For instance, if there are two arms with equal means $\mu_1 = \mu_2 = \mu$, one possible ϵ-Nash equilibrium is for them both to bid μ, and then for arm i^* to immediately defect after it is chosen. This is not a dominant strategy, since arm i^* surrenders its bonus for not defecting, but since this bonus is at most 1, this is still an ϵ-Nash equilibrium for any $\epsilon = o(T)$ which is larger than 1.

We can make Mechanism 3 more robust to strategies like this by increasing the size of the bonus with ϵ. If we additionally allow a tiny buffer between the current reported

average and w', this mechanism has the added property that it works even when there are a mixture of strategic and non-strategic arms (and the principal does not know which are which). In particular, this Mechanism 3 obtains low-regret in the classical stochastic multi-armed bandits setting, which implies that our negative results in Section 3.3 are really due to the adversarial nature of the low-regret guarantees.

As before, define μ_i as the mean of distribution D_i for $i = 1, ..., K$. Our mechanism takes in two parameters, B (representing the size of the bonus) and M (representing the size of the buffer). We will set $B = 2\epsilon^{1/4}T^{3/4}/\mu_{min}$ and $M = 8B^{-1/2}\ln(KT)$. In addition, we will define $u = -\log(\min_{i:\mu_i \neq 0} \mu_i) + 2 + M$. We assume $u = o(\frac{T}{BK})$.

Algorithm 3 Truthful mechanism for strategic/non-strategic arms in the tacit model

Play each arm B times (i.e. play arm 1 in the first B rounds, arm 2 in the next B rounds, etc.). Let $\bar{w}_i$ be the average value arm i reported in its B rounds.
Let $i^* = \arg\max \bar{w}_i$ (breaking ties lexicographically), and let $w' = \max_{i \neq i^*} \bar{w}_i$.
Tell arm i^* the value of w'. Play arm i^* for $R = T - (u+3)BK$ rounds. If arm i^* ever reports values with average less than $w' - M$ in any round after B rounds in this step, stop playing it immediately. If arm i^* gives average no less than $w' - M$, play it for B bonus rounds (ignoring the value it reports).
For each arm i such that $i \neq i^*$, play it for B rounds.
For each arm i satisfying $u + \log(\bar{w}_i - M) \geq 0$, play it $B\lfloor(u + \log(\bar{w}_i - M))\rfloor$ times. Then, with probability $u + \log(\bar{w}_i - M) - \lfloor u + \log(\bar{w}_i - M)\rfloor$, play arm i for B more rounds.

We begin by characterizing the dominant strategy for Mechanism 3. Similarly as in Lemma 3.4.1, we show that this dominant strategy involves each arm reporting their true mean in the beginning rounds.

Lemma 3.4.4. *The following strategy is the dominant strategy for arm i in Mechanism 3:*

1. *(line 1 of Mechanism 3) For the first B rounds, report a total sum of* $(\mu_i + M)B$.
2. *(lines 3,4 of Mechanism 3) If* $i = i^*$, *for the R rounds that the principal expects to see reported value* w', *report the value* $w' - M$. *For the B bonus rounds, report 0. If* $i \neq i^*$, *report 0.*

3. *(line 5 of Mechanism 3) For all other rounds, report* 0.

We use this to show that under any $o(T)$-Nash equilibrium, the principal receives $\mu' T - o(T)$ revenue under Mechanism 3.

Corollary 3.4.5. *Under Mechanism 3, the principal will receive revenue at least $\mu' T - o(T)$ whenever arms play according to an ϵ-Nash equilibrium, where μ' is the second largest mean in the set of means μ_i and $\epsilon = o(T)$.*

The dominant strategy in Lemma 3.4.4, as written, requires the arms to know their own means μ_i (in particular for step 1). However, if the arms don't initially know their means, they can instead simply report their value (plus M) each round, and still report a total sum of $(\mu_i + M)B$ in expectation. This no longer results in a strictly dominant strategy, but instead an $o(T)$-dominant strategy.

Lemma 3.4.6. *The following strategy is a prior-independent $o(T)$-dominant strategy for arm i in Mechanism 3:*

1. *(line 1 of Mechanism 3) For each round t in the first B rounds, report $v_{i,t} + M$.*
2. *(lines 3,4 of Mechanism 3) If $i = i^*$, for the R rounds that the principal expects to see reported value w', report the value $w' - M$. For the B bonus rounds, report 0. If $i \neq i^*$, report 0.*
3. *(line 5 of Mechanism 3) For all other rounds, report* 0.

It is an interesting question whether a more clever stochastic bandit algorithm can be embedded without destroying dominant strategies, and also whether a solution exists in exact dominant strategies for this model.

Similarly, the dominant strategy in Lemma 3.4.4 assumes we are in the unrestricted payment regime, because sometimes the value you must report (whether it is $\mu_i + M$ or $w' - M$) might be larger than the value received in that round. However,

again it is possible to adapt the mechanism (by setting $M = 0$) and dominant strategy in Lemma 3.4.4 to work for arms in the restricted payment regime at the cost of transforming it into a $o(T)$-dominant strategy. To do this, arms (as in the previous paragraph) simply report their value each round in the first phase of the mechanism. In the second phase of the mechanism, instead of reporting w' each round, they again report their full value, until they have reported a total of Rw' (at which point they start reporting 0 for the rest of the game).

Finally, we consider the case when some arms are strategic and other arms are non-strategic. Importantly, the principal does not know which arms are strategic and which are non-strategic. We show in this case that the principal can get (per round) the larger of the largest mean of the non-strategic arms and the second largest mean of the strategic arms.

Theorem 3.4.7. *If the strategic arms all play according to in Lemma 3.4.4, then the principal will get at least* $\max(\mu_s, \mu_n)T - o(T)$ *with probability* $1 - o(1/T)$. *Here* μ_s *is the second largest mean of the strategic arms and* μ_n *is the largest mean of the non-strategic arms.*

3.5 Conclusions and Future Directions

We consider the multi-armed bandit problem with strategic arms: arms obtain a reward when pulled and may pass any of it onto the principal. Our first main result shows that treating this purely as a learning problem results in undesirable approximate Nash equilibria for the principle (guaranteeing only $o(T)$ reward over T rounds). Our second main result shows that a careful combination of auctions, learning, and scoring rules provides a learning algorithm such that every approximate Nash equilibrium guarantees the principal $\Omega(T)$ reward (and even better - the arms have a dominant strategy). Still, we are far from understanding the complete picture of

multi-armed bandit problems in strategic settings. Many questions remain, both in our model and related models.

One limitation of our negative results is that they only show there exists some 'bad' approximate Nash equilibrium for the arms, i.e., one where any low-regret principal receives little revenue. This, however, says nothing about the space of all approximate Nash equilibria. Does there exist a low-regret mechanism for the principal along with an approximate Nash equilibria for the arms where the principal extracts significant utility? An affirmative answer to this question would raise hope for the possibility of a mechanism that can perform well in both the adversarial and strategic setting, whereas a negative answer would strengthen our claim that these two settings are fundamentally at odds.

One limitation of our positive results is that all of the learning takes place at the beginning of the protocol. As a result, our mechanism fails in cases where the arms' distributions can change over time. Is it possible to design good mechanisms for such settings? Ideally, any good mechanism should learn the arms' values continually throughout the T rounds, but accommodating this would require novel tools to handle incentives.

Throughout this chapter, whenever we consider strategic bandits we assume their rewards are stochastically generated. Can we say anything about strategic bandits with adversarially generated rewards? The key barrier here seems to be defining what a strategic equilibrium is in this case - arms need some underlying priors to reason about their future expected utility.

Finally, there are other quantities one may wish to optimize instead of the utility of the principal. For example, is it possible to design an efficient principal, who almost always picks the best arm (even if the arm passes along little to the principal)? Theorem B.1.3 implies the answer is no if the principal also has to be efficient in

the adversarial case, but are there other models where we can answer this question affirmatively?

Part II

Learning how to price

Chapter 4

Contextual Search via Intrinsic Volumes

This chapter is joint work with Renato Paes Leme [97].

4.1 Introduction

Consider the classical problem of binary search, where the goal is to find a hidden real number $x \in [0, 1]$, and where feedback is limited to guessing a number p and learning whether $p \leq x$ or whether $p > x$. One can view this as an online learning problem, where every round t a learner guesses a value $p_t \in [0, 1]$, learns whether or not $p_t < x$, and incurs some loss $\ell(x, p_t)$ (for some loss function $\ell(\cdot, \cdot)$). The goal of the learner is to minimize the total loss $\sum_{t=1}^{T} \ell(x, p_t)$ which can alternatively be thought of as the learner's *regret*. For example, for the loss function $\ell(x, p_t) = |x - p_t|$, the learner can achieve total regret bounded by a constant via the standard binary search algorithm.

In this chapter, we consider a contextual, multi-dimensional generalization of this problem which we call the *contextual search problem*. Now, the learner's goal is to learn the value of a hidden vector $v \in [0, 1]^d$. Every round, an adversary provides a context u_t, a unit vector in $\mathbb{R}^d$, to the learner. The learner must now guess a

value p_t, upon which they incur loss $\ell(\langle u_t, v\rangle, p_t)$ and learn whether or not $p_t \leq \langle u_t, v\rangle$. Geometrically, this corresponds to the adversary providing the learner with a hyperplane; the learner may then translate the hyperplane however they wish, and then learn which side of the hyperplane v lies on. Again, the goal of the learner is to minimize their total loss $\sum_{t=1}^{T} \ell(\langle u_t, v\rangle, p_t)$.

This framework captures a variety of problems in contextual decision-making. Most notably, it captures the well-studied problem of *contextual dynamic pricing* [8, 44, 80]. In this problem, the learner takes on the role of a seller of a large number of differentiated products. Every round t the seller must sell a new product with features summarized by some vector $u_t \in [0,1]^d$. They are selling this item to a buyer with fixed values $v \in [0,1]^d$ for the d features (that is, this buyer is willing to pay up to $\langle u, v\rangle$ for an item with feature vector u). The seller can set a price p_t for this item, and observes whether or not the buyer buys the item at this price. If a sale is made, the seller receives revenue p_t; otherwise the seller receives no revenue. The goal of the seller is to maximize their revenue over a time horizon of T rounds.

The dynamic pricing problem is equivalent to the contextual search problem with loss function ℓ satisfying $\ell(\theta, p) = \theta - p$ if $\theta \geq p$ and $\ell(\theta, p) = \theta$ otherwise. The one-dimensional variant of this problem was first introduced by Kleinberg and Leighton [90], who presented an $O(\log \log T)$ regret algorithm for this problem and showed that this was tight. Amin, Rostamizadeh and Syed [8] introduce the problem in its contextual, multi-dimensional form, but assume iid contexts. Cohen, Lobel, and Paes Leme [44] study the problem with adversarial contexts and improve the $\tilde{O}(\sqrt{T})$-regret obtainable from general purpose contextual bandit algorithms [1] to $O(d^2 \log T)$-regret, based on approximating the current knowledge set (possible values for v) with ellipsoids. This was later improved to $O(d \log T)$ in [99].

In this chapter we present algorithms for the contextual search problem with improved regret bounds (in terms of their dependence on T). More specifically:

1. For the symmetric loss function $\ell(\theta, p) = |\theta - p|$, we provide an algorithm that achieves regret $O(\mathsf{poly}(d))$. In contrast, the previous best-known algorithms for this problem (from the dynamic pricing literature) incur regret $O(\mathsf{poly}(d) \log T)$.

2. For the dynamic pricing problem, we provide an algorithm that achieves regret $O(\mathsf{poly}(d) \log \log T)$. This is tight up to a polynomial factor in d, and improves exponentially on the previous best known bounds of $O(\mathsf{poly}(d) \log T)$.

Both algorithms can be implemented efficiently in randomized polynomial time (and achieve the above regret bounds with high probability).

Techniques from Integral Geometry Classical binary search involves keeping an interval of possible values (the "knowledge set") and repeatedly bisecting it to decrease its length. In the one-dimensional case length can both be used as a potential function to measure the progress of the algorithm and as a bound for the loss. When generalizing to higher dimensions, the knowledge set becomes a higher dimensional convex set and the natural measure of progress (the volume) no longer directly bounds the loss in each step.

To address this issue we use concepts from the field of *integral geometry*, most notably the notion of *intrinsic volumes*. The field of integral geometry (also known as geometric probability) studies measures on convex subsets of Euclidean space which remain invariant under rotations/translations of the space. One of the fundamental results in integral geometry is that in d dimensions there are $d+1$ essentially distinct different measures, of which surface area and volume are two. These $d+1$ different measures are known as *intrinsic volumes*, and each corresponds to a dimension between 0 and d (for example, surface area and volume are the $(d-1)$-dimensional and d-dimensional intrinsic volumes respectively).

A central idea in our algorithm for the symmetric loss function is to choose our guess p_t so as to divide one of the d different intrinsic volumes in half. The choice of which intrinsic volume to divide in half depends crucially on the geometry of the current knowledge set. When the knowledge set is well-rounded and ball-like, we can get away with simply dividing the knowledge set in half by volume. As the knowledge set becomes thinner and more pointy, we must use lower and lower dimensional intrinsic volumes, until finally we must divide the one-dimensional intrinsic volume in half. By performing this division carefully, we can ensure that the total sum of all the intrinsic volumes of our knowledge set (appropriately normalized) decreases by at least the loss we incur each round.

Our algorithm for the dynamic pricing problem builds on top of the ideas developed for the symmetric loss together with a new technique for charging progress based on an isoperimetric inequality for intrinsic volumes that can be obtained from the Alexandrov-Fenchel inequality. This new technique allows us to combine the doubly-exponential buckets technique of Kleinberg and Leighton with our geometric approach to the symmetric loss and obtain an $O_d(\log \log T)$ regret algorithm for the pricing loss.

One can ask whether simpler algorithms can be obtained for this setting using only the standard notions of volume and width. We analyze simpler halving algorithms and show that while they obtain $O_d(1)$ regret for the symmetric loss, the dependency on the dimension d is exponentially worse. While the simple halving algorithms are defined purely in terms of standard geometric notions, our analysis of them still requires tools from intrinsic geometry. For the pricing loss case, we are not aware of any simpler technique just based on standard geometric notions that can achieve $O_d(\log \log T)$ regret.

Finally, we would like to mention that to the best of our knowledge this is the first application of intrinsic volumes to theoretical algorithm design.

Applications and Other Related Work The main application of our result is to the problem of contextual dynamic pricing. The dynamic pricing problem has been extensively studied with different assumptions on contexts and valuation. Our model is the same as the one in Amin et al [8], Cohen et al [43] and Lobel et al [99] who provide regret guarantees of $O(\sqrt{T})$, $O(d^2 \log T)$ and $O(d \log T)$ respectively. The problem was also studied with stochastic valuation and additional structural assumptions on contexts in Javanmard and Nazerzadeh [80], Javanmard [79] and Qiang and Bayati [119]. This line of work relies on techniques from statistic learning, such as greedy least squares, LASSO and regularized maximum likelihood estimators. The guarantees obtained there also have $\log T$ dependency on the time horizon.

The contextual search problem was also considered with the loss function $\ell(\theta, p) = \mathbf{1}\{|\theta - p| > \epsilon\}$. For this loss function, Lobel et al [99] provide the optimal regret guarantee of $O(d \log(1/\epsilon))$. The geometric techniques developed in this line of work were later applied by Gillen et al [67] in the design of online algorithms with an unknown fairness objective. Another important application of contextual search is the problem of personalized medicine studied by Bastani and Bayati [22] in which the algorithms is presented with patients who are described in terms of feature vectors and needs to decide on the dosage of a certain medication. The right dosage for each patient might depend on age, gender, medical history along with various other features. After prescribing a certain dosage, the algorithm only observes if the patient was underdosed or overdosed.

Chapter organization The remainder of the chapter is organized as follows. In Section 4.2 we define the contextual search problem and related notions. In Section 4.3 we review what is known about this problem in one dimension (where contexts are meaningless), specifically the $O(\log \log T)$ regret algorithm of Leighton and Kleinberg for the dynamic pricing problem and the corresponding $\Omega(\log \log T)$ lower bound. In

Section 4.4, we present our algorithms for the specific case where $d = 2$, where the relevant intrinsic volumes are just the area and perimeter, and where the proofs of correctness require no more than elementary geometry (and the 2-dimensional isoperimetric inequality). In Section 4.5, we define intrinsic volumes formally and introduce all relevant necessary facts. In Section 4.6, we present our two main algorithms in their general form, prove upper bounds on their regret, and argue that they can be implemented efficiently in randomized polynomial time. In Section 4.7, we consider simple halving algorithms (such as those that always halve the width or volume of the current knowledge set) and analyze their regret using our tools from integral geometry. Finally in Section 4.8 we discuss how to generalize our algorithms to other loss functions.

4.2 Preliminaries

4.2.1 Contextual Search

We define the *contextual search problem* as a game between between a *learner* and an *adversary*. The adversary begins by choosing a point $v \in [0,1]^d$. Then, every round for T rounds, the adversary chooses a context represented by an unit vector $u_t \in \mathbb{R}^d$ and gives it to the learner. The learner must then choose a value $p_t \in \mathbb{R}$, whereupon the learner accumulates regret $\ell(\langle u_t, v\rangle, p_t)$ (for some loss function $\ell(\cdot,\cdot)$) and learns whether $p_t \le \langle u_t, v\rangle$ or $p_t \ge \langle u_t, v\rangle$. The goal of the learner is to minimize their total regret, which is equal to the sum of their losses over all time periods: $\mathsf{Reg} = \sum_t \ell(\langle u_t, v\rangle, p_t)$.

We primarily consider two loss functions:

Symmetric loss. The symmetric loss measures the absolute value between the guess and the actual dot product, i.e,

$$\ell(\theta, p) = |\theta - p|.$$

Alternatively, $\ell(\theta, p)$ can be thought of as the distance between the learner's hyperplane $H_t := \{x \in \mathbb{R}^d; \langle u_t, x\rangle = p_t\}$ and the adversary's point v.

Pricing loss. The pricing loss corresponds to the revenue loss by pricing an item at p when the buyer's value is θ. If a price $p \leq \theta$ the product is sold with revenue p, so the loss with respect to the optimal revenue θ is $\theta - p$. If the price is $p > \theta$, the product is not sold and the revenue is zero, generating loss θ. In other words,

$$\ell(\theta, p) = \theta - p\mathbf{1}\{p \leq \theta\}.$$

The pricing loss function is highly asymmetric: underpricing by ϵ can only cause the revenue to decrease by ϵ while overpricing by ϵ can cause the item not to be sold generating a large loss.

4.2.2 Notation and framework

The algorithms we consider will keep track of a *knowledge set* $S_t \subseteq S_1 := [0, 1]^d$, which will be the set of vectors v consistent with all observations so far. In step t if the context is u_t and the guess is p_t, the algorithm will update S_{t+1} to $S_t^+(p_t; u_t)$ or $S_t^-(p_t; u_t)$ depending on the feedback obtained, where:

$$S_t^+(p_t; u_t) := \{x \in S_t; \langle u_t, x\rangle \geq p_t\} \quad \text{and } S_t^-(p_t; u_t) := \{x \in S_t; \langle u_t, x\rangle \leq p_t\}$$

Since S_1 is originally a convex set and since S_{t+1} is always obtained from S_t by intersecting it with a halfspace, our knowledge set S_t will remain convex for all t.

Given context u_t in round t, we let $\underline{p}_t$ and $\overline{p}_t$ be the minimum and maximum (respectively) of the dot product $\langle u_t, x \rangle$ that is consistent with S_t:

$$\underline{p}_t = \min_{x \in S_t} \langle u_t, x \rangle \quad \text{and} \quad \overline{p}_t = \max_{x \in S_t} \langle u_t, x \rangle$$

Finally, given a set S and an unit vector u we will define the width in the direcion u as

$$\mathsf{width}(S; u) = \max_{x \in S} \langle u, x \rangle - \min_{x \in S} \langle u, x \rangle.$$

We will consider strategies for the learner that map the current knowledge set S_t and context u_t to guesses p_t. In Algorithm 4 we summarize our general setup.

Algorithm 4 Contextual search framework

1: Adversary selects $v \in S_1 = [0, 1]^d$
2: **for** $t = 1$ to T **do**
3: Learner receives a unit vector $u_t \in \mathbb{R}^d$, $\|u_t\| = 1$.
4: Learner selects $p_t \in \mathbb{R}$ and incurs loss $\ell(\langle u_t, v \rangle, p_t)$.
5: Learner receives feedback and learns the sign of $\langle u_t, x \rangle - p_t$.
6: Learner updates S_{t+1} to $S_t^+(p_t; u_t)$ or $S_t^-(p_t; u_t)$ accordingly.
7: **end for**

Oftentimes, we will want to think of d as fixed, and consider only the asymptotic dependence on T of some quantity (e.g. the regret of some algorithm). We will use the notation $O_d(\cdot)$ and $\Omega_d(\cdot)$ to hide the dependency on d.

4.3 One dimensional case and lower bounds

In the one dimensional case, contexts are meaningless and the adversary only gets to choose the unknown parameter $v \in [0, 1]$. Here algorithms which achieve optimal regret (up to constant factors) are known for both the symmetic loss and the pricing

loss. We review them here both as a warmup for the multi-dimensional version and as a way to obtain lower bounds for the multi-dimensional problem.

For the symmetric loss function, binary search gives constant regret. If the learner keeps an interval S_t of all the values of v that are consistent with the feedback received and in each step guesses the midpoint, then the loss $\ell_t \leq |S_t| = 2^{-t}|S_1|$. Therefore the total regret $\mathsf{Reg} = \sum_t \ell_t = O(1)$.

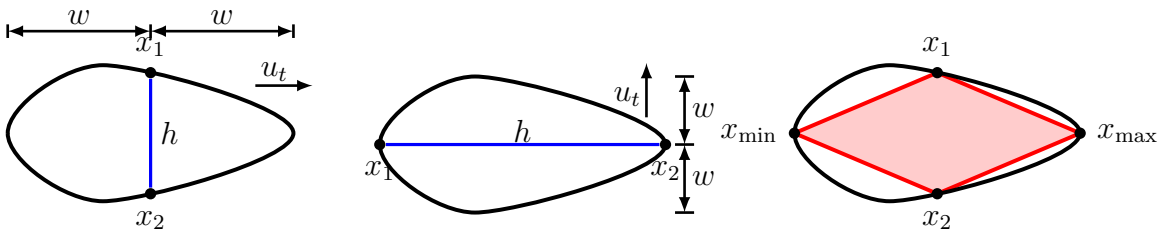
For the pricing loss, one reasonable algorithm is to perform $\log T$ steps of binary search, obtain an interval containing v of length $1/T$ and price at the lower end of this interval. This algorithm gives the learner regret $O(\log T)$. Kleinberg and Leighton [90] provide a surprising algorithm that exponentially improves upon this regret. Their policy biases the search towards lower prices to guarantee that if at some point the price p_t is above v, then the length of the interval S_t decreases by a large factor.

Kleinberg and Leighton's algorithm works as follows. At all rounds, they maintain a knowledge set $S_t = [a_t, a_t + \Delta_t]$. If $\Delta_t > 1/T$, they choose the price $p_t = a_t + 1/2^{2^{k_t}}$ where $k_t = \lfloor 1 + \log_2 \log_2 \Delta_t^{-1} \rfloor$ (this is approximately equivalent to choosing $p_t = a_t + \Delta_t^2$). Otherwise (if $\Delta_t \leq 1/T$), they set their price equal to a_t. (In Appendix C.1 we present their analysis of this algorithm.) Moreover, they show that this bound is tight up to constant factors:

Theorem 4.3.1 (Kleinberg and Leighton [90])**.** *The optimal regret for the contextual search problem with pricing loss in one dimension is* $\Theta(\log \log T)$.

Their result implies a lower bound for the d-dimensional problem. If the adversary only uses coordinate vectors $e_i = (0 \dots 010 \dots 0)$ as contexts, then the problem reduces to d independent instances of the one dimensional pricing problem.

Corollary 4.3.2. *Any algorithm for the d-dimensional contextual search problem with pricing loss must incur $\Omega(d \log \log T)$ regret.*



(a) $w \geq h$: midpoint cut (b) $w < h$: half area cut (c) area $\geq wh$

Figure 4.1

4.4 Two dimensional case

We start by showing how to obtain optimal regret for both loss functions in the two dimensional case. We highlight this special case since it is simple to visualize and conveys the geometric intuition for the general case. Moreover, it can be explained using only elementary plane geometry.

4.4.1 Symmetric loss

Our general approach will be to maintain a potential function of the current knowledge set which decreases each round by an amount proportional to the loss. Since at each time t, the loss is bounded by the width $\mathsf{width}(S_t, u_t)$ of the knowledge S_t in direction u_t, it suffices to show that our potential function decreases each round by some amount proportional to the width of the current knowledge set.

What should we pick as our potential function? Inspired by the one-dimensional case, where one can take the potential function to be the length of the current interval, a natural candidate for the potential function is the area of the current knowledge set. Unfortunately, this does not work; if the knowledge set is long in the direction of u_t and skinny in the perpendicular direction (e.g. Figure 4.1a), then it can have large width but arbitrarily small area.

Ultimately we want to make the width of the knowledge set as small as possible in any given direction. This motivates a second choice of potential function: the average

width of the knowledge set, i.e., $\frac{1}{2\pi}\int_0^{2\pi} \mathsf{width}(S_t, u_\theta)d\theta$ where $u_\theta = (\cos\theta, \sin\theta)$. A result of Cauchy (see Section 5.5 in [89]) shows that the average width of a convex 2-dimensional shape is proportional to the perimeter, so this potential function can alternately be thought of as the perimeter of the knowledge set.

Unfortunately, this too does not quite work. Now, if the set S_t is thin in the direction u_t and long in the perpendicular direction (e.g. Figure 4.1b), any cut will result in a negligible decrease in perimeter (in particular, the perimeter decreases by $O(w^2)$ instead of $\Theta(w)$).

This motivates us to consider an algorithm that keeps track of two potential functions: the perimeter P_t, and the square root of the area $\sqrt{A_t}$. Each iteration, the algorithm will (depending on the shape of the knowledge set) choose one of these two potentials to make progress in. If S_t is long in the u_t direction, cutting it through the midpoint will allow us to decrease the perimeter by an amount proportional to the loss incurred (Figure 4.1a). If S_t is thin in the u_t direction, then we can charge the loss to the square root of the area (Figure 4.1b).

In Algorithm 5 we describe how to compute the guess p_t from the knowledge set S_t and u_t. Recall that the full setup together with how knowledge sets are updated is defined in Algorithm 4.

Algorithm 5 2D-SymmetricSearch

$w = \frac{1}{2}(\overline{p}_t - \underline{p}_t)$ and $p_t^{\mathsf{mid}} = \frac{1}{2}(\overline{p}_t + \underline{p}_t)$
$h =$ length of the segment $S_t \cap \{x; \langle u_t, x\rangle = p_t^{\mathsf{mid}}\}$
if $w \geq h$ **then**
    set $p_t = p_t^{\mathsf{mid}}$
else
    set p_t such that $\mathsf{Area}(S_t^+) = \mathsf{Area}(S_t^-)$
end if

Theorem 4.4.1. *The 2D-SymmetricSearch algorithm (Algorithm 5) has regret bounded by* $8 + 2\sqrt{2}$ *for the symmetric loss.*

Proof. We will keep track of the perimeter P_t and the area A_t of the knowledge set S_t and consider the potential function $\Phi_t = P_t + \sqrt{A_t}/C$, where the constant $C = (1 - \sqrt{1/2})/2$. We will show that every round this potential decreases by at least the regret we incur that round:

$$\Phi_t - \Phi_{t+1} \geq |\langle u_t, v\rangle - p_t|.$$

This implies that the total regret is bounded by $\mathsf{Reg} \leq \Phi_1 = 4 + 2/(1 - \sqrt{1/2}) = 8 + 2\sqrt{2}$. We will write ℓ_t as shorthand for the loss $\ell(\langle u_t, v\rangle, p_t) = |\langle u_t, v\rangle - p_t|$ at time t.

We first note that both P_t and A_t are decreasing in t. This follows from the fact that S_{t+1} is a convex subset of S_t. We will show that when $w \geq h$, P_t decreases by at least ℓ_t, whereas when $w < h$, $\sqrt{A_t}$ decreases by at least ℓ_t.

Case $w \geq h$. In this case, $p_t = p_t^{\mathsf{mid}}$. We claim here that $P_t - P_{t+1} \geq w$. To see this, let x_1 and x_2 be the two endpoints of the line segment $S_t \cap \{x; \langle u_t, x\rangle = p_t^{\mathsf{mid}}\}$ (so that $h = \|x_1 - x_2\|$). Without loss of generality, assume $S_{t+1} = S_t^-$ (the other case is analogous).

Note that the boundary of S_{t+1} is the same as the boundary of S_t, with the exception that the segment of the boundary of S_t in the half-space $\{x; \langle u_t, x\rangle \geq p_t^{\mathsf{mid}}\}$ has been replaced with the line segment $\overline{x_1 x_2}$. The part of boundary of S_t in the halfspace $\{\langle u_t, x\rangle \geq p_{mid}\}$, reach some point on the line $\langle u_t, x\rangle = \overline{p}_t$, and return to x_2. Since $\overline{p}_t - p_t^{\mathsf{mid}} = w$, any such path must have length at least $2w$. From the fact that $w \geq h$, it follows that:

$$P_t - P_{t+1} \geq 2w - h \geq 2w - w \geq w \geq \ell_t.$$

Case $w < h$. We set p_t such that $A_{t+1} = A_t/2$, so

$$\Phi_t - \Phi_{t+1} \geq (\sqrt{A_t} - \sqrt{A_{t+1}})/C \geq \sqrt{A_t}(1 - \sqrt{1/2})/C = 2\sqrt{A_t}.$$

If we argue that $2\sqrt{A_t} \geq 2w \geq \ell_t$ we are done. To show this, define x_1 and x_2 as before, and let $x_{\min}$ be a point in S_t that satisfies $\langle u_t, x_{min} \rangle = \underline{p}_t$, and likewise let $x_{\max}$ be a point in S_t that satisfies $\langle u_t, x_{max} \rangle = \overline{p}_t$. Since S_t is convex, it contains the two triangles with endpoints (x_1, x_2, x_{max}) and (x_1, x_2, x_{min}), see Figure 4.1c. But these two triangles are disjoint and each have area at least $\frac{1}{2}wh$, so $A_t \geq wh \geq w^2$, since $h > w$. It follows that $\sqrt{A_t} \geq w$.

□

4.4.2 Pricing loss

To minimize regret for pricing loss, we want to somehow combine our above insight of looking at both the area and the perimeter with Leighton and Kleinberg's bucketing procedure for the 1D case. At first we might try to do this bucketing procedure just with the area.

That is, if the area of the current knowledge set belongs to the interval $[\Delta^2, \Delta]$, choose a price that carves off a subset of total area Δ^2. Now, if you overprice, you incur $O(1)$ regret, but the area of your knowledge set shrinks to Δ^2 (and belongs to a new "bucket"). On the other hand, if you underprice, your area decreases by at most Δ, so you can underprice at most Δ^{-1} times. If the regret you incurred this round was at most $O(\Delta)$, then this means that you would incur at most $O(1)$ total regret underpricing while your area belongs to this interval.

This would be true if the regret per round was at most the area of the knowledge set. Unfortunately, as noted earlier, this is not true; the regret per round scales with the width of the knowledge set, not the area, and you can have knowledge sets

with large width and small area. The trick, as before, is to look at both area and perimeter, and argue that at each step the bucketization argument for at least one of these quantities holds.

More specifically, let P_t again be the perimeter of the knowledge set at time t, A_t be the area of the knowledge set at time t, and let $A'_t = 2\sqrt{\pi A_t}$ be a normalization of A_t. Note that by the isoperimetric inequality for dimensions (which says that of all shapes with a given area, the circle has the least perimeter), we have that $P_t \geq A'_t$.

Let $\ell_k = \exp(-1.5^k)$. Our buckets will be the intervals $(\ell_{k+1}, \ell_k]$. We will define the function $\mathsf{bkt}(x) = k$ if $x \in (\ell_{k+1}, \ell_k]$. Our algorithm is described in Algorithm 6. The analysis follows.

Algorithm 6 2D-PricingSearch

$w = (\overline{p}_t - \underline{p}_t)$
$h_{max} =$ maximum length of a segment of the form $S_t \cap \{x; \langle u_t, x\rangle = p\}$
if $w < 1/T$ **then**
    choose $p_t = \underline{p}_t$.
else if $\mathsf{bkt}(A'_t) = \mathsf{bkt}(P_t)$ **then**
    set p_t such that $\mathsf{Area}(S_t^-) = \ell^2_{\mathsf{bkt}(A'_t)+1}/4\pi$.
else if $\mathsf{bkt}(A'_t) > \mathsf{bkt}(P_t)$ and $w < h_{max}$ **then**
    set p_t such that $\mathsf{Area}(S_t^-) = \ell^2_{\mathsf{bkt}(A'_t)+1}/4\pi$.
else if $\mathsf{bkt}(A'_t) > \mathsf{bkt}(P_t)$ and $w \geq h_{max}$ **then**
    set p_t such that $\mathsf{Perimeter}(S_t) - \mathsf{Perimeter}(S_t^+) = \frac{1}{2}\ell_{\mathsf{bkt}(P_t)+1}$.
end if

Theorem 4.4.2. *The 2D-PricingSearch algorithm (Algorithm 6) has regret bounded by $O(\log\log T)$ for the pricing loss.*

Proof. We will divide the behavior of the algorithm into four cases (depending on which branch of the if statement in Algorithm 6 is taken), and argue the total regret sustained in each case is at most $O(\log\log T)$.

- *Case 1: $w \leq 1/T$.* Whenever this happens, we pick the minimum possible price in our convex set, so we definitely underprice and sustain regret at most $1/T$. The total regret sustained in this case over T rounds is therefore at most 1.

- *Case 2:* $\mathsf{bkt}(A'_t) = \mathsf{bkt}(P_t)$. Let $\mathsf{bkt}(P_t) = \mathsf{bkt}(A'_t) = k$. Note that $k \leq O(\log \log T)$, or else we would be in Case 1 (if $P_t < 1/T$, then $w < 1/T$). Therefore, fix k; we will show that the total regret we incur for this value of k is at most $O(1)$. Summing over all $O(\log \log T)$ values of k gives us our upper bound.

  If we overprice and do not make a sale, this contributes regret at most 1, but then $A'_{t+1} = \ell_{k+1}$ and we leave this bucket. If we underprice, our regret is at most $w \leq P_t \leq \ell_k$, and the decrease in area $A_{t+1} - A_t = \ell_{k+1}^2/4\pi$. It follows that we can underprice at most $4\pi\ell_k^2/\ell_{k+1}^2$ before leaving this bucket, and therefore we incur total regret at most

$$4\pi \frac{\ell_k^2}{\ell_{k+1}^2} \ell_k = 4\pi \exp(-3 \cdot 1.5^k + 2 \cdot 1.5^{k+1}) = 4\pi = O(1).$$

- *Case 3:* $\mathsf{bkt}(A'_t) > \mathsf{bkt}(P_t)$ *and* $w < h_{max}$. Let $\mathsf{bkt}(A'_t) = r$ and $\mathsf{bkt}(P_t) = k$. As in case 2, we will fix r and argue that total regret we incur for this r is at most $O(1)$. As before, if we overprice, A'_{t+1} becomes ℓ_{r+1}, so we incur total regret at most $O(1)$ from overpricing.

  Now, note that since $w < h_{max}$, S_t contains two disjoint triangles with base h_{max} and combined height w (see Figure 4.1c), so $A_t \geq w h_{max} > w^2$, and therefore $w < \sqrt{A_t} = A'_t/2\sqrt{\pi}$. Therefore, if we underprice, we incur regret at most $A'_t/2\sqrt{\pi} \leq \ell_r$. As before, the area decreases by at least $A_{t+1} - A_t = \ell_{r+1}^2/4\pi$ if we underprice, so we underprice at most $4\pi\ell_k^2/\ell_{k+1}^2$ before leaving this bucket, and therefore we incur total regret at most

$$4\pi \frac{\ell_r^2}{\ell_{r+1}^2} \ell_r = 4\pi \exp(-3 \cdot 1.5^r + 2 \cdot 1.5^{r+1}) = 4\pi = O(1).$$

- *Case 4:* $\mathsf{bkt}(A'_t) > \mathsf{bkt}(P_t)$ *and* $w \geq h_{max}$. Let $\mathsf{bkt}(A'_t) = r$ and $\mathsf{bkt}(P_t) = k$, and fix k. Now $A_t \geq w h_{max} \geq h_{max}^2$, so $h_{max} < A'_t/2\sqrt{\pi} \leq \ell_r/2\sqrt{\pi}$.

  First, note that it is in fact possible to set p_t so that $\mathsf{Perimeter}(S_t) - \mathsf{Perimeter}(S_t^+) = \frac{1}{2}\ell_{k+1}$, since the total perimeter is at least ℓ_{k+1}, so this corresponds to just cutting off a chunk (S_t^+) of S_t of perimeter $\mathsf{Perimeter}(S_t) - \frac{1}{2}\ell_{k+1}$ (which is possible since this is nonnegative and less than $\mathsf{Perimeter}(S_t)$).

  If we overprice, then the perimeter of the new region (S_t^-) is equal to $\mathsf{Perimeter}(S_t) - \mathsf{Perimeter}(S_t^+) = \ell_{k+1}/2$, plus the length of the segment formed by the intersection of S_t with $\{x; \langle u_t, x\rangle = p_t\rangle$. The length of this segment is at most h_{max}, so the perimeter is at most $\ell_{k+1}/2 + \ell_r/2\sqrt{\pi} \leq \ell_{k+1}/2 + \ell_{k+1}/2\sqrt{\pi} \leq \ell_{k+1}$. This means we can overprice at most once before the perimeter changes buckets, and we thus incur at most $O(1)$ regret due to overpricing.

  If we underprice, then we incur regret at most $w \leq P_t \leq \ell_k$, and the perimeter of the new region decreases by at least $\ell_{k+1}/2 - \ell_r/2\sqrt{\pi} \geq \ell_{k+1}/5$. This means we can underprice at most $5\ell_k/\ell_{k+1}$ times before we switch buckets, and we incur total regret at most

$$5\frac{\ell_k}{\ell_{k+1}}\ell_k = 4\exp(-2 \cdot 1.5^r + 1.5^{r+1}) \leq 4 = O(1).$$

□

4.5 Interlude: Intrinsic Volumes

The main idea in the two dimensional case was to balance between making progress in a two-dimensional measure of the knowledge set (the area) and in a one-dimensional measure (the perimeter). To generalize this idea to higher dimensions we will keep

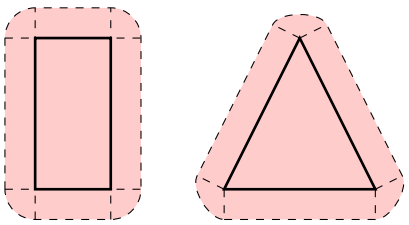


Figure 4.2: Steiner's formula for 2D: $\mathsf{Area}(\mathsf{K}+\epsilon\mathsf{B}) = \mathsf{Area}(\mathsf{K}) + \mathsf{Perimeter}(\mathsf{K})\cdot\epsilon + \pi\epsilon^2$

track of d potential functions, each corresponding to a j-dimensional measure of the knowledge set for each j in $\{1, 2, \ldots d\}$.

Luckily for us, one of the central objects of study in *integral geometry* (also known as *geometric probability*) corresponds exactly to a j-dimensional measure of a d-dimensional object. Many readers are undoubtedly familiar with two of these measures, namely volume (the d-dimensional measure) and surface area (the $(d-1)$-dimensional area) but it is less clear how to define the 1-dimensional measure of a three-dimensional convex set (indeed, Shanuel [126] jokingly calls his lecture notes on the topic "What is the length of a potato?"). These measures are known as *intrinsic volumes* and they match our intuition for how a j-dimensional measure of a d-dimensional set should behave (in particular, reducing to the regular j-dimensional volume as the set approaches a j-dimensional object).

We now present a formal definition of intrinsic volumes and summarize their most important properties. We refer to the excellent book by Klain and Rota [89] for a comprehensive introduction to integral geometry.

Intrinsic volumes can be defined as the coefficients that arise in Steiner's formula for the the volume of the (Minkowski) sum of a convex set $K \subseteq \mathbb{R}^d$ and an unit ball B. Steiner [126] shows that the $\mathsf{Vol}(K+\varepsilon B)$ is a polynomial in ϵ and the intrinsic volumes $V_j(K)$ are the (normalized) coefficients of this polynomial:

$$\mathsf{Vol}(K+\varepsilon B) = \sum_{j=0}^{d} \kappa_{d-j} V_j(K) \varepsilon^{d-j} \tag{4.1}$$

where κ_{d-j} is the volume of the $(d-j)$-dimensional unit ball. An useful exercise to get intuition about intrinsic volumes is to directly compute the intrinsic volumes of the parallelotope $[0, a_1] \times [0, a_2] \times [0, a_d]$. It is easy to check for $d = 2$ and 3 (see Figure 4.2) that $V_d = a_1 a_2 \ldots a_d$, $V_1 = a_1 + a_2 + \ldots + a_d$ and $V_0 = 1$. More generally V_j correspnds to the symmetric polynomial of degree j: $V_j = \sum_{S \subseteq [d]; |S|=j} a_S$ where $a_S = \prod_{s \in S} a_s$. In particular for $[0,1]^d$ the j-th intrinsic volume is $V_j = \binom{d}{j}$.

Definition 4.5.1 (Valuations)**.** *Let* Conv_d *be the class of compact convex bodies in* $\mathbb{R}^d$*. A* valuation *is a map* $\nu : \mathsf{Conv}_d \to \mathbb{R}$ *such that* $\nu(\emptyset) = 0$ *and for every* $S_1, S_2 \in \mathsf{Conv}_d$ *satisfying* $S_1 \cup S_2 \in \mathsf{Conv}_d$ *it holds that*

$$\nu(S_1 \cup S_2) + \nu(S_1 \cap S_2) = \nu(S_1) + \nu(S_2).$$

A valuation is said to be monotone *if* $\nu(S) \leq \nu(S')$ *whenever* $S \subseteq S'$*. A valuation is said to be* non-negative *is* $\nu(S) \geq 0$*. Finally, a valuation is* rigid *if* $\nu(S) = \nu(T(S))$ *for every rigid motion (i.e. rotations and translations)* T *of* $\mathbb{R}^d$*.*

To define what it means for a valuation to be continuous, we need a notion of distance betweeen convex sets. We define the Hausdorff distance $\delta(K, L)$ between two sets $K, L \in \mathsf{Conv}_d$ to be the the minimum ϵ such that $K + \epsilon B \subseteq L$ and $L + \epsilon B \subseteq K$ where B is the unit ball. This notion of distance allows us to define limits: a sequence $K_t \in \mathsf{Conv}_d$ converges to $K \in \mathsf{Conv}_d$ (we write this as $K_t \to K$) if $\delta(K_t, K) \to 0$. Continuity can now be defined in the natural way.

Definition 4.5.2 (Continuity)**.** *A valuation function* ν *is* continuous *if whenever* $K_t \to K$ *then* $\nu(K_t) \to \nu(K)$*.*

Theorem 4.5.3. *The intrinsic volumes are non-negative monotone continuous rigid valuations.*

In fact the intrinsic volumes are quite special since they form a basis for the set of all valuations with this property. This constitutes the fundamental result of the field of integral geometry:

Theorem 4.5.4 (Hadwiger)**.** *If ν is a continuous rigid valuation of Conv_d, then there are constants $c_0, \ldots, c_d$ such that $\nu = \sum_{i=0}^{d} c_i V_i$.*

Next we describe a few important properties of intrinsic valuations that will be useful in the analysis of our algorithms:

Theorem 4.5.5 (Homogeneity)**.** *The map V_j is j-homogenous, i.e., $V_j(\alpha K) = \alpha^j V_j(K)$ for any $\alpha \in \mathbb{R}_{\geq 0}$.*

Theorem 4.5.6 (Ambient independence)**.** *Intrinsic volumes are independent of the ambient space, i.e, if $K \in \mathsf{Conv}_d$ and K' is a copy of K embedded in a larger dimensional space*

$$K' = T(\{(x, 0_k); x \in K, 0_k \in \mathbb{R}^k\} \in \mathsf{Conv}_{d+k}$$

for a rigid transformation T, then for any $j \leq d$, we have $V_j(K) = V_j(K')$.

We now provide an inequality between intrinsic volumes which we will use later to derive an isoperimetric inequality for intrinsic volumes. The following inequality is a consequence of the Alexandrov-Fenchel inequality due to McMullen [105].

Theorem 4.5.7 (Inequality on intrinsic volumes)**.** *If $S \in \mathsf{Conv}_d$ and any $i \geq 1$ then*

$$V_i(S)^2 \geq \frac{i+1}{i} V_{i-1}(S) V_{i+1}(S).$$

One beautiful consequence of Hadwiger's theorem is a probabilistic interpretation of intrinsic volumes as the expected volume of the projection of a set onto a random subspace. To make this precise, define the Grassmannian $\mathrm{Gr}(d, k)$ as the collection of all k-dimensional linear subspaces of $\mathbb{R}^d$. The *Haar measure* on the Grassmannian

is the unique probability measure on $\mathrm{Gr}(d,k)$ that is invariant under rotations in $\mathbb{R}^d$ (i.e., $SO(\mathbb{R}^d)$).

Theorem 4.5.8 (Random Projections)**.** *For any $K \in \mathsf{Conv}_d$, the j-th intrinsic volume*

$$V_j(K) = \mathbb{E}_{H \sim \mathrm{Gr}(d,k)}[\mathsf{Vol}(\pi_H(K))]$$

where $H \sim \mathrm{Gr}(d,k)$ is a k-dimensional subspace H sampled according to the Haar measure, π_H is the projection on H and $\mathsf{Vol}(\pi_H(K))$ is the usual (k-dimensional) volume on H.

A remark on notation: we use Vol to denote the standard notion of volume and V_j to denote intrinsic volumes. When analyzing an object in a d-dimensional (sub)space, then $\mathsf{Vol} = V_d$.

4.6 Higher dimensions

In this section, we generalize our algorithms from Section 4.4 from the two-dimensional case to the general multi-dimensional case.

Both results require as a central component lower bounds on the intrinsic volumes of high dimensional cones. These bounds relate the intrinsic volume of a cone to the product of the cone's height and the intrinsic volume of the cone's base (a sort of "Fubini's theorem" for intrinsic volumes).

More formally, a *cone* S in $\mathbb{R}^{d+1}$ is the convex hull of a d-dimensional convex set K and a point $p \in \mathbb{R}^{d+1}$. If the distance from p to the affine subspace containing K is h, we say the cone has *height* h and *base* K. The lemma we require is the following.

Lemma 4.6.1 (Cone Lemma)**.** *Let K be a convex set in $\mathbb{R}^d$, and let S be a cone in $\mathbb{R}^{d+1}$ with base K and height h. Then, for all $0 \leq j \leq d$,*

$$V_{j+1}(S) \geq \frac{1}{j+1} h V_j(K).$$

In the two-dimensional case, this lemma manifests itself when we use the fact that the perimeter of a convex set with height h is at least h. We note that when $j = d$, Lemma 4.6.1 holds with equality and is a simple exercise in elementary calculus. On the other hand, when $0 \leq j < d$, there is no straightforward formula for the $(j+1)$-th intrinsic volume of a set in terms of the j-th intrinsic volume of its cross sections.

We begin by taking the Cone Lemma as true, and discuss how to use it to generalize our contextual search algorithms to higher dimensions in Sections 4.6.1 (for symmetric loss) and 4.6.2 (for pricing loss). We then prove the Cone Lemma in Section 4.6.3. Finally, in Section 4.6.4, we argue that both algorithms can be implmented efficiently.

4.6.1 Symmetric loss

In this section we present a $O_d(1)$ regret algorithm for the contextual search problem with symmetric loss in d dimensions. The algorithm, which we call SymmetricSearch, is presented in Algorithm 7.

Recall that in two dimensions, we always managed to choose p_t so that the loss from that round is bounded by the decrease in either the perimeter or the square root of the area. The main idea of Algorithm 7 is to similarly choose p_t such that the loss is bounded by the decrease in one of the intrinsic volumes, appropriately normalized. As before, if the width is large enough, we bound the loss by the decrease in the average width (i.e. the one-dimensional intrinsic volume $V_1(S)$). As the width gets smaller, we charge the loss to progressively higher-dimensional intrinsic volumes.

Constants c_0 through c_{d-1} in Algorithm 7 are defined so that $c_0 = 1$ and $c_i/c_{i-1} = \frac{1}{2i}$. In other words, $c_i = \frac{1}{2^{i-1}i!}$. Constant c_0 is only used in the analysis.

Algorithm 7 SymmetricSearch

1: $w = \frac{1}{2}\mathsf{width}(S_t; u_t)$
2: **for** $i = 1$ to d **do**
3: define $p_i \in \mathbb{R}$ such that $V_i(S_t^+(p_i; u_t)) = V_i(S_t^-(p_i; u_t))$.
4: define $K_i = \{x \in S_t; \langle u_t, x\rangle = p_i\}$.
5: define $L_i = (V_i(K_i)/c_i)^{1/i}$ (set $L_0 = \infty$).
6: **end for**
7: find j such that $L_{j-1} \geq w \geq L_j$
8: set $p_t = p_j$.

We first argue that this algorithm is well-defined:

Lemma 4.6.2. *SymmetricSearch (Algorithm 7) is well-defined, i.e., there is always a choice of p_i and j that satisfies the required properties.*

Proof. We begin by arguing that there exists a p_i such that $V_i(S_t^+(p_i; u_t)) = V_i(S_t^-(p_i; u_t))$. To see this, note that the functions $\phi^+(x) = V_i(S_t^+(x; u_t))$ and $\phi^-(x) = V_i(S_t^-(x; u_t))$ are continuous on $[\underline{p}_t, \overline{p}_t]$ since the intrinsic volumes are continuous with respect to Hausdorff distance (Definition 4.5.2 and Theorem 4.5.3). Moreover, since intrinsic volumes are monotone (Theorem 4.5.3), $\phi^+(x)$ is decreasing and ϕ^- is increasing on this interval. Finally, since $\phi^+(\underline{p}_t) = \phi^-(\overline{p}_t)$, it follows from the Intermediate Value Theorem that there exists a p_i where $\phi^+(p_i) = \phi^-(p_i)$, as desired.

To see that there exists a j such that $L_{j-1} \geq w \geq L_j$, note that $L_d = 0$ since K_d is in a $d-1$-dimensional hyperplane, so the segments $[L_j, L_{j-1})$ for $L_j < L_{j-1}$ cover the entire $[0, \infty)$. It follows that one such interval must contain w. □

Before we proceed to the regret bound, we will show the following two lemmas. The first lemma shows that if we pick j in this manner, then $V_j(S_t)^{1/j}$ will be at least $\Omega(w)$.

Lemma 4.6.3. $V_j(S_t) \geq \frac{1}{j} c_{j-1} w^j$.

Proof. For $j = 1$, note that S_t contains a segment of length $2w$, so $V_1(S_t)$ is at most the 1-dimensional intrinsic volume of that segment, which is exactly $2w$ (Theorem 4.5.6).

For $j > 1$, we known that S_t contains a cone with base K_{j-1} and height w (since the width of S_t is $2w$, there is a point at least distance w from the plane $H_{j-1} = \{x; \langle u_t, x \rangle = p_{j-1}\}$). By Theorem 4.6.1 and the fact that V_j is monotone, this implies that:

$$V_j(S_t) \geq \frac{1}{j} w V_{j-1}(K_{j-1}).$$

Since $w < L_{j-1} = (V_{j-1}(K_{j-1})/c_{j-1})^{1/j-1}$, we have that $V_{j-1}(K_{j-1}) \geq c_{j-1} w^{j-1}$. Substituting this into the previous expression, we obtain the desired result. □

The second lemma shows that if we pick j in this manner, then $V_j(S_{t+1})/V_j(S_t)$ is bounded above by a constant strictly less than 1.

Lemma 4.6.4. $V_j(S_{t+1}) \leq \frac{3}{4} V_j(S_t)$.

Proof. The set S_{t+1} is equal to either $S^+ = S_t^+(p_j; u_t)$ or $S^- = S_t^-(p_j; u_t)$. Our choice of p_j is such that $V_j(S^-) = V_j(S^+)$. Therefore:

$$2V_j(S_{t+1}) = V_j(S^+) + V_j(S^-) = V_j(S^- \cap S^+) + V_j(S^- \cup S^+) = V_j(K_j) + V_j(S_t)$$

To bound $V_j(K_j)$ in terms of $V_j(S_t)$ we observe that $w \geq L_j = (V_j(K_j)/c_j)^{1/j}$ so $V_j(K_j) \leq c_j w^j$. Plugging the previous lemma we get $V_j(K_j) \leq j \frac{c_j}{c_{j-1}} V_j(S_t) = \frac{1}{2} V_j(S_t)$ by the choice of constants. Substituting this inequality into the above equation gives us the desired result. □

Together, these lemmas let us argue that each round, the sum of the normalized intrinsic volumes $V_i(S_t)^{1/i}$ decreases by at least $\Omega(w)$ (and hence the total regret is constant).

Theorem 4.6.5. *The SymmetricSearch algorithm (Algorithm 7) has regret bounded by $O(d^4)$ for the symmetric loss.*

Proof. We will show that for the potential function $\Phi_t = \sum_{i=1}^{d} i^2 V_i(S_t)^{1/i}$ we can always charge the loss to the decrease in potential, i.e., $\Phi_t - \Phi_{t+1} \geq \Omega(w) \geq \Omega(\ell_t)$ and therefore, $\mathsf{Reg} \leq \sum_{t=1}^{\infty} \ell_t \leq O(\Phi_1)$. The initial potential is

$$\Phi_1 = \sum_{i=1}^{d} i^2 V_i([0,1]^d)^{1/i} = \sum_{i=1}^{d} i^2 \binom{d}{i}^{1/i} \leq \sum_{i=1}^{d} i^2 O(d) = O(d^4)$$

Since $V_j(S_t) \geq V_j(S_{t+1})$ by monotonicity, we can bound the potential change by $\Phi_t - \Phi_{t+1} \geq j^2[V_j(S_t)^{1/j} - V_j(S_{t+1})^{1/j}]$. We now show that this last term is $\Omega(w)$:

$$\begin{aligned}
j^2[V_j(S_t)^{1/j} - V_j(S_{t+1})^{1/j}] &\geq j^2 \left(1 - \left(\frac{3}{4}\right)^{1/j}\right) V_j(S_t)^{1/j} \\
&\geq j^2 \left(1 - \left(\frac{3}{4}\right)^{1/j}\right) \left(\frac{c_{j-1}}{j}\right)^{1/j} w \\
&\geq j^2 \left(1 - (1 - \log(4/3)/j)\right) \left(\frac{1}{2^{j-2} j!}\right)^{1/j} w \\
&\geq j^2 \left(\frac{\log(4/3)}{j}\right) \left(\frac{e}{2j}\right) w \\
&\geq \Omega(w).
\end{aligned}$$

Here the first inequality follows from Lemma 4.6.4 and the second from Lemma 4.6.3.

□

4.6.2 Pricing loss

In the 2-dimensional version of the dynamic pricing problem, we decomposed the range of each potential into $O(\log \log T)$ buckets and used the isoperimetric inequality $\sqrt{4\pi A_t} \leq P_t$ to argue that (when suitably normalized), the area always belonged to

a higher bucket than the perimeter. To apply the same idea here, we will apply our inequality on intrinsic volumes (Theorem 4.5.7) to obtain an isoperimetric inequality for intrinsic volumes:

Lemma 4.6.6 (Isoperimetric inequality)**.** *For any $S \in \mathsf{Conv}_d$ and any $i \geq 1$ it holds that*

$$(i!V_i(S))^{1/i} \geq ((i+1)!V_{i+1}(S))^{1/(i+1)}.$$

Proof. We proceed by induction. For $i = 1$, note that Theorem 4.5.7 gives us that $V_1(S)^2 \geq 2V_0(S)V_2(S)$. Since $V_0(S)$ equals 1 for any convex set S, this reduces to $V_1(S) \geq \sqrt{2!V_2(S)}$.

Now assume via the inductive hypothesis that we have proven the claim for all $j \leq i$. From Theorem 4.5.7 we have that

$$\begin{aligned} V_i(S)^2 &\geq \frac{i+1}{i} V_{i-1}(S)V_{i+1}(S) = \frac{i+1}{i!}((i-1)!V_{i-1}(S))V_{i+1}(S) \\ &\geq \frac{i+1}{i!}(i!V_i(S))^{(i-1)/i}V_{i+1}(S) = \frac{1}{i!^{(i+1)/i}}V_i(S)^{(i-1)/i}(i+1)!V_{i+1}(S). \end{aligned}$$

Rearranging, this reduces to $(i!V_i(S))^{(i+1)/i} \geq (i+1)!V_{i+1}(S)$, and therefore $(i!V_i(S))^{1/i} \geq ((i+1)!V_{i+1}(S))^{1/(i+1)}$. □

Inspired by the isoperimetric inequality we will keep track of the following "potentials" (these vary with t, but we will omit the subscript for notational convenience):

$$\varphi_i = (i!V_i(S_t))^{1/i}$$

Since $S_1 = [0,1]^d$, their initial values will be given by $\varphi_i = (i! \cdot \binom{d}{i})^{1/i} < di \leq d^2$. Since those quantities are monotone non-increasing, they will be in the interval $[0, d^2)$. We will divide this interval in ranges of doubly-exponentially decreasing length (as in one

and two dimensions). The ranges will be $(\ell_{k+1}, \ell_k]$ where

$$\ell_k = d^2 \exp(-\alpha^k) \quad \text{for} \quad \alpha = 1 + 1/d$$

To keep track of which range each of our potentials ϕ_i belongs to, define k_i so that $\varphi_i \in (\ell_{k_i+1}, \ell_{k_i}]$. By the isoperimetric inequality we know that:

$$\varphi_1 \geq \varphi_2 \geq \ldots \geq \varphi_d \qquad k_1 \leq k_2 \leq \ldots \leq k_d$$

Recall that in the 2-dimensional case, whenever the perimeter and the area were in the same range, we chose to make progress in the area. To extend this idea to higher dimensions, whenever many ϕ_i belong to the same range and we decide to make progress on that range, we will always choose the largest such φ_i:

$$M(i) = \max\{j; k_i = k_j\}$$

The complete method is summarized in Algorithm 8. As before, constants c_0 through c_{d-1} in Algorithm 8 are defined so that $c_0 = 1$ and $c_i/c_{i-1} = \frac{1}{2i}$. In other words, $c_i = \frac{1}{2^{i-1} i!}$.

We begin by arguing that our algorithm is well-defined. We ask the reader to recall the notation $\underline{p}_t = \min_{x \in S_t} \langle u_t, x \rangle$ and $\overline{p}_t = \max_{x \in S_t} \langle u_t, x \rangle$.

Lemma 4.6.7. *PricingSearch (Algorithm 8) is well defined, i.e., it is always possible to choose p_i and j with the desired properties.*

Proof. For the choice of p_i, if $V_i(S_t) - V_i(S_t^+(\overline{p}_t; u_t)) > \ell_{k_i+1}^i/(2 \cdot i!)$, then the function $\phi_i : [\underline{p}_t, \overline{p}_t] \to \mathbb{R}$, $\phi_i(p) = V_i(S_t) - V_i(S_t^+(p; u_t))$ is continuous and monotone with $\phi_i(\underline{p}_t) = 0$ and $\phi_i(\overline{p}_t) > \frac{1}{2} i! \cdot \ell_{k_i+1}^i$ so this guarantees the existence of such p_i.

For the choice of j, let $0 = i_0 < i_1 < \ldots < i_a = d$ be the indices i such that $M(i) = i$. Notice that the intervals $[L_{i_{s+1}}, L_{i_s})$ are of the form $[L_{M(i)}, L_{i-1})$ for

Algorithm 8 PricingSearch

$w = \frac{1}{2}\mathsf{width}(S_t; u_t)$
for $i = 1$ to d **do**
  let $\varphi_i = (i! \cdot V_i(S_t))^{1/i}$ and k_i such that $\varphi_i \in (\ell_{k_i+1}, \ell_{k_i}]$
  if $V_i(S_t) - V_i(S_t^+(\overline{p}_t; u_t)) > \ell^i_{k_i+1}/(2 \cdot i!)$ **then**
    choose p_i such that $V_i(S_t) - V_i(S_t^+(p_i; u_t)) = \ell^i_{k_i+1}/(2 \cdot i!)$
  else
    choose $p_i = \overline{p}_t$
  end if
  define $K_i = \{x \in S_t; \langle u_t, x\rangle = p_i\}$
  define $L_i = (V_i(K_i)/c_i)^{1/i}$ (define $L_0 = \infty$)
end for
if $w < 1/T$ **then**
  set $p_t = \underline{p}_t$.
else
  let $M(i) = \max\{j; k_i = k_j\}$
  find a j such that $L_{j-1} \geq w \geq L_{M(j)}$
  let $J = M(j)$ and set $p_t = p_J$.
end if

$i = i_s + 1$. Finally notice that the intervals $[L_{i_{s+1}}, L_{i_s})$ cover the entire interval $[L_d, L_0) = [0, \infty)$ so one of them must contain w. □

Before we proceed to the main analysis, we begin by proving a couple of lemmas regarding the ranges $(\ell_{k+1}, \ell_k]$ of the intrinsic volumes before and after each iteration. The first lemma says that if we overprice (i.e. S_t^- is chosen) the quantity φ_J jumps from the range $[\ell_{k_J+1}, \ell_{k_J})$ to the next range $(\ell_{k_J+2}, \ell_{k_J+1}]$.

Lemma 4.6.8. $[J! \cdot V_J(S_t^-(p_J; u_t))]^{1/J} \leq \ell_{k_J+1}$

Proof. We abbreviate $S_t^-(p_J; u_t)$ and $S_t^+(p_J; u_t)$ by S^- and S^+ respectively. Using the fact that V_J is a valuation and that $S^- \cap S^+ = K_J$ we have that:

$$V_J(S^-) = V_J(S_t) - V_J(S^+) + V_J(K_J) \leq \ell^J_{k_J+1}/(2 \cdot J!) + V_J(K_J)$$

It remains to show that $V_J(K_J) \leq \ell_{k_J+1}^J/(2 \cdot J!)$. To do this, we will again use the Cone Lemma to obtain the following inequalities:

$$\frac{1}{J+1} V_J(K_J) w \leq V_{J+1}(S_t) \leq \frac{1}{(J+1)!} \ell_{k_{(J+1)}}^{J+1} \leq \frac{1}{(J+1)!} \ell_{(k_J)+1}^{J+1}$$

The first inequality is the Cone Lemma (Lemma 4.6.1) applied to the fact that S_t contains a cone of base K_J and height at least w. The second inequality comes from the definition of k_J and the third inequality comes from the fact that $J = M(J)$ so $k_{J+1} \geq k_J + 1$.

Finally, observe that because of our choice of J, $w \geq L_J = (V_J(K_J)/c_J)^{1/J}$. Substituting in the previous equation we obtain:

$$\frac{1}{J+1} V_J(K_J)^{(J+1)/J} (c_J)^{-1/J} \leq \frac{1}{(J+1)!} \ell_{(k_J)+1}^{J+1}$$

Substituting the definition of c_J and simplifying, we get the desired bound of $V_J(K_J) \leq \ell_{k_J+1}^J/(2 \cdot J!)$. □

We next show that, for our chosen J, if we underprice, then the Jth intrinsic volume of our knowledge set decreases by at least $\ell_{k_J+1}^J$. This will allow us to bound the number of times we can potentially underprice before k_J changes (in particular, it is at most $2\ell_{k_J}^J/\ell_{k_J+1}^J$).

Lemma 4.6.9. $V_J(S_t) - V_J(S_t^+(p_J; u_t)) = \ell_{k_J+1}^J/(2 \cdot J!)$

Proof. Note that this equality is guaranteed by the algorithm's choice of p_J, except when $V_J(S_t) - V_J(S_t^+(\overline{p}_t; u_t)) < \ell_{k_J+1}^J/(2 \cdot J!)$ and $p_J = \overline{p}_t$. However, in this case, $S_t^-(\overline{p}_t; u_t) = S_t$ by the definition of $\overline{p}_t$. Lemma 4.6.8 then implies that $[J! \cdot V_J(S_t)]^{1/J} \leq \ell_{k_J+1}$, but this contradicts the definition of k_J. □

We now show that in each round, the width of the knowledge set (and thus our loss) is at most $2\ell_{k_J}$.

Lemma 4.6.10. $w \leq 2\ell_{k_J}$

Proof. We will derive both an upper and lower bound on $V_j(S_t)$. For the upper bound we again apply the Cone Lemma (Lemma 4.6.1).

$$V_j(S_t) \geq \frac{1}{j} V_{j-1}(K_{j-1}) w \geq \frac{1}{j}(c_{j-1} w^{j-1}) w$$

If $j > 1$, then the first inequality holds since S_t contains a cone of base K_{j-1} and height w, and the second inequality follows from the fact that $w \leq L_{j-1}$. If $j = 1$, then we observe that S_t contains a segment of length w, so $V_1(S_t) \leq w$.

To get a lower bound on $V_j(S_t)$, simply note that

$$V_j(S_t) \leq \ell_{k_j}^j/(j!) = \ell_{k_J}^j/(j!)$$

where the first inequality follows from the definition of k_j and the second from the fact that $k_j = k_J$ since $J = M(j)$.

Together the bounds imply that $c_{j-1} w^j/j \leq \ell_{k_J}^j/(j!)$. Substituting in the value of c_{j-1} and simplifying we obtain that $w \leq 2\ell_{k_J}$. □

Finally, we argue that if w is large enough (at least $1/T$), then k_J is at most $O_d(\log \log T)$. Once w is at most $1/T$, we can always price at $\underline{p}_t$ and incur at most $O(1)$ additional regret, so this provides a bound for the number of times we can e.g. overprice.

Lemma 4.6.11. *In iterations where $w \geq 1/T$, then $k_J \leq O(d \log \log(dT))$.*

Proof. It follows directly from Lemma 4.6.10: $1/T \leq w \leq 2\ell_{k_J} = 2d^2 \exp(-\alpha^{k_J})$. Simplifying the expression we get $k_J \leq O(d \log \log(dT))$ □

We are now ready to prove our main result:

Theorem 4.6.12. *The total loss of PricingSearch (Algorithm 8) is bounded by* $O(d^4 \log\log(dT))$.

Proof. We sum the loss in different cases. The first is when $w < 1/T$ and the algorithm prices at $\underline{p}_t$. In those occasions the algorithm always sells and the loss is at most $2w \leq 2/T$, so the total loss is at most 2.

The second case is when the algorithm overprices and doesn't sell. If the algorithm doesn't sell, then by Lemma 4.6.8, then ϕ_J goes from range $(\ell_{k_J+1}, \ell_{k_J}]$ to the next range $(\ell_{k_J+2}, \ell_{k_J+1}]$. Since $k_J \leq O(d \log\log(dT))$ by Lemma 4.6.11 this can happen at most this many times for each index J. Since there are d such indices and the loss of each event is at most 1, the total loss is bounded by $O(d^2 \log\log(dT))$.

The final case is when the algorithm underprices. The loss in this case is bounded by the width $2w$. We sum the total loss of events in which the algorithm overprices. We fix the selected index J and k_J. The loss in such a case is at most $2w \leq 4\ell_{k_J}$ by Lemma 4.6.10. Whenever this happens $S_{t+1} = S_t^+(p_J; u_t)$ so the J-th intrinsic volume decreases by $\ell^J_{k_J+1}/(2J!)$ since $V_J(S_t) - V_J(S_{t+1}) = V_J(S_t) - V_J(S_t^+) = \ell^J_{k_J+1}/(2J!)$ by Lemma 4.6.9. Since $V_J(S_t) \leq \ell^J_{k_J}/(J!)$. Therefore the total number of times it can happen is: $2\ell^J_{k_J}/\ell^J_{k_J+1}$. The total loss is at most the number of times the event can happen multiplied by the maximum loss for an event, which is:

$$\frac{2\ell^J_{k_J}}{\ell^J_{k_J+1}} \cdot (4\ell_{k_J}) = 8\frac{\ell^{J+1}_{k_J}}{\ell^J_{k_J+1}} = 8d^2 \exp(J\alpha^{k_J+1} - (J+1)\alpha^{k_J}) \leq 8d^2 \exp(\alpha^{k_J}(d\alpha - (d+1)) = 8d^2$$

since $\alpha = 1 + 1/d$. By summing over all d possible values of J and all $O(d \log\log(dT))$ values of k_J we obtain a total loss of $O(d^4 \log\log(dT))$. □

4.6.3 Proof of the Cone Lemma

We will prove the Cone Lemma in three steps. We start by proving some geometric lemmas about how linear transformations affect intrinsic volumes. We then use these

lemmas to bound the intrinsic volumes of cylinders. Finally, by approximating a cone as a stack of thin cylinders, we apply these bounds to prove the Cone Lemma.

Geometric lemmas

Define an *α-stretch* of $\mathbb{R}^d$ as a linear transformation which contracts $\mathbb{R}^d$ along some axis by a factor of α, leaving the remaining axes untouched (in other words, there is some coordinate system in which an α-stretch T_α sends $(x_1, x_2, \ldots, x_d)$ to $(\alpha x_1, x_2, \ldots, x_d)$).

A contraction is a linear transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $\|Tx\| \leq \|x\|$ for all $x \in \mathbb{R}^d$. An α-stretch is a contraction whenever $\alpha \in [0, 1]$.

Lemma 4.6.13. *Let H and H' be two (d-dimensional) hyperplanes in $\mathbb{R}^{d+1}$, whose normals are separated by angle θ. Let K be a convex body contained in H, and let K' be the projection of K onto H'. Then K' is (congruent to) a $(\cos\theta)$-stretch of K.*

Proof. Without loss of generality, let H' be the hyperplane with orthonormal basis $e_1, e_2, \ldots, e_d$, and let H be the hyperplane with orthonormal basis $e'_1 = (\cos\theta)e_1 + (\sin\theta)e_{d+1}, e'_2 = e_2, \ldots, e'_d = e_d$. Note that a point $a_1e'_1 + a_2e'_2 + \ldots a_ne'_n$ in H, projects to the point $(\cos\theta)a_1e_1 + a_2e_2 + \cdots + a_ne_n$ in H'. This is the definition of a $(\cos\theta)$-stretch. □

The next lemma bounds the change in the d-th volume of a $(d+1)$-dimensional object when it is transformed by a contraction. The analysis will be based on the fact that for a $(d+1)$-dimensional convex set S, $V_d(S)$ corresponds to half of the surface area. This fact can be derived either from Hadwiger's theorem (Theorem 4.5.4) or from Cauchy's formula for the surface area together with Theorem 4.5.8.

It is simpler to reason about the surface area of polyhedral convex sets (i.e. sets that can be described as a finite intersection of half-spaces). The boundary of a polyhedral convex set in $\mathbb{R}^{d+1}$ can be described as a finite collection of facets, which

are convex sets of dimension d. The surface area corresponds to the sum of the d-dimensional volume of the facets. For a 2-dimensional polytope the surface area correspond to the perimeter. For a 3-dimensional polytope the surface area corresponds to the sum of the area (the 2-dimensional volume) of the facets. For a general convex set K, the surface area can be computed as the limit of the surface area of K_t where K_t are polyhedral sets that converge (in the Hausdorff sense) to K. This is equivalent to the usual definition of the surface area as the surface integral of a volume element.

Given the discussion in the previous paragraph, to reason about how the surface area transforms after a linear transformation, it is enough to reason how the volume of d-dimensional convex sets (the facets) transform when the ambient $\mathbb{R}^{d+1}$ space is transformed by a linear transformation.

Lemma 4.6.14. *Let $K \in \mathsf{Conv}_{d+1}$ and T be a contraction, then*

$$V_d(T(K)) \geq \det T \cdot V_d(K)$$

Proof. By the previous discussion, $V_d(K)$ is proportional to the surface area of K. By taking finer and finer approximations of K by polytopes, it suffices to prove the result for a polyhedral set. We only need to argue how the d-dimensional volume of the facets is transfomed by T. The change in volume of a facet corresponds to the determinant of the transformation induced by T on the tangent space of that facet[1]. More precisely, given vectors linearly independent vectors $v_1, \ldots, v_d \in \mathbb{R}^{d+1}$, let P be the parallelepiped generated by them and let $V_d(P)$ be its volume. Let also n be the unit vector orthogonal to affine subspace containing P and N an unit segment in that direction, i.e., the set of points of the form tn for $t \in [0, 1]$, then:

$$V_{d+1}(T(P + N)) = (\det T) \cdot V_d(P + N) = (\det T) \cdot V_{d-1}(P)$$

[1]The tangent space of a facet is the space of all vectors that are parallel to that facet

where the first equality follows from how the (standard) volume transforms and the second since N is orthogonal to P and has size 1.

Now, since $T(P+N) = T(P)+T(N)$, the volume $V_{d+1}(T(P+N))$ can be written as $V_{d-1}(T(P))$ times the projection of N in the orthogonal direction of $T(P)$, which is $\langle Tn, n' \rangle \leq \|Tn\| \cdot \|n'\| \leq 1$ where n' is the orthogonal vector to $T(P)$ and $\|Tn\| \leq 1$ follows from the fact that T is a contraction. Therefore:

$$V_{d-1}(T(P)) \geq V_{d+1}(T(P+N)) = (\det T) \cdot V_{d-1}(P)$$

□

Intrinsic volumes of cylinders

Given a convex set K in $\mathbb{R}^d$, an *orthogonal cylinder* with base K and height w is the convex set in $\mathbb{R}^{d+1}$ formed by taking the Minkowski sum of K (embedded into $\mathbb{R}^{d+1}$) and a line segment of length w orthogonal to K.

Lemma 4.6.15. *Let K be a convex set in $\mathbb{R}^d$, and let S be an orthogonal cylinder with base K and height h. Then, for all $0 \leq j \leq d$,*

$$V_{j+1}(S) = V_{j+1}(K) + hV_j(K).$$

Proof. Embed K into $\mathbb{R}^{d+1}$ so that it lies in the hyperplane $x_{d+1} = 0$, and let L be the line segment from 0 to he_{d+1}, so that $S = K + L$ is an orthogonal cylinder with base K and height h. We will compute $\mathsf{Vol}_{d+1}(S + \varepsilon B_{d+1})$. Recall that Vol refers to the standard volume. Whenever we add subscripts (e.g. Vol_d) we do so to highlight that we are talking about the standard volume of a convex set in a d-dimensional (sub)space.

We claim we can decompose $S + \varepsilon B_{d+1}$ into two parts; one with total volume $\mathsf{Vol}_{d+1}(K + \varepsilon B_{d+1})$, and one with total volume $h\mathsf{Vol}_d(K + \varepsilon B_d)$. To begin, consider

the intersection of $S + \varepsilon B_{d+1}$ with $\{x_{d+1} \in [0, h]\}$. We claim this set has volume at least $h\mathsf{Vol}_d(K + \varepsilon B_d)$. In particular, note that (since S is an orthogonal cylinder) every cross-section of the form $(S + \varepsilon B_{d+1}) \cap \{x_{d+1} = t\}$ for $t \in [0, h]$ is congruent to the set $K + \varepsilon B_d$. It follows that the volume of this region is $h\mathsf{Vol}_d(K + \varepsilon B_d)$.

Next, consider the intersection of $S + \varepsilon B_{d+1}$ with the set $\{x_{d+1} \notin [0, h]\}$. This intersection has two components: a component S^+, the intersection of $S + \varepsilon B_{d+1}$ with the set $\{x_{d+1} \geq h\}$, and a component S^-, the intersection of $S + \varepsilon B_{d+1}$ with the set $\{x_{d+1} \leq 0\}$ (see Figure 4.3a). Now, define K^+ to be the intersection of $K + \varepsilon B_{d+1}$ with $\{x_{d+1} \geq 0\}$, and let K^- be the intersection of $K + \varepsilon B_{d+1}$ with $\{x_{d+1} \leq 0\}$. It is straightforward to verify that K^+ is congruent to S^+ and that K^- is congruent to S^-, and therefore the volume of this region is equal to $\mathsf{Vol}(K^+) + \mathsf{Vol}(K^-) = \mathsf{Vol}_{d+1}(K + \varepsilon B_{d+1})$.

We therefore have that $\mathsf{Vol}_{d+1}(S + \varepsilon B_{d+1}) = \mathsf{Vol}_{d+1}(K + \varepsilon B_{d+1}) + h\mathsf{Vol}_d(K + \varepsilon B_d)$. Expanding out all parts via Steiner's formula (4.1), we have that:

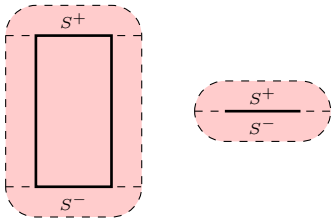
$$\sum_{j=0}^{d+1} \kappa_{d+1-j} V_j(S) \varepsilon^{d+1-j} = \sum_{j=0}^{d} \kappa_{d+1-j} V_j(K) \varepsilon^{d+1-j} + h \sum_{j=0}^{d} \kappa_{d-j} V_j(K) \varepsilon^{d-j}.$$

Equating coefficients of ε^{d-j}, we find that

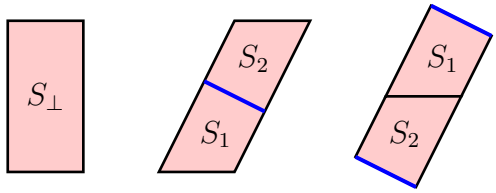
$$V_{j+1}(S) = V_{j+1}(K) + hV_j(K).$$

□

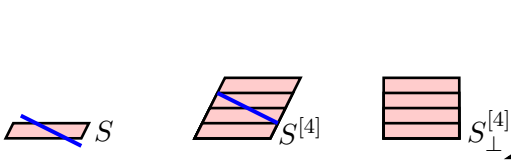
An *oblique cylinder* in $\mathbb{R}^{d+1}$ is formed by taking the Minkowski sum of a convex set $K \subset \mathbb{R}^d$ and a line segment L not necessarily perpendicular to K. The *height* of an oblique cylinder is equal to the length of the component of L orthogonal to the affine subspace containing K.



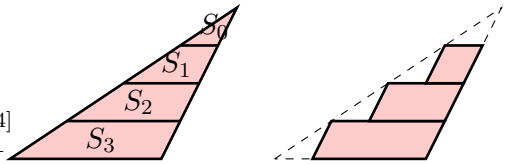
(a) Proof of Lemma 4.6.15: we decompose $K+L+\epsilon B$ (left) and $K+\epsilon B$ (right).



(b) First step in the proof of Lemma 4.6.16, we cut and re-assemble a cylinder



(c) Stacking thin oblique and orthogonal cylinders in Lemma 4.6.16



(d) Approximating a cone by thin oblique cylinders

Figure 4.3: Illustration of the cylinder and cone proofs. In all cases, the x-axis is a d-dimensional space and the y-axis a 1-dimensional space

Lemma 4.6.16. *Let K be a convex set in $\mathbb{R}^d$. If S is an oblique cylinder with base K and height h, and $S_\perp$ is an orthogonal cylinder with base K and height h, then (for all $1 \le j \le d+1$)*

$$V_j(S) \ge V_j(S_\perp).$$

Proof. Note that when $j = d+1$, $V_j(S) = V_j(S_\perp)$ as they are related by a linear transformation with determinant 1. For the remaining cases, we will first prove for $j = d$ and then reduce all other cases to $j = d$.

Case j = d (tall cylinder). Write $S = K + L$, where L is a line segment of length ℓ (with orthogonal component h with respect to K). We will begin by choosing a hyperplane H perpendicular to L that intersects S along its lateral surface, dividing it into two sections S_1 and S_2 (see Figure 4.3b). Note that this is only possible if the height h of this cylinder is large enough with respect to the diameter of K and angle

L makes with K. We address the case of the short cylinder in the next case. Let $K' = H \cap S$. By Theorem 4.5.3, we know that $V_d(S) = V_d(S_1) + V_d(S_2) - V_d(K')$.

Note that it is possible to reassemble S_1 and S_2 by gluing them along their copies of K to form an orthogonal cylinder with base K' and height ℓ. Call this cylinder S'. Again by Theorem 4.5.3, we have that $V_d(S') = V_d(S_1)+V_d(S_2)-V_d(K)$, and therefore $V_d(S) = V_d(S')+V_d(K)-V_d(K')$. But by Lemma 4.6.15, $V_d(S') = V_d(K')+\ell V_{d-1}(K')$, so $V_d(S) = V_d(K)+\ell V_{d-1}(K')$. On the other hand (also by Lemma 4.6.15), $V_d(S_\perp) = V_d(K) + hV_{d-1}(K)$. Therefore, to show that $V_d(S) \geq V_d(S_\perp)$, it suffices to show that $\ell V_{d-1}(K') \geq hV_{d-1}(K)$.

Now, note that K' is the projection of K onto the hyperplane H. The normal to H is parallel to L. Since L has length ℓ and orthogonal component h with respect to K, the angle between L and the normal to K equals $\arccos(h/\ell)$, from which it follows from Lemma 4.6.13 that K' is an (h/ℓ) stretch of K. By Lemma 4.6.14, it follows that $V_{j-1}(K') \geq (h/\ell)V_{j-1}(K)$, from which the desired inequality follows.

Case $\mathbf{j = d}$ (short cylinder). Finally, what if the original cylinder was not tall enough to divide into two components in the desired manner? To deal with this, let $S^{[n]}$ denote n copies of S stacked on top of each other (i.e. $S^{[n]} = K + nL$), and let $S_\perp^{[n]}$ denote n copies of $S_\perp$ stacked on top of each other (i.e. an orthogonal cylinder with base K and height nh). Repeatedly applying Theorem 4.5.3, we have that $V_d(S^{[n]}) = nV_d(S) + (n-1)V_d(K)$, and that $V_d(S_\perp^{[n]}) = nV_d(S_\perp) + (n-1)V_d(K)$. Therefore, to show that $V_d(S) \geq V_d(S_\perp)$, it suffices to show that $V_d(S^{[n]}) \geq V_d(S_\perp^{[n]})$. For some n, $S^{[n]}$ will be tall enough to divide as desired, which completes the proof.

Reducing $\mathbf{j < d}$ to $\mathbf{j = d}$. We can without loss of generality assume that K (the base of the cylinder) is in the plane spanned by the first d coordinate vectors. Also, let $\pi_d : \mathbb{R}^{d+1} \to \mathbb{R}^d$ be the projection in the first d coordinates.

Recall that the jth intrinsic volume $V_j(S)$ is equal to the expected volume of the projection of S onto a randomly chosen j-dimensional subspace of $\mathbb{R}^{d+1}$, where the distribution over subspaces is given by the Haar measure over $\mathrm{Gr}(d+1,k)$ (see Theorem 4.5.8).

Therefore, choose H according to this measure and let $P = \pi_d(H)$. Note that (almost surely) P is an element of $\mathrm{Gr}(d,j)$ and P is distributed according to the Haar measure of this Grassmannian. By the law of total expectation, we can write

$$V_j(S) \;=\; \mathbb{E}_{H\sim\mathrm{Gr}(d+1,j)}\left[V_j(\Pi_H S)\right] = \mathbb{E}_{P\sim\mathrm{Gr}(d,j)}\left[\mathbb{E}_{H\sim\mathrm{Gr}(d+1,j)}\left[V_j(\Pi_H S)\,|\,\pi_d(H)=P\right]\right]$$

Let P' be the element of $\mathrm{Gr}(d+1,j+1)$ spanned by P and e_{d+1}. Note that since $H \subset P'$, $\Pi_H S = \Pi_H \Pi_{P'} S$. We therefore claim that

$$\mathbb{E}_{H\sim\mathrm{Gr}(d+1,j)}\left[V_j(\Pi_H S)\,|\,\pi_d(H)=P\right] = V_j(\Pi_{P'} S).$$

Indeed, conditioned on $\pi_d(H) = P$, H is a (Haar-)uniform subspace of dimension j of the $j+1$-dimensional space P', from which the above equality follows. Therefore, we have that

$$V_j(S) = \mathbb{E}_{P\sim\mathrm{Gr}(d,j)}\left[V_j(\Pi_{P'} S)\right]$$

and similarly

$$V_j(S_\perp) = \mathbb{E}_{P\sim\mathrm{Gr}(d,j)}\left[V_j(\Pi_{P'} S_\perp)\right].$$

Now, since e_{d+1} belongs to P', if $S_\perp$ is an orthogonal cylinder with base K and height h in $\mathbb{R}^{d+1}$, then $\Pi_{P'} S_\perp$ is an orthogonal cylinder with base $\Pi_{P'} K$ and height h in P'. Likewise, $\Pi_{P'} S$ is an oblique cylinder with base $\Pi_{P'} K$ and height h in P'. Since P' is $j+1$ dimensional, it follows the previous cases that: $V_j(\Pi_{P'} S) \geq V_j(\Pi_{P'} S_\perp)$ so

$$V_j(S) = \mathbb{E}_{P \sim \mathrm{Gr}(\cdot, j)} \left[V_j(\Pi_{P'} S) \right] \geq \mathbb{E}_{P \sim \mathrm{Gr}(d, j)} \left[V_j(\Pi_{P'} S_\perp) \right] \geq V_j(S_\perp)$$

□

Intrinsic volumes of cones

A *cone* in $\mathbb{R}^{d+1}$ is the convex hull of a d-dimensional convex set K and a point $p \in \mathbb{R}^{d+1}$. If the distance from p to the affine subspace containing K is h, we say the cone has *height* h and *base* K.

Lemma 4.6.17. *Let K be a convex set in $\mathbb{R}^d$, and let S be a cone in $\mathbb{R}^{d+1}$ with base K and height h. Then, for all $0 \leq j \leq d$,*

$$V_{j+1}(S) \geq \frac{1}{j+1} h V_j(K).$$

Proof. Choose a positive integer n, and divide S into n parts via the hyperplanes $H_i = \{x_{d+1} = \frac{n-i}{n} h\}$ (for $0 \leq i \leq n$). For $0 \leq i < n$, let K_i be the intersection of H_i with S, and let and let S_i be the region of S bounded between hyperplanes H_i and H_{i+1} (see Figure 4.3d). Note that each S_i is a frustum with bases K_i and K_{i+1} and height h/n, and furthermore that each K_i is congruent to $\frac{i}{n} K$.

By repeatedly applying Theorem 4.5.3, we know that

$$V_{j+1}(S) = \sum_{i=0}^{n-1} V_{j+1}(S_i) - \sum_{i=1}^{n-1} V_{j+1}(K_i).$$

Note that each set S_i contains an oblique cylinder with base K_i (since K_i is a contraction of K_{i+1}, some translate of K_i is strictly contained inside K_{i+1}) and height h/n. It follows from Lemmas 4.6.15 and 4.6.16 that $V_{j+1}(S_i) \geq V_{j+1}(K_i) + \frac{h}{n} V_j(K_i)$. It follows that

$$V_{j+1}(S) \geq \sum_{i=1}^{n-1} \frac{h}{n} V_j(K_i) = \sum_{i=1}^{n-1} \frac{h}{n} V_j\left(\frac{i}{n}K\right) = \sum_{i=1}^{n-1} \frac{h}{n}\left(\frac{i}{n}\right)^j V_j(K) = \left(\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)^j \frac{1}{n}\right) h V_j(K).$$

As n goes to infinity, this sum approaches $\int_0^1 x^j dx = \frac{1}{j+1}$, and therefore we have that $V_{j+1}(S) \geq \frac{1}{j+1} h V_j(K)$.

□

4.6.4 Efficient implementation

We have thus far ignored issues of computational efficiency. In this subsection, we will show that algorithms SymmetricSearch (Algorithm 7) and PricingSearch (Algorithm 8) can be implemented in polynomial time by a randomized algorithm that succeeds with high probability.

The main primitive we require to implement both algorithms is a way to efficiently compute the intrinsic volumes of a convex set (and in particular a convex polytope, since our knowledge set starts as $[0,1]^d$ and always remains a convex polytope). Unfortunately, even computing the ordinary volume of a convex polytope (presented as an intersection of half-spaces) is known to be $\#P$-hard [20]. Fortunately, there exist efficient randomized algorithms to compute arbitrarily good multiplicative approximations of the volume of a convex set.

Theorem 4.6.18 (Dyer, Frieze, and Kannan [59])**.** *Let K be a convex subset of $\mathbb{R}^d$ with an efficient membership oracle (which given a point, returns whether or not $x \in K$). Then there exists a randomized algorithm which, given input $\varepsilon > 0$, runs in time $\mathsf{poly}(d, \frac{1}{\varepsilon})$ and outputs an ε-approximation to $\mathsf{Vol}(K)$ with high probability.*

We will show how we can extend this to efficiently compute (approximately, with high probability) the intrinsic volumes of a convex polytope presented as an intersection of half-spaces.

Theorem 4.6.19. *Let K be a polytope in $\mathbb{R}^d$ defined by the intersection of n half-spaces and contained in $[0,1]^d$. Then there exists a randomized algorithm which, given input $\varepsilon > 0$ and $1 \leq i \leq d$, runs in time $\mathsf{poly}(d, n, \frac{1}{\varepsilon})$, and outputs an ε-approximation to $V_i(K)$ with high probability.*

Proof. We use the fact (Theorem 4.5.8) that $V_i(K)$ is the expected volume of the projection of K onto a randomly chosen i-dimensional subspace (sampled according to the Haar measure). Since K is contained inside $[0,1]^d$, any i-dimensional projection of K will be contained within an i-dimensional projection of $[0,1]^d$, whose i-dimensional volume is at most $\mathsf{poly}(d)$. By Hoeffding's inequality, we can therefore obtain an ε-approximation to $V_i(K)$ by taking the average of $\mathsf{poly}(d, \frac{1}{\varepsilon})$ $(\varepsilon/2)$-approximations for volumes of projections of K onto i-dimensional subspaces.

To approximately compute the volume of a projection of K onto an i-dimensional subspace S, we will apply Theorem 4.5.8. Note that we can check whether a point belongs in the projection of K into S by solving an LP (the point adds i additional linear constraints to the constraints defining K). This can be done efficiently in polynomial time, and therefore we have a polynomial-time membership oracle for this subproblem. □

We now briefly argue that Theorem 4.6.19 allows us to implement efficient randomized variants of SymmetricSearch and PricingSearch which succeed with high probability. To do this, it suffices to note that all of the analysis of both algorithms is robust to tiny perturbations in computations of intrinsic volumes. For example, in SymmetricSearch the analysis carries through even if instead of K_i dividing S_t into two regions such that $V_i(S^+) = V_i(S^-)$, it divides them into regions satisfying $V_i(S^+) \in [(1-\varepsilon)V_i(S^-), (1+\varepsilon)V_i(S^-)]$ for some constant ε.

The only remaining implementation detail is how to hyperplanes K_i that divide the ith intrinsic volume of S_i equally (or in the case of PricingSearch, divide off a

fixed amount of intrinsic volume). Since intrinsic volumes are monotone (Theorem 4.5.3), this can be accomplished via binary search.

4.7 Halving algorithms

There are many simple algorithms one can try for the contextual search problem, like algorithms that always halve the width or volume of the current knowledge set. One natural question is whether these simple algorithms suffice to give the same sort of regret bounds as our algorithms based on intrinsic volumes (e.g. SymmetricSearch).

In this section, we show that while these algorithms also obtain $O_d(1)$ regret for the contextual search problem with symmetric loss, the dependence on d is exponential rather than polynomial (and moreover this dependence is tight, at least for the algorithm which always divides the width in half). Moreover, we show that even for these simpler algorithms, looking at how intrinsic volumes of the knowledge set change is a valuable technique for bounding the total regret.

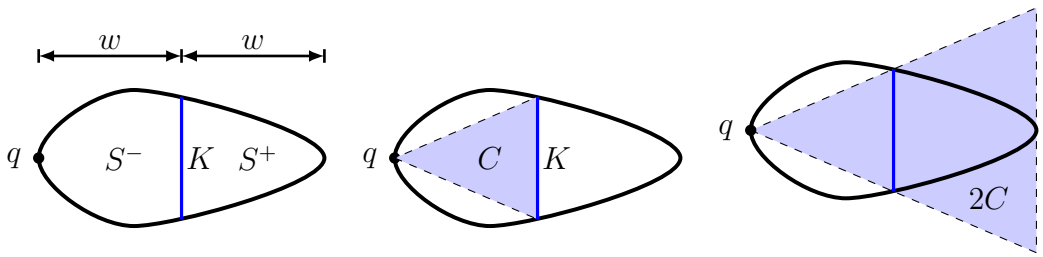
4.7.1 Dividing the width in half

In this section, we will analyze the algorithm which always cuts the width in half; that is, always guesses $p_t = p_t^{\mathsf{mid}} = \frac{1}{2}(\overline{p}_t + \underline{p}_t)$. We will show that for the symmetric loss function, this strategy achieves $2^{O(d)} = O_d(1)$ regret.

Our analysis will proceed similarly to the proof of Theorem 4.6.5. We will rely on the following lemma, which shows that dividing the width in half guarantees that the ratio of the intrinsic volume of the smaller half to that of the larger half is still lower-bounded by some function of d.

Lemma 4.7.1. *If $p_t = p_t^{\mathsf{mid}}$, then*

$$V_j(S^-) \geq 2^{-j} V_j(S^+) \quad \textit{and} \quad V_j(S^+) \geq 2^{-j} V_j(S^-)$$



(a) midpoint cut (b) $C \subseteq S^-$ (c) $S^+ \subseteq 2C$

Figure 4.4

Proof. Let $w = \frac{1}{2}(\overline{p}_t - \underline{p}_t)$ and K be the intersection of the hyperplan $\langle x, u_t \rangle = p_t$ with S_t. Choose a point q in S_t such that $\langle q, u_t \rangle = \underline{p}_t$ as depicted in Figure 4.4a. Consider the cone C formed by the convex hull of q and K (Figure 4.4b). Since S^- is convex, C is contained in S^-, and thus $V_j(S^-) \geq V_j(C)$.

Now, consider the dilation of the cone C by a factor of 2 about the point q (Figure 4.4c). This results in a new cone $2C$. We claim that this cone contains S^+. To see this, it suffices to note that the contraction of S^+ by a factor of $1/2$ about q lies within S^-. This follows from the fact that S is convex, and the width of S^- is equal to the width of S^+ (so any segment connecting q to some point $q' \in S^+$ has at least as much length in S^- than S^+).

It follows that $2^j V_j(C) = V_j(2C) \geq V_j(S^+)$. Combining this with our earlier inequality, the first result follows. The second result follows symmetrically. □

Theorem 4.7.2. *The algorithm that always sets $p_t = p_t^{\mathsf{mid}}$ has regret bounded by $2^{O(d)}$ for the symmetric loss.*

Proof. We will proceed similarly to the analysis of the SymmetricSearch algorithm. For any fixed round t, let $w = \frac{1}{2}\mathsf{width}(S_t; u_t)$ and K be the intersection of the hyperplane $\langle x, u_t \rangle = p_t^{\mathsf{mid}}$ with S_t.

Define a sequence of constants c_i so that $c_0 = 1$ and $c_i/c_{i-1} = 2^{-(i+1)}/i$ (in other words, $c_i = 2^{-(i+1)(i+2)/2}/i!$). For $1 \leq i \leq d$, define $L_i = (V_i(K)/c_i)^{1/i}$, and let $L_0 = \infty$. Choose j so that $L_{j-1} \geq w \geq L_j$ (it is always possible to do this by the

same logic as in Theorem 4.6.5). Note that the proof of Lemma 4.6.3 carries over verbatim to show that $V_j(S_t) \geq \frac{1}{j} c_{j-1} w^j$. We now proceed in two steps:

Step 1 First we show that $V_j(S_{t+1}) \leq (1 - 2^{-(j+2)}) V_j(S_t)$.

The set S_{t+1} is either S^+ or S^-. By Lemma 4.7.1, we therefore have that

$$V_j(S_{t+1}) \leq \frac{1}{1 + 2^{-j}} (V_j(S^+) + V_j(S^-)).$$

Now, since V_j is a valuation, $V_j(S^+) + V_j(S^-) = V_j(S_t) + V_j(K)$. Now, $V_j(K) = c_j L_j^j \leq c_j w^j$ by the choice of j. Combining this with the fact $V_j(S_t) \geq \frac{1}{j} c_{j-1} w^j$, we observe that $V_j(K) \leq j \frac{c_j}{c_{j-1}} V_j(S_t)$, and therefore that

$$V_j(S^+) + V_j(S^-) \leq \left(1 + j \frac{c_j}{c_{j-1}}\right) V_j(S_t) = (1 + 2^{-(j+1)}) V_j(S_t).$$

It follows that

$$V_j(S_{t+1}) \leq \frac{1}{1 + 2^{-j}} \cdot (1 + 2^{-(j+1)}) V_j(S_t) \leq (1 - 2^{-(j+2)}) V_j(S_t).$$

Step 2: Next we consider the potential function $\Phi(t) = \sum_{i=1}^{d} V_i(S_t)^{1/i}$. Note that $\Phi(0) = \text{poly}(d)$. We will show that each round, $\Phi(t)$ decreases by at least $2^{-O(d)} w$. Since the loss each round is upper bounded by w, this proves our theorem.

$$\begin{aligned} V_j(S_t)^{1/j} - V_j(S_{t+1})^{1/j} &\geq (1 - (1 - 2^{-(j+2)})^{1/j}) V_j(S_t)^{1/j} \geq \frac{1}{j} 2^{-(j+2)} V_j(S_t)^{1/j} \\ &\geq \frac{1}{j} 2^{-(j+2)} \left(\frac{c_{j-1}}{j}\right)^{1/j} w \geq 2^{-O(d)} w. \end{aligned}$$

□

The exponential dependency on the dimension is tight, as it is shown in an example by Cohen et al [43].

Theorem 4.7.3. *There is an instance of the contextual search problem with symmetric loss such that the algorithm that always sets $p_t = p_t^{mid}$ incurs regret $2^{\Omega(d)}$.*

Proof. See [43]. The instance they describe is as follows: let the dimension d be a multiple of 8 and $v = 0 \in [0,1]^d$. Let X_t be iid random subsets of $[d]$ of size $d/4$. Now, consider feature vectors of the form $u_t = \mathbf{1}\{X_t\}/\sqrt{d/4}$ where $\mathbf{1}\{X_t\}$ is the indicator vector of X_t. By standard concentration bounds we have that with high probability for any $s < t < 2^{\Omega(d)}$ we will have $|X_s \cap X_t| \leq d/8$. Therefore for such $s < t$, $\langle \mathbf{1}\{X_t\}, u_s \rangle \leq \frac{1}{2}\langle \mathbf{1}\{X_s\}, u_s \rangle$ and hence $\mathbf{1}\{X_t\} \in S_t$ where S_t is the knowledge set in step t. It implies that the loss in the t-th step is at least $\Omega(1)$. Since there are $2^{\Omega(d)}$ such steps, the loss grows exponentially in d. □

4.7.2 Dividing the volume in half

In this section we will consider the algorithm which always divides the volume of our current knowledge set in half. More specifically, this algorithm always chooses p_t so that $\mathsf{Vol}(S^+) = \mathsf{Vol}(S^-)$. We will show that for the symmetric loss function, this strategy also achieves $2^{O(d)} = O_d(1)$ regret.

Like in the previous subsection, we will argue that splitting the volume in half guarantees that the other intrinsic volumes are split in some ratio bounded away from 0 and 1. To do this, we will first show that splitting the volume in half imposes constraints on the ratio of the widths of S^+ and S^-, and then adapt the proof of Lemma 4.7.1. The proof follows the same scheme depicted in Figure 4.4 but with unequal widths for S^+ and S^-.

Lemma 4.7.4. *Assume p_t is chosen so that $\mathsf{Vol}(S^+) = \mathsf{Vol}(S^-)$. Then if $w^+ = \mathsf{width}(S^+; u_t)$ and $w^- = \mathsf{width}(S^-; u_t)$,*

$$w^- \geq (2^{1/d} - 1)w^+ \quad \textit{and} \quad w^+ \geq (2^{1/d} - 1)w^-.$$

Proof. Choose a point q in S^- that is distance w^- from the hyperplane $\langle x, u_t \rangle = p_t$. Let K be the intersection of this hyperplane with the original set S_t. Consider the cone C formed by the convex hull of q and K. Since S^- is convex, C is contained in S^-, and therefore $\mathsf{Vol}(S^-) \geq \mathsf{Vol}(C)$.

Now, consider the dilation of the cone C by a factor of $\alpha = (w^+ + w^-)/w^-$ about the point q. By similar logic as in the proof of Lemma 4.7.1, this cone αC contains S^+. In fact, since C is contained in αC (since $\alpha > 1$) and since C is contained in S^- (which has zero volume intersection with S^+), S^+ is contained in $\alpha C \setminus C$. We therefore have that

$$\mathsf{Vol}(\alpha C) - \mathsf{Vol}(C) \geq \mathsf{Vol}(S^+) = \mathsf{Vol}(S^-) \geq \mathsf{Vol}(C).$$

Since $\mathsf{Vol}(\alpha C) = \alpha^d \mathsf{Vol}(C)$, this implies that $\alpha^d \geq 2$, and therefore that $(w^+ + w^-)/w^- = \alpha \geq 2^{1/d}$ and $w^+/w^- \geq 2^{1/d} - 1$, as desired. The other inequality follows by symmetry. □

Lemma 4.7.5. *Let $w^+ = \mathsf{width}(S^+; u_t)$ and $w^- = \mathsf{width}(S^-; u_t)$. Assume p_t is chosen so that $w^+ \geq \alpha w^-$ and $w^- \geq \alpha w^+$, for some $\alpha > 0$. Then, for all $1 \leq j \leq d$,*

$$V_j(S^-) \geq \left(1 + \frac{1}{\alpha}\right)^{-j} V_j(S^+) \quad \textit{and} \quad V_j(S^+) \geq \left(1 + \frac{1}{\alpha}\right)^{-j} V_j(S^-).$$

Proof. We follow the argument in the proof of Lemma 4.7.1. The only difference is that we now must consider the dilation of the cone C by a factor of $1 + \frac{1}{\alpha}$ about q, as the ratio of the width of S_t to the width of S^- (or S^+) is at most $1 + \frac{1}{\alpha}$. □

Corollary 4.7.6. *If p_t is chosen so that $\mathsf{Vol}(S^+) = \mathsf{Vol}(S^-)$, then for all $1 \leq j \leq d$,*

$$V_j(S^-) \geq \left(1+\frac{1}{\alpha}\right)^{-j} V_j(S^+) \quad \textit{and} \quad V_j(S^+) \geq \left(1+\frac{1}{\alpha}\right)^{-j} V_j(S^-),$$

where $\alpha = 2^{1/d} - 1 = \Theta(d^{-1})$.

Proof. Follows from Lemmas 4.7.4 and 4.7.5. □

Theorem 4.7.7. *The algorithm that always sets* p_t *such that* $\mathsf{Vol}(S_t^+) = \mathsf{Vol}(S_t^-)$ *has regret bounded by* $2^{O(d \log d)}$ *for the symmetric loss.*

Proof. We will proceed similarly to the analysis of the SymmetricSearch algorithm. Consider a fixed round t, let $w = \frac{1}{2}\mathsf{width}(S_t; u_t)$, and let K be the intersection of the hyperplane $\langle x, u_t \rangle = p_t$ with S_t.

Let $\alpha = 2^{1/d} - 1$, and let $\lambda = 1 + \frac{1}{\alpha}$. Note that $\lambda \geq 2$ and $\lambda = \Theta(d)$. Define a sequence of constants c_i so that $c_0 = 1$ and $c_i/c_{i-1} = \lambda^{-(i+1)}/i$ (in other words, $c_i = \lambda^{-(i+1)(i+2)/2}/i!$). For $1 \leq i \leq d$, define $L_i = (V_i(K)/c_i)^{1/i}$, and let $L_0 = \infty$. Choose j so that $L_{j-1} \geq w \geq L_j$ (it is always possible to do this by the same logic as in Theorem 4.6.5). The proof of Lemma 4.6.3 again carries over verbatim to show that $V_j(S_t) \geq \frac{1}{j} c_{j-1} w^j$. We again proceed in two steps similarly to the proof of Theorem 4.7.2.

Step 1: We first show that $V_j(S_{t+1}) \leq (1 - \lambda^{-(j+2)}) V_j(S_t)$.

The set S_{t+1} is either S^+ or S^-. By Corollary 4.7.6, we therefore have that $V_j(S_{t+1}) \leq \frac{1}{1+\lambda^{-j}}(V_j(S^+) + V_j(S^-))$.

Now, since V_j is a valuation, $V_j(S^+) + V_j(S^-) = V_j(S_t) + V_j(K)$. Since $w \geq L_j$, $V_j(K) \leq c_j w^j$. Combining this with the fact $V_j(S_t) \geq \frac{1}{j} c_{j-1} w^j$, we observe that $V_j(K) \leq j \frac{c_j}{c_{j-1}} V_j(S_t)$, and therefore that

$$V_j(S^+) + V_j(S^-) \leq \left(1 + j\frac{c_j}{c_{j-1}}\right) V_j(S_t) = (1 + \lambda^{-(j+1)}) V_j(S_t).$$

It follows that (since $\lambda \geq 2$)

$$V_j(S_{t+1}) \leq \frac{1}{1+\lambda^{-j}} \cdot (1+\lambda^{-(j+1)})V_j(S_t) \leq (1-\lambda^{-(j+2)})V_j(S_t).$$

Step 2: We next consider the potential function $\Phi(t) = \sum_{i=1}^{d} V_i(S_t)^{1/i}$. Note that $\Phi(0) = \text{poly}(d)$. We will show that each round, $\Phi(t)$ decreases by at least $2^{-O(d \log d)}w$. Since the loss each round is upper bounded by w, this proves our theorem.

In particular, note that

$$\begin{aligned} V_j(S_t)^{1/j} - V_j(S_{t+1})^{1/j} &\geq (1-(1-\lambda^{-(j+2)})^{1/j})V_j(S_t)^{1/j} \geq \frac{1}{j}\lambda^{-(j+2)}V_j(S_t)^{1/j} \\ &\geq \frac{1}{j}\lambda^{-(j+2)}\left(\frac{c_{j-1}}{j}\right)^{1/j} w \geq 2^{-O(d\log d)}w. \end{aligned}$$

□

4.8 General loss functions

Throughout this chapter we have focused on the special cases of the symmetric loss function and the pricing loss function. In this subsection we briefly explore the landscape of other possible loss functions and what regret bounds we can obtain for them.

For simplicity, we restrict ourselves to loss functions of the form $\ell(\langle u_t, v\rangle, p_t) = F(\langle u_t, v\rangle - p_t)$. Note that while some functions (e.g. the pricing loss function) may not be of this form, they may be dominated by some function of this form (e.g. $F(x) = x$ for $x \geq 0$ and $F(x) = 1$ for $x \leq 0$), and hence any regret bound that holds for this simplified loss function holds for the original loss function.

We begin by showing that if $F(x)$ goes to 0 polynomially quickly from both sides (i.e. if $F(x) \leq |x|^\beta$ for some $\beta > 0$), then SymmetricSearch still achieves constant regret.

Theorem 4.8.1. *If $F(x) = |x|^\beta$, for $\beta > 0$, then SymmetricSearch (Algorithm 7) achieves regret $O_{d,\beta}(1)$ for the contextual search problem with this loss function.*

Proof. We modify the proof of Theorem 4.6.5 to look at the potential function $\Phi_t = \sum_{i=1}^{d} V_i(S_t)^{\beta/i}$. The change in potential in each round is now at least (for some $j \in [d]$):

$$\begin{aligned} V_j(S_t)^{\beta/j} - V_j(S_{t+1})^{\beta/j} &\geq \left(1 - \left(\frac{3}{4}\right)^{\beta/j}\right) V_j(S_t)^{\beta/j} \\ &\geq \left(1 - \left(\frac{3}{4}\right)^{\beta/j}\right) \left(\frac{c_{j-1}}{j}\right)^{\beta/j} w^\beta \geq O_{d,\beta}(1)\ell_t. \end{aligned}$$

It follows that the total regret of SymmetricSearch is $O_{d,\beta}(1)$. □

Similarly, we can show that for functions F which are discontinuous on one side and converge to zero polynomially quickly on the other side, the PricingSearch algorithm (with a slightly different choice of parameters) achieves $O_{d,\alpha}(1)$ regret.

Theorem 4.8.2. *Let $\alpha > 0$ be a constant, and let $F(x) = |x|^\beta$, for $x \geq 0$ and let $F(x) = 1$ for $x < 0$. PricingSearch (Algorithm 8) with parameter $\alpha = 1 + \frac{\beta}{d}$ achieves regret $O_{d,\beta}(\log\log T)$ for the contextual search problem with this loss function.*

Proof. Again, we modify the proof of Theorem 4.6.12. Lemmas 4.6.8, 4.6.9, 4.6.10, and 4.6.11 hold as written. The only necessary change is in the underpricing case of the proof of Theorem 4.6.12, where the maximum loss for an event is now $(4\ell_{k_J})^\beta$, and so the total loss from underpricing (for a fixed value of J and k_J) is at most

$$\begin{aligned} \frac{2\ell_{k_J}^J}{\ell_{k_J+1}^J} \cdot (4\ell_{k_J})^\beta &= 2^{1+2\beta} d^{2\beta} \exp(J\alpha^{k_J+1} - (J+\beta)\alpha^{k_J}) \\ &\leq 2^{1+2\beta} d^{2\beta} \exp(\alpha^{k_J}(d\alpha - (d+\beta)) \\ &= 2^{1+2\beta} d^{2\beta} = O_{d,\beta}(1). \end{aligned}$$

It follows that the total regret of PricingSearch is $O_{d,\beta}(\log \log T)$. $\square$

Part III

Learning how to rank

Chapter 5

Condorcet-consistent and approximately strategyproof tournament rules

This chapter is joint work with Ariel Schvartzman and Matthew Weinberg [128].

5.1 Introduction

In recent years, numerous scandals have unfolded surrounding match fixing and throwing at the highest levels of competitive sports (e.g. Olympic Badminton [85], Professional Tennis [47], European Football [127], and even eSports [134]). In some instances, the motivation behind these scandals was gambling profits, and no amount of clever tournament design can possibly mitigate this. In others, however, the surprising motivation was an improved performance *at that same tournament.* For instance, four Badminton teams (eight players) were disqualified from the London 2012 Olympics for throwing matches. Interestingly, the reason teams wanted to lose their matches was in order to *improve* their probability of winning an Olympic medal. Olympic Badminton (like many other sports) conducts a two-phase tournament. In the first

stage, groups of four play a round-robin tournament, with the top two teams advancing. In the second stage, the advancing teams participate in a single elimination tournament, seeded according to their performance in the group stage. An upset in one group left one of the world's top teams with a low seed, so many teams actually preferred to receive a *lower* seed coming out of the group stage to face the tougher opponent as late as possible.

While much of the world blames the teams for their poor sportsmanship, researchers in voting theory have instead critiqued the poor tournament design that punished teams for trying to maximize their chances of winning a medal. Specifically, the two-phase tournament lacks the basic property of *monotonicity*, where no competitor can unilaterally improve their chances of winning by throwing a match that they otherwise could have won. Thus, recent work has addressed the question of whether tournament structures exist that are both fair, in that they select some notion of a qualified winner, and strategyproof, in that teams have no incentive to do anything but play their best in each match.

One minimal notion of fairness studied is *Condorcet-consistence*, which just guarantees that whenever one competitor wins *all* of their matches (and is what's called a *Condorcet winner*), they win the event with probability 1. Designing Condorcet-consistent, monotone rules is simple: any single elimination bracket suffices. Popular voting rules such as the Copeland Rule or the Random Condorcet Removal Rule are also Condorcet-consistent and monotone, but two-phase tournaments with an initial group play aren't [117].

Still, monotonicity only guarantees that no team wishes to unilaterally throw a match to improve their chances of winning, whereas one might also hope to guarantee that no two teams could fix the outcome of their match in order to improve the probability that one of them wins. While we have to go back further in history to find a clear instance of this kind of match-fixing, it did indeed result in a historical

scandal. In the 1982 FIFA World Cup (again a two-stage tournament), Austria, West Germany, and Algeria were in the same group of four where two would advance. Algeria had already won two matches and lost one, Austria was 2-0, West Germany was 1-1, and the only remaining game was Austria vs. West Germany. Due to tie-breakers and the specific outcomes of previous matches, Austria would have been eliminated by a large West German victory, and West Germany would have been eliminated by a loss or draw. Once West Germany scored an early goal, *both* teams essentially threw the rest of the match, allowing both of them to advance at Algeria's expense [136]. While the incident was never formally investigated, many fans were confident the teams had colluded beforehand, and the event is remembered as the "disgrace of Gijón." Before being eliminated, Algeria had become the first African team to beat a European team at the World Cup, and also the first to win two games. West Germany went on to become the runners-up of the tournament.

Motivated by events like this, it is important also to design tournaments where no two teams can fix the outcome of their match and improve the probability that one of them wins. Altman and Kleinberg terms this property 2-Strongly Nonmanipulable (2-SNM), and showed that no tournament rule is both Condorcet-consistent and 2-SNM [4] (it was previously shown by Altman et. al. that no *deterministic* rule is both Condorcet-consistent and 2-SNM [5]).

In light of this, both works relax the notion of Condorcet-consistency and design tournament rules that are at least *non-imposing* (could possibly select each competitor as a winner) and 2-SNM [5], or α-Condorcet-consistent (if there is a Condorcet winner, she wins with probability at least α) and 2-SNM. While these relaxations are well-motivated for settings where pair-wise comparisons are only *implicitly* made, and not even necessarily learned in the end (e.g. elections), it is hard to imagine a successful sports competition format where a competitor could win all their matches and still leave empty handed. This happened during the 2008 NCAA Football Sea-

son. Utah went undefeated (#2, 13-0) in their region but were not invited to the bowl game because critics deemed their schedule weak. They were eventually ranked second nation-wide and beat Alabama (#6, 12-2) in the Sugar Bowl, while Florida (#1, 13-1) beat Oklahoma (#5, 12-2) for the National Championship. This event prompted organizers to reconsider the process by which teams are invited to the National Championship game.

Motivated by match-based applications such as sporting events, where the outcome of pair-wise matches is *explicitly* learned and used to select a winner, we consider instead the design of tournament rules that are exactly Condorcet-consistent, but only approximately 2-SNM. Specifically, we say that a tournament rule is 2-SNM-α if it is *never* possible for two teams i and j to fix their match such that the probability that the winner is in $\{i, j\}$ improves by at least α. The idea behind this relaxation is that whatever motivates j to throw the match (perhaps j and i are teammates, perhaps i is paying j some monetary bribe, etc.), the potential gains scale with α. So it is easier to disincentivize manipulation (either through investigations and punishments, reputation, or just feeling morally lousy) in tournaments that are less manipulable.

5.1.1 Our Results

Our main result is a matching upper and lower bound of 1/3 on attainable values of α for Condorcet-consistent 2-SNM-α tournament rules. The optimal rule that attains this upper bound is actually quite simple: a random single elimination bracket. Specifically, each competitor is randomly placed into one of $2^{\lceil \log_2 n \rceil}$ seeds, along with $2^{\lceil \log_2 n \rceil} - n$ byes, and then a single elimination tournament is played.

Proving a lower bound of 1/3 is straight-forward: imagine a tournament with three players, A, B and C, where A beats B, B beats C, and C beats A. Then some pair must win with combined probability at most 2/3. Yet, any pair could create a Condorcet winner by colluding, who necessarily wins with probability 1 in

any Condorcet-consistent rule. Embedding this within examples for arbitrary n is also easy: just have A, B, and C each beat all of the remaining $n-3$ competitors[1].

On the other hand, proving that a random single elimination bracket is optimal is tricky, but our proof is still rather clean. For any i, j in any tournament, we directly show that i can improve her probability of winning by at most $1/3$ when j throws their match using a coupling argument. For every deterministic single elimination bracket where i and j could potentially gain from manipulation (because i would be the champion if i beat j, but j would *not* be the champion even if j beat i), we construct *two* deterministic single elimination brackets where no potential exists (possibly because one of them will lose before facing each other, or because the winner would be in $\{i, j\}$ no matter the outcome of their match). For our coupling to be valid, we not only need each mapping to be invertible, but also for their images to be disjoint. Our coupling is necessarily somewhat involved in order to obtain this property, but otherwise we believe our proof is likely as simple as possible. Because the probability that j wins cannot possibly go up by throwing a match to i, this immediately proves that a random single elimination bracket is 2-SNM-1/3.

We also show that the Copeland rule, a popular rule that chooses the team with the most wins, is asymptotically 2-SNM-1, the *worst* possible. Essentially, the problem is that if all teams have the same number of wins, then any two can collude to guarantee that one of them wins, no matter the tie-breaking rule. We further show that numerous other formats, (the Random Voting Caterpillar, the Iterative Condorcet Rule, and the Top Cycle Rule) are all at best 2-SNM-1/2. The same example is bad for all three formats: there is one superman who beats $n-2$ of the remaining players, and one kryptonite, who beats only the superman (but loses to the other $n-2$ players).

[1]Interestingly, this lower-bound example is far from pathological and occurs at even the highest levels of professional sports (see [118], for instance).

Our results extend to settings where the winner of each pairwise match is not deterministically known, but randomized (i.e. all partipants know that i will beat j with probability p_{ij}). Specifically, we show that any rule that is 2-SNM-α when all $p_{ij} \in \{0, 1\}$ is also 2-SNM-α for arbitrary p_{ij}. Clearly, any lower bound using integral p_{ij} also provides a lower bound for arbitrary p_{ij}, so as far as upper/lower bounds are concerned the models are equivalent. Of course, the randomized model is much more realistic, so it is convenient that we can prove theorems in this setting by only studying the deterministic setting, which is mathematically much simpler.

Finally, we consider manipulations among coalitions of $k > 2$ participants. We say that a rule is k-SNM-α if no set S of size $\leq k$ can *ever* manipulate the outcomes of matches between players in S to improve the probability that the winner is in S by more than α. We prove a simple lower bound of $\alpha = \frac{k-1}{2k-1}$ on all Condorcet-consistent rules, and conjecture that this is tight.

5.1.2 Related Works

The mathematical study of tournament design has a rich literature, ranging from social choice theory to psychology. The overarching goal in these works is to design tournament rules that satisfy various properties a designer might find desirable. Examples of such properties might be that all players are treated equally, that a winner is chosen without a tiebreaking procedure, or that a "most qualified" winner is selected [63, 124, 57, 121, 146, 110, 129]. See [96] for a good review of this literature and its connections to other fields as well.

Most related to our work are properties involving *strategic manipulation*. In the more general field of Voting Theory, there is a rich literature on the design of strategyproof mechanisms dating back to Arrow's Impossibility Theorem [13] and the Gibbard-Satterthwaite Theorem [65, 125, 66]. While tournaments are a very special case (voters are indifferent among outcomes where they do not win, voters can only

"lie" in specific ways, etc.), tournament design indeed seems to inherit much of the impossibility associated with strategyproof voting procedures [4], [5].

Specifically, Altman et. al. proved that no deterministic tournament rule is 2-SNM and Condorcet-consistent, and Altman and Kleinberg proved that no randomized tournament rule is 2-SNM and Condorcet-consistent either [5, 4]. More recently, Pauly studied the specific two-stage tournament rule used by the World Cup (and Olympic Badminton, etc.) [117]. There, it is shown essentially that the problem lies in the first round group stage: no changes to the second phase can possibly result in a strategyproof [2] tournament.

To cope with their impossibility results, Altman et. al. propose a relaxation of Condorcet-consistence called *non-imposing*. A rule r is non-imposing if for all i, there exists a T such that player i wins with probability 1. They design a clever recursive rule that is non-imposing and 2-SNM for all $n \neq 3$. Interestingly, they also show that for $n = 3$ no such rule exists. Altman and Kleinberg consider a different relaxation called *α-Condorcet-consistent*. A rule r is α-Condorcet-consistent if whenever i is a Condorcet winner in T, we have their probability of winning T is at least α. They design a rule that is $2/n$-Condorcet-consistent and 2-SNM (in fact it is also k-SNM for all k), but conjecture that much better is attainable.

The two works above are most similar to ours in spirit: motivated by the non-existence of Condorcet-consistent and 2-SNM tournament rules, we relax one of the notions. These previous works relax Condorcet-consistency while maintaining 2-SNM exactly, and are most appropriate in settings where pairwise comparisons of players are only learned *implicitly* (or perhaps not at all) through the outcome and not *explicitly* as the result of matches. Instead, we relax the notion of 2-SNM and maintain the notion of Condorcet-consistency exactly. In settings like sports competitions where pairwise comparisons of players are learned explicitly through matches played,

[2] See [117] for the specific notion of strategyproofness studied.

Condorcet-consistency is a non-negotiable desideratum. Therefore, we believe our approach is more natural in such settings.

Another line of work introduced by [21] considers a different kind of strategyproofness: how much control does the designer of a single-elimination tournament have over the winner? Can the designer efficiently find a bracket in such a way to maximize the likelihood that a player of their choice wins the tournament? The models in this area assume that the designer is given the probabilities p_{ij} that team i beats team j and the problem is known in the literature as *agenda control* when p_{ij} are real numbers and Tournament Fixing Problem (TFP) when all probabilities are 0 or 1.

On the negative side, it is known that for n-player tournaments it is NP-hard to decide whether or not there exists a seeding such that the probability of team k winning is at least δ, given k, δ, even if $p_{ij} \in \{0, 0.5, 1\}$ for all i, j [143]. [137] show that the hardness results persist even for the TFP when the given team k is a king (for every team j, either k beats j or k beats a team that beats j) with at least $n/4$ wins, or a 3−king (is at most 3 "wins" away from every team) that wins at least half of their games. Follow up work [88] shows that in the case of balanced single elimination brackets, it is still NP-hard to find a bracket that favors team k when the designer is allowed to bribe at most $(1-\varepsilon)\log n$ of the teams to throw their respective matches.

On the positive side, there exist structural results that dictate when it is computationally efficient to find a tournament that favors a given team. [137] show conditions under which, for large enough tournaments, any sufficiently good team can be favored by the tournament seeding. Other results [88, 87] show conditions under which 3−kings can be made into winners of single-elimination tournaments.

A large body of literature exists regarding manipulation and bribery in the context of voting rules. For an introduction, we recommend the reader consult chapter 7 of the handbook [111].

5.1.3 Conclusions and Future Work

Our work contributes to a recent literature on incentive compatible tournament design. While most previous works insist on strong incentive properties and relaxed fairness properties, such rules are inadequate for sporting events. Instead, we insist at least that events maintain Condorcet-consistency, and aim to relax strategyproofness as minimally as possible.

At a high level, our work suggests (similar to previous works), that single elimination brackets are desirable whenever incentive issues come into play. However, previous desiderata (such as those considered in [4]) don't necessarily rule out other tournament formats, like the Copeland rule, which is ubiquitous in tournaments (both as a complete format and as subtournaments in a two-phase format). In comparison, our work identifies single elimination brackets (2-SNM-1/3) as having significantly better strategic properties versus the Copeland rule (2-SNM-1).

Our work also identifies two practical suggestions when match-fixing is a concern that aren't explained by prior benchmarks. First, when hosting a single elimination tournament, it might be desirable to release the exact bracket as late as possible. The idea is that as soon as the exact bracket is known, competitors have greater incentive to fix matches (in our model, up to three times as much), which presumably takes some time and organization. Obviously, there are more tradeoffs at play: a later release inconveniences athletes and fans, and (perhaps more importantly to the designers) could negatively impact ticket sales. But our work does at least identify match-fixing as a part of this tradeoff. Note that some Olympic events (such as Taekwondo) contest the entire competition in a single day at a single venue, so a delayed release may indeed be practical. We also note that a similar "fix" was applied after the 1982 World Cup: the last two matches in each group are now played at the same time to minimize the amount of information teams have when making potentially strategic decisions.

Additionally, our work suggests that even in the optimal tournament, hefty punishments for cheaters might be necessary in order to discourage match-fixing (even without taking gambling into consideration). In many sports, winning an Olympic gold can make a career. Unfortunately, our work suggests that punishments roughly on this order might be necessary in order to properly deter match-fixing.

Finally, we propose two directions for future work. First, while we obtain tight results for Condorcet-consistent 2-SNM-α rules, we only prove a lower bound of k-SNM-$\frac{k-1}{2k-1}$ for Condorcet-consistent rules and $k > 2$. We conjecture that this is tight, but unfortunately simulations indicate that all of the formats studied in our work do *not* achieve this bound. So it is an interesting open question to design a rule that does. Even partial results (of the form identified below) would require a new tournament format than those considered in this work.

Open Question 1. *Does there exist a tournament rule that is Condorcet-consistent and k-SNM-$\frac{k-1}{2k-1}$ for all k? What about a family of rules $\mathcal{F}$ such that for all k, F_k is k-SNM-$\frac{k-1}{2k-1}$? What about a rule that is k-SNM-$1/2$ for all k?*[3]

It is also important to study what bounds are attainable in restricted versions of our probabilistic model (e.g. if for all i, j, the probability that i beats j lies in $[\epsilon, 1-\epsilon]$). Realistic instances at least have *some* non-zero probability of an upset in every match, but our lower bounds don't hold in this model. So it is interesting to see if better formats are possible.

Open Question 2. *Is a random single elimination bracket still optimal among Condorcet-consistent rules (w.r.t. 2-SNM-α) if for all i, j, the probability that i beats j lies in $[\epsilon, 1-\epsilon]$? How does the optimal attainable α for Condorcet-consistent, 2-SNM-α tournament formats change as a function of ϵ?*

[3]Note that $\frac{k-1}{2k-1} \to 1/2$ as $k \to \infty$.

5.2 Preliminaries and Notation

In this section, we present notation used throughout the remainder of this chapter. Where possible, we adopt notation from [4].

Definition 5.2.1. *A (round-robin) tournament T on n players is the set of outcomes of the $\binom{n}{2}$ games played between all pairs of distinct players. We write $T_{ij} = 1$ if player i beats player j and $T_{ij} = -1$ otherwise. We also let $\mathcal{T}_n$ denote the set of tournaments on n players.*

Definition 5.2.2. *For a subset $S \subseteq [n]$ of players, two tournaments T and T' are S-adjacent if they only differ on the outcomes of some subset of games played between members of S. In particular, two tournaments T and T' are $\{i, j\}$ adjacent if they only differ in the result of the game played between player i and player j.*

Definition 5.2.3. *A tournament rule (or winner determination rule) $r : \mathcal{T}_n \to \Delta([n])$ is a mapping from the set of tournaments on n players to probability distributions over these n players (representing the probability we choose a given player to be the winner). We will write $r_i(T) = \Pr[r(T) = i]$ to denote the probability that player i wins tournament T under rule r.*

Many tournament rules, while valid by the above definition, would be ill-suited for running an actual tournament; for example, the tournament rule which always crowns player 1 the winner. In an attempt to restrict ourselves to 'reasonable' tournament rules, we consider tournaments that obey the following two criteria.

Definition 5.2.4. *Player i is a Condorcet winner in tournament T if player i wins their match against all the other $n - 1$ players. A tournament rule r is Condorcet-consistent if $r_i(T) = 1$ whenever i is a Condorcet winner in T.*

Definition 5.2.5. *A tournament rule r is monotone if, for all i, $r_i(T)$ does not increase when i loses a game it wins in T. That is, if i beats j in T and T and T' are $\{i, j\}$ adjacent, then if r is monotone, $r_i(T) \geq r_i(T')$.*

Intuitively, this first criterion requires us to award the prize to the winner in the case of a clear winner (hence making the tournament a contest of skill), and the second criterion makes it so that players have an incentive to win their games. There are various other criteria one might wish a tournament rule to satisfy; many can be found in [4].

In this chapter, we consider the scenario where certain coalitions of players attempt to increase the overall chance of one of them winning by manipulating the outcomes of matches within players of the coalition. The simplest case of this is in the case of coalitions of size 2, where player j might throw their match to player i. If T is the original tournament and T' is the manipulated tournament where j loses to i, then player i gains $r_i(T')-r_i(T)$ from the manipulation, and player j loses $r_j(T)-r_j(T')$ (in terms of probability of winning). Therefore, as long as $r_i(T')-r_i(T) > r_j(T)-r_j(T')$, it will be in the players' interest to manipulate. Equivalently, if $r_i(T') + r_j(T') > r_i(T) + r_j(T)$ (i.e., the probability either player i or j wins increases upon throwing the match), there is incentive for i and j to manipulate.

Ideally, we would like to choose a tournament rule so that, regardless of the tournament, there will be no incentive to perform manipulations of the above sort. This is encapsulated in the following definition from [4].

Definition 5.2.6. *A tournament rule r is 2-strongly non-manipulable (2-SNM) if, for all pairs of $\{i, j\}$-adjacent tournaments T and T', $r_i(T) + r_j(T) = r_i(T') + r_j(T')$.*

Unfortunately, no tournament rules exist that are simultaneously Condorcet-consistent and 2-strongly non-manipulable (this is shown in [4] and also follows from our lower bound in Section 5.3.1). As tournament designers, one way around this obstacle is to discourage manipulation. This discouragement can take many forms, both explicit (if players are caught fixing matches, they are disqualified/fined) and implicit (it is logistically hard to fix matches, it is unsportsmanlike). The focus of this chapter is to quantify *how manipulable* certain tournament formats are (i.e. how

much can teams possibly gain by fixing matches), the idea being that it is easier to discourage manipulation in tournaments that are less manipulable.

Definition 5.2.7. *A tournament rule r is 2-strongly non-manipulable at probability α (2-SNM-α) if, for all i and j and pairs of $\{i,j\}$-adjacent tournaments T and T', $r_i(T') + r_j(T') - r_i(T) - r_j(T) \leq \alpha$.*

It is straightforward to generalize this definition to larger coalitions of colluding players.

Definition 5.2.8. *A tournament rule r is k-strongly non-manipulable at probability α (k-SNM-α) if, for all subsets S of players of size at most k, for all pairs of S-adjacent tournaments T and T', $\sum_{i \in S} r_i(T') - \sum_{i \in S} r_i(T) \leq \alpha$.*

5.2.1 The Random Single-Elimination Bracket Rule

Our main result concerns a specific tournament rule we call the *random single-elimination bracket rule*. This rule can be defined formally as follows.

Definition 5.2.9. *A single-elimination bracket (or bracket, for short) B on $n = 2^h$ players is a complete binary tree of height h whose leaves are labelled with some permutation of the n players. The outcome of a bracket B under a tournament T is the labelling of internal nodes of B where each node is labelled by the winner of its two children under T. The winner of B under T is the label of the root of B under this labelling.*

Definition 5.2.10. *The random single-elimination bracket rule r is a tournament rule on $n = 2^h$ players where $r_i(T)$ is the probability player i is the winner of B under T when B is chosen uniformly at random from the set of $n!$ possible brackets.*

If n is not a power of 2, we define the random single-elimination bracket rule on n players by introducing $2^{\lceil \log_2 n \rceil} - n$ dummy players who lose to all of the existing n players.

It is straightforward to check that the random single-elimination bracket rule is both Condorcet-consistent and monotone. Our main result (Theorem 5.3.3) shows that in addition to these properties, the random single-elimination bracket rule is 2-SNM-1/3 (which is the best possible, by Theorem 5.3.1).

We give some examples of other common tournament rules in Section 5.3.4. While many of these rules are both Condorcet-consistent and monotone, we do not know of any which are additionally 2-SNM-1/3.

5.3 Main Result

5.3.1 Lower bounds for k-SNM-α

We begin by showing that no tournament rule is 2-SNM-α for $\alpha < 1/3$. A similar theorem appears as Proposition 17 in [4] (which states that $\alpha = 0$ is impossible).

Theorem 5.3.1. *There is no Condorcet-consistent tournament rule on n players (for $n \geq 3$) that is 2-SNM-α for $\alpha < \frac{1}{3}$.*

Proof. Consider the tournament T on three players A, B, and C where A beats B, B beats C, and C beats A (illustrated in Figure 5.1). . Note that, while this tournament has no Condorcet winner, changing the result of any of the three games results in a Condorcet winner. For example, if A bribes C to lose to A, then A becomes the Condorcet winner.

If we have a tournament rule r that is 2-SNM-α, then combining this with the above fact gives rise to the following three inequalities.

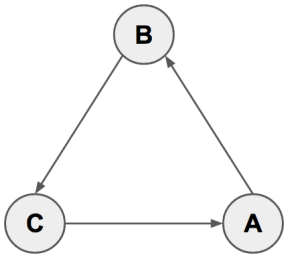


Figure 5.1: A tournament which attains the lower bound of $\alpha = 1/3$ for all tournament rules.

$$\begin{aligned} r_A(T) + r_B(T) &\geq 1 - \alpha \\ r_B(T) + r_C(T) &\geq 1 - \alpha \\ r_C(T) + r_A(T) &\geq 1 - \alpha \end{aligned}$$

Together these imply $r_A(T) + r_B(T) + r_C(T) \geq \frac{3}{2}(1 - \alpha)$. But $r_A(T) + r_B(T) + r_C(T) = 1$; it follows that $\alpha \geq \frac{1}{3}$, as desired.

We can extend this counterexample to $n > 3$ players by introducing $n - 3$ dummy players who all lose to A, B, and C; the argument above continues to hold.

□

We can use similar logic to prove lower bounds for the more general case of k-SNM-α.

Theorem 5.3.2. *There is no Condorcet-consistent tournament rule on n players (for $n \geq 2k - 1$) that is k-SNM-α for $\alpha < \frac{k-1}{2k-1}$.*

Proof. Consider the following tournament T on the $2k - 1$ players labelled 1 through $2k-1$. Each player i wins their match versus the $k-1$ players $i+1, i+2, \ldots, i+(k-1)$, and loses their match versus the $k-1$ players $i-1, i-2, \ldots, i-(k-1)$ (indices taken modulo $2k - 1$). Note that the coalition of players $S_i = \{i, i-1, \ldots, i-(k-1)\}$ of

size k can cause i to become a Condorcet winner if all players in the coalition agree to lose their games with i. If we have a tournament rule r that is k-SNM-α, then this implies the following $2k-1$ inequalities (one for each $i \in [2k-1]$):

$$\sum_{j \in S_i} r_j(T) \geq 1 - \alpha \tag{5.1}$$

Summing these $2k-1$ inequalities, we obtain

$$k \sum_{j=1}^{2k-1} r_j(T) \geq (2k-1)(1-\alpha) \tag{5.2}$$

Since $\sum_{j=1}^{2k-1} r_j(T) \leq 1$, this implies that $\alpha \geq \frac{k-1}{2k-1}$, as desired. Again, it is possible to extend this example to any number of players $n \geq 2k-1$ by introducing dummy players who lose to all $2k-1$ of the above players. □

5.3.2 Random single elimination brackets are 2-SNM-1/3

We now show that the random single elimination bracket rule is optimal against coalitions of size 2. The proof idea is simple; for every bracket B that contributes to the incentive to manipulate $r_i(T') + r_j(T') - r_i(T) - r_j(T)$ we will show that there are two that do not (in other words, for every scenario where team i benefits from the manipulation, there exist two other scenarios where the maniuplation does not benefit either team).

Theorem 5.3.3. *The random single elimination bracket rule is 2-SNM-1/3.*

Proof. Let $\mathcal{B}$ be the set of $n!$ different possible brackets amongst the n players. For a given tournament T and a given player i, write $\mathbb{1}(B, T, i)$ to represent the indicator variable which is 1 if i wins bracket B under the outcomes in T and 0 otherwise. Then we can write

$$r_i(T) = \frac{1}{|\mathcal{B}|} \sum_{B \in \mathcal{B}} \mathbb{1}(B, T, i).$$

Assume i loses to j in T. Then, if we let T' be the tournament that is $\{i, j\}$ adjacent to T, we can write the increase in utility resulting from j throwing to i

$$\frac{1}{|\mathcal{B}|} \sum_{B \in \mathcal{B}} \left(\mathbb{1}(B, T', i) + \mathbb{1}(B, T', j) - \mathbb{1}(B, T, i) - \mathbb{1}(B, T, j) \right). \tag{5.3}$$

Our goal is to show that this sum is at most $1/3$. Now, note that if i does not end up playing j in bracket B under T, i also does not play j in B under T' (and vice versa). In these brackets, $\mathbb{1}(B, T', i) = \mathbb{1}(B, T, i)$ and $\mathbb{1}(B, T', j) = \mathbb{1}(B, T, j)$, so these brackets contribute nothing to the sum in Equation 5.3. On the other hand, in a bracket B where i does play j, we are guaranteed that $\mathbb{1}(B, T, i) = 0$ and $\mathbb{1}(B, T', j) = 0$ (since i loses to j in T and j loses to i in T'). Therefore, letting $\mathcal{B}_{ij}$ be the subset of $\mathcal{B}$ of brackets where i meets j, we can rewrite Equation 5.3 as

$$\frac{1}{|\mathcal{B}|} \sum_{B \in \mathcal{B}_{ij}} \left(\mathbb{1}(B, T', i) - \mathbb{1}(B, T, j) \right).$$

Since $\mathbb{1}(B, T', i) \leq 1$, this is at most

$$\frac{1}{|\mathcal{B}|} \sum_{B \in \mathcal{B}_{ij}} \left(1 - \mathbb{1}(B, T, j) \right).$$

This final sum counts exactly the number of brackets B where i and j meet (under T, so j beats i) but j does not win the tournament. Call such brackets *bad*, and call the remaining brackets *good*. We will exhibit two injective mappings σ_i and σ_j from bad brackets to good brackets such that the ranges of σ_i and σ_j are disjoint. This implies that there are at least twice as many good brackets as bad brackets, and thus that the sum above is at most $1/3$, completing the proof.

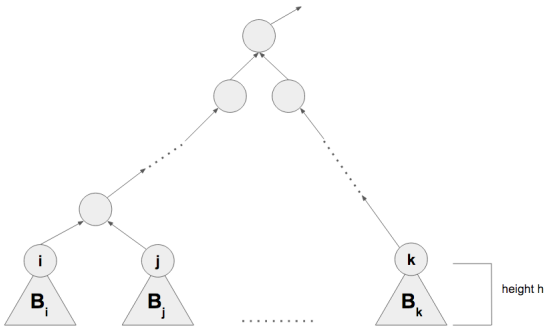


Figure 5.2: An example of a bad bracket B.

For both mappings, we will need the following terminology. Consider a bad bracket B, and consider the path from j up to the root of this tree. The nodes of this path are labelled by players that j would face if they got that far. More specifically, j has some opponent in the first round. Should j win, j would face some opponent in the second round, then the third round, etc. all the way to the finals, and these opponents do not depend on the outcomes of any of j's matches. Then since B is a bad bracket, j does not win, and at least one of the players on this path can beat j. Choose the **latest** such player (i.e. the closest to the root) and call this player k. Note that k might *not* be the player that knocks j out of the tournament (that is the *first* player along this path who would beat j).

Suppose that i and j meet at height h of the bracket (i.e. in the h^{th} round). Let B_i, B_j, B_k be the subtrees of height h that contain i, j, and k respectively. An example is shown in Figure 5.2.

We first describe the simpler of the two maps, σ_i. Define $\sigma_i(B)$ by swapping the subtrees B_i and B_k as shown in Figure 5.3. In this bracket j will lose to k before ever meeting i, so $\sigma_i(B)$ is good. Moreover σ_i is injective since we can construct its inverse. In $\sigma_i(B)$, j certainly would lose to k at height h before reaching i. Furthermore, because we didn't change B_j at all, j still wins all of its first $h-1$ matches and makes it to k (because we started from a B where j makes it to i at height h). So we can

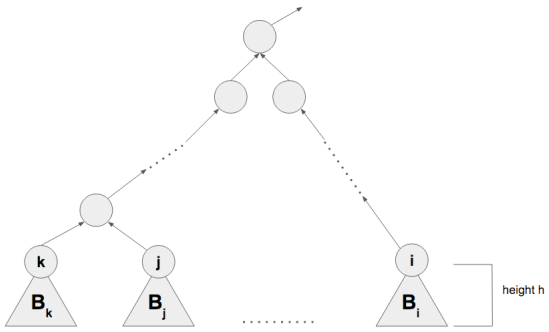


Figure 5.3: $\sigma_i(B)$.

identify k as the first player who beats j in $\sigma_i(B)$, learn the height h, and undo the swap of B_k and B_i.

We now describe the second map, σ_j. To construct $\sigma_j(B)$, begin by swapping the subtrees B_j and B_k (see Figure 5.4). Note that the bracket formed in this way is good; since we chose k to be the latest player on j's path to victory that can beat j, if j meets i, j will also beat all subsequent players and win the tournament (note that it is of course possible that j doesn't even make it to i, in which case $\sigma_j(B)$ is still good. But it is clear that *if* j meets i, then j will win the tournament, so $\sigma_j(B)$ is good in either case). Unfortunately, this map as stated is not injective; in particular, we cannot recover the height h to undo the swap as in the previous case.

The only reason we cannot uniquely identify k in the same way as when we invert σ_i is that i might meet some player k' at height $h' < h$ in B_i who also could beat j. So, intuitively, we would like to swap such players out with players who lose to j. Since j beats all of its opponents in B_j, B_j is an ample source of such players. We will therefore perform some additional 'subswap' operations, swapping subtrees of B_j and B_k so as to uniquely identify k as the first player i meets in $\sigma_j(B)$ who can beat j.

Specifically, for $0 \leq h' < h$, let $a(h')$ be the opponent i plays at height h' in B_i, and let $B_i(h')$ be the subtree of B_i with root $a(h')$ (note that the player that i meets at height h' is the root of a subtree of height $h' - 1$, and that all these subtrees are

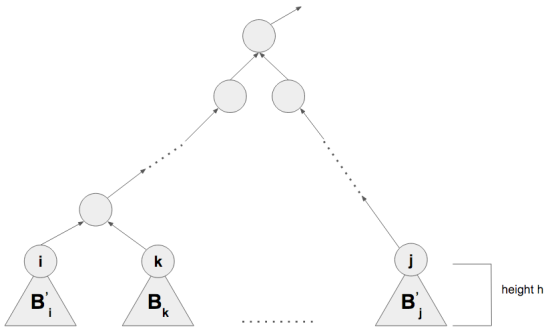


Figure 5.4: $\sigma_j(B)$.

disjoint). Similarly, let $b(h')$ be the opponent j plays at height h' in B_j, and let $B_j(h')$ be the subtree of B_j with root $b(h')$. To construct $\sigma_j(B)$ from B, first swap B_j and B_k. Then for each $h' \in [0, h)$ such that $a(h')$ would beat j, swap the subtrees $B_i(h')$ and $B_j(h')$. See Figure 5.5 for an illustration of a subswap operation.

Note that $\sigma_j(B)$ is still good; it is still the case that if j meets i, j will beat all subsequent players (all we have done in that part of the bracket is perhaps alter whether or not j will indeed meet i). On the other hand, since j makes it to height h in B_j, j can beat player $b(h')$ for all h', so k is now the first player i would encounter in $\sigma_j(B)$ who can beat j. From this, we can recover k and thus h, and undo the swap of B_i and B_j. To undo the subswaps, observe that because we started with a bad bracket B, that j must have beaten all opponents it faces in the first h rounds. Since all opponents on j's path who beat j at height less than h were necessarily put there by our subswap operations, we can just find all such opponents and swap them back out. This process inverts σ_j, thus proving that σ_j is injective.

Finally, note that in $\sigma_i(B)$, k must play j before either plays i, whereas in $\sigma_j(B)$, k must play i before either plays j. Therefore the ranges of σ_i and σ_j are disjoint, and this completes the proof.

For the reader aiming to understand our coupling argument better, Appendix D.1 contains some specific examples. □

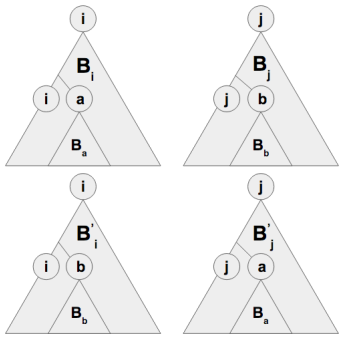


Figure 5.5: Subswap operation for σ_j.

5.3.3 Extension to randomized outcomes

Thus far we have been assuming that all match results are deterministic and known to the players in advance. Of course, this is not true in general; in real life, the outcomes of games are inherently unpredictable. It is perhaps imaginable that this unpredictability could increase the incentive to manipulate. In this section we show that this is not the case; a simple application of linearity of expectation shows that results about deterministic tournaments still hold for their randomized counterparts. We begin by defining a randomized tournament as follows.

Definition 5.3.4. *A randomized tournament $\mathcal{T}$ is a random variable whose values range over (deterministic) tournaments T. As shorthand, we will write $\mathbb{P}_{\mathcal{T}}(T)$ to represent the probability that $\mathcal{T} = T$.*

Note that this definition accounts for the most straightforward generalization of tournament outcomes from deterministic to randomized, where for each match between players i and j we assign a probability p_{ij} to the probability that i beats j. This definition further allows for the possibility of correlation between matches (e.g., with some probability player i has a good day and wins all his matches, and with some probability he has a bad day and loses all his matches).

Manipulations in this randomized model are similar to manipulations in the deterministic model in that they effectively force the result of a match to a win or a loss.

Formally, let $\sigma_{ij}(T)$ for a (deterministic) tournament T be the tournament formed by T but where i beats j (if i beats j in T, then $\sigma_{ij}(T) = T$). A tournament rule r is 2-SNM-α if for all i and j,

$$\mathbb{E}_{\mathcal{T}}\left[r_i(\sigma_{ij}(\mathcal{T})) + r_j(\sigma_{ij}(\mathcal{T})) - r_i(\mathcal{T}) - r_j(\mathcal{T})\right] \leq \alpha \tag{5.4}$$

We then have the following theorem:

Theorem 5.3.5. *If a rule r is 2-SNM-α in the deterministic tournament model, it is also 2-SNM-α in the randomized tournament model.*

Proof. Note that we can write the expectation in Equation 5.4 as

$$\sum_{T} \mathbb{P}_{\mathcal{T}}(T)\left(r_i(\sigma_{ij}(T)) + r_j(\sigma_{ij}(T)) - r_i(T) - r_j(T)\right)$$

If r is 2-SNM-α for deterministic tournaments, then each term in this sum is at most $\mathbb{P}_{\mathcal{T}}(T)\alpha$. It follows that this sum is at most α, and therefore r is also 2-SNM-α for randomized tournaments. □

It is straightforward to generalize the above definitions and result to the case of k-SNM-α.

5.3.4 Other tournament formats

Finally, there are many other tournament formats that are either used in practice or have been previously studied. In this section we show that many of these formats are more susceptible to manipulation than the random single elimination bracket rule; in particular, all of the following formats are at best 2-SNM-1/2.

By far the most common tournament rule for round robin tournaments is some variant of a 'scoring' rule, where the winner is the player who has won the most games (with ties broken in some fashion if multiple players have won the same maximum

number of games). In voting theory, this rule is often called Copeland's rule, or Copeland's method [46].

Definition 5.3.6. *A tournament rule r is a Copeland rule if the winner is always selected from the set of players with the maximum number of wins.*

We begin by showing that no Copeland rule can be 2-SNM-α for any $\alpha < 1$ (regardless of how the rule breaks ties).

Theorem 5.3.7. *There is no Copeland rule on n players that is 2-SNM-α for $\alpha < 1 - \frac{2}{n-1}$.*

Proof. Assume to begin that $n = 2k + 1$ is odd, and let r be a Copeland rule on n players. Let T be the tournament where each player i beats the k players $\{i+1, i+2, \ldots, i+k\}$ but loses to the k players $\{i-1, i-2, \ldots, i-k\}$, with indices taken modulo n (similar to the tournament in the proof of Theorem 5.3.2).

Since $\sum_{i=1}^{n} r_i(T) = 1$, there must be some i such that $r_{i-1}(T) + r_i(T) \leq \frac{2}{n}$. On the other hand, if player $i-1$ throws their match to player i, then player i becomes the unique Copeland winner (winning $k+1$ games) and $r_i(T') = 1$. It follows that, for such a rule, if r is 2-SNM-α, then $\alpha \geq 1 - \frac{2}{n}$.

If n is even, then we can embed the above example for $n-1$ by assigning one player to be a dummy player that loses to all teams. This immediately implies $\alpha \geq 1 - \frac{2}{n-1}$ in this case. □

In [4], Altman and Kleinberg provide three examples of tournament rules that are Condorcet-consistent and monotone: the top cycle rule, the iterative Condorcet rule, and the randomized voting caterpillar rule. We prove lower bounds on α for each of these in turn. Interestingly, the same tournament provides all three lower bounds.

Definition 5.3.8. *The* superman-kryptonite *tournament on n players has i beat j whenever $i < j$, except that player n beats player 1. That is, player 1 beats everyone except for player n, who loses to everyone except for player 1.*

Now we show that the superman-kryptonite tournament provides lower bounds against the tournament rules considered in [4].

Definition 5.3.9. *The top cycle of a tournament T is the minimal set of players who never lose to any other player. The top cycle rule is a tournament rule which assigns the winner to be a uniformly random element of this set.*

Theorem 5.3.10. *The top cycle rule on n players is not 2-SNM-α for any $\alpha < 1 - \frac{2}{n}$.*

Proof. Let T be the superman-kryptonite tournament on n players. The top cycle in T contains all the players, so $r_1(T) + r_n(T) = \frac{2}{n}$. However, if player n throws their match to player 1, player 1 becomes a Condorcet winner and $r_1(T') = 1$. It follows that $\alpha \geq 1 - \frac{2}{n}$. □

Definition 5.3.11. *The iterative Condorcet rule is a tournament rule that uniformly removes players at random until there is a Condorcet winner, and then assigns that player to be the winner.*

Theorem 5.3.12. *The iterative Condorcet rule on n players is not 2-SNM-α for any $\alpha < \frac{1}{2} - \frac{1}{n(n-1)}$.*

Proof. Let T be the superman-kryptonite tournament on n players. Note that no Condorcet winner will appear until either player 1 is removed, player n is removed, or all other $n-2$ players are removed. If all the other $n-2$ players are removed before players 1 or n (which occurs with probability $\frac{2}{n(n-1)}$), then player n wins. If this does not happen and player n is removed before player 1 (which occurs with probability $\frac{1}{2}\left(1 - \frac{2}{n(n-1)}\right) = \frac{1}{2} - \frac{1}{n(n-1)}$), then player 1 becomes the Condorcet winner and wins. Otherwise, player 1 will be removed before player n, while some players in 2 through $n-1$ remain, and one of them will become the Condorcet winner (the remaining player in $\{2, \ldots, n-1\}$ with lowest index). It follows that $r_1(T) = \frac{1}{2} - \frac{1}{n(n-1)}$ and $r_n(T) = \frac{2}{n(n-1)}$, so $r_1(T) + r_n(T) = \frac{1}{2} + \frac{1}{n(n-1)}$.

On the other hand, if player n throws their match to player 1, then again player 1 becomes a Condorcet winner and $r_1(T') = 1$. It follows that $\alpha \geq \frac{1}{2} - \frac{1}{n(n-1)}$. $\square$

Definition 5.3.13. *The randomized voting caterpillar rule is a tournament rule which chooses a winner as follows. Choose a random permutation π of $[n]$. Start by matching $\pi(1)$ and $\pi(2)$, and choose a winner according to T. Then for all $i \geq 3$ match $\pi(i)$ with the winner of the most recent match. The player that wins the last match (against $\pi(n)$) is declared the winner.*

Theorem 5.3.14. *The randomized voting caterpillar rule on n players is not 2-SNM-α for any $\alpha < \frac{1}{2} - \frac{n-3}{n(n-1)}$.*

Proof. Let T be the superman-kryptonite tournament on n players. The only way player 1 loses is if either player n occurs later in π than player 1 (which happens with probability $\frac{1}{2}$) or if $\pi(n) = 1$ and $\pi(1) = 2$ and they play in the first round (which happens with probability $\frac{1}{n(n-1)}$). The only way player n can win is if $\pi(n) = n$ (i.e., they only play the very last game), in which case they will play player 1 and win (this happens with probability $\frac{1}{n}$). It follows that $r_1(T) = \frac{1}{2} - \frac{1}{n(n-1)}$ and $r_n(T) = \frac{1}{n}$, so $r_1(T) + r_n(T) = \frac{1}{2} + \frac{n-2}{n(n-1)}$.

On the other hand, if player n throws their match to player 1, then again player 1 becomes a Condorcet winner and $r_1(T') = 1$. It follows that $\alpha \geq \frac{1}{2} - \frac{n-2}{n(n-1)}$. $\square$

Chapter 6

Optimal instance adaptive algorithm for the top-K ranking problem

This chapter is joint work with Xi Chen, Sivakanth Gopi, and Jieming Mao [40].

6.1 Introduction

The problem of inferring a ranking over a set of n items, such as documents, images, movies, or URL links, is an important problem in machine learning and finds many applications in recommender systems, web search, social choice, and many other areas. One of the most popular forms of data for ranking is pairwise comparison data, which can be easily collected via, for example, crowdsourcing, online games, or tournament play. The problem of ranking aggregation from pairwise comparisons has been widely studied and most work aims at inferring a total ordering of all the items (see, e.g., [114]). However, for some applications with a large number of items (e.g., rating of restaurants in a city), it is only necessary to identify the set of top K items. For these applications, inferring the total global ranking order unnecessarily increases the

complexity of the problem and requires significantly more samples. Typically, the sample complexity of recovering the set of top K items is inversely related to the gap between item K and item $K+1$. On the other hand, the sample complexity of recovering the global ranking order might depend on the the minimum of the gaps between two consecutive items.

In the basic setting for this problem, there is a set of n items with some true underlying ranking. For possible pair (i, j) of items, an analyst is given r noisy pairwise comparisons between those two items, each independently ranking i above j with some probability p_{ij}. From this data, the analyst wishes to identify the top K items in the ranking, ideally using as few samples r as is necessary to be correct with sufficiently high probability. The noise in the pairwise comparisons (i.e., the probabilities p_{ij}) is constrained by the choice of noise model. Many existing models - such as the Bradley-Terry-Luce model (BTL) [25, 101], the Thurstone model [140], and their variants - are *parametric* comparison models, in that each probability p_{ij} is of the form $f(s_i, s_j)$, where s_i is a 'score' associated with item i. While these parametric models yield many interesting algorithms with provable guarantees [41, 78, 138], the models enforce strong assumptions on the probabilities of incorrect pairwise comparisons that might not hold in practice [52, 104, 141, 19].

A more general class of pairwise comparison model is the strong stochastic transitivity (SST) model, which subsumes the aforementioned parameter models as special cases and has a wide range of applications in psychology and social science (see, e.g., [52, 104, 62]). The SST model only enforces the following coherence assumption: if i is ranked above j, then $p_{il} \geq p_{jl}$ for all other items l. [130] pioneered the algorithmic and theoretical study of ranking aggregation under SST models. For top-K ranking problems, [132] proposed a counting-based algorithm under a very general noise model that includes SST as a special case. The algorithm simply orders the items by the total number of pairwise comparisons won. For a certain class of instances, this

algorithm is in fact optimal; any algorithm with a constant probability of success on these instances needs roughly at least as many samples as this counting algorithm. However, this does not rule out the existence of other instances where the counting algorithm performs asymptotically worse than some other algorithm (see the example in Eq. (6.1)).

Under the SST model, we study algorithms for the top-K problem from the standpoint of *instance-specific analysis* (a.k.a. *competitive analysis* in the computer science). This is in spirit very similar to the notion "instance optimal"[61]. We give an algorithm which, on any instance, needs at most $\tilde{O}(\sqrt{n})$ times as many samples as the best possible algorithm for that instance to succeed with the same probability. We further show this result is tight (up to polylogarithmic factors): for any algorithm, there are instances where that algorithm needs at least $\tilde{\Omega}(\sqrt{n})$ times as many samples as the best possible algorithm. In contrast, the counting algorithm of [132] sometimes requires $\Omega(n)$ times as many samples as the best possible algorithm, even when the probabilities p_{ij} are bounded away from 1.

Our main technical tool is the introduction of a new decision problem we call *domination*, which captures the difficulty of solving the top-K problem while being simpler to directly analyze via information theoretic techniques. The domination problem can be thought of as a restricted one-dimensional variant of the top-K problem, where the analyst is only given the outcomes of pairwise comparisons that involve item i or j, and wishes to determine whether i is ranked above j. Our proof of the above claims proceeds by proving analogous competitive ratio results for the domination problem, and then carefully embedding the domination problem as part of the top-K problem. To establish the competitive ratio for the domination, we start from a simple case where the comparison probabilities are bounded away from zero and one. We first show that a popular counting algorithm developed by [132] has a suboptimal competitive ratio of $\tilde{\Theta}(n)$. The main reason is that the counting algorithm

treats samples from different coordinates of comparison probability vector equally. To address the issue of the counting algorithm, another *maximum algorithm* is first proposed. However, the maximum algorithm still leads to a sub-optimal competitive ratio and it fails whenever the counting algorithm performs well. Therefore, we develop techniques to combine the counting and maximum algorithms together, which give the optimal competitive ratio of $\tilde{O}(\sqrt{n})$. More detailed description of this idea is provided in Section 6.3.1.

6.1.1 Related Work

The problem of sorting a set of items from a collection of pairwise comparisons is one of the most classical problems in computer science and statistics. Many works investigate the problem of recovering the total ordering under noisy comparisons drawn from some parametric model. For the BTL model, Negahban et al. [114] propose the *Rank-Centrality* algorithm, which serves as the building block for many spectral ranking algorithms. Lu and Boutilier [100] give an algorithm for sorting in the Mallows model. Rajkumar and Agarwal [120] investigate which statistical assumptions (BTL models, generalized low-noise condition, etc.) guarantee convergence of different algorithms to the true ranking.

More recently, the problem of top-K ranking has received a lot of attention. Chen and Suh [41], Jang et al. [78], and Suh et al. [138] all propose various spectral methods for the BTL model or a mixture of BTL models. Eriksson [60] considers a noisy observation model where comparisons deviating from the true ordering are *i.i.d.* with bounded probability. In [132], Shah and Wainwright propose a counting-based algorithm, which motivates our work. However, their algorithm is not instance adaptive and we provide a simple instance (see Eq. (6.1)) illustrating that the sample complexity in [132] is sub-optimal on that instance. The top-K ranking problem is also related to the best K arm identification in multi-armed bandit [33, 76, 147].

However, in the latter problem, the samples are *i.i.d.* random variables rather than pairwise comparisons and the goal is to identify the top K distributions with largest means.

This paper and the above references all belong to the *non-active* setting: the set of data provided to the algorithm is fixed, and there is no way for the algorithm to adaptively choose additional pairwise comparisons to query. In several applications, this property is desirable, specifically if one is using a well-established dataset or if adaptivity is costly (e.g., on some crowdsourcing platforms). Nonetheless, the problems of sorting and top-K ranking are incredibly interesting in the adaptive setting as well. Several works [2, 77, 86, 27] consider the adaptive noisy sorting problem with (noisy) pairwise comparisons and explore the sample complexity to recover an (approximately) correct total ordering in terms of some distance function (e.g,., Kendall's tau). In [144], Wauthier et al. propose simple weighted counting algorithms to recovery an approximate total ordering from noisy pairwise comparisons. Dwork et al. [58] and Ailon et al. [3] consider a related *Kemeny optimization* problem, where the goal is to determine the total ordering that minimizes the sum of the distances to different permutations. More recently, the top-K ranking problem in the active setting has been studied by Braverman et al. [26] where they consider the tradeoff between the sample complexity of algorithms and the number of rounds of adaptivity. All of this work takes place in much more constrained noise models than the SST model. A very recent work by Heckel et. al. [74] investigates the active ranking under a general class of nonparametric models and also establishes a lower bound on the number of comparisons for parametric models. However, developing an active instance-adaptive ranking algorithm under the SST model still remains an interesting open problem.

The instance adaptivity has been widely studied in many statistical estimation problems. For example, the adaptive estimation is an important topic in nonparametric shape-restricted regression (see, e.g., [72, 38, 39, 71]). Shah et. al. [131]

study the adaptive estimation problem for estimating comparison probabilities in a SST model. The concept of instance adaptivity is also closely related to the oracle inequality, which relates the performance of a constructed estimator with that of an "oracle" estimator with the information about local structure of the parameter space (see the survey paper [36] and the book [91] and references therein).

[132] discussed the approximate recovery of top items. The approximate recovery would be suitable for many practical applications. In their paper, they showed that this approximate relaxation allows a less constrained separation threshold. For our algorithms, it is not clear that the approximate relaxation can significantly improve the competitive ratios. It is an interesting open question to extend our work to see if the approximate recovery can result in better competitive ratios.

6.2 Preliminaries and Problem Setup

6.2.1 The Top-K problem

Consider the following problem. An analyst is given a collection of n items, labelled 1 through n. These items have some true ordering defined by a permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that for $1 \leq u < v \leq n$, the item labelled $\pi(u)$ has a better rank than the item labelled $\pi(v)$ (i.e., the item with label i has a better rank than the item j if and only if $\pi^{-1}(i) < \pi^{-1}(j)$). The analyst's goal is to determine the set of the top K items, i.e., $\{\pi(1), \ldots, \pi(k)\}$.

The analyst receives r samples. Each sample consists of pairwise comparisons between all pairs of items. All the pairwise comparisons are independent with each other. The outcomes of the pairwise comparison between any two items is characterized by the probability matrix $\mathbf{P} \in [0,1]^{n \times n}$. For a pair of items (i, j), let $X_{i,j} \in \{0, 1\}$ be the outcome of the comparison between the item i and j, where $X_{i,j} = 1$ means i is preferred to j (denoted by $i \succ j$) and $X_{i,j} = 0$ otherwise. Further, let $\mathcal{B}(z)$

denote the Bernoulli random variable with mean $z \in [0,1]$. The outcome $X_{i,j}$ follows $\mathcal{B}(\mathbf{P}_{\pi^{-1}(i),\pi^{-1}(j)})$, i.e.,

$$\Pr(X_{i,j} = 1) = \Pr(i \succ j) = \mathbf{P}_{\pi^{-1}(i),\pi^{-1}(j)}.$$

The probability matrix $\mathbf{P}$ is said to be strong stochastic transitive (SST) if it satisfies the following definition.

Definition 6.2.1. *The $n \times n$ probability matrix $\mathbf{P} \in [0,1]^{n \times n}$ is strong stochastic transitive (SST) if*

1. *For $1 \leq u < v \leq n$, $\mathbf{P}_{u,l} \geq \mathbf{P}_{v,l}$ for all $l \in [n]$.*
2. *$\mathbf{P}$ is shifted-skew-symmetric (i.e., $\mathbf{P} - 0.5$ is skew-symmetric) where $\mathbf{P}_{v,u} = 1 - \mathbf{P}_{u,v}$ and $\mathbf{P}_{u,u} = 0.5$ for $u \in [n]$.*

The first condition claims that when the item i has a higher rank than item j (i.e., $\pi^{-1}(i) < \pi^{-1}(j)$), for any other item k, we have

$$\Pr(i \succ k) = \mathbf{P}_{\pi^{-1}(i),\pi^{-1}(k)} \geq \Pr(j \succ k) = \mathbf{P}_{\pi^{-1}(j),\pi^{-1}(k)}.$$

Remark 6.2.1. *Many classical parametric models such that BTL [25, 101] and Thurstone (Case V) [140] models are special cases of SST. More specifically, parametric models assume a score vector $w_1 \geq w_2 \geq \ldots \geq w_n$. They further assume that the comparison probability $\mathbf{P}_{u,v} = F(w_u - w_v)$, where $F : \mathbb{R} \to [0,1]$ is a non-decreasing function and $F(t) = 1 - F(-t)$ (e.g., $F(t) = 1/(1 + \exp(-t))$ in BTL models). By the property of F, it is easy to verify that $\mathbf{P}_{u,v} = F(w_u - w_v)$ satisfy the conditions in Definition 6.2.1.*

Under the SST models, we can formally define the top-K ranking problem as follows. The top-K ranking problem takes the inputs n, k, r that are known to the algorithm and the SST probability matrix $\mathbf{P}$ that is unknown to the algorithm.

Definition 6.2.2. TOP-K$(n, k, \mathbf{P}, r)$ *is the following algorithmic problem:*

1. *A permutation* π *of* $[n]$ *is uniformly sampled.*

2. *The algorithm is given samples* $X_{i,j,l}$ *for* $i \in [n], j \in [n], l \in [r]$*, where each* $X_{i,j,l}$ *is sampled independently according to* $\mathcal{B}(\mathbf{P}_{\pi^{-1}(i),\pi^{-1}(j)})$*. The algorithm is also given the value of* k*, but not* π *or the matrix* $\mathbf{P}$*.*

3. *The algorithm succeeds if it correctly outputs the set of labels* $\{\pi(1), ..., \pi(k)\}$ *of the top* k *items.*

Remark 6.2.2. *We note that [132] considers a slightly different observation model in which each pair is queried* r *times. For each query, one can obtain a comparison result with the probability* $p_{obs} \in (0, 1]$ *and with probability* $1-p_{obs}$*, the query is invalid. In this model, each pair will be compared* $r \cdot p_{obs}$ *times on expectation. When* $p_{obs} = 1$*, it reduces to our model in Definition 6.2.2, where we observe exactly* r *comparisons for each pair. Our results can be easily extended to deal with the observation model in [132] by replacing* r *with the effective sample size,* $r \cdot p_{obs}$*. We omit the details for the sake of simplicity.*

Our primary metric of concern is the *sample complexity* of various algorithms; that is, the minimum number of samples an algorithm A requires to succeed with a given probability. To this end, we call the triple $S = (n, k, \mathbf{P})$ an *instance* of the TOP-K problem, and write $r_{min}(S, A, p)$ to denote the minimum value such that for all $r \geq r_{min}(S, A, p)$, A succeeds on instance S with probability p when given r samples. When p is omitted, we will take $p = \frac{3}{4}$; i.e., $r_{min}(S, A) = r_{min}(S, A, \frac{3}{4})$. It is worthwhile to note that, by repeating the algorithm constant number of times and taking the majority output, solving the problem for any constant error translates to a solution with polynomially decaying error, and the sample complexity will increase only by a multiplicative logarithmic factor.

6.2.2 The Domination problem

To solve the problem of Top-K, we study a key sub-problem called Domination, which captures the core of the difficulty of Top-K. In particular, Domination captures the dominance relation between two consecutive rows of a SST probability matrix. Domination is formally defined as follows.

Definition 6.2.3. Domination$(n, \mathbf{p}, \mathbf{q}, r)$ *is the following algorithmic problem:*

1. $\mathbf{p} = (p_1, \cdots, p_n)$ *and* $\mathbf{q} = (q_1, \cdots, q_n)$ *are two vectors of probabilities that satisfy* $1 \geq p_i \geq q_i \geq 0$ *for all* $i \in [n]$. $\mathbf{p}, \mathbf{q}$ *are not given to the algorithm.*

2. *A random bit* B *is sampled from* $\mathcal{B}(\frac{1}{2})$. *Samples* $X_{i,j}, Y_{i,j}$ *(for* $i \in [n], j \in [r]$*) are generated as follows:*

   (a) *Case* $B = 0$*: each* $X_{i,j}$ *is independently sampled according to* $\mathcal{B}(p_i)$ *and each* $Y_{i,j}$ *is independently sampled according to* $\mathcal{B}(q_i)$.

   (b) *Case* $B = 1$*: each* $X_{i,j}$ *is independently sampled according to* $\mathcal{B}(q_i)$ *and each* $Y_{i,j}$ *is independently sampled according to* $\mathcal{B}(p_i)$.

   The algorithm is given the samples $X_{i,j}$ *and* $Y_{i,j}$*, but is not given the bit* B *or the values of* $\mathbf{p}$ *and* $\mathbf{q}$.

3. *The algorithm succeeds if it correctly outputs the value of the hidden bit* B.

From Definition 6.2.1, it is clear for any pair of rows (or columns) of a SST probability matrix $\mathbf{P}$, one row (or column) will dominate another. As before, we are interested in the sample complexity of algorithms for Domination. We call the triple $C = (n, \mathbf{p}, \mathbf{q})$ an instance of Domination, and write $r_{min}(C, A, p)$ to be the minimum value such that for all $r \geq r_{min}(C, A, p)$, algorithm A succeeds at solving Domination$(n, \mathbf{p}, \mathbf{q}, r)$ with probability at least p. Moreover, for notational simplicity, let $r_{min}(C, A) = r_{min}(C, A, \frac{3}{4})$.

There are at least two main approaches one can take to analyze the sample complexity of problems like TOP-K or DOMINATION. The first (and more common) approach is to bound the value of $r_{min}(S, A)$ by some explicit function $f(S)$ of a TOP-K instance S. This is the approach taken by [132]. They show that for some simple function f (roughly, the square of the reciprocal of the absolute difference of the sums of the k-th and $(k+1)$-th rows of the matrix $\mathbf{P}$ i.e. $1/\|\mathbf{P}_k - \mathbf{P}_{k+1}\|_1^2$), there is an algorithm A such that for all instances S, $r_{min}(S, A) = O(f(S))$; moreover this is optimal in the sense that there exists an instance S such that for all algorithms A, $r_{min}(S, A) = \Omega(f(S))$. While this is a natural approach, it leaves open the question of what the correct choice of f should be; indeed, different choices of f give rise to different 'optimal' algorithms A which outperform each other on different instances.

In this paper, we take the second approach, which is to compare the sample complexity of an algorithm on an instance to the sample complexity of the best possible algorithm on that instance. Formally, let $r_{min}(S, p) = \inf_A r_{min}(S, A, p)$ and let $r_{min}(S) = r_{min}(S, \frac{3}{4})$. An ideal algorithm A would satisfy $r_{min}(S, A) = \Theta(r_{min}(S))$ for all instances S of TOP-K; more generally, we are interested in bounding the ratio between $r_{min}(S, A)$ and $r_{min}(S)$. We call this ratio the *competitive ratio* of the algorithm, and say that an algorithm is $f(n)$-competitive if $r_{min}(S, A) \leq f(n) r_{min}(S)$. We likewise define the corresponding notions for DOMINATION.

6.3 Main Results

In our main upper bound result, we give a linear-time algorithm for TOP-K which is $\tilde{O}(\sqrt{n})$-competitive (restatement of Corollary 6.7.5):

Theorem 6.3.1. *There is an algorithm A for* Top-K *such that A runs in time $O(n^2 r)$ and on every instance S of* Top-K *on n items,*

$$r_{min}(S, A) \leq O(\sqrt{n} \log n) r_{min}(S).$$

In our main lower bound result, we show that up to logarithmic factors, this $\sqrt{n}$ competitive ratio is optimal (restatement of Theorem 6.8.1):

Theorem 6.3.2. *For any algorithm A for* Top-K*, there exists an instance S of* Top-K *on n items such that*

$$r_{min}(S, A) \geq \Omega\left(\frac{\sqrt{n}}{\log n}\right) r_{min}(S).$$

In comparison, for the counting algorithm A' of [132], there exist instances S such that $r_{min}(S, A') \geq \tilde{\Omega}(n) r_{min}(S)$. For example, consider the instance $S = (n, k, \mathbf{P})$ with

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}+\varepsilon & \cdots & \cdots & \frac{1}{2}+\varepsilon \\ \frac{1}{2}-\varepsilon & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \frac{1}{2}+\varepsilon \\ \frac{1}{2}-\varepsilon & \cdots & \cdots & \frac{1}{2}-\varepsilon & \frac{1}{2} \end{bmatrix} \tag{6.1}$$

It is straightforward to show that with $\Theta(\log n/\varepsilon^2)$ samples, we can learn all pairwise comparisons correctly with high probability by taking a majority vote, and therefore even sort all the elements correctly. This implies that $r_{min}(S) = O(\log n/\varepsilon^2)$. On the other hand, we show in Corollary 6.5.4 that $r_{min}(S, A') = \Omega(n/\varepsilon^2)$ when $\varepsilon < 1/10$.

6.3.1 Main Techniques and Overview

We prove our main results by first proving similar results for DOMINATION which we defined in Definition 6.2.3. Intuitively DOMINATION captures the main hardness of TOP-K while being much simpler to analyze. Once we prove upper bound and lower bounds for the sample complexity of DOMINATION, we will use reductions to prove analogous results for TOP-K.

We begin in Section 6.4, by proving a general lower bound on the sample complexity of domination. Explicitly, for a given instance $C = (n, \mathbf{p}, \mathbf{q})$ of DOMINATION, we show that $r_{min}(C) \geq \Omega(1/\mathbb{I}(\mathbf{p}, \mathbf{q}))$ where $\mathbb{I}(\mathbf{p}, \mathbf{q})$ is the amount of information we can learn about the bit B from one sample of pairwise comparison in each of the coordinates.

In Section 6.5, we proceed to design algorithms for DOMINATION restricted to instances $C = (n, \mathbf{p}, \mathbf{q})$ where $\delta \leq p_i, q_i \leq 1 - \delta$ for some constant $0 < \delta \leq 1/2$. In this regime $\mathbb{I}(\mathbf{p}, \mathbf{q}) = \Theta(\|\mathbf{p} - \mathbf{q}\|_2^2)$, which makes it easier to argue our algorithms are not too bad compared with the optimal one. We first consider an algorithm we call the counting algorithm $\mathcal{A}_{count}$ (Algorithm 9), which is a DOMINATION analogue of the counting algorithm proposed by [132]. We show that $\mathcal{A}_{count}$ has a competitive ratio of $\tilde{\Theta}(n)$. Intuitively, the main reason $\mathcal{A}_{count}$ fails is that $\mathcal{A}_{count}$ tries to consider samples from different coordinates equally important even when they are sampled from a very unbalanced distribution (for example, $p_1 \neq q_1, p_2 = q_2, ..., p_n = q_n$). We then consider another algorithm we call the max algorithm $\mathcal{A}_{max}$ (Algorithm 10) which simply finds $i' = \max_i |\sum_{j=1}^{r}(X_{i,j} - Y_{i,j})|$ and outputs B according the sign of $\sum_{j=1}^{r}(X_{i',j} - Y_{i',j})$. We show $\mathcal{A}_{max}$ also has a competitive ratio of $\tilde{\Theta}(n)$. Interestingly, $\mathcal{A}_{max}$ fails for a different reason from $\mathcal{A}_{count}$, namely that $\mathcal{A}_{max}$ does not use the information fully from all coordinates when the samples are sampled from a very balanced distribution. In fact, $\mathcal{A}_{count}$ performs well whenever $\mathcal{A}_{max}$ fails and vice versa. We therefore show how combine $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ in two different ways to get

two new algorithms: $\mathcal{A}_{comb}$ (Algorithm 11) and $\mathcal{A}_{cube}$ (Algorithm 12). We show that both of these new algorithms have a competitive ratio of $\tilde{O}(\sqrt{n})$, which is tight by Theorem 6.8.2. While $\mathcal{A}_{cube}$ has a slightly better competitive ratio ($O(\sqrt{n})$ versus $O(\sqrt{n \log n})$), the method introduced in $\mathcal{A}_{comb}$ is more general and allows one to combine any two algorithms for DOMINATION and to obtain the better one of the two performances on any instance.

In Section 6.6, we extend $\mathcal{A}_{comb}$ to design an efficient algorithm for DOMINATION in the general regime. In this regime, $\mathbb{I}(\mathbf{p}, \mathbf{q})$ can be much larger than $\|\mathbf{p} - \mathbf{q}\|_2^2$, particularly for values of p_i and q_i very close to 0 or 1. In these corner cases, the counting algorithm $\mathcal{A}_{count}$ and max algorithm $\mathcal{A}_{max}$ can fail very badly; we will show that even for fixed n, their competitive ratios can grow arbitrarily large (Lemma 6.6.6 and Lemma 6.6.7). One main reason for this failure is that, even when $|p_i - q_i| < |p_j - q_j|$, samples from coordinate i could convey much more information than the samples from coordinate j (consider, for example, $p_i = \varepsilon/2, q_i = 0$, and $p_j = 1/2 + \varepsilon, q_j = 1/2$). Taking this into account, we design a new algorithm $\mathcal{A}_{coup}$ (Algorithm 13) which has a competitive ratio of $O(\sqrt{n} \log n)$ in the general regime. The new algorithm builds off $\mathcal{A}_{coup}$ and still combines features from both $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$, but also better estimates the importance of each coordinate. To estimate how much information each coordinate has, the new algorithm divides the samples into $\Theta(\log n)$ groups and checks how often samples from coordinate i are consistent with themselves. If one coordinate has a large proportion of the total information, it uses samples from that coordinate to decide B, otherwise it takes a majority vote on samples from all coordinates.

In Section 6.7, we return to TOP-K and present an algorithm that has a competitive ratio of $\tilde{O}(\sqrt{n})$, thus proving Theorem 6.3.1. Our algorithm works by reducing the TOP-K problem to several instances of the DOMINATION problem (see Theorem 6.6.5). At a high level, the algorithm tries to find the top k rows by pairwise comparisons of rows, each of which can be thought of as an instance of DOMINATION.

We use algorithm $\mathcal{A}_{coup}$ to solve these DOMINATION instances. Since we only need to make at most n^2 comparisons, if $\mathcal{A}_{coup}$ outputs the correct answer with at least $1 - \frac{\varepsilon}{n^2}$ probability for each comparison, then by union bound all the comparisons will be correct with probability at least $1 - \varepsilon$. However, to find the top k rows, we do not actually need to compare all the rows to each other; Lemma 6.7.1 shows that we can find the top k rows with high probability while making only $O(n)$ comparisons. Using this lemma, we get a linear time algorithm (linear in the size of the input, i.e. $\Theta(n^2 r)$) for solving TOP-K. Finally in Lemma 6.7.4, we extend the lower bound for DOMINATION proved in Lemma 6.4.2 to show a lower bound on the number of samples any algorithm would need on a specific instance of TOP-K. Combining these results, we prove Theorem 6.3.1.

Finally, in Section 6.8, we show that the algorithms for both DOMINATION and TOP-K presented in the previous sections have the optimal competitive ratio (up to polylogarithmic factors). Specifically, we show that for any algorithm A solving DOMINATION, there exists an instance C of domination where $r_{min}(C, A) \geq \tilde{\Omega}(\sqrt{n}) r_{min}(C)$ (Theorem 6.8.2). We accomplish this by constructing a distribution $\mathcal{C}$ over instances of DOMINATION such that each instance in the support of this distribution can by solved by an algorithm with low sample complexity (Theorem 6.8.5) but any algorithm that succeeds over the entire distribution requires $\tilde{\Omega}(\sqrt{n})$ times more samples (Theorem 6.8.7). We then embed DOMINATION in TOP-K (similarly as in Section 6.7) to show an analogous $\tilde{\Omega}(\sqrt{n})$ lower bound for TOP-K (Theorem 6.8.1).

6.4 Lower bounds on the sample complexity of domination

We start by establishing lower bounds on the number of samples $r_{min}(C)$ needed by any algorithm to succeed with constant probability on a given instance $C = (n, \mathbf{p}, \mathbf{q})$ of DOMINATION. This is controlled by the quantity $\mathbb{I}(\mathbf{p}, \mathbf{q})$, which is the amount of information we can learn about the bit B given one sample of pairwise comparison between each of the coordinates of $\mathbf{p}$ and $\mathbf{q}$.

Definition 6.4.1. *Given* $0 \leq p, q \leq 1$*, define*

$$\mathbb{I}(p, q) = (p(1-q) + q(1-p)) \left(1 - H\left(\frac{p(1-q)}{p(1-q) + q(1-p)}\right)\right).$$

Given $\mathbf{p} = (p_1, \cdots, p_n) \in [0,1]^n, \mathbf{q} = (q_1, \cdots, q_n) \in [0,1]^n$*, define* $\mathbb{I}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \mathbb{I}(p_i, q_i)$.

Lemma 6.4.2. *Let* $C = (n, \mathbf{p}, \mathbf{q})$ *be an instance of* DOMINATION*. Then* $r_{min}(C) \geq 0.05/\mathbb{I}(\mathbf{p}, \mathbf{q})$.

Proof. The main idea is to bound the mutual information between the samples and the correct output, and then apply Fano's inequality. Let $\mathbf{p} = (p_1, \cdots, p_n)$ and $\mathbf{q} = (q_1, \cdots, q_n)$. Recall that B indicates the correct output and that $X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}$ are the samples given to the algorithm. By Fact E.1.1, $I(B; X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) = I(B; X_{1,1}Y_{1,1}) + I(B; X_{1,2}, ..., X_{n,r}, Y_{1,2}, ..., Y_{n,r} | X_{1,1}Y_{1,1})$. When $\mathbf{p}$, $\mathbf{q}$ and B are given, each sample ($X_{i,j}$ or $Y_{i,j}$) is independent of the other samples, and thus $I(X_{1,1}Y_{1,1}; X_{1,2}, ..., X_{n,r}, Y_{1,2}, ..., Y_{n,r} | B) = 0$. By Fact E.1.2, we then have $I(B; X_{1,2}, ..., X_{n,r}, Y_{1,2}, ..., Y_{n,r} | X_{1,1}Y_{1,1}) \leq I(B; X_{1,2}, ..., X_{n,r}, Y_{1,2}, ..., Y_{n,r})$ and therefore $I(B; X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) \leq I(B; X_{1,1}Y_{1,1}) + I(B; X_{1,2}, ..., X_{n,r}, Y_{1,2}, ..., Y_{n,r})$. Repeating this, we get $I(B; X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) \leq \sum_{i=1}^{n} \sum_{j=1}^{r} I(B; X_{i,j}Y_{i,j})$.

By Fact E.1.3, we have

$$\begin{aligned} & I(B; X_{i,j}Y_{i,j}) \\ = \; & \Pr[B=0] \cdot D(X_{i,j}Y_{i,j}|B=0 \| X_{i,j}Y_{i,j}) \\ & + \Pr[B=1] \cdot D(X_{i,j}Y_{i,j}|B=1 \| X_{i,j}Y_{i,j}) \\ = \; & (p_i(1-q_i) + q_i(1-p_i)) \\ & \cdot \left(1 - H\left(\frac{p_i(1-q_i)}{p_i(1-q_i)+q_i(1-p_i)}\right)\right) \\ = \; & \mathbb{I}(p_i, q_i). \end{aligned}$$

It follows that $I(B; X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) \leq \sum_{i=1}^{n} \sum_{j=1}^{r} I(B; X_{i,j}Y_{i,j}) = r \cdot \sum_{i=1}^{n} \mathbb{I}(p_i, q_i) = r\mathbb{I}(\mathbf{p}, \mathbf{q})$. For any algorithm, let p_e be its error probability on DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$. By Fano's inequality, we have that

$$\begin{aligned} H(p_e) \; & \geq \; H(B|X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) \\ & = \; H(B) - I(B; X_{1,1}, X_{1,2}, ..., X_{n,r}, Y_{1,1}, ..., Y_{n,r}) \\ & = \; 1 - r\mathbb{I}(\mathbf{p}, \mathbf{q}) \geq 0.95. \end{aligned}$$

Since $H(p_e) \geq 0.95$, we find that $p_e \geq 1/4$, as desired. □

In the following section, we will concern ourselves with instances $C = (n, \mathbf{p}, \mathbf{q})$ that satisfy $\delta \leq p_i, q_i \leq 1 - \delta$ for some constant δ for all i. For such instances, we can approximate $\mathbb{I}(p, q)$ by the ℓ_2 distance between $\mathbf{p}$ and $\mathbf{q}$.

Lemma 6.4.3. *For some $0 < \delta \leq \frac{1}{2}$, let $\delta \leq p, q \leq 1 - \delta$. Then*

$$\frac{1}{4 \ln 2}(p-q)^2 \leq \mathbb{I}(p, q) \leq \frac{1}{\delta \ln 2}(p-q)^2.$$

Proof. Let $x = p(1-q)$ and $y = q(1-p)$. Then $\mathbb{I}(p,q) = (x+y)(1 - H(\frac{x}{x+y}))$ and $p - q = x - y$. We need to show that

$$(x+y)\left(1 - H\left(\frac{x}{x+y}\right)\right) \leq \frac{1}{\delta \ln 2}(x-y)^2.$$

By Fact E.1.4,

$$\frac{1}{\ln 2}z^2 \leq 1 - H\left(\frac{1}{2}+z\right) = D\left(\frac{1}{2}+z\middle\|\frac{1}{2}\right) \leq \frac{4}{\ln 2}z^2,$$

and therefore

$$\frac{1}{4\ln 2}\frac{(x-y)^2}{(x+y)} \leq (x+y)\left(1 - H\left(\frac{x}{x+y}\right)\right) \leq \frac{1}{\ln 2}\frac{(x-y)^2}{(x+y)}.$$

Since $x+y = p(1-q)+q(1-p) \geq 2\sqrt{p(1-p)q(1-q)} \geq 2\delta(1-\delta) \geq \delta$, this implies the desired upper bound. The lower bound also holds since, $x+y = p(1-q)+q(1-p) \leq \sqrt{p^2+(1-p)^2}\cdot\sqrt{q^2+(1-q)^2} \leq \delta^2 + (1-\delta)^2 \leq 1$. □

Corollary 6.4.4. *Let $C = (n, \mathbf{p}, \mathbf{q})$ be an instance of* Domination *satisfying $\delta \leq \mathbf{p}_i, \mathbf{q}_i \leq 1-\delta$ for all $i \in [n]$. Then*

$$r_{min}(C) \geq 0.05\ln(2)\cdot\frac{\delta}{\|\mathbf{p}-\mathbf{q}\|_2^2}.$$

Proof. By Lemma 6.4.3, $\mathbb{I}(\mathbf{p},\mathbf{q}) \leq \|\mathbf{p}-\mathbf{q}\|_2^2/(\delta\ln 2)$. The result then follows from Lemma 6.4.2. □

6.5 Domination in the well-behaved regime

We now proceed to the problem of designing algorithms for Domination which are competitive on all instances. As a warmup, we begin by considering only instances $C = (n, \mathbf{p}, \mathbf{q})$ of Domination satisfying $\delta \leq p_i, q_i \leq 1-\delta$ for all $i \in [n]$ where $0 < \delta \leq$

1/2 is some fixed constant. This regime of instances captures much of the interesting behavior of DOMINATION, but with the added benefit that the mutual information between the samples and B behaves nicely in this regime: in particular $\mathbb{I}(\mathbf{p}, \mathbf{q}) = \Theta(\|\mathbf{p} - \mathbf{q}\|_2^2)$ (see Lemma 6.4.3). By Corollary 6.4.4, we have $r_{min} \geq \Omega(\frac{1}{\|\mathbf{p}-\mathbf{q}\|_2^2})$. This fact will make it easier to design algorithms for DOMINATION which are competitive in this regime.

In Section 6.5.1, we give two simple algorithms (counting algorithm and max algorithm) which can solve DOMINATION given $\tilde{O}(n/\|\mathbf{p} - \mathbf{q}\|_2^2)$ samples which gives them a competitive ratio of $\tilde{O}(n)$. We will then show that this is tight, i.e. their competitive ratio is $\tilde{\Theta}(n)$ in Lemma 6.5.3 and Lemma 6.5.5. While the sample complexities of these two algorithms are not optimal, they have the nice property that whenever one performs badly, the other performs well. In Section 6.5.2, we show how to combine the counting algorithm and the max algorithm to give two different algorithms which can solve DOMINATION using only $\tilde{O}(\sqrt{n}/\|\mathbf{p} - \mathbf{q}\|_2^2)$ samples i.e. they have a competitive ratio of $\tilde{O}(\sqrt{n})$. According to Theorem 6.8.2, this is the best we can do up to polylogarithmic factors.

6.5.1 Counting algorithm and max algorithm

We now consider two simple algorithms for DOMINATION$(n, \mathbf{p}, \mathbf{q})$, which we call the *counting algorithm* (Algorithm 9) and the *max algorithm* (Algorithm 10) denoted by $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ respectively. We show that both algorithms require $\tilde{O}(\frac{n}{\|\mathbf{p}-\mathbf{q}\|_2^2})$ samples to solve DOMINATION (Lemmas 6.5.1 and 6.5.2). By Corollary 6.4.4, we have $r_{min} \geq \Omega(\frac{1}{\|\mathbf{p}-\mathbf{q}\|_2^2})$, leading to a $\tilde{O}(n)$ competitive ratio for these algorithms. We show in Lemma 6.5.3 and Lemma 6.5.5 that this is tight up to polylogarithmic factors i.e. their competitive ratio is $\tilde{\Theta}(n)$.

Both the counting algorithm and the max algorithm begin by computing (for each coordinate i) the differences between the number of ones in the $X_{i,j}$ samples and $Y_{i,j}$

Algorithm 9 The counting algorithm $\mathcal{A}_{count}$ for DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$

for $i = 1$ to n **do**
  $S_i = \sum_{j=1}^{r}(X_{i,j} - Y_{i,j})$
end for
$Z = \sum_{i=1}^{n} S_i$
If $Z > 0$, output $B = 0$. If $Z < 0$, output $B = 1$. If $Z = 0$, output $B = 0$ with probability $1/2$ and output $B = 1$ with probability $1/2$.

Algorithm 10 The max algorithm $\mathcal{A}_{max}$ for DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$

for $i = 1$ to n **do**
  $S_i = \sum_{j=1}^{r}(X_{i,j} - Y_{i,j})$
end for
$i' = \arg\max |S_i|$
$Z = S_{i'}$
If $Z > 0$, output $B = 0$. If $Z < 0$, output $B = 1$. If $Z = 0$, output $B = 0$ with probability $1/2$ and output $B = 1$ with probability $1/2$.

samples; i.e., we compute the values $S_i = \sum_{j=1}^{r}(X_{i,j} - Y_{i,j})$. The counting algorithm $\mathcal{A}_{count}$ decides whether to output $B = 0$ or $B = 1$ based on the sign of $\sum_i S_i$, whereas the max algorithm decides its output based on the sign of the S_i with the largest absolute value. See Algorithms 9 and 10 for detailed pseudocode for both $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$.

We omit proofs in this subsection. They can be found in Appendix E.2.

We begin by proving upper bounds for the sample complexities of both $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$. In particular, both $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ need at most $\tilde{O}(n)$ times as many samples as the best possible algorithm for any instance in this regime.

Lemma 6.5.1. *Let $C = (n, \mathbf{p}, \mathbf{q})$ be an instance of* DOMINATION. *Then*

$$r_{min}(C, \mathcal{A}_{count}, 1 - \alpha) \leq \frac{2n \ln(\alpha^{-1})}{\|\mathbf{p} - \mathbf{q}\|_1^2}.$$

If C further satisfies $\delta \leq p_i, q_i \leq 1 - \delta$ for all i for some constant $\delta > 0$, then

$$r_{min}(C, \mathcal{A}_{comb}) \leq O(n) r_{min}(C).$$

Lemma 6.5.2. *Let $C = (n, \mathbf{p}, \mathbf{q})$ be an instance of* Domination. *Then*

$$r_{min}(C, \mathcal{A}_{max}, 1-\alpha) \leq \frac{8\ln(2n\alpha^{-1})}{\|\mathbf{p}-\mathbf{q}\|_\infty^2}$$

If C further satisfies $\delta \leq p_i, q_i \leq 1-\delta$ for all i for some constant δ, then

$$r_{min}(C, \mathcal{A}_{comb}) \leq O(n \log n) r_{min}(C).$$

We now show that the upper bounds we proved above are essentially tight. In particular, we demonstrate instances where both $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ need $\tilde{\Omega}(n)$ times as many samples as the best possible algorithms for those instances. Interestingly, on the instance where $\mathcal{A}_{count}$ suffers, $\mathcal{A}_{max}$ performs well, and vice versa. This fact will prove useful in the next section.

Lemma 6.5.3. *For each $\varepsilon < \frac{1}{10}$ and each sufficiently large n, there exists an instance $C = (n, \mathbf{p}, \mathbf{q})$ of* Domination *such that the following two statements are true:*

1. $r_{min}(C, \mathcal{A}_{max}, 1-\frac{2}{n}) \leq \frac{16\ln n}{\varepsilon^2}$.
2. $r_{min}(C, \mathcal{A}_{count}) \geq \frac{n}{128\varepsilon^2}$.

It is not hard to observe that in certain cases, the counting algorithm of [132] for Top-K reduces to the algorithm $\mathcal{A}_{count}$ for Domination. It follows that there also exists an $\Omega(n)$ multiplicative gap between the sample complexity of their counting algorithm and the sample complexity of the best algorithm on some instances.

Corollary 6.5.4. *Let A' be the* Top-K *algorithm of [132], and let $S = (n, k, \mathbf{P})$ be a* Top-K *instance, with $\mathbf{P}$ as described in Section 6.3. Then, for sufficiently large n and $\varepsilon < 1/10$, $r_{min}(S, A') \geq \Omega(\frac{n}{\varepsilon^2})$.*

We will now show that $\mathcal{A}_{max}$ has a competitive ratio of $\tilde{\Omega}(n)$.

Lemma 6.5.5. *For each sufficiently large n, there exists an instance $C = (n, \mathbf{p}, \mathbf{q})$ of* Domination *such that the following two statements are true:*

1. $r_{min}(C, \mathcal{A}_{count}, 1 - \frac{1}{n}) \leq 2n^3 \ln n$.

2. $r_{min}(C, \mathcal{A}_{max}, \frac{4}{5}) \geq \frac{n^4}{2^{14} \ln n}$.

6.5.2 $\tilde{O}(\sqrt{n})$-competitive algorithms

We will now demonstrate two algorithms for DOMINATION that use at most $\tilde{O}(\sqrt{n})$ times more samples than the best possible algorithm for each instance. According to Theorem 6.8.2, this is the best we can do up to polylogarithmic factors.

Note that the counting algorithm $\mathcal{A}_{count}$ tends to work well when the max algorithm $\mathcal{A}_{max}$ fails, and vice versa (e.g., Lemmas 6.5.3 and 6.5.5). Therefore, intuitively, combining both algorithms in some way should lead to better performance.

Both of the algorithms we present in this section share this intuition. We begin (in Lemma 6.5.6) by demonstrating a very general method for combining any two algorithms for DOMINATION. Applying this to $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$, we obtain an algorithm $\mathcal{A}_{comb}$ that satisfies $r_{min}(C, \mathcal{A}_{comb}) \leq O(\sqrt{n \log n}) \cdot r_{min}(C)$ (Corollary 6.5.7) for instances C in this regime. We then show an alternate algorithm with slightly better performance than $\mathcal{A}_{comb}$, which we call the *sum of cubes* algorithm $\mathcal{A}_{cube}$. This algorithm satisfies $r_{min}(C, \mathcal{A}_{cube}) \leq O(\sqrt{n}) \cdot r_{min}(C)$ for instances C in this regime (Theorem 6.5.10).

Combining counting and max

We first show how to combine any two algorithms for DOMINATION to get an algorithm that always does at least as well as the better of the two algorithms. Call an algorithm $\mathcal{A}$ for DOMINATION *stable* if it always outputs the correct answer with probability at least 1/2 (i.e. it always does at least as well as a random guess). Note that $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ are both stable. We have the following lemma.

Lemma 6.5.6. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two stable algorithms for* DOMINATION. *Then there exists an algorithm $\mathcal{A}_{comb}$ such that for all instances C of* DOMINATION,

$$r_{min}(C, \mathcal{A}_{comb}, 1-\alpha) \leq 32 \ln(\alpha^{-1}) \cdot \min\left(r_{min}(C, \mathcal{A}_1), r_{min}(C, \mathcal{A}_2)\right)$$

Proof. See Algorithm 11 for a description of $\mathcal{A}_{comb}$. Assume without loss of generality that $B = 0$, and let $r = 32 \log(n\alpha^{-1}) \min\left(r_{min}(C, \mathcal{A}_1), r_{min}(C, \mathcal{A}_2)\right)$. We will show that $\mathcal{A}_{comb}$ outputs $B = 0$ correctly with probability at least $1 - \alpha$.

Let $r' = \frac{r}{32 \ln n}$; note that either $r' \geq r_{min}(C, \mathcal{A}_1)$ or $r_{min}(C, \mathcal{A}_2)$. Assume first that $r' \geq r_{min}(C, \mathcal{A}_1)$. Then, $\mathcal{A}_1$ will output $B = 0$ in each of its $16 \ln \alpha^{-1}$ groups with probability at least $\frac{3}{4}$. On the other hand, since it is stable, $\mathcal{A}_2$ will output $B = 0$ in each of its groups with probability at least $\frac{1}{2}$. Therefore

$$\mathbb{E}\left[\frac{Z_1 + Z_2}{2}\right] \leq \frac{1}{8} + \frac{1}{4} \leq \frac{3}{8}.$$

Since $\frac{Z_1+Z_2}{2}$ is the average of $32 \ln \alpha^{-1}$ random variables, by Hoeffding's inequality, the probability that $\frac{Z_1+Z_2}{2} \geq \frac{1}{2}$ is at most $\exp\left(-2(32 \ln \alpha^{-1})(\frac{1}{8})^2\right) \leq \alpha$.

Similarly, if $r' \geq r_{min}(C, \mathcal{A}_2)$, the probability that $\frac{Z_1+Z_2}{2} \geq \frac{1}{2}$ is also at most α. This concludes the proof. □

Algorithm 11 Combining two algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$ for DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$

1: Divide the samples into $32 \ln \alpha^{-1}$ groups.
2: Run $\mathcal{A}_1$ on each of the first $16 \ln \alpha^{-1}$ groups and let Z_1 be the average of the outputs.
3: Run $\mathcal{A}_2$ on each of the last $16 \ln \alpha^{-1}$ groups and let Z_2 be the average of the outputs.
4: If $\frac{Z_1+Z_2}{2} \leq \frac{1}{2}$ output $B = 0$, else output $B = 1$.

Corollary 6.5.7. *Let $\mathcal{A}_{comb}$ be the algorithm we obtain by combining $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ in the manner of Lemma 6.5.6. Then for any instance $C = (n, \mathbf{p}, \mathbf{q})$ of* DOMINATION,

$$r_{min}(C, \mathcal{A}_{comb}) \leq O\left(\frac{\sqrt{n \log n}}{\|\mathbf{p} - \mathbf{q}\|_2^2}\right).$$

If C further satisfies $\delta \leq p_i, q_i \leq 1 - \delta$ for all i for some constant δ, then

$$r_{min}(C, \mathcal{A}_{comb}) \leq O(\sqrt{n \log n}) r_{min}(C).$$

Proof. This follows from Lemmas 6.5.1, 6.5.2, 6.5.6, and the following observation:

$$\begin{aligned} \min\left(\frac{n}{\|\mathbf{p}-\mathbf{q}\|_1^2}, \frac{\log n}{\|\mathbf{p}-\mathbf{q}\|_\infty^2}\right) &\leq \sqrt{\frac{n}{\|\mathbf{p}-\mathbf{q}\|_1^2} \cdot \frac{\log n}{\|\mathbf{p}-\mathbf{q}\|_\infty^2}} \\ &\leq \frac{\sqrt{n \log n}}{\|\mathbf{p}-\mathbf{q}\|_2^2}. \end{aligned}$$

The last inequality follows from the fact that for any vector $\mathbf{x}$, $\|\mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_1 \cdot \|\mathbf{x}\|_\infty$. The second part of the corollary then follows directly from Corollary 6.4.4. □

The sum of cubes algorithm

We now give a different algorithm for DOMINATION which we call the *sum of cubes algorithm*, $\mathcal{A}_{cube}$. If we let $S_i = \sum_j (X_i - Y_i)$, then intuitively, whereas $\mathcal{A}_{count}$ decides its output based on the signed ℓ_1 norm of the S_i and whereas $\mathcal{A}_{max}$ decides its output based on the signed ℓ_∞ norm of the S_i, $\mathcal{A}_{cube}$ decides its output based on the signed ℓ_3 norm of the S_i. See Algorithm 12 for a detailed description of the algorithm.

Algorithm 12 Sum of cubes algorithm $\mathcal{A}_{cube}$ for DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$

1: $T_{i,j} = 1$ with probability $\frac{1}{2} + \frac{(X_{i,j} - Y_{i,j})}{2}$ and $T_{i,j} = -1$ with probability $\frac{1}{2} - \frac{(X_{i,j} - Y_{i,j})}{2}$

2: $S_i = \sum_{j=1}^{r} T_{i,j}$
3: $Z = \sum_{i=1}^{n} S_i^3$
4: If $Z \geq 0$, output $B = 0$. If $Z < 0$, output $B = 1$.

To analyze the performance of $\mathcal{A}_{cube}$, we begin by analyzing statistical properties of the random variable S.

Lemma 6.5.8. *Let $S = \sum_{j=1}^{r} X_j$ where $X_1, \cdots, X_r$ are i.i.d $\{-1, 1\}$-valued random variables with mean $\varepsilon \geq 0$ and $r \geq 8$. Let $Z = S^3$. Then*

$$\mathbb{E}[Z] \geq 2r^2\varepsilon + \frac{1}{2}r^3\varepsilon^3$$
$$\mathrm{Var}[Z] \leq 15r^3 + 36r^4\varepsilon^2 + 9r^5\varepsilon^4.$$

Proof. By applying the multinomial theorem and using the fact that $X_i^2 = 1$ for each i, we can write multilinear expressions for S^3 and S^6. We can now use linearity of expectation and the independence among the X_i's to compute the mean and variance exactly.

$$\begin{aligned}
\mathbb{E}[Z] = \mathbb{E}[S^3] &= (-2r + 3r^2)\varepsilon + (2r - 3r^2 + r^3)\varepsilon^3 \\
&\geq 2r^2\varepsilon + \frac{1}{2}r^3\varepsilon^3 \\
\mathrm{Var}[Z] = E[S^6] - \mathbb{E}[S^3]^2 &= (16r - 30r^2 + 15r^3) \\
&+ (-136r + 282r^2 - 183r^3 + 36r^4)\varepsilon^2 \\
&+ (240r - 522r^2 + 381r^3 - 108r^4 + 9r^5)\varepsilon^4 \\
&+ (-120r + 270r^2 - 213r^3 + 72r^4 - 9r^5)\varepsilon^6 \\
&\leq 15r^3 + 36r^4\varepsilon^2 + 9r^5\varepsilon^4
\end{aligned}$$

□

Lemma 6.5.9. *Let $S_i = \sum_{j=1}^{r} X_{i,j}$ where for each $i \in [n]$, $X_{i,1}, \cdots, X_{i,r}$ are i.i.d $\{-1, 1\}$-valued random variables with mean ε_i, along with the condition that either all $\varepsilon_i \geq 0$ or all $\varepsilon_i \leq 0$. Let $Z = \sum_{i=1}^{n} S_i^3$. If $r \geq 8$ and $r \geq \eta\sqrt{n}/(\sum_{i=1}^{n} \varepsilon_i^2)$ for some $\eta \geq 1$ then, $\mathbb{E}[Z]^2 \geq \frac{\eta}{36}\mathrm{Var}[Z]$.*

Proof. Without loss of generality, we can assume that $\varepsilon_i \geq 0$ for all $i \in [n]$. By Lemma 6.5.8,

$$\mathbb{E}[Z]^2 \geq 4r^4(\sum_i \varepsilon_i)^2 + \frac{1}{4}r^6(\sum_i \varepsilon_i^3)^2 + 2r^5(\sum \varepsilon_i)(\sum_i \varepsilon_i^3) \tag{6.2}$$

$$\mathrm{Var}[Z] \leq 15nr^3 + 36r^4 \sum_i \varepsilon_i^2 + 9r^5 \sum_i \varepsilon_i^4. \tag{6.3}$$

We will show that each term in the Equation 6.3 is dominated by some term in Equation 6.2.

$$nr^3 = r^5\frac{n}{r^2} \leq \frac{1}{\eta^2}r^5(\sum_i \varepsilon_i^2)^2 \leq \frac{1}{\eta^2}r^5(\sum_i \varepsilon_i)(\sum_i \varepsilon_i^3)$$

(Cauchy-Schwarz inequality)

$$r^4(\sum_i \varepsilon_i^2) \leq \frac{1}{\eta\sqrt{n}}r^5(\sum_i \varepsilon_i^2)^2 \leq \frac{1}{\eta\sqrt{n}}r^5(\sum_i \varepsilon_i)(\sum_i \varepsilon_i^3)$$

$$\begin{aligned} r^5(\sum_i \varepsilon_i^4) &\leq r^6\frac{1}{\eta\sqrt{n}}(\sum_i \varepsilon_i^2)(\sum_i \varepsilon_i^4) \\ &\leq r^6\frac{1}{\eta\sqrt{n}}\left(\sqrt{n}\cdot(\sum_i \varepsilon_i^4)^{1/2}\right)(\sum_i \varepsilon_i^4) && \text{(Cauchy-Schwarz inequality)} \\ &= \frac{r^6}{\eta}(\sum_i \varepsilon_i^4)^{3/2} \leq \frac{r^6}{\eta}(\sum_i \varepsilon_i^3)^2 && \text{(monotonicity of } \ell_p \text{ norms)} \end{aligned}$$

Adding the above inequalities, we get $\mathrm{Var}[Z] \leq \frac{36}{\eta}\,\mathbb{E}[Z]^2$. □

Theorem 6.5.10. *If $C = (n, \mathbf{p}, \mathbf{q})$ is any instance of* Domination*, then*

$$r_{min}(C, \mathcal{A}_{cube}) \leq \max\left(\frac{144\sqrt{n}}{\|\mathbf{p}-\mathbf{q}\|_2^2}, 8\right).$$

If C satisfies $\delta \leq p_i, q_i \leq 1-\delta$ for all i for some constant δ, then

$$r_{min}(C, \mathcal{A}_{cube}) \leq O(\sqrt{n})r_{min}(C).$$

Proof. Assume without loss of generality that $B = 0$. We have $S_i = \sum_{j=1}^{r} T_{i,j}$ and $Z = \sum_{i=1}^{n} S_i^3$. Note that for each i, the $T_{i,j}$ are i.i.d. $\{-1, 1\}$ random variables with mean $\mathbb{E}[T_{i,j}] = p_i - q_i$. Applying Lemma 6.5.9 with $\eta = 144$, if $r \geq \max\left(\frac{144\sqrt{n}}{\|\mathbf{p}-\mathbf{q}\|_2^2}, 8\right)$ we have that $\mathbb{E}[Z]^2 \geq \frac{144}{36}\mathrm{Var}[Z] = 4\mathrm{Var}[Z]$. Since the algorithm makes an error (i.e. outputs $B = 1$) when $Z < 0$, we can use Chebyshev's inequality to bound the probability that $Z < 0$.

$$\Pr[Z < 0] \leq \Pr[|Z - \mathbb{E}[Z]| \geq \mathbb{E}[Z]] \leq \frac{\mathrm{Var}[Z]}{\mathbb{E}[Z]^2} \leq \frac{1}{4}.$$

The second part of the theorem then follows directly from Corollary 6.4.4. □

6.6 Domination in the general regime

In this section, we consider Domination in the general regime. Unlike in the previous section, it is no longer true that $I(X_{i,j}Y_{i,j}; B) = \mathbb{I}(p_i, q_i) = \Theta((p_i - q_i)^2)$. In particular, when p_i and q_i are both very small, $\mathbb{I}(p_i, q_i)$ can be much bigger than $(p_i - q_i)^2$; as a result, the algorithms designed in the previous section can fail under these circumstances.

In Section 6.6.1, we present a new algorithm $\mathcal{A}_{coup}$ which is $\tilde{O}(\sqrt{n} \cdot r_{min})$-competitive. According to Theorem 6.8.2, this is the best we can do up to polylogarithmic factors. In Section 6.6.2, we then demonstrate that the general regime is indeed harder than the restricted regime in Section 6.5. In particular, we give instances where the algorithms presented in the previous section fail; we show that the competitive ratio of these algorithms is unbounded (even for fixed n).

6.6.1 An $\tilde{O}(\sqrt{n})$-competitive algorithm

Here we give an algorithm that only needs $O(\sqrt{n}\log(n)/\mathbb{I}(\mathbf{p}, \mathbf{q}))$ samples to solve Domination (Theorem 6.6.5). By Lemma 6.4.2, this is only $\tilde{O}(\sqrt{n})$ times as many

samples as the optimal algorithm needs. Intuitively, the algorithm works as follows: if for some coordinate i, $X_{i,1}Y_{i,1}...X_{i,r}, Y_{i,r}$ conveys enough information about B, we will only use samples from coordinate i to determine B. Otherwise, the information about B must be well-spread throughout all the coordinates, and a majority vote will work.

We begin by bounding the probability we can determine the answer from a single fixed coordinate. The following four lemmas will be used to prove Theorem 6.6.5 and their proofs can be found in Appendix E.3.

Lemma 6.6.1 (Sanov's theorem)**.** *Let $\mathcal{P}(\Sigma)$ denote the space of all probability distributions on some finite set Σ. Let $R \in \mathcal{P}(\Sigma)$ and let $Z_1, \cdots, Z_k$ be i.i.d random variables with distribution R. For every $x \in \Sigma^k$, we can define an empirical probability distribution $\hat{P}_x$ on Σ as*

$$\forall \sigma \in \Sigma \quad \hat{P}_x(\sigma) = \frac{|\{i \in [k] : x_i = \sigma\}|}{k}.$$

Let C be a closed convex subset of $\mathcal{P}(\Sigma)$ such that for some $P \in C$, $D(P||R) < \infty$. Then

$$\Pr\left[\hat{P}_{(Z_1,\cdots,Z_k)} \in C\right] \leq \exp\left(-k(\ln 2)D(Q^*||R)\right)$$

where $Q^* = \operatorname{argmin}_{Q \in C} D(Q||R)$ is unique. In the case when $D(Q||R) = \infty$ for all $Q \in C$, $\Pr\left[\hat{P}_{(Z_1,\cdots,Z_k)} \in C\right] = 0$.

Proof. See exercise 2.7 and 3.20 in [48]. □

Sanov's theorem allows us to bound the following probability that we incorrectly rank two Bernoulli variables (e.g., X_i and Y_i for a fixed coordinate i) from k independent samples.

Lemma 6.6.2. *Let $0 \leq q < p \leq 1$ and let $X_1, \cdots, X_k$ be i.i.d $\mathcal{B}(p)$ and $Y_1, \cdots, Y_k$ be i.i.d $\mathcal{B}(q)$. Then $\Pr\left[\sum_{i=1}^k (X_i - Y_i) \leq 0\right] \leq \exp\left(-2(\ln 2)k \log\left(\frac{1}{\sqrt{pq}+\sqrt{(1-p)(1-q)}}\right)\right)$.*

We can in turn relate the upper bound in Lemma 6.6.2 to the quantity $\mathbb{I}(p, q)$.

Lemma 6.6.3.

$$2 \log \left(\frac{1}{\sqrt{pq} + \sqrt{(1-p)(1-q)}} \right) \geq \frac{1}{2} \mathbb{I}(p, q).$$

Combining Lemma 6.6.2 and Lemma 6.6.3, we can show the following corollary which says that given $\Omega(1/\mathbb{I}(p, q))$ samples, we can correctly rank two Bernoulli variables with constant probability.

Corollary 6.6.4. *In* DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$*, for any* $i \in [n]$*, if* $r > 6/\mathbb{I}(p_i, q_i)$*, then*

$$\Pr \left[\text{sign} \left(\sum_{j=1}^{r} (X_{i,j} - Y_{i,j}) \right) = (-1)^B \right] > 5/6.$$

Proof. Assume we are in the $B = 0$ case, the other case is similar. Fix an $i \in [n]$. By Lemma 6.6.2,

$$\begin{aligned}
&\Pr \left[\sum_{j=1}^{r} (X_{i,j} - Y_{i,j}) \leq 0 \right] \\
&\leq \exp \left(-r(\ln 2) \log \left(\frac{1}{\sqrt{p_i q_i} + \sqrt{(1-p_i)(1-q_i)}} \right) \right) \\
&\leq \exp \left(-r(\ln 2) I_i / 2 \right) && \text{(By Lemma 6.6.3)} \\
&= 2^{-rI_i/2} < 1/8.
\end{aligned}$$

□

We now introduce what we call the *general coupling algorithm* $\mathcal{A}_{coup}$ for DOMINATION. A detailed description of the algorithm can be found in Algorithm 13; more briefly the algorithm works as follows:

1. Split the r samples for each of the n coordinates into $\ell = 18 \log(2n\alpha^{-1})$ equally-sized segments where α is the error parameter. For each coordinate i and segment j, set $S_{i,j} = 1$ if more samples from X equal 1 than samples from Y,

and -1 otherwise. This can be thought of as running a miniature version of the counting algorithm on each segment; $S_{i,j} = 1$ is evidence that $B = 0$, and $S_{i,j} = -1$ is evidence that $B = -1$.

2. Let i' be the coordinate i which maximizes $\left|\sum_{j=1}^{\ell} S_{i,j}\right|$ (i.e. the coordinate that is "most consistently" either 1 or -1). If $\left|\sum_{j=1}^{\ell} S_{i',j}\right| \geq \ell/3$ (i.e. at least $2\ell/3$ of the segments for this coordinate agree on the value of B), output B according to the sign of $\sum_{j=1}^{\ell} S_{i',j}$.

3. Otherwise, for each segment, take the majority of the votes from each of the n coordinates; that is, for each $1 \leq j \leq \ell$, set $T_j = \text{sign}(\sum_{i=1}^{n} S_{i,j})$. Then take another majority over the segments, by setting $Z_2 = \text{sign}(\sum_{j=1}^{\ell} T_j)$. Finally, if $Z_2 > 0$ output $B = 0$; otherwise, output $B = 1$.

Theorem 6.6.5. *If $C = (n, \mathbf{p}, \mathbf{q})$ is any instance of* Domination*, then*

$$r_{min}(C, \mathcal{A}_{coup}, 1-\alpha) \leq \frac{2592\sqrt{n}\ln(2n\alpha^{-1})}{\mathbb{I}(\mathbf{p}, \mathbf{q})}$$

and thus

$$r_{min}(C, \mathcal{A}_{coup}) \leq O(\sqrt{n}\log n) \cdot r_{min}(C).$$

Proof. Let $I_i = \mathbb{I}(p_i, q_i)$, $r = 2592\sqrt{n}\log(2n\alpha^{-1})/\mathbb{I}(\mathbf{p}, \mathbf{q})$ and $\ell = 18\ln(2n\alpha^{-1})$. There are two cases to consider:

1. **Case 1**: There exists an i' such that $24\sqrt{n}I_{i'} \geq \sum_{k=1}^{n} I_k$.

   By symmetry, we can assume that $B = 0$. In this case, we have that $\frac{r}{\ell} \geq \frac{24\cdot 6\sqrt{n}}{\sum_{k=1}^{n} I_k} \geq \frac{6}{I_{i'}}$. By Corollary 6.6.4, for each $j = 1, \ldots, \ell$, $\Pr[S_{i',j} = 1] \geq 5/6$. Therefore we have

$$\mathbb{E}\left[\sum_{j=1}^{l} S_{i',j}\right] \geq \ell \cdot (5/6 - 1/6) = 2\ell/3.$$

Algorithm 13 General coupling algorithm $\mathcal{A}_{coup}$ for DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$

$\ell = 18\log(2n\alpha^{-1})$.
for $i = 1$ to n **do**
  for $j = 1$ to ℓ **do**
    $S_{i,j} = \text{sign}(\sum_{t=(j-1)*(r/\ell)+1}^{jr/\ell} X_{i,t} - Y_{i,t})$
    If $S_{i,j} = 0$, let $S_{i,j} = 1$ with probability $1/2$ and let $S_{i,j} = -1$ with probability $1/2$.
  end for
end for
$i' = \arg\max_i |\sum_{j=1}^{\ell} S_{i,j}|$
$Z_1 = \sum_{j=1}^{\ell} S_{i',j}$
if $|Z_1| \geq \ell/3$ **then**
  If $Z_1 > 0$ output $B = 0$, else output $B = 1$.
else
  for $j = 1$ to l **do**
    $T_j = \text{sign}(\sum_{i=1}^{n} S_{i,j})$.
    If $T_j = 0$, let $T_j = 1$ with probability $1/2$ and let $T_j = -1$ with probability $1/2$.
  end for
  $Z_2 = \text{sign}(\sum_{j=1}^{\ell} T_j)$.
  If $Z_2 = 0$, let $Z_2 = 1$ with probability $1/2$ and let $Z_2 = -1$ with probability $1/2$.
  If $Z_2 > 0$ output $B = 0$, else output $B = 1$.
end if

Since $S_{i',1}, ..., S_{i',l}$ are independent when B is given, by the Chernoff bound, we have that $\Pr\left[\sum_{j=1}^{l} S_{i',j} \geq \ell/3\right] \geq 1 - \exp(-\ell \cdot (1/3)^2 \cdot (1/2)) \geq 1 - \frac{\alpha}{2n}$. For $i \neq i'$, since $p_i \geq q_i$, we still have $\Pr[S_{i,j} = 1] \geq 1/2$. By a similar argument, we get $\Pr\left[\sum_{j=1}^{l} S_{i,j} \geq -\ell/3\right] \geq 1 - \exp(-\ell \cdot (1/3)^2 \cdot (1/2)) \geq 1 - \frac{\alpha}{2n}$. Let W be the event that $\sum_{j=1}^{\ell} S_{i',j} \geq \ell/3$ and for $i \neq i'$, $\sum_{j=1}^{l} S_{i,j} \geq -\ell/3$. By the union bound, we have that $\Pr[W] \geq 1 - n \cdot \frac{\alpha}{2n} = 1 - \frac{\alpha}{2}$. Moreover, when W happens, we know that $Z_1 \geq \ell/3$ and $\mathcal{A}_{coup}$ outputs $B = 0$. Therefore, in Case 1, the probability that $\mathcal{A}_{coup}$ outputs B correctly is at least $1 - \frac{\alpha}{2}$.

2. **Case 2**: For all $i \in \{1, \ldots, n\}$, $24\sqrt{n}I_i < \sum_{k=1}^{n} I_k$.

Similarly as in Case 1, since $\Pr[S_{i,j} = (-1)^B] \geq 1/2$, the probability that $|Z_1| \geq \ell/3$ and our algorithm outputs wrongly is at most $\frac{\alpha}{2}$. For the rest of Case 2, assume $|Z_1| < \ell/3$.

Now fix a coordinate i. Our plan is to first lower bound the amount of information samples from coordinate i have about B by using Corollary 6.6.4 and the subadditivity of information. Let $s = r/\ell$, and let $s' = s \cdot \lceil \frac{6}{sI_i} \rceil$. Imagine that we have s' new samples, $U_{i,1}, V_{i,1}, ..., U_{i,s'}, V_{i,s'}$, where each $(U_{i,j}, V_{i,j})$ $(j = 1, \ldots, s')$ is generated independently according to the same distribution as $(X_{i,1}, Y_{i,1})$. Since $s' \geq 6/I_i$, by Corollary 6.6.4, we have that

$$\Pr\left[\text{sign}\left(\sum_{j=1}^{s'}(U_{i,j} - V_{i,j})\right) = (-1)^B\right] > 5/6.$$

Write $(U_iV_i)^{[a,b]}$ as shorthand for the sequence $((U_{i,a}, V_{i,a}), \ldots (U_{i,b}, V_{i,b}))$, and define $(X_iY_i)^{[a,b]}$ analogously. By Fano's inequality, we have that

$$\begin{aligned} & I\left((U_iV_i)^{[1,s']}; B\right) \\ = \; & H(B) - H(B|(U_iV_i)^{[1,s']}) \\ \geq \; & H(\tfrac{1}{2}) - H(1 - \tfrac{5}{6}) = 1 - H(\tfrac{1}{6}) \geq 1/3. \end{aligned}$$

Since $I((U_iV_i)^{[1,s]}; (U_iV_i)^{[s+1,s']}|B) = 0$ (our new samples are independent given B), we have

$$\begin{aligned} & I((U_iV_i)^{[1,s']}; B) \\ = \; & I((U_iV_i)^{[1,s]}; B|(U_iV_i)^{[s+1,s']}) \\ & + I((U_iV_i)^{[s+1,s']}; B) \\ \leq \; & I((U_iV_i)^{[1,s]}; B) + I((U_iV_i)^{[s+1,s']}; B) \\ & \text{(by Fact E.1.2)} \end{aligned}$$

Repeating this procedure, we get

$$I((U_iV_i)^{[1,s']};B) \leq \sum_{u=1}^{\lceil \frac{6}{sI_i} \rceil} I((U_iV_i)^{[(u-1)s+1,us]};B).$$

Since we know that for any $u = 1, ..., \lceil \frac{6}{sI_i} \rceil$,

$$I((U_iV_i)^{[(u-1)s+1,us]};B) = I((X_iY_i)^{[1,s]};B),$$

we get

$$I((X_iY_i)^{[1,s]};B) \geq I((U_iV_i)^{[1,s']};B) \cdot \frac{1}{\lceil \frac{6}{sI_i} \rceil} \geq \frac{sI_i}{6 \cdot 6}.$$

The last inequality is true because $\frac{6}{sI_i} = \frac{\sum_{k=1}^{n} I_k}{24\sqrt{n}I_i} \geq 1$.

After we lower bound $I((X_iY_i)^{[1,s]};B)$, we are going to show that we can output B correctly with reasonable probability based on samples only from coordinate

i.

$$
\begin{aligned}
& \frac{sI_i}{6 \cdot 6} \\
\leq\ & I((X_iY_i)^{[1,s]}; B) \\
=\ & \sum_x \Pr[(X_iY_i)^{[1,s]} = x] \\
& \cdot D(B|(X_iY_i)^{[1,s]} = x \| B) \\
\leq\ & \sum_x \Pr[(X_iY_i)^{[1,s]} = x] \\
& \cdot \big(2(\Pr[B = 0|(X_iY_i)^{[1,s]} = x] - 1/2)^2 \\
& +2(\Pr[B = 1|(X_iY_i)^{[1,s]} = x] - 1/2)^2\big) \\
& \qquad \text{(by Fact E.1.4)} \\
=\ & \sum_x \Pr[(X_iY_i)^{[1,s]} = x] \\
& \cdot (\Pr[B = 0|(X_iY_i)^{[1,s]} = x] \\
& - \Pr[B = 1|(X_iY_i)^{[1,s]} = x])^2 \\
\leq\ & \sum_x \Pr[(X_iY_i)^{[1,s]} = x] \\
& \cdot |\Pr[B = 0|(X_iY_i)^{[1,s]} = x] \\
& - \Pr[B = 1|(X_iY_i)^{[1,s]} = x]|.
\end{aligned}
$$

When $\sum_{j=1}^{s}(X_{i,j} - Y_{i,j}) > 0$, it is easy to check that

$$\Pr[B = 0|(X_iY_i)^{[1,s]}] > \Pr[B = 1|(X_iY_i)^{[1,s]}].$$

Therefore,

$$\begin{aligned}
& \Pr[S_{i,1} = (-1)^B] \\
= & \sum_x \Pr[(X_i Y_i)^{[1,s]} = x] \\
& \cdot \max(\Pr[B = 0 | (X_i Y_i)^{[1,s]} = x], \\
& \Pr[B = 1 | (X_i Y_i)^{[1,s]} = x]) \\
= & \frac{1}{2} + \frac{1}{2} \cdot \sum_x \Pr[(X_i Y_i)^{[1,s]} = x] \cdot \\
& |\Pr[B = 0 | (X_i Y_i)^{[1,s]} = x] \\
& - \Pr[B = 1 | (X_i Y_i)^{[1,s]} = x]| \\
\geq & \frac{1}{2} + \frac{sI_i}{12 \cdot 6} \\
\geq & \frac{1}{2} + \frac{\sqrt{n} I_i}{\sum_{k=1}^n I_k}.
\end{aligned}$$

Similarly, we can show for all $i = 1, ..., n$, $j = 1, ..., l$,

$$\Pr[S_{i,j} = (-1)^B] \geq \frac{1}{2} + \frac{\sqrt{n} I_i}{\sum_{k=1}^n I_k}.$$

Now without loss of generality assume that $B = 0$. We have that $\mathbb{E}\left[\sum_{i=1}^n S_{i,j}\right] \geq \sum_{i=1}^n \left(\frac{1}{2} + \frac{\sqrt{n} I_i}{\sum_{k=1}^n I_k} - \frac{1}{2} + \frac{\sqrt{n} I_i}{\sum_{k=1}^n I_k}\right) = 2\sqrt{n}$. Therefore, by the Chernoff bound,

$$\Pr[T_j = 1] \geq 1 - e^{-(1/n) \cdot (2\sqrt{n})^2 \cdot (1/2)} > 3/4.$$

By the Chernoff bound again,

$$\Pr[Z_2 > 0] \geq 1 - e^{-\ell \cdot (1/2)^2 \cdot (1/2)} \geq 1 - \frac{\alpha}{2n}.$$

Since we initially fail with probability at most $\frac{\alpha}{2}$, by the union bound, in Case 2 we fail with probability at most $\frac{\alpha}{2} + \frac{\alpha}{2n} < \alpha$. This concludes the proof.

□

6.6.2 $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ with unbounded competitive ratios even for constant n

In this section, we show that the competitive ratios of $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ are unbounded even when n is a constant. In other words, we cannot upper bound the competitive ratios of $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ by only a function of n. The competitive ratio also needs to depend on some parameters of the instance. We prove this by showing instances where the competitive ratios of $\mathcal{A}_{count}$ and $\mathcal{A}_{max}$ also depend on ε which is some parameter of the instances in Lemma 6.6.6 and Lemma 6.6.7. The result in Lemma 6.6.6 can be easily generalized to show that the counting algorithm of [132] for Top-K also has unbounded competitive ratio even when n is a constant. Proofs can be found in Appendix E.3.

Lemma 6.6.6. *For each sufficiently large n and for any $\varepsilon > 0$, there exists an instance $C = (n, \mathbf{p}, \mathbf{q})$ of* Domination *such that the following two statements are true:*

1. $r_{min}(C, \mathcal{A}_{coup}, 1 - \frac{2}{n}) \leq \frac{5184\sqrt{n}\log n}{\varepsilon}$
2. $r_{min}(C, \mathcal{A}_{count}) \geq \frac{n}{16\varepsilon^2}$.

Lemma 6.6.7. *For each sufficiently large n and any $0 < \varepsilon < 1/n^3$, there exists an instance $C = (n, \mathbf{p}, \mathbf{q})$ of* Domination *such that the following two statements are true.*

1. $r_{min}(C, \mathcal{A}_{coup}, 1 - \frac{2}{n}) \leq \frac{518400\sqrt{n}\ln n}{\varepsilon}$.
2. $r_{min}(C, \mathcal{A}_{max}, \frac{9}{10}) \geq \frac{1}{\varepsilon^2 2^{14}\ln n}$

6.7 Reducing top-K to domination

In this section, we will finally reduce Top-K to Domination, thus proving Theorem 6.3.1. First, we will give an algorithm for Top-K problem that uses $\mathcal{A}_{coup}$ for Domination as a subroutine. We need Lemma 6.7.1 and Lemma 6.7.2 for the algorithm. Their proof can be found in Appendix E.4. We begin by reducing Top-K to the following graph theoretic problem.

Lemma 6.7.1. *Let $G = ([n], E)$ be a directed complete graph on vertices $\{1, 2, \cdots, n\}$ i.e. for every distinct $i, j \in [n]$, either $(i, j) \in E$ or $(j, i) \in E$ but not both. Suppose there is a subset $S \subset [n]$ of size k such that $(i, j) \in E$ for every $i \in S$ and $j \notin S$. Then there is a randomized algorithm which runs in expected running time $O(n)$ and finds the set S given oracle access to the edges of G. Moreover there is some absolute constant $C > 0$ such that for every $\lambda \geq 1$, the probability that the algorithm runs in more than $C\lambda n$ time is bounded by $\exp(-\lambda)$.*

The following lemma shows that when $p \geq q$, $\mathbb{I}(p, q)$ is an increasing function of p and a decreasing function of q.

Lemma 6.7.2. *Let $0 \leq q' \leq q \leq p \leq p' \leq 1$, then $\mathbb{I}(p', q') \geq \mathbb{I}(p, q)$.*

We are now ready to give an algorithm for Top-K.

Theorem 6.7.3. *There exists an algorithm A for Top-K such that for any $\alpha > 0$ and any instance $S = (n, k, \mathbf{P})$, A runs in time $O(n^2 r \log(1/\alpha))$ and satisfies*

$$r_{min}(S, A, 1 - \alpha) \leq \frac{7776\sqrt{n}\log(2n\alpha^{-1})}{\mathbb{I}(\mathbf{P}_k, \mathbf{P}_{k+1})}$$

where $\mathbf{P}_k, \mathbf{P}_{k+1}$ are the k and $k+1$ rows of $\mathbf{P}$.

Proof. Let $\mathbf{P}_i$ denote the i^{th} row of $\mathbf{P}$, and let $\Delta = I(\mathbf{P}_k, \mathbf{P}_{k+1})$. Recall that $\mathcal{A}$ is given as input the three-dimensional array of samples $Z_{i,j,l}$, where for each $i, j \in [n]$

and $1 \leq l \leq r$, $Z_{i,j,l}$ is the result of the lth noisy comparison between item i and item j (sampled from $\mathcal{B}(\mathbf{P}_{\pi^{-1}(i),\pi^{-1}(j)})$). We will define a complete directed graph $G = ([n], E)$ as follows. For every $1 \leq i < j \leq n$ and $1 \leq h \leq n$, run $\mathcal{A}_{coup}$ with input $X_{h,l} = Z_{i,h,l}$ and $Y_{h,l} = Z_{j,h,l}$; if $\mathcal{A}_{coup}$ returns $B = 0$, then direct the edge from i towards j, and otherwise, direct the edge from j towards i.

Let $T = \{\pi(1), \pi(2), \ldots, \pi(k)\}$ be the set of labels of the top k items. We claim that if $i \in T$ and $j \notin T$, then with probability at least $1 - \frac{\alpha}{n^2}$, the edge is directed from i towards j. To see this, note that in the corresponding input to $\mathcal{A}_{coup}$, X is drawn from $\mathbf{P}_{\pi^{-1}(i)}$ and Y is drawn from $\mathbf{P}_{\pi^{-1}(j)}$. If $i \in T$ and $j \notin T$, then $\pi^{-1}(i) \leq k < \pi^{-1}(j)$. In particular, $\mathbf{P}_{\pi^{-1}(i)}$ dominates $\mathbf{P}_{\pi^{-1}(j)}$, and moreover by Lemma 6.7.2, $\mathbb{I}(\mathbf{P}_{\pi^{-1}(i)}, \mathbf{P}_{\pi^{-1}(j)}) \geq \Delta$. It follows from Theorem 6.6.5 that A_{coup} outputs $B = 0$ on this input with probability at least $1 - \frac{\alpha}{2n^2}$, since in general,

$$\begin{aligned} r_{min}(C, \mathcal{A}_{coup}, 1 - \tfrac{\alpha}{2n^2}) &\leq \frac{2592\sqrt{n}\log(4n^3\alpha^{-1})}{\mathbb{I}(\mathbf{p}, \mathbf{q})} \\ &\leq \frac{7776\sqrt{n}\log(2n\alpha^{-1})}{\mathbb{I}(\mathbf{p}, \mathbf{q})}. \end{aligned}$$

By the union bound, the probability that all of these comparisons are correct is at least $1 - \frac{\alpha}{2}$. Therefore, by the tail bounds in Lemma 6.7.1, we can find the subset T in $O(n\log(1/\alpha))$ oracle calls to $\mathcal{A}_{coup}$ with probability at least $1 - \frac{\alpha}{2}$. The probability of failure is at most $\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$. Each call to $\mathcal{A}_{coup}$ takes $O(nr)$ time, so the overall time of the algorithm is $O(n^2 r \log(1/\alpha))$.

□

To prove that this algorithm is competitive, we will conclude by proving a lower bound on $r_{min}(S)$ (again, by reduction to the appropriate lower bound for Domination).

Lemma 6.7.4. *Let* $S = (n, k, \mathbf{P})$ *be an instance of* Top-K. *Then* $r_{min}(S) \geq \frac{0.1}{\mathbb{I}(\mathbf{P}_k, \mathbf{P}_{k+1})}$.

Proof. We will proceed by contradiction. Suppose there exists an algorithm A which satisfies $r_{min}(S, A) \leq \frac{0.01}{\mathbb{I}(P_k, P_{k+1})}$. We will show how to convert this into an algorithm A' which solves the instance $C = (n, \mathbf{P}_k, \mathbf{P}_{k+1})$ of Domination with probability at least $\frac{3}{4}$ when given at least $2r = 0.05/\mathbb{I}(\mathbf{P}_k, \mathbf{P}_{k+1})$ samples, thus contradicting Lemma 6.4.2.

The algorithm A' is described in Algorithm 14; essentially, A' embeds the inputs X and Y to the Domination instance as rows/columns k and $k+1$ respectively of the Top-K instance. It is easy to check that the $Z_{i,j,l}$ for $i, j \in [n]$,$l \in [r]$ generated in A' are distributed according to the same distribution as the corresponding elements in the instance S of Top-K. Therefore A will output the top k items correctly with probability at least 3/4. In addition, if $B = 0$ the item labeled k will be in the top k items and if $B = 1$ the item labeled k will not be in the top k items. Therefore, A' succeeds to solve this instance of Domination with probability at least 3/4, leading to our desired contradiction.

Algorithm 14 Algorithm A' for the lower bound reduction

1: Get input $X_{i,l}$, $Y_{i,l}$ for $i \in [n]$ and $l \in [2r]$ from Domination$(n, \mathbf{P}_k, \mathbf{P}_{k+1}, 2r)$.
2: Generate a random permutation π on n elements s.t. $\pi(\{k, k+1\}) = \{k, k+1\}$.
3: **for** $i \in [n]$, $j \in [n], l \in [r]$ **do**
4: If $i = k$, set $Z_{i,j,l} = X_{j,l}$.
5: If $i = k+1$, set $Z_{i,j,l} = Y_{j,l}$.
6: If $i \notin \{k, k+1\}, j = k$, set $Z_{i,j,l} = X_{i,l+r}$.
7: If $i \notin \{k, k+1\}, j = k+1$, set $Z_{i,j,l} = Y_{i,l+r}$.
8: If $i \notin \{k, k+1\}, j \notin \{k, k+1\}$, sample $Z_{i,j,l}$ from $\mathcal{B}(\mathbf{P}_{\pi^{-1}(i),\pi^{-1}(j)})$.
9: **end for**
10: Run A on samples $Z_{i,j,l}$, $i, j \in [n]$, $l \in [r]$.
11: If A said k is amongst the top k items, output $B = 0$. Otherwise output $B = 1$.

□

We are now ready to prove our main upper bound result.

Corollary 6.7.5. *There is an algorithm A for* Top-K *such that A runs in time $O(n^2 r)$ and on every instance S of* Top-K *on n items,*

$$r_{min}(S, A) \leq O(\sqrt{n} \log n) r_{min}(S).$$

Proof. Let $S =$Top-K$(n, k, \mathbf{P}, \cdot)$ be an instance of Top-K. By Lemma 6.7.4,

$$r_{\min}(S) \geq \frac{0.1}{\mathbb{I}(\mathbf{P}_k, \mathbf{P}_{k+1})}.$$

If A is the algorithm in Theorem 6.7.3 with $\alpha = \frac{1}{4}$ then A runs in time $O(n^2 r)$ and

$$r_{\min}(S, A) \leq O\left(\frac{\sqrt{n} \log n}{\mathbb{I}(\mathbf{P}_k, \mathbf{P}_{k+1})}\right).$$

Combining these two inequalities, we obtain our result.

□

6.8 Lower bounds for domination and top-K

In the previous section we demonstrated an algorithm that solves Top-K on any distribution using at most $\tilde{O}(\sqrt{n})$ times more samples than the optimal algorithm for that distribution (see Corollary 6.7.5). In this section, we show this is tight up to logarithmic factors; for any algorithm, there exists some distribution where that algorithm requires $\tilde{\Omega}(\sqrt{n})$ times more samples than the optimal algorithm for that distribution. Specifically, we show the following lower bound.

Theorem 6.8.1. *For any algorithm A, there exists an instance S of* Top-K *of size n such that $r_{min}(S, A) \geq \Omega\left(\frac{\sqrt{n}}{\log n}\right) r_{min}(S)$.*

As in the previous sections, instead of proving this lower bound directly, we will first prove a lower bound for the domination problem, which we will then embed in a Top-K instance.

Theorem 6.8.2. *For any algorithm A, there exists an instance C of* Domination *of size n such that $r_{min}(C, A) \geq \Omega\left(\frac{\sqrt{n}}{\log n}\right) r_{min}(C)$.*

6.8.1 A hard distribution for domination

To prove Theorem 6.8.2, we will show that there exists a distribution over instances of the domination problem such that, while each instance in the support of this distribution can be solved by some algorithm with a small number of samples, any algorithm requires a large number of samples given an instance randomly sampled from this distribution.

Let $\mathcal{C}$ be a distribution over instances C of the domination problem of size n. We extend r_{min} to distributions by defining $r_{min}(\mathcal{C}, A, p)$ as the minimum number of samples algorithm A needs to successfully solve Domination with probability at least p over instances randomly sampled from $\mathcal{C}$, and let $r_{min}(\mathcal{C}, A) = r_{min}(\mathcal{C}, A, 3/4)$. The following lemma relates the distributional sample complexity to the single instance sample complexity.

Lemma 6.8.3. *For any $p > 1/2$, algorithm A and any distribution $\mathcal{C}$ over instances of the domination problem, there exists a C in the support of $\mathcal{C}$ such that $r_{min}(C, A, p) \geq r_{min}(\mathcal{C}, A, p)$.*

Proof. Let $\varepsilon(C, A, r)$ be the probability that algorithm A errs given r samples from C. By the definition of $r_{\min}(\mathcal{C}, A, p)$, we have that

$$\sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}}[C] \cdot \varepsilon(C, A, r_{min}(\mathcal{C}, A, p)) = 1 - p$$

It follows that there exists some $C^* \in \mathsf{supp}\mathcal{C}$ such that

$$\varepsilon(C^*, A, r_{min}(\mathcal{C}, A, p)) \geq 1 - p$$

Since $\varepsilon(C^*, A, r)$ is decreasing in r, this implies that $r_{min}(C^*, A, p) \geq r_{min}(\mathcal{C}, A, p)$, as desired. □

We will find it useful to work with distributions that are only mostly supported on easy instances. The following lemma lets us do that.

Lemma 6.8.4. *Let $\mathcal{C}$ be a distribution over instances of the domination problem, and let E be an event with $\Pr[E] = 1-\delta$. Then for any algorithm A and any $1-\delta > p > \frac{1}{2}$, $r_{min}(\mathcal{C}|E, A, p+\delta) \geq r_{min}(\mathcal{C}, A, p)$ (here $\mathcal{C}|E$ denotes the distribution $\mathcal{C}$ conditioned on event E occurring).*

Proof. By the definition of $r_{\min}(\mathcal{C}, A, p)$, we have that

$$\sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}}[C] \cdot \varepsilon(C, A, r_{min}(\mathcal{C}, A, p)) = 1 - p$$

Rewrite this as

$\Pr[\overline{E}] \cdot \sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}|\overline{E}}[C] \cdot \varepsilon(C, A, r_{min}(\mathcal{C}, A, p)) + \Pr[E] \cdot \sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}|E}[C] \cdot \varepsilon(C, A, r_{min}(\mathcal{C}, A, p)) = 1 - p$

Since $\sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}|\overline{E}}[C] = 1$ and $\Pr[\overline{E}] = \delta$, it follows that

$$\sum_{C \in \mathsf{supp}\mathcal{C}} \Pr_{\mathcal{C}|E}[C] \cdot \varepsilon(C, A, r_{min}(\mathcal{C}, A, p)) \geq 1 - p - \delta$$

from which it follows that $r_{min}(\mathcal{C}|E, A, p+\delta) \geq r_{min}(\mathcal{C}, A, p)$.

□

We can now define the hard distribution for the domination problem. Define $\gamma = \frac{1}{100\sqrt{n}}$. Let S_P be a random subset of $[n]$ where each $i \in [n]$ is independently chosen to belong to S_P with probability γ. Likewise, define S_Q the same way (independently of

S_P). Finally, fix n constants R_i all in the range $[\frac{1}{4}, \frac{3}{4}]$ (for now, it is okay to consider only the case where $R_i = \frac{1}{2}$ for all i; to extend this lower bound to the top-k problem, we will need to choose different values of R_i). Then the hard distribution $\mathcal{C}_{hard}$ is the distribution over instances $C(S_P, S_Q) = (n, \mathbf{p}, \mathbf{q})$ of DOMINATION where

$$p_i = \begin{cases} R_i(1+\varepsilon) & \text{if } i \in S_P \\ R_i & \text{if } i \notin S_P \end{cases}$$

and

$$q_i = \begin{cases} R_i(1-\varepsilon) & \text{if } i \in S_Q \\ R_i & \text{if } i \notin S_Q \end{cases}$$

We claim that the majority of the instances in the support of $\mathcal{C}_{hard}$ have an algorithm that requires few samples. Intuitively, if S_P and S_Q are fixed, then the best algorithm for that specific instance can restrict attention only to the indices in S_P and S_Q. In particular, if S_P is large enough (some constant times its expected size), then simply throwing away all indices not in S_P and counting which row has more heads is an efficient algorithm for recovering the dominant set.

Theorem 6.8.5. *Fix any S_P and S_Q such that $|S_P| \geq \frac{1}{10}n\gamma$. Then $r_{min}(C(S_P, S_Q), p) = O\left(\frac{\log(1-p)^{-1}}{\varepsilon^2\sqrt{n}}\right)$ for all $p < 1$.*

Proof. It suffices to demonstrate an algorithm A such that $r_{min}(C(S_P, S_Q), A, p) = O\left(\frac{\log(1-p)^{-1}}{\varepsilon^2\sqrt{n}}\right)$.

Any algorithm A receives two sets X, Y, each of r samples from n coins. Write $X = (X_1, X_2, \ldots, X_n)$, where each $X_i = (X_{i,1}, X_{i,2}, \ldots X_{i,r})$ is the collection of r samples from coin i (likewise, write $Y = (Y_1, Y_2, \ldots, Y_n)$, and $Y_i = (Y_{i,1}, Y_{i,2}, \ldots Y_{i,r})$). Consider the following algorithm: A computes the value

$$T = \sum_{i \in S_P} \sum_{j=1}^{r} (X_{i,j} - Y_{i,j})$$

and outputs that $B = 0$ if $T \geq 0$ and outputs $B = 1$ otherwise.

For each i, j, let $A_{i,j} = X_{i,j} - Y_{i,j}$. If $B = 0$, then $A_{i,j} \in [-1, 1]$, $\mathbb{E}[A_{i,j}] \geq \varepsilon R_i \geq \frac{\varepsilon}{4}$ and all the $A_{i,j}$ are independent. It follows from Hoeffding's inequality that in this case,

$$\begin{aligned}
\Pr[T < 0] &= \Pr[T - \mathbb{E}[T] < -\mathbb{E}[T]] \\
&\leq \exp\left(-\frac{2E[T]^2}{4|S_P|r}\right) \\
&= \exp\left(-\frac{|S_P|r\varepsilon^2}{32}\right) \\
&\leq \exp\left(-\frac{\gamma n \varepsilon^2 r}{320}\right) \\
&= \exp\left(-\frac{\sqrt{n}\varepsilon^2 r}{32000}\right)
\end{aligned}$$

Therefore, choosing $r = \frac{32000 \ln(1-p)^{-1}}{\sqrt{n}\varepsilon^2} = O\left(\frac{\log(1-p)^{-1}}{\sqrt{n}\varepsilon^2}\right)$ guarantees $\Pr[T < 0] \leq 1 - p$. Similarly, the probability that $T \geq 0$ if $B = 1$ is also at most $1 - p$ for this r. The conclusion follows. □

By a simple Chernoff bound, we also know that the event that S_P has size at least $\frac{1}{10}n\gamma$ occurs with high probability.

Lemma 6.8.6. $\Pr\left[|S_P| \geq \frac{1}{10}n\gamma\right] \geq 1 - e^{-\sqrt{n}/400}$.

In the following subsection, we will prove that for all A, $r_{min}(\mathcal{C}_{hard}, A)$ is large. More precisely, we will prove the following theorem.

Theorem 6.8.7. *For all algorithms A, $r_{min}(\mathcal{C}_{hard}, A, \frac{2}{3}) = \Omega\left(\frac{1}{\varepsilon^2 \log n}\right)$.*

Given that this theorem is true, we can complete the proof of Theorem 6.8.2.

Proof of Theorem 6.8.2. By Theorem 6.8.7, for any algorithm A, $r_{min}(\mathcal{C}_{hard}, A, \frac{2}{3}) = \Omega\left(\frac{1}{\varepsilon^2 \log n}\right)$. Let E be the event that $|S_P| \geq \frac{1}{10}n\gamma$. By Lemma 6.8.6, if $n \geq (400 \ln \frac{12}{11})^2$, $\Pr[E] \geq \frac{1}{12}$. It then follows from Lemma 6.8.4 that

$$\begin{aligned} r_{min}(\mathcal{C}_{hard}|E, A) &= r_{min}(\mathcal{C}_{hard}|E, A, 3/4) \\ &\geq r_{min}(\mathcal{C}_{hard}, A, 2/3) \\ &\geq \Omega\left(\frac{1}{\varepsilon^2 \log n}\right). \end{aligned}$$

It then follows by Lemma 6.8.3 that there is a specific instance $C = C(S_P, S_Q)$ with $|S_P|$ at least $\frac{1}{10}\gamma n$ such that $r_{min}(C, A) \geq \Omega\left(\frac{1}{\varepsilon^2 \log n}\right)$. On the other hand, by Theorem 6.8.5, for this C, $r_{min}(C) \leq O\left(\frac{1}{\varepsilon^2 \sqrt{n}}\right)$. It follows that for any algorithm A, there exists an instance C such that $r_{min}(C, A) \geq \Omega\left(\frac{\sqrt{n}}{\log n}\right) r_{min}(C)$, as desired.

□

6.8.2 Proof of lower bounds

In this subsection, we prove Theorem 6.8.7; namely, we will show that any algorithm needs at least $\Omega\left(\frac{1}{\varepsilon^2 \log n}\right)$ samples to succeed on $\mathcal{C}_{hard}$ with constant probability. Our main approach will be to bound the mutual information between the samples provided to the algorithm and the correct output (recall that B is the hidden bit that determines whether the samples in X are drawn from $\mathbf{p}$ or from $\mathbf{q}$).

Lemma 6.8.8. *If $I(XY; B) < 0.05$, then there is no algorithm that can succeed at identifying B with probability at least $\frac{2}{3}$.*

Proof. Fix an algorithm A, and let p_e be the probability that it errs at computing B. By Fano's inequality, we have that

$$\begin{aligned} H(p_e) &\geq H(B|XY) \\ &= H(B) - I(XY;B) \\ &= 1 - I(XY;B) \\ &> 0.95 \end{aligned}$$

Since $H(\frac{1}{3}) \leq 0.95$, it follows that A must err with probability at least $1/3$. $\square$

Via the chain rule, we can decompose $I(XY;B)$ into the sum of many smaller mutual informations.

Lemma 6.8.9. $I(XY;B) \leq \sum_{i=1}^{n} (I(X_i;B) + I(Y_i;B))$

Proof. Write $X^{<i}$ to represent the concatenation $X_1 X_2 \dots X_{i-1}$. By the chain rule, we have that

$$I(XY;B) = \sum_{i=1}^{n} I(X_i Y_i; B | X^{<i} Y^{<i})$$

We claim that $I(X_i Y_i; X^{<i} Y^{<i} | B) = 0$. To see this, note that given B, each coin in X_i is sampled from some $\mathcal{B}(p)$ distribution, where p only depends on whether $i \in S_P$ or $i \in S_Q$. Since each i is chosen to belong to S_P and S_Q independently with probability γ, this implies X_i (and similarly Y_i) are independent from $X^{<i}$ and $Y^{<i}$ given B. By Fact E.1.2, this implies that $I(X_i Y_i; B | X^{<i} Y^{<i}) \leq I(X_i Y_i; B)$, and therefore that

$$I(XY;B) \leq \sum_{i=1}^{n} I(X_i Y_i; B).$$

Likewise, we can write $I(X_iY_i;B) = I(X_i;B) + I(Y_i;B|X_i)$. Since $I(X_i;Y_i|B) = 0$ (since S_P and S_Q are chosen independently), again by Fact E.1.2 it follows that $I(Y_i;B|X_i) \leq I(Y_i;B)$ and therefore that

$$I(XY;B) \leq \sum_{i=1}^{n} \left(I(X_i;B) + I(Y_i;B)\right).$$

□

Lemma 6.8.10. *If $n \geq 400$ and $r = \frac{1}{100\varepsilon^2 \ln n}$, then for all i, $I(B;X_i) = I(B;Y_i) \leq \frac{1}{100n}$.*

Proof. By symmetry, $I(B;X_i) = I(B;Y_i)$. We will show that $I(B;X_i) \leq \frac{1}{100n}$.

Let $Z_i = \sum_j X_{i,j}$. Note that Z_i is a sufficient statistic for B, and therefore $I(B;X_i) = I(B;Z_i)$. By Fact E.1.3,

$$\begin{aligned} I(B;Z_i) &= \mathbb{E}_{Z_i}[D(B|Z_i\|B)] \\ &= \sum_{z=0}^{r} \Pr[Z_i = z] \cdot D(\Pr[B=0|Z_i=z]\|\tfrac{1}{2}). \end{aligned}$$

We next divide the range of z into two cases.

1. **Case 1**: $|z - rR_i| \leq 11r\varepsilon \ln n$.

   In this case, we will bound the size of $D(\Pr[B=0|Z_i=z]\|\frac{1}{2})$. Note that

$$\begin{aligned} & \left|\Pr[B=0|Z_i=z] - \frac{1}{2}\right| && (6.4) \\ = & \left|\frac{\Pr[Z_i=z|B=0]\cdot\Pr[B=0]}{\Pr[Z_i=z]} - \frac{1}{2}\right| \\ = & \left|\frac{\Pr[Z_i=z|B=0]}{\Pr[Z_i=z|B=0]+\Pr[Z_i=z|B=1]} - \frac{1}{2}\right| \\ = & \frac{|\Pr[Z_i=z|B=0] - \Pr[Z_i=z|B=1]|}{2(\Pr[Z_i=z|B=0]+\Pr[Z_i=z|B=1])} && (6.5) \end{aligned}$$

Now, note that

$$\begin{aligned}\Pr[Z_i = z|B = 0] &= (1-\gamma)\binom{r}{z}R_i^z(1-R_i)^{r-z} \\ &\quad + \gamma\binom{r}{z}(R_i(1+\varepsilon))^z(1-R_i(1+\varepsilon))^{r-z} \\ \Pr[Z_i = z|B = 1] &= (1-\gamma)\binom{r}{z}R_i^z(1-R_i)^{r-z} \\ &\quad + \gamma\binom{r}{z}(R_i(1-\varepsilon))^z(1-R_i(1-\varepsilon))^{r-z}\end{aligned}$$

We can therefore lower bound the denominator of (6.5) via

$$\begin{aligned}& 2(\Pr[Z_i = z|B = 0] + \Pr[Z_i = z|B = 1]) \\ \geq\ & 4(1-\gamma)\binom{r}{z}R_i^z(1-R_i)^{r-z} \\ \geq\ & 2\binom{r}{z}R_i^z(1-R_i)^{r-z}\end{aligned}$$

Likewise, we can write the numerator of (6.5) as

$$|\Pr[Z_i = z|B = 0] - \Pr[Z_i = z|B = 1]| = \gamma\binom{r}{z}R_i^z(1-R_i)^{r-z}M$$

where

$$
\begin{aligned}
M &= \Big|(1+\varepsilon)^z \left(\frac{1-R_i(1+\varepsilon)}{1-R_i}\right)^{r-z} \\
&\quad -(1-\varepsilon)^z \left(\frac{1-R_i(1-\varepsilon)}{1-R_i}\right)^{r-z} \Big| \\
&= \Big|(1+\varepsilon)^z \left(1-\frac{R_i}{1-R_i}\varepsilon\right)^{r-z} \\
&\quad -(1-\varepsilon)^z \left(1+\frac{R_i}{1-R_i}\varepsilon\right)^{r-z} \Big|.
\end{aligned}
$$

To bound M, note that (applying the inequality $1+x \le e^x$)

$$
\begin{aligned}
&(1+\varepsilon)^z \left(1-\frac{R_i}{1-R_i}\varepsilon\right)^{r-z} \\
\le\ & \exp\left(\varepsilon z - \varepsilon\frac{R_i}{1-R_i}(r-z)\right) \\
=\ & \exp\left(\varepsilon\frac{z-rR_i}{1-R_i}\right) \\
\le\ & \exp(4\varepsilon(z-rR_i)) \\
\le\ & \exp(44r\varepsilon^2 \ln n) \\
=\ & e^{0.44} \\
<\ & 2
\end{aligned}
$$

Similarly, $(1-\varepsilon)^z \left(1+\frac{R_i}{1-R_i}\varepsilon\right)^{r-z} \le 2$. It follows that $M \le 2$, and therefore that

$$
\begin{aligned}
& \left|\Pr[B=0|Z_i=z]-\frac{1}{2}\right| \\
= & \frac{|\Pr[Z_i=z|B=0]-\Pr[Z_i=z|B=1]|}{2(\Pr[Z_i=z|B=0]+\Pr[Z_i=z|B=1])} \\
\leq & \frac{\gamma\binom{r}{z}R_i^z(1-R_i)^{r-z}M}{2\binom{r}{z}R_i^z(1-R_i)^{r-z}} \\
= & \frac{\gamma M}{2} \\
\leq & \gamma
\end{aligned}
$$

By Fact E.1.4, this implies that

$$D(\Pr[B=0|Z_i=z]\|\tfrac{1}{2}) \leq \frac{4\gamma^2}{\ln 2}.$$

2. **Case 2**: $|z - rR_i| > 11r\varepsilon \ln n$.

   Let Z^+ be the sum of r i.i.d. $\mathcal{B}(R_i(1+\varepsilon))$ random variables. Note that since Z is the sum of r $\mathcal{B}(p)$ random variables for some $p \leq R_i(1+\varepsilon)$, $\Pr[Z^+ \geq x] \geq \Pr[Z \geq x]$ for all x. Therefore, by Hoeffding's inequality, we have that

$$
\begin{aligned}
& \Pr\left[Z - rR_i \geq 11r\varepsilon \ln n\right] \\
\leq & \Pr\left[Z^+ - rR_i \geq 11r\varepsilon \ln n\right] \\
\leq & \Pr\left[Z^+ - rR_i(1+\varepsilon) \geq r\varepsilon(11\ln n - R_i)\right] \\
\leq & \Pr\left[Z^+ - \mathbb{E}[Z^+] \geq 10r\varepsilon \ln n\right] \\
\leq & \exp\left(-\frac{2(10r\varepsilon \ln n)^2}{r}\right) \\
= & \exp(-2\ln n) \\
= & n^{-2}
\end{aligned}
$$

Likewise, we can show that

$$\Pr\left[Z - rR_i \leq -11r\varepsilon \ln n\right] \leq n^{-2}$$

so

$$\Pr\left[|Z - rR_i| \geq 11r\varepsilon \ln n\right] \leq 2n^{-2}$$

Combining these two cases, we have that (for $n \geq 400$)

$$\begin{aligned} I(B; Z_i) &= \sum_{z=0}^{r} \Pr[Z_i = z] \cdot D(\Pr[B = 0|Z_i = z] \| \tfrac{1}{2}) \\ &\leq \sum_{|\|z\| - r/2| > 11r\varepsilon \ln n} \Pr[Z_i = z] \cdot 1 \\ &\quad + \sum_{|\|z\| - r/2| \leq 11r\varepsilon \ln n} \Pr[Z_i = z] \cdot O(\gamma^2) \\ &\leq 2n^{-2} + \frac{4\gamma^2}{\ln 2} \\ &\leq \frac{1}{100n}. \end{aligned}$$

□

We can now complete the proof of Theorem 6.8.7.

Proof of Theorem 6.8.7. Combining Lemmas 6.8.9 and 6.8.10, we have that if $r = \frac{1}{100\varepsilon^2 \ln n}$, then (for $n \geq 400$) $I(XY; B) \leq 2nI(X_i; B) \leq 0.02$. Therefore by Lemma 6.8.8, there exists no algorithm A that, given this number of samples, correctly identifies B (and thus solves the domination problem) with probability at least 2/3. It follows that

$$r_{min}(\mathcal{C}_{hard}, A, \tfrac{2}{3}) \geq \frac{1}{100\varepsilon^2 \ln n} = \Omega\left(\frac{1}{\varepsilon^2 \log n}\right)$$

as desired.

□

6.8.3 Proving lower bounds for Top-K

We will now show how to use our hard distribution of instances of Domination to generate a hard distribution of instances of Top-K. Our goal will be to embed our Domination instance as rows k and $k+1$ of our SST matrix; hence, intuitively, deciding which of the two rows (k or $k+1$) belongs to the top k is as hard as solving the domination problem.

Unfortunately, the SST condition imposes additional structure that prevents us from directly embedding any instance of the domination problem. However, for appropriate choices of the constants R_i, all instances in the support of $\mathcal{C}_{hard}$ give rise to valid SST matrices.

Specifically, we construct the following distribution $\mathcal{S}_{hard}$ over Top-K instances S of size $n+2$. Consider the distribution $\mathcal{C}_{hard}$ over Domination instances of size n, where for $1 \leq i \leq n$, $R_i = \frac{1}{4} + \frac{i}{8n}$, and $\varepsilon = \frac{1}{100n^2}$. Now, consider the following map f from Domination instances $C = (\mathbf{p}, \mathbf{q})$ to Top-K instances $S = f(C) = (n+2, k, \mathbf{P})$: we choose $k = n+1$ (so that the problem becomes equivalent to identifying row $n+2$) and define the matrix $\mathbf{P}$ as follows:

$$\mathbf{P}_{ij} = \begin{cases} \mathbf{p}_j & \text{if } i = n+1 \text{ and } j \leq n \\ \mathbf{q}_j & \text{if } i = n+2 \text{ and } j \leq n \\ 1 - \mathbf{p}_i & \text{if } j = n+1 \text{ and } i \leq n \\ 1 - \mathbf{q}_i & \text{if } j = n+2 \text{ and } i \leq n \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

In general, for arbitrary $\mathbf{p}$ and $\mathbf{q}$, this matrix may not be an SST matrix. Note however that for this choice of R_i and ε, it is always the case that $R_i(1+\varepsilon) \leq R_{i+1}(1-\varepsilon)$, so for all i (regardless of sample C), $p_i < p_{i+1}$. In addition, all the R_i belong to $[1/4, 3/8]$, so for all i, p_i and q_i are less than $1/2$. From these two observations, it easily follows that if C belongs to the support of $\mathcal{C}_{hard}$, $\mathbf{P}$ is an SST matrix, and $f(C)$ is a valid instance of the top-k problem. We will write $\mathcal{S}_{hard} = f(\mathcal{C}_{hard})$ to denote the distribution of instances of top-k $f(C)$ where C is sampled from $\mathcal{C}_{hard}$. Likewise, for any event E (e.g. the event that $|S_P| \geq \frac{1}{10}n\gamma$) , we write $\mathcal{S}_{hard}|E$ to denote the distribution $f(\mathcal{C}_{hard}|E)$.

We will begin by showing that, if there exists a sample efficient algorithm for some DOMINATION instance C in the support of $\mathcal{C}_{hard}$, there exists a similarly efficient algorithm for the corresponding TOP-K instance $S = f(C)$.

Lemma 6.8.11. *If $C \in \mathsf{supp}\mathcal{C}_{hard}$ and $S = f(C)$, then*

$$r_{min}(S) \leq \max(r_{min}(C, \frac{4}{5}), 1000n^2(1 + \ln n)).$$

Proof. Let A be an algorithm that successfully solves the DOMINATION instance C with probability at least $\frac{4}{5}$ using $r_{min}(C, \frac{4}{5})$ samples. We will show how to use A to construct an algorithm A' that solves the TOP-K instance S with probability at least $3/4$ using $r = \max(r_{min}(C, \frac{4}{5}), 1000n^2(1 + \ln n))$ samples.

For each i, j, write $Z_{i,j} = \sum_{\ell=1}^{r} Z_{i,j,\ell}$. Our algorithm A' operates as follows.

1. We begin by finding the two rows with the smallest row sums $\sum_j Z_{i,j}$. Let these two rows have indices c and d. We claim that, with high probability, $\pi^{-1}(\{c, d\}) = \{n+1, n+2\}$.

   To see this, note that for all $i \notin \pi(\{n+1, n+2\})$, $\mathbf{P}_{i,j} \geq \frac{1}{2}$, so $\mathbb{E}\left[\sum_j Z_{i,j}\right] \geq \left(\frac{n}{2} + 1\right) r$. Thus, for any fixed $i \notin \pi(\{n+1, n+2\})$, it follows from Hoeffding's inequality that

$$\Pr\left[\sum_j Z_{i,j} \leq \left(\frac{7}{16}n+1\right)r\right] \leq \exp\left(-\frac{nr}{128}\right)$$

so by the union bound, the probability that there exists an $i \notin \pi^{-1}(\{n+1, n+2\})$ such that $\sum_j Z_{i,j} \leq \left(\frac{7}{16}n+1\right)r$ is at most $n\exp\left(-\frac{nr}{128}\right)$.

On the other hand, if $i \in \pi(\{n+1, n+2\})$ then $\mathbf{P}_{i,j} \leq \frac{3}{8}(1+\varepsilon)$ unless $j \in \pi(\{n+1, n+2\})$, where $\mathbf{P}_{i,j} = \frac{1}{2}$; it follows that in this case, $\mathbb{E}\left[\sum_j Z_{i,j}\right] \leq \left(\frac{3n}{8}(1+\varepsilon)+1\right)r$. Similarly, applying Hoeffding's inequality in this case, we find that for any fixed $i \in \pi^{-1}(\{n+1, n+2\})$, $\Pr\left[\sum_j Z_{i,j} \geq \left(\frac{7}{16}n+1\right)r\right] \leq \exp\left(-\frac{nr}{128(1+\varepsilon)^2}\right) \leq 1.5\exp\left(-\frac{nr}{128}\right)$ and thus the probability that there exists some $i \in \pi^{-1}(\{n+1, n+2\})$, such that $\sum_j Z_{i,j} \geq \left(\frac{7}{16}n+1\right)r$ is at most $3\exp\left(-\frac{nr}{128}\right)$. It follows that, altogether, the probability that $\pi^{-1}(\{c, d\}) \neq \{n+1, n+2\}$ is at most $(n+3)\exp\left(-\frac{nr}{128}\right)$. Since $r \geq 1000n^2 \ln n$, this is at most $4\exp(-1000/128) < 0.01$.

2. We next sort the values $Z_{c,j}$ for $j \in [n+2]\setminus\{c, d\}$ and obtain indices $j_1, j_2, \ldots, j_n$ so that $Z_{c,j_1} \leq Z_{c,j_2} \leq \cdots \leq Z_{c,j_n}$. We claim that, with high probability, for all a, $\pi^{-1}(j_a) = a$.

   For each i, let U_i be the interval $\left[R_i(1-\varepsilon) - \frac{1}{20n}, R_i(1+\varepsilon) + \frac{1}{20n}\right]$. Note that, by our choice of R_i and ε, all the intervals U_i are disjoint, with U_i less than U_{i+1} for all i. We will show that with high probability, $\frac{1}{r}Z_{c,\pi(i)} \in U_i$ for all i, thus implying the previous claim.

   Note that $Z_{c,\pi(i)}$ is the sum of r $\mathcal{B}(p)$ random variables, where p is either $(1+\varepsilon)R_i$, R_i, or $(1-\varepsilon)R_i$. By Hoeffding's inequality, it follows that

$$\begin{aligned} & \Pr\left[Z_{c,\pi(i)} \geq r\left(R_i(1+\varepsilon)+\frac{1}{20n}\right)\right] \\ \leq\ & \exp\left(-2\frac{(r/20n)^2}{r}\right) \\ =\ & \exp\left(-\frac{r}{200n^2}\right) \end{aligned}$$

Likewise,

$$\Pr\left[Z_{c,\pi(i)} \leq r\left(R_i(1-\varepsilon)-\frac{1}{20n}\right)\right] \leq \exp\left(-\frac{r}{200n^2}\right)$$

Thus, for any fixed i,

$$\Pr\left[\frac{Z_{c,\pi(i)}}{r} \notin U_i\right] \leq 2\exp\left(-\frac{r}{200n^2}\right)$$

and by the union bound, the probability this fails for some i is at most $2n\exp\left(-\frac{r}{200n^2}\right)$. Since $r \geq 1000n^2(1+\ln n)$, $\exp\left(-\frac{r}{200n^2}\right) \leq (ne)^{-5}$, so this probability is at most $2e^{-5} < 0.02$.

3. Finally, we give algorithm A as input $X_{i,\ell} = Z_{c,j_i,\ell}$ and $Y_{i,\ell} = Z_{d,j_i,\ell}$. Note that (conditioned on the above two claims holding), this input is distributed equivalently to input from the DOMINATION instance C. In particular, if $\pi^{-1}(c) = n+1$ and $\pi^{-1}(d) = n+2$, then each $X_{i,\ell}$ is distributed according to $\mathcal{B}(\mathbf{p}_i)$ and each $Y_{i,\ell}$ is distributed according to $\mathcal{B}(\mathbf{q}_i)$, and if $\pi^{-1}(c) = n+2$ and $\pi^{-1}(d) = n+1$, then each $X_{i,\ell}$ is distributed according to $\mathcal{B}(\mathbf{q}_i)$ and each $Y_{i,\ell}$ is distributed according to $\mathcal{B}(\mathbf{p}_i)$. Thus, if A returns $B = 0$, we return $[n+2] \setminus \{d\}$ as the top $n+1$ indices, and if A returns $B = 1$, we return $[n+2] \setminus \{c\}$ as the top $n+1$ indices.

The probability that A fails given that steps 1 and 2 succeed is at most 0.2, and the probability that either of the two steps fail to succeed is at most $0.01+0.02 = 0.03$. Since $0.2 + 0.03 < \frac{1}{4}$, A' succeeds with probability at least $\frac{3}{4}$, as desired.

□

Corollary 6.8.12. *Let E be the event that $|S_P| \geq \frac{1}{10}n\gamma$. If $C \in \mathsf{supp}(\mathcal{C}_{hard}|E)$ and $S = f(C)$, then $r_{min}(S) \leq O(n^{3.5})$.*

Proof. Recall that by Theorem 6.8.5, for any $C \in \mathsf{supp}(\mathcal{C}_{hard}|E)$, $r_{min}(C, \frac{4}{5}) \leq O\left(\frac{1}{\sqrt{n\varepsilon^2}}\right) = O(n^{3.5})$. By Lemma 6.8.11, $r_{min}(S) \leq \max(r_{min}(C, \frac{4}{5}), 1000n^2(1 + \ln n)) \leq O(n^{3.5})$. □

We next show that solving Top-K over the distribution $\mathcal{S}_{hard}|E$ is at least as hard as solving Domination over the distribution $\mathcal{C}_{hard}|E$.

Lemma 6.8.13. *For any algorithm A that solves* Top-K*, there exists an algorithm A' that solves domination such that $r_{min}(\mathcal{S}_{hard}, A, p) \geq \frac{1}{2}r_{min}(\mathcal{C}_{hard}, A', p)$.*

Proof. We will show more generally that for any distribution $\mathcal{C}$ of Domination instances, if $\mathcal{S} = f(\mathcal{C})$ is a valid distribution of Top-K instances, then $r_{min}(\mathcal{S}, A, p) \geq \frac{1}{2}r_{min}(\mathcal{C}, A', p)$.

We will construct A' by embedding the domination instance inside a top-k instance in much the same way that the function f does, and then using A to solve the top-k instance. We receive as input two sets of samples $X_{i,\ell}$ and $Y_{i,\ell}$ (where $1 \leq i, j \leq n$ and $1 \leq \ell \leq r$) from some Domination instance C drawn from $\mathcal{C}$. We then generate a random permutation π of $[n+2]$. We use our input and this permutation to generate a matrix $Z_{i,j,\ell}$ (where $1 \leq i, j \leq n+2$ and $1 \leq \ell \leq \frac{r}{2}$) of samples to input to A as follows.

For $1 \leq i, j \leq n$, set each $Z_{\pi(i),\pi(j),\ell}$ to be a random $\mathcal{B}(\frac{1}{2})$ random variable. Similarly, for $n+1 \leq i, j \leq n+2$, set each $Z_{\pi(i),\pi(j),\ell}$ to be a random $\mathcal{B}(\frac{1}{2})$ random

variable. Now, for all $1 \leq j \leq n$, set $Z_{\pi(n+1),\pi(j),\ell} = X_{j,\ell}$ and set $Z_{\pi(n+2),\pi(j),\ell} = Y_{j,\ell}$. Similarly, for all $1 \leq i \leq n$, set $Z_{\pi(i),\pi(n+1),\ell} = 1 - X_{i,\ell+r/2}$ and set $Z_{\pi(i),\pi(n+2),\ell} = 1 - Y_{i,\ell+r/2}$. Finally, set $k = n+1$ and ask A to solve the TOP-K instance defined by k and $Z_{i,j,\ell}$. If A returns that $\pi(n+1)$ is in the top $n+1$ indices, return $B = 0$, and otherwise return $B = 1$.

From our construction, if the r samples of X and Y are distributed according to a DOMINATION instance C, then the $r/2$ samples of Z are distributed according to the TOP-K instance $S = f(C)$. Since A succeeds with probability p on distribution $\mathcal{S}$ with $r_{min}(\mathcal{S}, A, p)$ samples, A' therefore succeeds with probability p on distribution $\mathcal{C}$ with $2r_{min}(\mathcal{S}, A, p)$ samples, thus implying that $r_{min}(\mathcal{S}, A, p) \geq \frac{1}{2} r_{min}(\mathcal{C}, A', p)$. □

Corollary 6.8.14. *For all algorithms A that solve* TOP-K, $r_{min}(\mathcal{S}_{hard}, A, \frac{2}{3}) = \Omega\left(\frac{n^4}{\log n}\right)$.

Proof. Theorem 6.8.7 tells us that for all algorithms A' that solve DOMINATION, $r_{min}(\mathcal{C}_{hard}, A, \frac{2}{3}) = \Omega\left(\frac{1}{\varepsilon^2 \log n}\right) = \Omega\left(\frac{n^4}{\log n}\right)$. Combining this with Lemma 6.8.13, we obtain the desired result. □

We can now prove Theorem 6.8.1 in much the same fashion as Theorem 6.8.2.

Proof of Theorem 6.8.1. By Corollary 6.8.14, $r_{min}(\mathcal{S}_{hard}, A, \frac{2}{3}) = \Omega\left(\frac{n^4}{\log n}\right)$. Let E be the event that $|S_P| \geq \frac{1}{10} n\gamma$ (in the original DOMINATION instance C). By Lemma 6.8.6, if $n \geq (400 \ln \frac{12}{11})^2$, $\Pr[E] \geq \frac{1}{12}$, and it follows from Lemma 6.8.4 that

$$\begin{aligned} r_{min}(\mathcal{S}_{hard}|E, A) &= r_{min}(\mathcal{S}_{hard}|E, A, \frac{3}{4}) \\ &\geq r_{min}(\mathcal{S}_{hard}, A, \frac{2}{3}) \\ &\geq \Omega\left(\frac{n^4}{\log n}\right) \end{aligned}$$

It therefore follows from 6.8.3 that there is a specific instance S in the support of $\mathcal{S}_{hard}|E$ such that $r_{min}(S, A) \geq \Omega\left(\frac{n^4}{\log n}\right)$. However, by Corollary 6.8.12, $r_{min}(S) \leq O(n^{3.5})$. It follows that for any algorithm A, there exists an instance S of Top-K such that $r_{min}(S, A) \geq \Omega\left(\frac{\sqrt{n}}{\log n}\right) r_{min}(S)$, as desired. □

Part IV

Appendices

Appendix A

Appendix for Chapter 2

A.1 Good no-regret algorithms for the buyer

In this section we show that there exists a (contextual) no-regret algorithm for the buyer which guarantees that the seller receives at most the Myerson revenue per round (i.e., $\mathsf{Mye}(\mathcal{D})T$ in total). As mentioned earlier, it does not suffice for the buyer to simply run the contextualization $\mathsf{cont}(M)$ for some no-regret learning algorithm M (and in fact, if M is mean-based, the seller can extract strictly more than $\mathsf{Mye}(\mathcal{D})T$, as we will see later). However, by modifying $\mathsf{cont}(M)$ so that it has not just no-regret with respect to the best stationary policy, but so that it additionally does not regret playing as if it had some other context, we obtain a no-regret algorithm for the buyer which guarantees the seller receives no more than $\mathsf{Mye}(\mathcal{D})$ per round.

The details of the algorithm are presented in Algorithm 15. Recall that the distribution $\mathcal{D}$ is supported over m values $v_1 < v_2 < \cdots < v_m$, where for each $i \in [m]$, v_i has probability q_i under $\mathcal{D}$. The algorithm takes a no-regret algorithm M for the classic multi-armed bandit problem, and runs M instances of it, one per possible value u. Each instance M_i of M learns not only over the possible K actions, but also over $i-1$ virtual actions corresponding to values v_1 through v_{i-1}. Picking the virtual

action associated with v_j corresponds to the buyer pretending they have value v_j, and playing accordingly (i.e., querying M_j).

This algorithm is very similar in structure to the construction of a low swap-regret bandits algorithm from a generic no-regret bandits algorithm (see [24]). The main difference is that whereas swap regret guarantees no-regret with respect to swapping actions (i.e. always playing action i instead of action j), this algorithm guarantees no-regret with respect to swapping *contexts* (i.e., always pretending you have context i when you actually have context j). In addition, the auction structure of our problem allows us to only consider contexts with valuations smaller than our current valuation v_i; this puts a limit of m on the number of recursive calls per round, as opposed to the low swap regret algorithm where one must solve for the stationary distribution of a Markov chain over m states each round.

Algorithm 15 No-regret algorithm for buyer (restatement of Algorithm 1).

Let M be a δ-no-regret algorithm for the classic multi-armed bandit problem, with $\delta = o(T)$. Initialize m copies of M, M_1 through M_m.
Instance M_i of M will learn over $K + i - 1$ arms.
The first K arms of M_i ("bid arms") correspond to the K possible menu options $b_1, b_2, \ldots, b_K$.
The last $i - 1$ arms of M_i ("value arms") correspond to the $i - 1$ possible values (contexts) $v_1, \ldots, v_{i-1}$.
for $t = 1$ to T **do**
    if buyer has value v_i **then**
        Use M_i to pick one arm from the $K + i - 1$ arms.
        if the arm is a bid arm b_j **then**
            Pick the menu option j (i.e. bid b_j).
        else if the arm is a value arm v_j **then**
            Sample an arm from M_j (but don't update its state). If it is a bid arm, pick the corresponding menu option. If it is a value arm, recurse.
        end if
        Update the state of algorithm M_i with the utility of this round.
    end if
end for

We now proceed to show that Algorithm 15 has our desired guarantees.

Theorem A.1.1. *Let $q_{min} = \min_i q_i$. If the buyer plays according to Algorithm 15 then the seller (even if they play an adaptive strategy) receives no more than $\mathsf{Mye}(\mathcal{D})T + \frac{m\delta}{q_{min}}$ revenue.*

Proof. For each $i \in [m]$, define h_i to be the expected number of rounds the buyer receives the item when they have value v_i. For each $i \in [m]$ define r_i to be the expected total payment from the buyer to the seller when the buyer has value v_i. Our goal is to upper bound $\sum_i r_i$, the total revenue the seller receives.

Recall that every strategy must contain a zero option in its menu, where the buyer pays nothing and doesn't receive the item (and hence receives zero utility). Since each M_i is a δ-no-regret algorithm, we know that the buyer does not regret always choosing the zero option when they have value v_i. It follows that, for all $i \in [m]$, we have that

$$v_i h_i - r_i \geq -\delta. \tag{A.1}$$

The following lemma shows that when $j > i$, the buyer does not regret pretending to have value v_i when they have value v_j.

Lemma A.1.2. *For all $1 \leq i < j \leq m$,*

$$(v_j h_j - r_j)/q_j \geq (v_j h_i - r_i)/q_i - \delta/q_j.$$

Proof. From the algorithm, we know that M_j does not regret always playing the value arm corresponding to v_i. We define the following notation. For all $i \in [m], t \in [T]$ and any history π of $t-1$ rounds (including for each round which option is chosen and the utility of that round), define $h_i(t, \pi)$ to be the probability of getting item in round t given history π when buyer has value v_i and define $r_i(t, \pi)$ to be the expected price paid in round t when the buyer has value v_i given history π.

Let Π_t be the distribution of histories at round t, for $t = 0, ..., T-1$. The no-regret guarantee tells us that

$$\sum_{t=1}^{T} q_j \cdot \mathbb{E}_{\pi\sim\Pi_{t-1}} \left[(h_j(t,\pi)v_j - r_j(t,\pi)) - (h_i(t,\pi)v_j - r_i(t,\pi))\right] \geq -\delta. \tag{A.2}$$

Note that

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}_{\pi\sim\Pi_{t-1}}[h_j(t,\pi)q_j] &= h_j, \\
\sum_{t=1}^{T} \mathbb{E}_{\pi\sim\Pi_{t-1}}[h_i(t,\pi)q_i] &= h_i, \\
\sum_{t=1}^{T} \mathbb{E}_{\pi\sim\Pi_{t-1}}[r_j(t,\pi)q_j] &= r_j, \\
\sum_{t=1}^{T} \mathbb{E}_{\pi\sim\Pi_{t-1}}[r_i(t,\pi)q_i] &= r_i.
\end{aligned}$$

Dividing (A.2) through by q_j and substituting in these relations, we arrive at the statement of the lemma. □

Now define $\lambda_i = \sum_{j\leq i} \frac{1}{q_j}$, and define

$$r'_i = \frac{r_i}{q_i} - \lambda_i\delta. \tag{A.3}$$

It follows from Lemma A.1.2 that for all $1 \leq i < j \leq m$,

$$\frac{v_j h_j}{q_j} - r'_j \geq \frac{v_j h_i}{q_i} - r'_i. \tag{A.4}$$

From (A.1), we also have for all $i \in [m]$,

$$\frac{v_i h_i}{q_i} - r'_i \geq 0. \tag{A.5}$$

We will argue from these constraints that $\sum_i q_i r'_i \leq \mathsf{Mye}(\mathcal{D})T$. To do this, we will construct a single-round mechanism for selling an item to a buyer with value distribution $\mathcal{D}$ such that this mechanism has expected revenue $\sum_i q_i r'_i/T$; the result then follows from the optimality of the Myerson mechanism ([112]).

To construct this mechanism, first find a sequence of indices $a_1, a_2, \ldots, a_l$ via the following algorithm.

1: $l \leftarrow 1$, $a_1 \leftarrow 1$.
2: **for** $i = 2$ to m **do**
3: **if** $r'_i \geq r'_{a_l}$ **then**
4: $l \leftarrow l + 1$, $a_l \leftarrow i$.
5: **end if**
6: **end for**

It is easy to verify that following this algorithm results in $r'_{a_1} \leq r'_{a_2} \leq \cdots \leq r'_{a_l}$. For any $a_i \leq j < a_{i+1}$ (assuming $a_{l+1} = m + 1$), $r'_j < r'_{a_i}$.

Lemma A.1.3. *For a bidder with value distribution $\mathcal{D}$, the following menu of l options will achieve revenue at least $\sum_{i=1}^{m} r'_i q_i/T$: for each $1 \leq i \leq l$, the buyer has the choice of paying r'_{a_i}/T, and receiving the item with probability $h_{a_i}/(q_{a_i}T)$.*

Proof. Consider some value v_j in $\mathcal{D}$. We will show that the buyer with value v_j will pay at least r'_j/T, thus proving the lemma. Assume $a_i \leq j \leq a_{i+1}$.

We have (from (A.5) and the monotonicity of v_i) that

$$\frac{v_j h_{a_i}}{q_{a_i}} - r'_{a_i} \geq \frac{v_{a_i} h_{a_i}}{q_{a_i}} - r'_{a_i} \geq 0.$$

This means the buyer with value u_j receives non-negative utility by choosing option i. For any $1 \leq i' < i$, we have (from (A.4)) that

$$\frac{v_{a_i} h_{a_i}}{q_{a_i}} - r'_{a_i} \geq \frac{v_{a_i} h_{a_{i'}}}{q_{a_{i'}}} - r'_{a_{i'}}.$$

Since $r'_{a_i} \geq r'_{a_{i'}}$, the above inequality implies that

$$\frac{h_{a_i}}{q_{a_i}} \geq \frac{h_{a_{i'}}}{q_{a_{i'}}}.$$

It follows that

$$v_j \left(\frac{h_{a_i}}{q_{a_i}} - \frac{h_{a_{i'}}}{q_{a_{i'}}} \right) \geq v_{a_i} \left(\frac{h_{a_i}}{q_{a_i}} - \frac{h_{a_{i'}}}{q_{a_{i'}}} \right) \geq r'_{a_i} - r'_{a_{i'}}.$$

This means the buyer with value v_j prefers option i to all options $i' < i$. Therefore this buyer will choose an option from $\{i, i+1, \ldots, l\}$. Since $r'_j \leq r'_{a_i} \leq r'_{a_{i+1}} \leq \cdots \leq r'_{a_l}$, we know that this buyer will pay at least r'_j/T, as desired. □

It follows from the optimality of the Myerson auction that $\sum_i q_i r'_i/T \leq \mathsf{Mye}(\mathcal{D})$, and therefore that $\sum_i q_i r'_i \leq \mathsf{Mye}(\mathcal{D})T$. Expanding out r'_i via (A.3), we have that

$$\begin{aligned} \sum_i q_i r'_i &= \sum_i r_i - \sum_i q_i \lambda_i \delta \\ &\geq \sum_i r_i - \delta \cdot \max_i \lambda_i \\ &\geq \sum_i r_i - \frac{m\delta}{q_{min}}, \end{aligned}$$

from which the theorem follows. □

We can remove the explicit dependence on q_{min} by filtering out all values which occur with small enough probability.

Corollary A.1.4 (Restatement of Theorem 2.3.2)**.** *There exists a no-regret algorithm for the buyer where the seller receives no more than* $\mathsf{Mye}(\mathcal{D})T + O(m\sqrt{\delta T})$ *revenue.*

Proof. Ignore all values v_i with $q_i \leq \sqrt{\delta/T}$ (whenever a round with this value arises, choose an arbitrary action for this round). There are m total values, so this happens with at most probability $m\sqrt{\delta/T}$, and therefore modifies the regret and revenue in expectation by at most $O(m\sqrt{\delta T}) = o(T)$.

The regret bound from Theorem A.1.1 then holds with $q_{min} \geq \sqrt{\delta/T}$, from which the result follows. □

A.1.1 Multiple bidders

Interestingly, we show that by slightly modifying Algorithm 15, we obtain an algorithm (Algorithm 16) that works for the case where there are *multiple bidders*. In the multiple bidder setting, there are B bidders with independent valuations for the item. Each round t, bidder ℓ receives a value $v_\ell(t)$ for the item drawn from a distribution $\mathcal{D}_\ell$ (independently of all other values). Each distribution $\mathcal{D}_\ell$ is supported over m_ℓ values, $v_{\ell,1} < v_{\ell,2} < \cdots < v_{\ell,m_\ell}$, where $v_{\ell,i}$ occurs under $\mathcal{D}_\ell$ with probability $q_{\ell,i}$. Every round each bidder ℓ submits a bid $b_\ell(t)$, and the auctioneer decides on an allocation rule $\mathbf{a}_t$, which maps ℓ-tuples of bids $(b_1(t), b_2(t), \ldots, b_B(t))$ to ℓ-tuples of probabilities $(a_1(t), a_2(t), \ldots, a_B(t))$ and a pricing rule $\mathbf{p}_t$, which maps ℓ-tuples of bids $(b_1(t), b_2(t), \ldots, b_B(t))$ to ℓ-tuples of prices $(p_1(t), p_2(t), \ldots, p_B(t))$. The allocation rule $\mathbf{a}_t$ must additionally obey the supply constraint that $\sum_\ell a_\ell(t) \leq 1$. Bidder ℓ wins the item with probability $a_\ell(t)$ and pays $p_\ell(t)$.

We show that if every bidder plays the no-regret algorithm Algorithm 16, then the auctioneer (even if playing adaptively) is guaranteed to receive no more than $\mathsf{Mye}(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_B)T + o(T)$ revenue, where $\mathsf{Mye}(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_B)$ is the optimal revenue obtainable by an auctioneer selling a single item to B bidders with valuations drawn independently from distributions $\mathcal{D}_\ell$. In other words, if every bidder plays according to Algorithm 16, the seller can do nothing better than running the single-round optimal Myerson auction every round.

The only difference between Algorithm 15 and Algorithm 16 is that instance M_i in Algorithm 16 has a value arm for every possible value, not only the values less than v_i. This means that the recursion depth of this algorithm is potentially unlimited, however it will still terminate in finite expected time since we insist that M has a positive probability of picking any arm (in particular, it will eventually pick a bid arm). We can optimize the runtime of step 11 of Algorithm 16 by eliciting a probability distribution over arms from each instance M_i, constructing a Markov chain, and solving for the stationary distribution. This takes $O((K+m)^3)$ time per step of this algorithm.

Algorithm 16 No-regret algorithm for a bidder (when there are multiple bidders).

1: Let M be a δ-no-regret algorithm for the classic multi-armed bandit problem (that always has some positive probability of choosing any arm), with $\delta = o(T)$. Initialize m copies of M, M_1 through M_m.
2: Instance M_i of M will learn over $K+m$ arms.
3: The first K arms of M_i ("bid arms") correspond to the K possible menu options $b_1, \ldots, b_K$.
4: The last m arms of M_i ("value arms") correspond to the m possible values (contexts) $v_1, \ldots, v_m$.
5: **for** $t = 1$ to T **do**
6: **if** buyer has value v_i **then**
7: Use M_i to pick one arm from the $K+m$ arms.
8: **if** the arm is a bid arm b_j **then**
9: Pick the menu option j (i.e. bid b_j).
10: **else if** the arm is a value arm v_j **then**
11: Sample an arm from M_j (but don't update its state). If it is a bid arm, pick the corresponding menu option. If it is a value arm, recurse.
12: **end if**
13: Update the state of algorithm M_i with the utility of this round.
14: **end if**
15: **end for**

Theorem A.1.5. *Let $q_{min} = \min_{\ell,i} q_{\ell,i}$. If every bidder plays according to Algorithm 16 then the auctioneer (even if they play an adaptive strategy) receives no more than $\mathsf{Mye}(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_B)T + O\left(\sqrt{\frac{\delta T}{q_{min}}}\right)$ revenue.*

Proof. Similarly as before, let $h_{\ell,i}$ equal the expected number of rounds bidder ℓ receives the item while having value $v_{\ell,i}$, and let $r_{\ell,i}$ equal the expected total amount bidder ℓ pays to the auctioneer while having value $v_{\ell,i}$. Again, our goal is to upper bound $\sum_\ell \sum_i r_{\ell,i}$, the total expected revenue the seller receives.

Note that, as before, since every strategy contains a zero option in its menu, we have that (for all $\ell \in [B]$ and $i \in [m_\ell]$)

$$v_{\ell,i} h_{\ell,i} - r_{\ell,i} \geq -\delta. \tag{A.6}$$

Repeating the argument of Lemma A.1.2 (which still holds in the multiple bidder setting), we additionally have that (for all $\ell \in [B]$ and $1 \leq i < j \leq m_\ell$),

$$\frac{v_{\ell,j} h_{\ell,j} - r_{\ell,j}}{q_{\ell,j}} \geq \frac{v_{\ell,j} h_{\ell,i} - r_{\ell,i}}{q_{\ell,i}} - \frac{\delta}{q_{\ell,j}}. \tag{A.7}$$

We will now (as in the proof of Theorem A.1.1) construct a mechanism for the single-round instance of the problem of an auctioneer selling a single item to B bidders with valuations independently drawn from $\mathcal{D}_\ell$. Our mechanism M will work as follows:

1. The auctioneer will begin by asking each of the bidders for their valuations. Assume that bidder ℓ reports valuation v'_ℓ (we will insist that v'_ℓ belongs to the support of $\mathcal{D}_\ell$).

2. The auctioneer will then sample a $t \in [T]$ uniformly at random.

3. For each bidder ℓ, the auctioneer will calculate $a_\ell(t)$ and $p_\ell(t)$, the expected allocation probability and price bidder ℓ has to pay in round t of the dynamic T-round mechanism, *conditioned on* $v_\ell(t) = v'_\ell$ *for all* ℓ.

4. The auctioneer will then give the item to bidder ℓ with probability $a_\ell(t)$, and charge bidder ℓ a price $p_\ell(t)$.

Note that since the allocation rules $\mathbf{a}_t$ must always satisfy the supply constraint, the probabilities $a_\ell(t)$ we sample also obey this supply constraint, and therefore this is a valid mechanism for the single-round problem. We will now show it is approximately incentive compatible.

Lemma A.1.6. *Mechanism M is $\frac{\delta}{q_{min}T}$-Bayesian incentive compatible and $\frac{\delta}{q_{min}T}$-ex-interim individually rational.*

Proof. To begin, we claim that in expectation, if bidder ℓ reports valuation $v_{\ell,i}$ (and everyone else reports truthfully), then the expected probability bidder ℓ receives the item (under this single-round mechanism) is equal to $h_{\ell,i}/Tq_{\ell,i}$. Likewise, we claim that, if bidder ℓ reports valuation $v_{\ell,i}$ (and everyone else reports truthfully), the expected payment bidder they pay is equal to $r_{\ell,i}/Tq_{\ell,i}$.

To see why this is true, let $h_\ell(t, i_1, i_2, \ldots, i_B)$ equal the probability bidder ℓ gets the item (in the multi-round mechanism) at time t conditioned on $v_\ell(t) = v_{\ell,i}$ for all $\ell \in [B]$. By construction, the probability $a'_{\ell,i}$ bidder ℓ receives the item (in mechanism M) after reporting valuation $v_{\ell,i}$ is equal to

$$a'_{\ell,i} = \frac{1}{T}\sum_t \sum_{\ell' \neq \ell, v_{\ell',i_{\ell'}} \in \mathsf{supp}\mathcal{D}_{\ell'}} \prod_{\ell' \neq \ell} q_{\ell',i_{\ell'}} h_\ell(t, i_1, i_2, \ldots, i_{\ell-1}, i, i_{\ell+1}, \ldots, i_B).$$

On the other hand, we can write $h_{\ell,i}$ in terms of our function h_ℓ as

$$h_{\ell,i} = \sum_t \sum_{\ell' \neq \ell, v_{\ell',i_{\ell'}} \in \mathsf{supp}\mathcal{D}_{\ell'}} q_{\ell,i} \prod_{\ell' \neq \ell} q_{\ell',i_{\ell'}} h_\ell(t, i_1, i_2, \ldots, i_{\ell-1}, i, i_{\ell+1}, \ldots, i_B).$$

It follows that $a'_{\ell,i} = \frac{h_{\ell,i}}{Tq_{\ell,i}}$. A similar calculation shows that if $p'_{\ell,i}$ is the expected payment of bidder ℓ (if they report valuation $v_{\ell,i}$ and everyone else reports truthfully), then $p'_{\ell,i} = \frac{r_{\ell,i}}{Tq_{\ell,i}}$.

Now, recall that a mechanism is ϵ-BIC if misreporting your value increases your expected utility by at most ϵ (assuming everyone else reports truthfully). To show that mechanism M is ϵ-BIC, it therefore suffices to show that for all $j \neq i$, that

$$a'_{\ell,j} v_{\ell,i} - p'_{\ell,j} \leq a'_{\ell,i} v_{\ell,i} - p'_{\ell,i} + \epsilon.$$

But for $\epsilon = \delta/(q_{min}T)$, this follows from equation (A.7). Similarly, M is ϵ-ex-interim IR if for all i,

$$a'_{\ell,i} v_{\ell,i} - p'_{\ell,i} \geq -\epsilon.$$

Again, this follows from equation (A.6), and the result therefore follows. □

We now apply the following lemma from [51], which lets us transform an ϵ-BIC mechanism M into a BIC mechanism M' at the cost of $O(\sqrt{\varepsilon})$ revenue.

Lemma A.1.7. *If M is an ϵ-BIC, ϵ-ex-interim IR mechanism for selling a single item to several bidders with independent valuations, then there exists a BIC, ex-interim IR mechanism M' for the same problem that satisfies $\mathsf{Rev}(M') \geq \mathsf{Rev}(M) - O(\sqrt{\epsilon})$.*

Proof. See Theorem 3.3 in [51]. □

Applying Lemma A.1.7 to our mechanism, we obtain a mechanism M' that satisfies $\mathsf{Rev}(M') \geq \mathsf{Rev}(M) - O(\sqrt{\frac{\delta}{q_{min}T}})$. Finally, note that since the Myerson auction is the optimal Bayesian-incentive compatible mechanism for this problem, $\mathsf{Rev}(M') \leq \mathsf{Mye}(\mathcal{D}_1, \ldots, \mathcal{D}_B)$. On the other hand, since (from the proof of Lemma A.1.6) the expected payment bidder ℓ pays under mechanism M when being truthful is equal to:

$$\sum_i q_{\ell,i} \cdot \frac{r_{\ell,i}}{T q_{\ell,i}} = \frac{1}{T} \sum_i r_{\ell,i}.$$

It follows that

$$\frac{1}{T}\sum_{\ell}\sum_{i} r_{\ell,i} \leq \mathsf{Mye}(\mathcal{D}_1,\ldots,\mathcal{D}_B) + O(\sqrt{\frac{\delta}{q_{min}T}}),$$

and thus that

$$\sum_{\ell}\sum_{i} r_{\ell,i} \leq \mathsf{Mye}(\mathcal{D}_1,\ldots,\mathcal{D}_B)T + O(\sqrt{\frac{\delta T}{q_{min}}}).$$

□

A.2 Achieving full welfare against non-conservative buyers

In this section, we will show that if the buyer uses a mean-based algorithm instead of Algorithm 15, the seller has a strategy which extracts the entire welfare from the buyer (hence leaving the buyer with zero utility).

Theorem A.2.1 (Restatement of Theorem 2.3.1)**.** *If the buyer is non-conservative and running a mean-based algorithm, for any constant $\varepsilon > 0$, there exists a strategy for the seller which obtains revenue at least $(1-\varepsilon)\mathsf{Val}(\mathcal{D})T - o(T)$.*

Proof. If every element in the support of $\mathcal{D}$ is at least $1-\varepsilon$, then the seller can simply always sell the item at price $1-\varepsilon$ (since $\mathcal{D}$ is supported on $[0,1]$, this ensures a $(1-\varepsilon)$ approximation to the buyer's welfare). From now on, we will assume that $\mathcal{D}$ is not entirely supported on $[1-\varepsilon, 1]$.

Recall that $\mathcal{D}$ is supported on m values $v_1 < v_2 < \cdots < v_m$, where v_i is chosen with probability q_i. Define $\rho = \min(v_m, 1-\varepsilon/2)$, and define $\delta = (1-\rho)/(1-v_1)$. Since $v_1 < 1-\varepsilon/2$ and $v_1 < v_m$, we know that $v_1 < \rho$ and therefore $\delta < 1$. Notice that here we can make the strategy independent of $\mathcal{D}$ if we just pick $\rho = 1-\varepsilon/2$ and $\delta = \varepsilon/2$ (but setting ρ and δ according to information about $\mathcal{D}$ can reduce the number of arms).

Consider the following strategy for the seller. In addition to the zero arm, the seller will offer $n = \frac{\log(\varepsilon/2)}{\log(1-\delta)}$ possible options, each with maximum bid value $b_i = 1$. We divide the timeline of each arm into three "sessions" in the following way:

1. $\emptyset$ **session:** For the first $(1-(1-\delta)^{i-1})T$ rounds, the seller charges 0 and does not give the item to the buyer (i.e. $(p_{i,t}, q_{i,t}) = (0,0)$).

2. **0 session:** For the next $(1-\delta)^{i-1}(1-\rho)T$ rounds, the seller charges 0 and gives the item to the buyer (i.e. $(p_{i,t}, q_{i,t}) = (0,1)$).

3. **1 session:** For the final $(1-\delta)^{i-1}\rho T$ rounds, the seller charges 1 and gives the item to the buyer (i.e. $(p_{i,t}, q_{i,t}) = (1,1)$).

Note that this strategy is monotone; if $i < j$, then $p_{i,t} \geq p_{j,t}$ and $a_{i,t} \geq a_{j,t}$.

Assume that the buyer is running a γ-mean-based algorithm, for some $\gamma = o(1)$. Define $A_j = (1-\rho(1-\delta)^{j-1})T$ and $B_j(v) = A_j + \frac{\min(v,\rho)}{1-v_1}(1-\rho)(1-\delta)^{j-1}T - \gamma T$. Note that A_j is the round where arm j starts its 1 session; we show in the following Lemma that (by the mean-based property), the buyer with value v will prefer arm j over any arm $j' < j$ over all rounds in the interval $[A_j, B_j(v)]$.

Lemma A.2.2. *For each $v_i \in \mathcal{D}$, $j \in \{1,\ldots,n-1\}$, and round $\tau \in [A_j, B_j(v_i)]$, $\sigma_{j,\tau}(v_i) > \sigma_{j',\tau}(v_i) + \gamma T$ for all $j' > j$.*

Proof. Note that arm j starts its 1 session at round $A_j \leq \tau$. It follows that

$$\begin{aligned}\sigma_{j,\tau}(v_i) &= v_i\left(\tau - (1-(1-\delta)^{j-1})T\right) - \left((1-\delta)^{j-1}T\rho - (T-\tau)\right) \\ &= (T-\tau) + (v_i-\rho)(1-\delta)^{j-1}T + v_i\tau - Tv_i.\end{aligned}$$

Now consider the cumulative utility of playing some arm $j' > j$. It is easy to verify that $B_j < A_{j+1}$, and therefore arm j' is still either in its $\emptyset$ session or its 0 session.

Since arm $j+1$ starts its 0 session the earliest, it follows that $\sigma_{j',\tau}(v_i) \leq \sigma_{j+1,\tau}(v_i)$, so from now on, assume without loss of generality that $j' = j+1$. There are two cases:

1. If $\tau < T(1-(1-\delta)^j)$, the utility is 0.
2. If $\tau \geq T(1-(1-\delta)^j)$, the utility is $(\tau - T(1-(1-\delta)^j))v_i$.

It suffices to show that

$$(T-\tau) + (v_i - \rho)(1-\delta)^{j-1}T + v_i\tau - Tv_i \geq \max(0, (\tau - T(1-(1-\delta)^j))v_i) + \gamma T.$$

We have that

$$\begin{aligned}
& (T-\tau) + (v_i-\rho)(1-\delta)^{j-1}T + v_i\tau - Tv_i - (\tau - T(1-(1-\delta)^j))v_i \\
= \;& v_i(1-\delta)^{j-1}\delta T + (T-\tau) - \rho(1-\delta)^{j-1}T \\
\geq \;& v_i(1-\delta)^{j-1}\delta T + (T - B_j(v_i)) - \rho(1-\delta)^{j-1}T \\
= \;& (1-\delta)^{j-1}T\left(v_i\delta - (1-\rho)\frac{\min(v_i,\rho)}{1-v_1}\right) + \gamma T \\
= \;& (1-\delta)^{j-1}T(1-\rho)\left(\frac{v_i - \min(v_i,\rho)}{1-v_1}\right) + \gamma T \\
\geq \;& \gamma T.
\end{aligned}$$

Similarly

$$\begin{aligned}
& (T-\tau) + (v_i-\rho)(1-\delta)^{j-1}T + v_i\tau - Tv_i \\
\geq \;& T - B_j(v_i) + (v_i-\rho)(1-\delta)^{j-1}T + v_iB_j(v_i) - Tv_i \\
= \;& (B_j(v_i) - T(1-(1-\delta)^{j-1}))v_i + (T - B_j(v_i) - \rho(1-\delta)^{j-1}T) \\
= \;& (B_j(v_i) - T(1-(1-\delta)^{j-1}))v_i - \min(v_i,\rho)\delta(1-\delta)^{j-1}T + \gamma T \\
\geq \;& (B_j(v_i) - T(1-(1-\delta)^{j-1}))v_i - v_i\delta(1-\delta)^{j-1}T + \gamma T \\
\geq \;& (B_j(v_i) - T(1-(1-\delta)^j))v_i + \gamma T \\
\geq \;& \gamma T.
\end{aligned}$$

□

It follows from the mean-based condition (Definition 2.2.2) that in the interval $[A_j, B_j(v_i)]$ the buyer with value v_i will, with probability at least $(1 - n\gamma)$, choose an arm currently in its 1-session (i.e. an arm with label at most j) and hence pay 1 each round. Since the buyer has value v_i for the item with probability q_i, the total contribution of the buyer with value v_i to the expected revenue of the seller is given by

$$\begin{aligned}
& q_i \sum_{j=1}^{n} (1-\gamma)(B_j(v_i) - A_j(v_i)) \\
= \; & q_i \sum_{j=1}^{n} (1-n\gamma) \left(\frac{\min(u,\rho)}{1-v_1}(1-\rho)(1-\delta)^{j-1}T - \gamma T \right) \\
= \; & (1-n\gamma) q_i T \left(-n\gamma + \frac{(1-\rho)\min(v_i,\rho)}{1-v_1} \sum_{j=1}^{n} (1-\delta)^{j-1} \right) \\
= \; & (1-n\gamma) q_i T \left(-n\gamma + \frac{(1-\rho)\min(v_i,\rho)(1-(1-\delta)^n)}{(1-v_1)\delta} \right) \\
= \; & (1-n\gamma) q_i T \left(-n\gamma + \min(v_i,\rho)(1-(1-\delta)^n) \right) \\
= \; & q_i T \min(v_i,\rho)(1-(1-\delta)^n) - o(T) \\
\geq \; & q_i T \left(1 - \frac{\varepsilon}{2}\right)^2 v_i - o(T) \\
\geq \; & (1-\varepsilon) q_i v_i T - o(T).
\end{aligned}$$

Here we have used the fact that $(1-(1-\delta)^n) = 1-\varepsilon/2$ (since $n = \log(\varepsilon/2)/\log(1-\delta)$) and $\min(v_i, \rho) \geq (1-\varepsilon/2)v_i$ (since if $\min(v_i,\rho) \neq v_i$, then $\rho = (1-\varepsilon/2) \geq (1-\varepsilon/2)v_i$. Summing this contribution over all $v_i \in \mathcal{D}$, we have that the expected revenue of the seller is at least

$$\begin{aligned}\sum_i \left((1-\varepsilon)q_i v_i T - o(T)\right) &= (1-\varepsilon)\left(\sum_i q_i v_i\right) T - o(T) \\ &= (1-\varepsilon)\,\mathbb{E}_{v\sim\mathcal{D}}[v]T \\ &= (1-\varepsilon)\mathsf{Val}(\mathcal{D})T.\end{aligned}$$

□

A.2.1 Switching-mean-based algorithms

One reason why we were able to exploit mean-based algorithms in the previous section (and in general) is that they do not adapt quickly enough to changes in the best arm. One way this is partially modelled in multi-armed bandits is through the concept of "switching regret". The $S-$*switching regret* (or just *switching regret*, when S is clear from context) of an algorithm $\mathcal{A}$ for the multi-armed bandits problem is the difference between the overall performance of $\mathcal{A}$ and the performance of the best algorithm in hindsight that switches arms at most S times. As before, we say that an algorithm is no-switching-regret if its expected switching regret is $o(T)$. No-switching-regret algorithms are easily seen to be *not* mean-based (for one, they are not fooled by the example given in the introduction). A natural question then arises: to be robust against such manipulation, does it suffice to simply have no-switching-regret for some S?

In [16], the authors present an algorithm (EXP3.S) for the multi-armed bandits algorithm with $\sqrt{SKT\log T}$. While this algorithm is not mean-based, it does have the following mean-based-like property. Let Π be the set of different policies which switch at most S times (note that for constant S there are $O(T^S)$ policies in this set). We say a policy $\pi \in \Pi$ is γ-*dominated* at time t if there exists another policy $\pi' \in \Pi$ such that the cumulative reward $\sigma_{\pi,t}$ of playing π until round t satisfies

$\sigma_{\pi,t} < \sigma_{\pi',t} - \gamma T$. Then an algorithm is *$\gamma$-switching-mean-based* if, each round, with probability at least $1-\gamma$, it plays according to a non-γ-dominated strategy, and an algorithm is *switching-mean-based* if it is γ-switching-mean-based for some $\gamma = o(1)$. Just as EXP3 is mean-based, EXP3.S is switching-mean-based (the proof follows that of Theorem A.4.3).

We will show here that it is possible to extend our counterexample in the previous section to achieve full-welfare against a switching-mean-based buyer.

Theorem A.2.3. *If the buyer is non-conservative and running a switching-mean-based algorithm (with $S = O(1)$ switches), then for any constant $\varepsilon > 0$, there exists a strategy for the seller which obtains revenue at least $(1-\varepsilon)\mathsf{Val}(\mathcal{D})T - o(T)$.*

Proof. We will use the example in Theorem 2.3.1 as a blackbox. Divide the time horizon into $P > S$ phases of length T/P each. From Theorem 2.3.1, we can construct a set of arms which achieves $(1-\varepsilon')\mathsf{Val}(\mathcal{D})(T/P)$ welfare in a time horizon of length T/P using some number n of arms. Our example will have nP total arms, with each phase having n arms assigned to it. The n arms for phase i out of P will have the following payout structure:

1. In phases $j < i$, all arms charge nothing and do not give the item.
2. In phase i, arms behave according to the payout structure of our example from Theorem 2.3.1.
3. In phases $j > i$, all arms charge 1 and give the item.

Since n arms in the example in Theorem 2.3.1 can be assigned monotone bids, this payout structure allows this set of nP arms to be assigned monotone bids (in particular, if $j > i$, then all the arms in phase j should have lower associated bids than those in phase i). Because arms are monotone decreasing, we can without loss of generality restrict ourselves to looking at strategies which only ever switch to arms

with lower bids (i.e. arms active in later phases, or arms in the current phase but with a lower bid). We will label the ith arm in the pth phase as $a_{i,p}$.

Consider a buyer running a no-switching-regret strategy with S switches. Assume we are at the tth round in the pth phase (so the $\tau = (t + pT/P)$th round overall), where $p > S$, and $t > \gamma T = o(T)$. Further assume the buyer has some fixed value v. Let a_i be the action with the highest cumulative utility for a buyer with this value (i.e. the action a mean-based buyer is most likely to play) in the tth round of our example, and denote this utility by U. We claim the following.

Lemma A.2.4. *If $U \geq (\epsilon'/P + \gamma)T$, then for all non-dominated $\pi \in \Pi$, π will play an arm at round τ of the form $a_{i',p}$ with cumulative utility (in the example) of at least $U - \gamma T$.*

Proof. To show this, we first argue that any policy that plays two different arms in phase p is dominated. To do this, let δT be the length of the longest 0 session in our example, and let $\delta' T$ be the length of the second longest 0 session in our example. Note that any policy that switches at most s times achieves cumulative utility up to time τ at most $s\delta T$ (since the largest utility you can receive from any given arm is δT). Moreover, as long as τ belongs to the $(S+1)$th phase or later, there is a policy which switches s times and receives utility $s\delta T$ (simply switch from the arm with the longest 0 session in phase i to the arm with the longest 0 session in phase $i+1$ for the first S phases, finally switching to the zero arm at the end). On the other hand, any policy that plays two different arms in phase p receives utility at most $(s-1)\delta T + \delta' T$, which is less than $s\delta T - \gamma T$ for large enough T (since γ is $o(1)$). It follows that such policies are dominated.

We next argue that if $U \geq \gamma T$ (where U is the cumulative utility of playing action a_i until round t in our subexample), then any policy which plays an arm of the form $a_{j,p'}$ with $p' \neq p$ (i.e. an arm belonging to a different phase) for more than γT rounds in phase p is dominated. To see this, consider the last switch the policy makes:

- If the policy switches to the zero arm (or an arm in a phase $p' > p$, which behaves identically to the zero arm for this time range), then it can increase its utility by at least U by instead switching to $a_{i,p}$.

- If the policy switches to an arm $a_{i,p'}$ with $p' < p$, then this switch will result in at most additional $\epsilon' T/P$ utility, since our example has the property that playing any fixed arm from round t' to the final round T results in at most $\epsilon' T/P$ utility. Again it follows that the policy can increase its utility by at least $U - \epsilon' T/P \geq \gamma T$ by switching to $a_{i,p}$.

- Finally, if the policy switches to an arm of the desired form $a_{j,p}$, but at a time t' at least γT rounds after the start of phase p, it can increase its utility by at least γT by switching at the beginning of phase p.

It follows that all such policies are dominated. Finally, any policy that plays some arm $a_{j,p}$ (with cumulative utility in our example less than $U - \gamma T$) can increase its utility by at least γT by switching to $a_{i,p}$ instead of $a_{j,p}$. Our claim is thus proven. □

As a consequence of this claim, for each phase

□

A.3 Optimal revenue against conservative buyers

In Theorem 2.3.1, we demonstrated a mechanism for the seller that extracts full welfare from a buyer running a mean-based learning algorithm. This mechanism, while in some sense as good as possible (it is impossible to extract more than welfare from any buyer running a no-regret strategy), has several drawbacks. One general drawback is that it is extremely unlikely the mechanism in Section A.2 would arise naturally as the allocation rule for any sort of auction that might arise in practice. A more specific drawback is that this mechanism assumes buyers are learning over all

possible bids, instead of just bids less than their value; indeed, all arms essentially cost the maximum possible price per round, and their only difference is when they give the item away for free and when they charge for it.

In this section, we address the second drawback by studying this problem for *conservative buyers*; buyers who are constrained to only submit bids less than their current value for the item. We characterize via a linear program the optimal revenue attainable for the seller when playing against conservative buyers running a mean-based learning algorithm over their set of allowable bids. We show that, while we can no longer achieve the full welfare as in Section A.2, we can still achieve strictly more than the Myerson revenue. Interestingly, our optimal mechanism has a natural interpretation as a repeated first-price auction with gradually decreasing reserve, thus also partially addressing the first drawback. Notably, this auction is a *critical auction*. Since clever buyers act conservatively in critical auctions, this mechanism is simultaneously the optimal critical auction against clever buyers.

A.3.1 Characterizing the optimal revenue

$$\begin{aligned} \textbf{maximize} \quad & \sum_{i=1}^{m} q_i(v_i x_i - u_i) \\ \textbf{subject to} \quad & u_i \geq (v_i - v_j) \cdot x_j, \quad \forall\, i, j \in [m] : i > j \\ & u_i \geq 0, 1 \geq x_i \geq 0, \quad \forall\, i \in [m] \end{aligned}$$

Figure A.1: The mean-based revenue LP (same as Figure 2.1).

We begin by describing the optimal strategy for the seller against mean-based conservative buyers. Fix some small constant $\varepsilon > 0$. Recall that the buyer's value distribution $\mathcal{D}$ is supported on the m values $0 \leq v_1 < v_2 < \cdots < v_m \leq 1$, with $\Pr[v_i] = q_i$. The seller will offer m options, one for each possible value. Option i

(corresponding to bidding $b_i = v_i$) will charge 0 and not allocate the item for the first $(1-x_i)T$ rounds, and charge $b_i - \varepsilon$ and allocate the item for the remaining $x_i T$ rounds. The values x_i are computed by finding an optimal solution to the above LP (Figure A.1), which we call the *mean-based revenue LP*. We will call the value of this LP the *mean-based revenue* of $\mathcal{D}$, and write this as $\mathsf{MBRev}(\mathcal{D})$. Our goal in this subsection will be to show that this strategy achieves approximately $\mathsf{MBRev}(\mathcal{D})T$ total revenue against a conservative buyer running a mean-based algorithm, and that this is tight; no other strategy for a non-adaptive seller can obtain more than $\mathsf{MBRev}(\mathcal{D})T$ revenue.

To show that this is a valid strategy for the seller, we need to show that the values x_i are monotone increasing. Luckily, this follows simply from the structure of the mean-based revenue LP.

Lemma A.3.1. *Let $x_1, x_2, \ldots, x_m, u_1, u_2, \ldots, u_m$ be an optimal solution to the mean-based revenue LP. Then for all $i < j$, $x_i < x_j$.*

Proof. We proceed by contradiction. Suppose that the sequence of x_i are not monotone; then there exists an $1 \leq i \leq m-1$ such that $x_i > x_{i+1}$. Now consider another solution of the LP, where we increase x_{i+1} to x_i, keeping the value of all other variables the same. This new solution does not violate any constraints in the LP since for all $j > i+1$, $u_j \geq (v_j - v_i) \cdot x_i \geq (v_j - v_{i+1}) \cdot x_i$. However this change increases the value of the objective by $v_{i+1} q_{i+1}(x_i - x_{i+1}) > 0$, thus contradicting the fact that $x_1, \ldots, x_m, u_1, ..., u_m$ was an optimal solution of the mean-based revenue LP. □

We begin by showing that this strategy achieves revenue at least $\mathsf{MBRev}(\mathcal{D})T - o(T)$ when the buyer is using a mean-based algorithm.

Theorem A.3.2 (Restatement of Theorem 2.3.6)**.** *The above strategy for the seller gets revenue at least $(\mathsf{MBRev}(\mathcal{D}) - \varepsilon)T - o(T)$ against a conservative buyer running a mean-based algorithm. In addition, this strategy is critical.*

Proof. First of all, by Lemma A.3.1, it is easy to check the strategy is critical.

To prove the rest, we will show that: i) the buyer with value v_i receives the item for at least $x_iT - o(T)$ turns (receiving $v_ix_iT - o(T)$ total utility from the items), and ii) this buyer's net utility is at most $(u_i + \varepsilon)T + o(T)$. This implies that this buyer pays the seller at least $x_iv_iT - (u_i + \varepsilon)T - o(T)$ over the course of the T rounds; taking expectation over all v_i completes the proof.

Assume the buyer is running a γ-mean-based learning algorithm. Consider the buyer when they have value v_i. Note that

$$\sigma_{j,t}(v_i) = (v_i - v_j + \varepsilon) \cdot \max(0, t - (1 - x_j)T).$$

We first claim that after round $(1 - x_i)T + \gamma T/\varepsilon$, the buyer will buy the item (i.e., choose an option that results in him getting the item) each round with probability at least $1 - m\gamma$. To see this, first note that $\sigma_{i,t}(v_i) \geq \gamma T$ when $t \geq (1 - x_i)T + \gamma T/\varepsilon$. Then, since the cumulative utility of any arm is 0 until it starts offering the item, it follows from the mean-based condition that the buyer will pick a specific arm that is not offering the item with probability at most γ, and therefore choose some good arm with probability at least $1 - m\gamma$. It follows that, in expectation, the buyer with value v_i receives the item for at least $(1 - m\gamma)(x_iT - \gamma T/\varepsilon) = x_iT - o(T)$ turns.

We now proceed to upper bound the overall expected utility of the buyer. For each index $j \leq i$, let S_j be the set of t where $\sigma_{j,t}(v_i) > \sigma_{j',t}(v_i)$ for all other j'. Note that since each $\sigma_{j,t}(v_i)$ is a linear function in t (when positive), each S_j is either the empty set or an interval (y_jT, z_jT). Since all the v_i are distinct, note that these intervals partition the interval $((1 - x_i)T, T)$ (with the exception of up to m endpoints of these intervals); in particular, $\sum_{j \geq i}(z_j - y_j) = x_i$.

Let $\varepsilon' = \min_j(v_{j+1} - v_j)$. Note that, if $t \in (y_jT + \gamma T/\varepsilon', z_jT - \gamma T/\varepsilon')$, then for all $j' \neq j$, $\sigma_{j,t}(v_i) > \sigma_{j',t}(v_i) + \gamma T$. This follows since $\sigma_{j,t}(v_i) - \sigma_{j',t}(v_i)$ is linear in t with slope $v_j - v_{j'}$, and $|v_j - v_{j'}| > \varepsilon'$. It follows that if t is in this interval, then the

buyer will choose option j with probability at least $1 - m\gamma$ (by a similar argument as before).

Define $j(t) = \arg\max_j \sigma_{j,t}(v_i)$ to be the index of the arm with the current largest cumulative reward, and let $\sigma_{max,t}(v_i) = \sum_{s=1}^{t} r_{j(s),s}(v_i)$ be the cumulative utility of always playing the arm with the current highest cumulative reward for the first t rounds. The following lemma shows that $\sigma_{max,T}(v_i)$ is close to $\max_j \sigma_{j,T}(v_i)$. (In other words, playing the best arm every round and playing the best-at-the-end arm every round have similar payoffs if the historically best arm does not change often).

Lemma A.3.3. $|\sigma_{max,T}(v_i) - \max_j \sigma_{j,T}(v_i)| \leq m$.

Proof. Let $W = |\{t|j(t) \neq j(t+1)\}|$ equal the number of times the best arm switches values; note that since each $\sigma_{j,t}(v_i)$ is linear, W is at most m. Let $t_1 < t_2 < \cdots < t_W$ be the values of t such that $j(t) \neq j(t+1)$. Additionally define $t_0 = 1$ and $t_{W+1} = T$. Then, dividing the cumulative reward $\sigma_{max,t}$ into intervals by these t_i, we get that

$$\begin{aligned}
\sigma_{max,t}(v_i) &= \sum_{s=1}^{t} r_{j(s),s}(v_i) \\
&= \sum_{i=1}^{W+1} (\sigma_{j(t_i),t_i}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) \\
&= \sigma_{j(T),T}(v_i) + \sum_{i=1}^{W+1} (\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) \\
&= \max_j \sigma_{j,t}(v_i) + \sum_{i=1}^{W+1} (\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i))
\end{aligned}$$

It therefore suffices to show that $|\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)| \leq 1$ for all i. To see this, note that (by the definition of $j(t)$), $\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i) > 0$, and that $\sigma_{j(t_{i-1}),t_{i-1}+1}(v_i) - \sigma_{j(t_i),t_{i-1}+1}(v_i) < 0$. However,

$$(\sigma_{j(t_{i-1}),t_{i-1}+1}(v_i) - \sigma_{j(t_i),t_{i-1}+1}(v_i)) =$$
$$(\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)) + (r_{j(t_{i-1}),t_{i-1}+1}(v_i) - r_{j(t_i),t_{i-1}+1}(v_i))$$

Since $0 \leq r_{j,t}(u) \leq 1$, it follows that $|\sigma_{j(t_{i-1}),t_{i-1}}(v_i) - \sigma_{j(t_i),t_{i-1}}(v_i)| \leq 1$. This completes the proof. □

Let $\sigma_T(v_i) = \sum_{t=1}^{T} \mathbb{E}[r_{I_t,t}(v_i)]$ denote the expected cumulative utility of this buyer at time T. We claim that $\sigma_T \leq \max_j \sigma_{j,T}(v_i) + o(T)$. To see this, recall that, for $t \in (y_j T + \gamma T/\varepsilon', z_j T - \gamma T/\varepsilon')$, $\Pr[I_t \neq j] \leq m\gamma$, and therefore $\mathbb{E}[r_{I_t,t}] \leq r_{j,t} + m\gamma$. Furthermore, note that for $t \in S_j$, $j(t) = j$, so $r_{j,t} = r_{j(t),t}$ and $\mathbb{E}[r_{I_t,t}] \leq r_{j(t),t} + m\gamma$. It follows that

$$
\begin{aligned}
\sigma_T(v_i) &= \sum_{t=1}^{T} \mathbb{E}[r_{I_t,t}(v_i)] \\
&\leq \sum_{t=(1-x_i)T}^{T} \mathbb{E}[r_{I_t,t}(v_i)] \\
&= \sum_{j=1}^{i} \sum_{t=y_jT}^{z_jT} \mathbb{E}[r_{I_t,t}(v_i)] \\
&\leq \sum_{j=1}^{i} \left(\frac{2\gamma T}{\varepsilon'} + \sum_{t=y_jT+\gamma T/\varepsilon'}^{z_jT-\gamma T/\varepsilon'} \mathbb{E}[r_{I_t,t}(v_i)] \right) \\
&\leq \sum_{j=1}^{i} \left(\frac{2\gamma T}{\varepsilon'} + \sum_{t=y_jT+\gamma T/\varepsilon'}^{z_jT-\gamma T/\varepsilon'} (r_{j(t),t}(v_i) + m\gamma) \right) \\
&\leq \frac{2m\gamma T}{\varepsilon'} + m\gamma T + \sum_{t=1}^{T} r_{j(t),t}(v_i) \\
&= \frac{2m\gamma T}{\varepsilon'} + m\gamma T + \sigma_{max,T}(v_i) \\
&\leq \frac{2m\gamma T}{\varepsilon'} + m\gamma T + m + \max_j \sigma_{j,T}(v_i) \\
&= \max_j \sigma_{j,T}(v_i) + o(T).
\end{aligned}
$$

Finally, note that

$$
\begin{aligned}
\max_j \sigma_{j,T}(v_i) &= \max_{j<i}(v_i - v_j + \varepsilon)x_j T \\
&\leq (\max_{j<i}(v_i - v_j)x_j + \varepsilon)T \\
&= (u_i + \varepsilon)T
\end{aligned}
$$

It follows that $\sigma_T(v_i) \leq (u_i + \varepsilon)T + o(T)$, as desired.

□

We now proceed to show that this bound is in fact optimal; no strategy for the seller (even an adaptive one) can achieve better revenue against a no-regret , conservative buyer.

Theorem A.3.4 (Restatement of Theorem 2.3.4)**.** *Any strategy for the seller achieves revenue at most* $\mathsf{MBRev}(\mathcal{D})T + o(T)$ *against a conservative buyer running a no-regret algorithm.*

Proof. Assume the buyer is running a δ-no-regret algorithm, for some $\delta = o(T)$. Consider an arbitrary strategy for the seller with K arms, where arm j is labelled with maximum bid b_j. We begin by claiming that the following LP (Figure A.2) provides an upper bound on the revenue obtainable by this strategy against our no-regret buyer.

$$
\begin{aligned}
\textbf{maximize} \quad & \sum_{i=1}^{m} q_i(v_i x_i - u_i) \\
\textbf{subject to} \quad & u_i \geq v_i y_j - \overline{p}_j - \delta/T, \; i \in [m], j \in [K] : v_i \geq b_j \\
& \overline{p}_j \leq b_j y_j, \qquad\qquad j \in [K] \\
& x_i = y_j, \qquad\qquad i \in [m], j = \arg \max_{j \in [K] : b_j \leq v_i} b_j \\
& \overline{p}_j \geq 0, 1 \geq y_j \geq 0, \quad j \in [K]
\end{aligned}
$$

Figure A.2: LP', with variables x_i, u_i, y_j, and $\overline{p}_j$

Lemma A.3.5. *Let V' be the optimal value of LP' (see Figure A.2). Then the expected revenue of the seller is at most $V'T$.*

Proof. Given our strategy for the seller, we will assign values to variables in the following way. Fix a strategy for the buyer, and let $y_j = \frac{1}{T}\,\mathbb{E}\left[\sum_t a_{j,t}\right]$ be the expected average probability that arm j gives the item and let $\overline{p}_j = \frac{1}{T}\,\mathbb{E}\left[\sum_t p_{j,t}\right]$ be the expected average price charged by arm j. We will define x_i through the third constraint, and

set $u_i = \max_j(v_i y_j - \overline{p}_j - \delta/T)$. We will show that this assignment of variables satisfies all the constraints, and that the objective function evaluated on this assignment of variables is at least the seller's revenue using this strategy.

The first and third constraints are satisfied via our choices of x_i and u_i. The constraint $\overline{p}_j \leq b_j y_j$ is satisfied since $p_{j,t} \leq b_j a_{j,t}$ for all t. Finally, $0 \leq y_j \leq 1$ is satisfied since y_j is an average probability.

We now must show that the seller's revenue is at most $q_i(v_i x_i - u_i)$. We begin by claiming that x_i is an upper bound for the expected fraction of the time that the buyer receives the item when he has value v_i. To see this, note first that the buyer is conservative, and therefore will not bid on any arm with bid value larger than v_i. Choose j so that b_j is maximized over all $b_j \leq v_i$; note that since the seller's strategy is monotone, $a_{j,t} > a_{j',t}$ for any $j' < j$, so the buyer will receive the item at most $\mathbb{E}\left[\frac{1}{T}\sum_t a_{j,t}\right] = y_j$ of the time in expectation. But by our third constraint, $x_i = y_j$, so x_i is an upper bound on the average probability that the buyer with value v_i gets the item, and therefore $\sum_{i=1}^m q_i v_i x_i$ is an upper bound on the average welfare of the buyer.

We next claim that $\sum_i q_i u_i$ is a lower bound for the average utility of the buyer. To see this, note that since the buyer is using a δ-no-regret algorithm, when the value is v_i, the buyer should not regret always playing some arm j with $w_j \leq v_i$. Therefore the average surplus of value v_i should satisfy the constraint on u_i, and so $\sum_{i=1}^m q_i \cdot u_i$ is a lower bound on the average surplus of the buyer.

Finally, note that the seller's revenue is just the buyer's welfare minus the buyer's surplus. Combining the upper bound on the buyer's welfare and the lower bound on the buyer's surplus, we get our desired upper bound on the seller's revenue. □

We will now show how to transform a solution of this LP into a solution to the mean-based revenue LP while ensuring that its value does not decrease by more than δ/T. To begin, it is easy to see that there exists an optimal solution of LP' that

satisfies $\overline{p}_j = y_j \cdot w_j$ for all $j \in [K]$. We can thus increase each u_i by δ/T, since this will decrease the value of the LP by at most δ/T as $\sum_{i=1}^{m} q_i = 1$. This solution now satisfies $u_i \geq (v_i - b_j)y_j$ for all $i \in [m], j \in [K] : v_i \geq b_j$. Finally, for each $i, j \in [m] : i > j$, note that for $\ell = \arg\max_{\ell \in [K]: b_\ell \leq v_j} b_\ell$, we have that $b_\ell \leq v_j$. It follows that $u_i \geq (v_i - v_j)y_\ell = (v_i - v_j)x_j$, and therefore that this solution is a valid solution of the mean-based revenue LP.

From the above argument, we can conclude that $V_1 \leq R_{mb}(\mathcal{D}) + \delta/T$. It follows from Lemma A.3.5 that the total revenue is upper bounded by $T(\mathsf{MBRev}(\mathcal{D}) + \delta/T) = R_{mb}(\mathcal{D})T + o(T)$, as desired. □

Note that the proof of Lemma A.3.5 relies on the fact that our allocation rule is monotone. We can show that this constraint is necessary; with non-monotone strategies, the seller can extract up to the full welfare of a conservative buyer playing a mean-based strategy. The proof of this fact can be found in Appendix **??**.

A.3.2 Bounding $\mathsf{MBRev}(\mathcal{D})$

In this section, we compare the mean-based revenue $\mathsf{MBRev}(\mathcal{D})$ to our two benchmarks: the Myerson revenue for the item, $\mathsf{Mye}(\mathcal{D})$, and the buyer's expected value for the item, $\mathsf{Val}(\mathcal{D})$. It is not too hard to see that $\mathsf{MBRev}(\mathcal{D}) \leq \mathsf{Val}(\mathcal{D})$ (the value of the mean-based revenue LP is clearly at most $\sum_i q_i v_i = \mathsf{Val}(\mathcal{D})$) and that $\mathsf{MBRev}(\mathcal{D}) \geq \mathsf{Mye}(\mathcal{D})$ (the seller can achieve $\mathsf{Mye}(\mathcal{D})$ by just always selling the item the Myerson price). We show here that $\mathsf{MBRev}(\mathcal{D})$ is not a constant factor approximation to either $\mathsf{Mye}(\mathcal{D})$ or $\mathsf{Val}(\mathcal{D})$, and thus lies strictly between our two benchmarks in general.

We will begin by showing that $\mathsf{MBRev}(\mathcal{D})$ is monotone with respect to stochastic dominance. We will break from notation somewhat by considering distributions $\mathcal{D}$ supported on $[1, H]$ rather than $[0, 1]$; since $\mathsf{Mye}(\mathcal{D})$, $\mathsf{MBRev}(\mathcal{D})$, and $\mathsf{Val}(\mathcal{D})$ are all linear in the values v_i, dividing all values through by H results restores the condition

that $\mathcal{D}$ is supported on $[0,1]$ while preserving the multiplicative gaps between these quantities.

Definition A.3.6. *A distribution $\mathcal{D}$ stochastically dominates distribution $\mathcal{D}'$ if for all t, $\Pr_{u\sim\mathcal{D}}[u \geq t] \geq \Pr_{u\sim\mathcal{D}'}[u \geq t]$.*

Lemma A.3.7. *If distribution $\mathcal{D}$ stochastically dominates distribution $\mathcal{D}'$, then $\mathsf{MBRev}(\mathcal{D}) \geq \mathsf{MBRev}(\mathcal{D}')$.*

Proof. Note that we can write $\mathsf{MBRev}(\mathcal{D})$ in the form

$$\mathsf{MBRev}(\mathcal{D}) = \max_x \mathbb{E}_{v_i\sim\mathcal{D}}\left[v_i x_i - \max_j (v_i - v_j)x_j\right]$$

To show $\mathsf{MBRev}(\mathcal{D}) \geq \mathsf{MBRev}(\mathcal{D}')$, it suffices to show that for all increasing x (i.e. $x_i \geq x_j$ for $i \geq j$), that

$$\mathbb{E}_{v_i\sim\mathcal{D}}\left[v_i x_i - \max_j (v_i - v_j)x_j\right] \geq \mathbb{E}_{v_i\sim\mathcal{D}'}\left[v_i x_i - \max_j (v_i - v_j)x_j\right]$$

Note that if $\mathcal{D}$ stochastically dominates distribution $\mathcal{D}'$, then for any increasing function f, $\mathbb{E}_{u\sim\mathcal{D}}[f(u)] \geq \mathbb{E}_{u\sim\mathcal{D}'}[f(u)]$. It suffices to show that $f(v_i) = v_i x_i - \max_j (v_i - v_j)x_j$ is increasing in i (and hence in v_i). In particular, we wish to show that, for $i' > i$,

$$v_{i'} x_{i'} - \max_j (v_{i'} - v_j)x_j \geq v_i x_i - \max_j (v_{i'} - v_j)x_j$$

or equivalently,

$$\min_j \left(v_{i'} x_{i'} - (v_{i'} - v_j)x_j\right) \geq \min_j \left(v_i x_i - (v_i - v_j)x_j\right).$$

To show this, it suffices to show that for each j,

$$v_{i'} x_{i'} - (v_{i'} - v_j)x_j \geq v_i x_i - (v_i - v_j)x_j$$

or equivalently,

$$v_{i'}x_{i'} - v_i x_i \geq (v_{i'} - v_i)x_j.$$

This follows since

$$\begin{aligned} v_{i'}x_{i'} - v_i x_i &\geq v_{i'}x_i - v_i x_i \\ &= (v_{i'} - v_i)x_i \\ &\geq (v_{i'} - v_i)x_j. \end{aligned}$$

Here we have used the fact that $x_{i'} \geq x_i \geq x_j$. This concludes the proof.

□

For ease of analysis, we will also switch to considering continuous distributions $\mathcal{D}$. The definitions of $\mathsf{Mye}(\mathcal{D})$ and $\mathsf{Val}(\mathcal{D})$ still hold for continuous $\mathcal{D}$. Since the mean-based revenue LP implies that, in the optimal solution, $u_i = \max_j (v_i - v_j)x_j$, we can write $\mathsf{MBRev}(\mathcal{D})$ for a continuous $\mathcal{D}$ supported on $[1, H]$ with pdf $q(v)$ as

$$\mathsf{MBRev}(\mathcal{D}) = \max_{x(v)} \int_1^H q(v)(vx(v) - \max_{w<u}(v - w)x(w))dv.$$

By discretizing appropriately, all gaps we prove for continuous $\mathcal{D}$ extend to discrete values of $\mathcal{D}$.

Definition A.3.8. *The* equal revenue curve *is the (continuous) distribution $\mathcal{D}_{ERC}$ supported on $[1, \infty)$ with CDF $F(v) = 1 - \frac{1}{v}$. The* equal revenue curve truncated at H *is the distribution distribution $\mathcal{D}_{ERC}(H)$ supported on $[1, H]$ with CDF $F(v) = 1 - \frac{1}{v}$ for $v \leq H$ and $F(v) = 0$ for $v > H$.*

Note that $\mathsf{Mye}(\mathcal{D}_{ERC}) = 1$ (since $v(1 - F(v)) = 1$ for all $v \geq 1$). Likewise, $\mathsf{Mye}(\mathcal{D}_{ERC}(H)) = 1$.

Lemma A.3.9. *Let $\mathcal{D}_{ERC}(H)$ be the equal revenue curve truncated at H. Let $\mathcal{D}$ be any distribution supported on $[1, H]$ with $\mathsf{Mye}(D) = 1$. Then $\mathcal{D}_{ERC}(H)$ stochastically dominates $\mathcal{D}$.*

Corollary A.3.10. *The distribution $\mathcal{D}$ supported on $[1, H]$ that maximizes $\mathsf{MBRev}(\mathcal{D})$ subject to $\mathsf{Mye}(\mathcal{D}) = 1$ is the truncated equal revenue curve $\mathcal{D}_{ERC}(H)$.*

Theorem A.3.11. $\mathsf{MBRev}(\mathcal{D}_{ERC}(H)) \geq \Omega(\log \log H)$.

Proof. Note that for $\mathcal{D}_{ERC}(H)$, the pdf $q(v)$ is given by $q(v) = \frac{1}{v^2}$, so

$$\begin{aligned}
\mathsf{MBRev}(\mathcal{D}_{ERC}(H)) &\geq \max_{x(v)} \int_1^H q(v)(vx(v) - \max_{w<v}(v-w)x(w))dv \\
&= \max_{x(v)} \int_1^H \frac{1}{v}\left(x(v) - \max_{w<v}\left(1 - \frac{w}{v}\right)x(w)\right)dv.
\end{aligned}$$

Here the maximum of $x(v)$ is taken over all increasing functions from $[1, H]$ to $[0, 1]$. Consider the function $x(v) = \frac{\log v}{\log H}$. In this case, $(v - w)x(w)$ is maximized when:

$$\begin{aligned}
\frac{d}{dv}\left((v-w)x(w)\right) &= 0 \\
(v-w)x'(w) - x(w) &= 0 \\
(v-w)\frac{1}{w \log H} - \frac{\log w}{\log H} &= 0 \\
w + w \log w &= v.
\end{aligned}$$

If we choose w so that the above inequality holds, then note that $dv = (2 + \log w)dw$. It follows that

$$
\begin{aligned}
& \mathsf{MBRev}(\mathcal{D}_{ERC}(H)) \\
\geq\ & \frac{1}{\log H}\int_1^H \frac{1}{w+w\log w}\left(\log(w+w\log w)-\left(1-\frac{w}{w+w\log w}\right)\log w\right)(2+\log w)dw \\
=\ & \frac{1}{\log H}\int_1^H \frac{(2+\log w)}{w+w\log w}\left(\log(w+w\log w)-\log w+\frac{\log w}{1+\log w}\right)dw \\
\geq\ & \frac{1}{\log H}\int_1^H \frac{(2+\log w)}{w+w\log w}\log(1+\log w)dw \\
\geq\ & \frac{1}{\log H}\int_1^H \frac{\log(1+\log w)}{w}dw \\
=\ & \frac{\log(H)\log(1+\log H)-\log(1+\log H)-\log H}{\log H} \\
=\ & \Omega(\log\log H)
\end{aligned}
$$

□

Theorem A.3.12. $\mathsf{MBRev}(\mathcal{D}_{ERC}(H)) \leq O(\log\log H)$.

Proof. Note that, up to a point mass at H which contributes at most $H(1/H) = 1$ to the mean-based revenue, $\mathsf{MBRev}(\mathcal{D}_{ERC}(H))$ is given by

$$\max_{x(v)} \int_1^H \frac{1}{v}\left(x(v) - \max_{w<v}\left(1-\frac{w}{v}\right)x(w)\right)dv.$$

Let $f(v) : [1,\infty) \to [1,\infty)$ be a function that satisfies $f(v) < v$ for all $v \in [1,\infty)$. By choosing $w = f(v)$, we have that

$$
\begin{aligned}
& \mathsf{MBRev}(\mathcal{D}_{ERC}(H)) \\
\leq\ & \max_{x(v)} \left(\int_1^H \frac{1}{v} \left(x(v) - \left(1 - \frac{f(v)}{v} \right) x(f(v)) \right) dv \right) \\
=\ & \max_{x(v)} \left(\int_1^H \frac{x(v)}{v} dv - \int_1^H \frac{1}{v} \left(1 - \frac{f(v)}{v} \right) x(f(v)) dv \right) \\
=\ & \max_{x(v)} \left(\int_{f(H)}^H \frac{x(v)}{v} dv + \int_1^{f(H)} \frac{x(v)}{v} dv - \int_1^H \left(\frac{1}{v} - \frac{f(v)}{v^2} \right) x(f(v)) dv \right) \\
=\ & \max_{x(v)} \left(\int_{f(H)}^H \frac{x(v)}{v} dv + \int_1^H \frac{x(f(v))f'(v)}{f(v)} dv - \int_1^H \left(\frac{1}{v} - \frac{f(v)}{v^2} \right) x(f(v)) dv \right) \\
=\ & \max_{x(v)} \left(\int_{f(H)}^H \frac{x(v)}{v} dv + \int_1^H \left(\frac{f'(v)}{f(v)} + \frac{f(v)}{v^2} - \frac{1}{v} \right) x(f(v)) dv \right).
\end{aligned}
$$

Choose $f(v) = \frac{v}{1+\log v}$. Note that, for this choice of f,

$$
f'(v) = \frac{\log v}{(1+\log v)^2},
$$

and so

$$
\begin{aligned}
\frac{f'(v)}{f(v)} + \frac{f(v)}{v^2} - \frac{1}{v} &= \frac{\log v}{v(1+\log v)} + \frac{1}{v(1+\log v)} - \frac{1}{v} \\
&= 0.
\end{aligned}
$$

It follows that (since $x(v) \in [0,1]$ for all v)

$$\begin{aligned}
\mathsf{MBRev}(\mathcal{D}_{ERC}(H)) &\leq \max_{x(v)} \int_{f(H)}^{H} \frac{x(v)}{v} dv \\
&\leq \int_{H/(\log H+1)}^{H} \frac{dv}{v} \\
&= \log(\log H + 1) \\
&= O(\log \log H).
\end{aligned}$$

□

Corollary A.3.13 (Restatement of Theorem 2.3.8)**.** *The gap* $\mathsf{MBRev}(\mathcal{D})/\mathsf{Mye}(\mathcal{D})$ *can grow arbitrarily large. For distributions* $\mathcal{D}$ *supported on* $[1, H]$*, this gap can be as large as* $\Omega(\log \log H)$ *(and this is tight). Similarly, the gap* $\mathsf{Val}(\mathcal{D})/\mathsf{MBRev}(\mathcal{D})$ *can grow arbitrarily large. For distributions* $\mathcal{D}$ *supported on* $[1, H]$*, this gap can be as large as* $\Omega(\log H / \log \log H)$.

A.4 Mean-based learning algorithms

In this appendix we will show that Multiplicative Weights and EXP3 - the most common adversarial no-regret algorithms for the experts and bandits case respectively - are mean-based, as per Definition 2.2.1. We expect that many variants of these algorithms along with other no-regret learning algorithms are also mean-based, and can be shown to be mean-based via similar methods of proof.

We begin by showing that Multiplicative Weights (Algorithm 17) is mean-based. Multiplicative Weights, also known as Hedge (see survey [12] for more details) is a simple no-regret learning algorithm for the full-information setting. It proceeds by maintaining a weight w_i for each option. Every round, Multiplicative Weights chooses an option with probability proportional to w_i, and then updates each weight w_i by

multiplying it by $e^{\varepsilon r_i}$, where ε is a parameter of the algorithm and r_i is the reward from option i this round.

Algorithm 17 Multiplicative Weights algorithm.

1: Choose $\varepsilon = \sqrt{\frac{\log K}{T}}$. Initialize K weights, letting $w_{i,t}$ be the value of the ith weight at round t. Initially, set all $w_{i,0} = 1$.
2: **for** $t = 1$ to T **do**
3: Choose option i with probability $p_{i,t} = w_{i,t-1} / \sum_j w_{j,t-1}$.
4: **for** $j = 1$ to K **do**
5: Set $w_{j,t} = w_{j,t-1} \cdot e^{\varepsilon r_{j,t}}$.
6: **end for**
7: **end for**

Theorem A.4.1. *The Multiplicative Weights algorithm (Algorithm 17) is mean-based.*

Proof. Define $\gamma = 2(T\varepsilon)^{-1} \log(T\varepsilon)$. We will show that Multiplicative Weights is γ-mean-based. Note that since $\varepsilon = \sqrt{\frac{\log K}{T}}$, $\gamma = o(1)$ and therefore Multiplicative Weights is mean-based.

Note that $w_{i,t} = e^{\varepsilon \sigma_{i,t}}$. Therefore, if $\sigma_{i,t} - \sigma_{j,t} < -\gamma T$, we have $\sigma_{i,t-1} - \sigma_{j,t-1} < -\gamma T + 1 < -\gamma T/2$, it follows that

$$\begin{aligned} p_{i,t} &= \frac{w_{i,t-1}}{\sum_j w_{j,t-1}} \\ &\leq \frac{w_{i,t-1}}{w_{j,t-1}} \\ &= e^{\varepsilon(\sigma_{i,t-1} - \sigma_{j,t-1})} \\ &< e^{-\varepsilon\gamma T/2} \\ &= e^{-\log(T\varepsilon)} = 1/(T\varepsilon) \leq \gamma. \end{aligned}$$

It follows that Multiplicative Weights is γ-mean-based. □

We now show the Follow-the-Perturbed-Leader algorithm (Algorithm 18) is mean-based.

Algorithm 18 Follow-the-Perturbed-Leader algorithm.

1: Choose $\varepsilon = \sqrt{\frac{\log K}{T}}$.
2: **for** $t = 1$ to T **do**
3: For each arm, sample $per_i \geq 0$ independently from exp. distribution $d\mu(x) = \varepsilon e^{-\varepsilon x}$.
4: Choose option i with largest $\sigma_{i,t-1} + per_i$.
5: **end for**

Theorem A.4.2. *The Follow-the-Perturbed-Leader algorithm (Algorithm 18) is mean-based.*

Proof. Let $\gamma = \sqrt{\frac{1}{T}} \cdot \log(T)$. When $\sigma_{i,t} < \sigma_{j,t} - \gamma T$, the probability option i is chosen at round i is at most

$$Pr[per_i > \sigma_{i,t-1} - \sigma_{j,t-1}] = e^{-\varepsilon(\sigma_{i,t-1} - \sigma_{j,t-1})} \leq e^{-\varepsilon\gamma T/2} < \sqrt{\frac{1}{T}} < \gamma.$$

Therefore the Follow-the-Perturbed-Leader algorithm (Algorithm 18) is γ-mean-based. □

We will now show that EXP3 (Algorithm 19) is mean-based. EXP3 can be thought of as an extension of Multiplicative Weights to the incomplete information (bandits) setting. Since we no longer observe every option's reward each round, we cannot perform the same weight update rule as in Multiplicative Weights. Instead, if we choose option i, we update weight w_i by multiplying it with $e^{\varepsilon r_i/p_i}$, where p_i is the probability of picking this option this round (i.e. $w_i/\sum w_j$), and leave all other weights unmodified. Since $\mathbb{E}[\frac{r_{i,t}}{p_{i,t}} \mathbb{1}_{I_t=i}] = r_{i,t}$, this accomplishes in expectation (in some sense) the same update rule as Multiplicative Weights. It is known that (for fixed K) if $\varepsilon = T^{-\alpha}$ for some $\alpha \in (0, 1)$, then EXP3 is no-regret ([16]). This regret is minimized when $\alpha = 1/2$, but for convenience of analysis we will show that EXP3 is mean-based when $\alpha = 1/4$. EXP3 is still no-regret when $\alpha = 1/4$.

Theorem A.4.3. *The EXP3 algorithm (Algorithm 19) is mean-based.*

Algorithm 19 EXP3 algorithm.

1: Choose a parameter $\varepsilon \in (0,1)$. Initialize K weights, letting $w_{i,t}$ be the value of the ith weight at round t. Initially, set all $w_{i,0} = 1$.
2: **for** $t = 1$ to T **do**
3: Choose option i with probability $p_{i,t} = (1 - K\varepsilon)\frac{w_{i,t-1}}{\sum_j w_{j,t-1}} + \varepsilon$.
4: Set $w_{i,t} = w_{i,t-1} \cdot e^{\varepsilon r_{i,t}/p_{i,t}}$.
5: **end for**

Proof. We will set $\varepsilon = T^{-1/4}$ and $\gamma = 2(2\sqrt{2}+1)T^{-1/4}\log T$. We will show that EXP3 is γ-mean-based.

Define $\hat{\sigma}_{i,t} = \sum_{s=1}^{t} \frac{r_{i,s}}{p_{i,s}} \cdot \mathbb{1}_{I_s=i}$. Note that $\hat{\sigma}_{i,t} - \sigma_{i,t}$ is a martingale in t; indeed, conditioned on the actions from time 1 up to time $t-1$, $\mathbb{E}\left[\frac{r_{i,s}}{p_{i,s}} \cdot \mathbb{1}_{I_s=i}\right] = r_{i,s}$. In addition, note that $\left|\frac{\varepsilon r_{i,s}}{p_{i,s}} \cdot \mathbb{1}_{I_s=i} - \varepsilon r_{i,s}\right| \leq \frac{1}{p_{i,s}} \leq 1/\varepsilon$, since $p_{i,s} \geq \varepsilon$ by definition. It follows from Azuma's inequality that, for any $1 \leq i \leq K$, $1 \leq t \leq T$, and $M > 0$,

$$\Pr\left[|\hat{\sigma}_{i,t} - \sigma_{i,t}| \geq M\right] \leq 2\exp\left(-\frac{M^2\varepsilon^2}{2T}\right).$$

We will choose M so that $M\varepsilon = \sqrt{2T\log T}$; for this M, it follows that

$$\Pr\left[|\hat{\sigma}_{i,t} - \sigma_{i,t}| \geq M\right] \leq \frac{2}{T}.$$

Now, note that $w_{i,t} = e^{\varepsilon\hat{\sigma}_{i,t}}$. If $\sigma_{i,t} - \sigma_{j,t} < -\gamma T$, we have $\sigma_{i,t-1} - \sigma_{j,t-1} < -\gamma T + 1 < -\gamma T/2$, it then follows that

$$\begin{aligned}
p_{i,t} &= (1 - K\varepsilon)\frac{w_{i,t-1}}{\sum_j w_{j,t-1}} + \varepsilon \\
&\leq \min\left(\frac{w_{i,t-1}}{w_{j,t-1}}, 1\right) + \varepsilon \\
&= \min(e^{\varepsilon(\hat{\sigma}_{i,t-1} - \hat{\sigma}_{j,t-1})}, 1) + \varepsilon \\
&\leq e^{\varepsilon(\sigma_{i,t-1} - \sigma_{j,t-1}) + 2M\varepsilon} + \frac{2}{T} + \varepsilon \\
&< e^{-\varepsilon\gamma T/2 + 2\sqrt{2T\log T}} + \frac{2}{T} + \varepsilon \\
&\leq e^{-\sqrt{T}\log T} + \frac{2}{T} + T^{-1/4} \\
&\leq \gamma.
\end{aligned}$$

□

Finally, we prove Theorem 2.2.3, showing that the contextualization of a mean-based algorithm is still mean-based. In particular, the contextualizations of the above three algorithms (Multiplicative Weights, Follow the Perturbed Leader, and EXP3) are all mean-based algorithms for the contextual bandits problem.

Theorem A.4.4 (Restatement of Theorem 2.2.3)**.** *If an algorithm for the experts problem or multi-armed bandits problem is mean-based, then its contextualization is also a mean-based algorithm for the contextual bandits problem.*

Proof. Assume M is a γ-mean-based algorithm. We will show M' is $\frac{1}{\min_c \Pr[c]}\left(\gamma + \frac{2\sqrt{\log(mKT)}}{T^{1/2}}\right)$-mean-based.

First define $\hat{\sigma}_{i,t}(c) = \sum_{s:s\leq t,\, c_s = c} r_{i,s}(c)$ to be the total reward given by arm i on rounds where the context is c. Since M is γ-mean-based, whenever $\hat{\sigma}_{i,t}(c) < \hat{\sigma}_{j,t}(c) - \gamma T$, then the probability $p_{i,t}(c)$ that the algorithm pulls arm i on round t if it has context c satisfies $p_{i,t}(c) < \gamma$.

We will proceed to show that $\hat{\sigma}_{i,t}(c) < \hat{\sigma}_{j,t}(c) - \gamma T$ with sufficiently large probability. It is easy to check that $\mathbb{E}[\hat{\sigma}_{i,t}(c)] = \sigma_{i,t}(c) \cdot \Pr[c]$. By the Chernoff bound, we have that

$$\Pr\left[|\hat{\sigma}_{i,t}(c) - \sigma_{i,t}(c) \cdot \Pr[c]| \geq \sqrt{T \log(mKT)}\right] \leq 2\exp(-2T\log(mKT)/t) \leq \frac{2}{T^2 m^2 K^2}.$$

By the union bound, with probability at least $\frac{2}{Tm^2K^2}$, we have $|\hat{\sigma}_{i,t}(c) - \sigma_{i,t}(c) \cdot \Pr[c]| \geq \sqrt{T\log(mKT)}$ for all i,t, and c. In this case we have that $\sigma_{i,t}(c) < \sigma_{j,t}(c) - \frac{1}{\Pr[c]}(\gamma T + 2\sqrt{T\log(mKT)})$ implies that $\hat{\sigma}_{i,t}(c) < \hat{\sigma}_{j,t}(c) - \gamma T$.

Therefore, if $\sigma_{i,t}(c) < \sigma_{j,t}(c) - \frac{1}{\Pr[c]}(\gamma T + 2\sqrt{T\log(mKT)})$ and the context of round t is c, then $p_{i,t}(c) < \gamma + \frac{2}{Tm^2K^2} \leq (\frac{1}{\min_c \Pr[c]}(\gamma + \frac{1}{T^{1/2}}))$. □

Appendix B

Appendix for Chapter 3

B.1 Negative Results

In this section, we show that algorithms that achieve low-regret in the multi-armed bandits problem with adversarial values perform poorly in the strategic multi-armed bandits problem. Throughout this section, we will assume we are working in the restricted payment model (i.e., arms can only pass along a value $w_{i,t}$ that is at most $v_{i,t}$), but all proofs also work in the unrestricted payment model (and in fact are much easier there).

B.1.1 Tacit Observational Model

We begin by showing that in the tacit observational model, where arms cannot see the amounts passed on by other arms, it is still possible for the arms to collude and leave the principal with $o(T)$ revenue.

We begin by proving this result for the case of two arms, where the proof is slightly simpler.

Theorem B.1.1. *Let mechanism M be a (ρ, δ)-low regret algorithm for the multi-armed bandit problem with two arms, where $\rho \leq T^{-2}$ and $\delta \geq \sqrt{T \log T}$. Then in*

the strategic multi-armed bandit problem under the tacit observational model, there exist distributions D_1, D_2 and an $O(\sqrt{T\delta})$-Nash Equilibrium where a principal using mechanism M gets at most $O(\sqrt{T\delta})$ revenue.

Proof. Let D_1 and D_2 be distributions with means μ_1 and μ_2 respectively, such that $|\mu_1 - \mu_2| \leq \max(\mu_1, \mu_2)/2$. Additionally, assume both D_1 and D_2 are supported on $[\sqrt{\delta/T}, 1]$. We now describe the equilibrium strategy S^* (the below description is for arm 1; S^* for arm 2 is symmetric):

1. Set parameters $B = 6\sqrt{T\delta}$ and $\theta = \sqrt{\frac{\delta}{T}}$.
2. Define $c_{1,t}$ to be the number times arm 1 is pulled in rounds $1, ..., t$. Similarly define $c_{2,t}$ to be the number times arm 2 is pulled in rounds $1, ..., t$.
3. For $t = 1, \ldots, T$:
   (a) If there exists a $t' \leq t-1$ such that $c_{1,t'} < c_{2,t'} - B$, set $w_{1,t} = v_{1,t}$.
   (b) If the condition in (a) is not true, let $p_{1,t}$ be the probability that the principal will pick arm 1 in this round conditioned on the history (assuming player 2 is also playing S^*), and let $p_{2,t} = 1 - p_{1,t}$. Then:
      i. If $c_{1,t-1} < c_{2,t-1}$ and $p_{1,t} < p_{2,t}$, set $w_{1,t} = \theta$.
      ii. Otherwise, set $w_{1,t} = 0$.

We will now show that (S^*, S^*) is an $O(\sqrt{T\delta})$-Nash equilibrium. To do this, for any deviating strategy S', we will both lower bound $u_1(M, S^*, S^*)$ and upper bound $u_1(M, S', S^*)$, hence bounding the net utility of deviation.

We begin by proving that $u_1(M, S^*, S^*) \geq \frac{\mu_2 T}{2} - O(\sqrt{T\delta})$. We need the following lemma.

Lemma B.1.2. *If both arms are using strategy S^*, then with probability $\left(1 - \frac{4}{T}\right)$, $|c_{1,t} - c_{2,t}| \leq B$ for all $t \in [T]$.*

Proof. Assume that both arms are playing the strategy S^* with the modification that they never defect (i.e. condition (a) in the above strategy is removed). This does not change the probability that $|c_{1,t} - c_{2,t}| \leq B$ for all $t \in [T]$.

Define $R_{1,t} = \sum_{s=1}^{t} w_{1,s} - \sum_{s=1}^{t} w_{I_s,s}$ be the regret the principal experiences for not playing only arm 1. Define $R_{2,t}$ similarly. We will begin by showing that with high probability, these regrets are bounded both above and below. In particular, we will show that with probability at least $1 - \frac{2}{T}$, $R_{i,t}$ lies in $[-2\theta\sqrt{T \log T} - \delta, \delta]$ for all $t \in [T]$ and $i \in \{1, 2\}$.

To do this, note that there are two cases where the regrets $R_{1,t}$ and $R_{2,t}$ can possibly change. The first is when $p_{1,t} > p_{2,t}$ and $c_{1,t} > c_{2,t}$. In this case, the arms offer $(w_{1,t}, w_{2,t}) = (0, \theta)$. With probability $p_{1,t}$ the principal chooses arm 1 and the regrets update to $(R_{1,t+1}, R_{2,t+1}) = (R_{1,t}, R_{2,t} + \theta)$, and with probability $p_{2,t}$ the principal chooses arm 2 and the regrets update to $(R_{1,t+1}, R_{2,t+1}) = (R_{1,t} - \theta, R_{2,t})$. It follows that $\mathbb{E}[R_{1,t+1} + R_{2,t+1} | R_{1,t} + R_{2,t}] = R_{1,t} + R_{2,t} + (p_{1,t} - p_{2,t})\theta \geq R_{1,t} + R_{2,t}$.

In the second case, $p_{1,t} < p_{2,t}$ and $c_{2,t} < c_{1,t}$, and a similar calculation shows again that $\mathbb{E}[R_{1,t+1} + R_{2,t+1} | R_{1,t} + R_{2,t}] = R_{1,t} + R_{2,t} + (p_{2,t} - p_{1,t})\theta \geq R_{1,t} + R_{2,t}$. It follows that $R_{1,t} + R_{2,t}$ forms a submartingale.

From the above analysis, it is also clear that $|(R_{1,t+1} + R_{2,t+1}) - (R_{1,t} + R_{2,t})| \leq \theta$. It follows from Azuma's inequality that, for any fixed $t \in [T]$,

$$\Pr\left[R_{1,t} + R_{2,t} \leq -2\theta\sqrt{T \log T}\right] \leq \frac{1}{T^2}$$

Applying the union bound, with probability at least $1 - \frac{1}{T}$, $R_{1,t} + R_{2,t} \geq -2\theta\sqrt{T \log T}$ for all $t \in [T]$. Furthermore, since the principal is using a (T^{-2}, δ)-low-regret algorithm, it is also true that with probability at least $1 - T^{-2}$ (for any fixed t) both $R_{1,t}$ and $R_{2,t}$ are at most δ. Applying the union bound again, it is true that $R_{1,t} \leq \delta$ and $R_{2,t} \leq \delta$ for all t with probability at least $1 - \frac{1}{T}$. Finally, combining this with the earlier inequality (and applying union bound once more), with probability

at least $1 - \frac{2}{T}$, $R_{i,t} \in \left[-2\theta\sqrt{T \log T} - \delta, \delta\right]$, as desired. For the remainder of the proof, condition on this being true.

We next proceed to bound the probability that (for a fixed t) $c_{1,t} - c_{2,t} \leq B$. Define the random variable τ to be the largest value $s \leq t$ such that $c_{1,\tau} - c_{2,\tau} = 0$ – note that if $c_{1,t} - c_{2,t} \geq 0$, then $c_{1,s} - c_{2,s} \geq 0$ for all s in the range $[\tau, t]$. Additionally let Δ_s denote the ± 1 random variable given by the difference $(c_{1,s} - c_{2,s}) - (c_{1,s-1} - c_{2,s-1})$. We can then write

$$\begin{aligned} c_{1,t} - c_{2,t} &\leq \sum_{s=\tau+1}^{t} \Delta_s \\ &\leq \sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} > p_{2,s}} + \sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} \leq p_{2,s}} \end{aligned}$$

Here the first summand corresponds to times s where one of the arms offers θ (and hence the regrets change), and the second summand corresponds to times where both arms offer 0. Note that since $c_{1,s} \geq c_{2,s}$ in this interval, the regret $R_{2,s}$ increases by θ whenever $\Delta_s = 1$ (i.e., arm 1 is chosen), and furthermore no choice of arm can decrease $R_{2,s}$ in this interval. Since we know that $R_{2,s}$ lies in the interval $\left[-2\theta\sqrt{T \log T} - \delta, \delta\right]$ for all s, this bounds the first sum by

$$\sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} > p_{2,s}} \leq \frac{2\delta + 2\theta\sqrt{T \log T}}{\theta} = \frac{2\delta}{\theta} + 2\sqrt{T \log T}$$

On the other hand, when $p_{1,s} \leq p_{2,s}$, then $\mathbb{E}[\Delta_s] = p_{1,s} - p_{2,s} \leq 0$. By Hoeffding's inequality, it then follows that with probability at least $1 - \frac{1}{T^2}$,

$$\sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} \leq p_{2,s}} \leq 2\sqrt{T \log T}$$

Altogether, this shows that with probability at least $1 - \frac{1}{T^2}$,

$$c_{1,t} - c_{2,t} \leq \frac{2\delta}{\theta} + 4\sqrt{T \log T} \leq 6\sqrt{T\delta} = B$$

The above inequality therefore holds for all t with probability at least $1 - \frac{1}{T}$. Likewise, we can show that $c_{2,t} - c_{1,t} \leq B$ also holds for all t with probability at least $1 - \frac{1}{T}$. Since we are conditioned on the regrets $R_{i,t}$ being bounded (which is true with probability at least $\frac{2}{T}$), it follows that $|c_{1,t} - c_{2,t}| \leq B$ for all t with probability at least $1 - \frac{4}{T}$.

□

By Lemma B.1.2, we know that with probability $1 - \frac{4}{T}$, $|c_{1,t} - c_{2,t}| \leq B$ throughout the mechanism. In this case, arm 1 never uses step (a), and $c_{1,T} \geq (T - B)/2$. Therefore

$$\begin{aligned} u_1(M, S^*, S^*) &\geq \left(1 - \frac{4}{T}\right) \cdot (\mu_1 - \theta) \cdot (T - B)/2 \\ &\geq \frac{\mu_1 T}{2}\left(1 - \frac{4}{T} - \frac{\theta}{\mu_1} - \frac{B}{T}\right) \\ &= \frac{\mu_1 T}{2} - 2\mu_1 - \frac{\theta T}{2} - \frac{B\mu_1}{2} \\ &\geq \frac{\mu_1 T}{2} - O(\sqrt{T\delta}). \end{aligned}$$

Now we will show that $u_1(M, S', S^*) \leq \frac{\mu_1 T}{2} + O(\sqrt{T\delta})$. Without loss of generality, we can assume S' is deterministic. Let M_R be the deterministic mechanism when M's randomness is fixed to some outcome R. Consider the situation when arm 1 is using strategy S', arm 2 is using strategy S^* and the principal is using mechanism M_R. There are two cases:

1. $c_{1,t} - c_{2,t} \leq B$ is true for all $t \in [T]$. In this case, we have

$$u_1(M_R, S', S^*) \leq c_{1,T} \cdot \mu_1 \leq \mu_1(T + B)/2.$$

2. There exists some t such that $c_{1,t} - c_{2,t} > B$: Let $\tau_R + 1$ be the smallest t such that $c_{1,t} - c_{2,t} > B$. We know that $c_{1,\tau_R} - c_{2,\tau_R} \leq B$. Therefore we have

$$\begin{aligned}
& u_1(M_R, S', S^*) \\
= & \sum_{t=1}^{T} (\mu_1 - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
= & \sum_{t=1}^{T} (\mu_1 - w_{2,t}) \cdot \mathbb{1}_{I_t=1} + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq & \; c_{1,\tau_R}\mu_1 + \mu_1 + (T - \tau_R - 1)\max(\mu_1 - \mu_2, 0) + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq & \; \mu_1(\tau_R + B)/2 + \mu_1 + (T - \tau_R - 1)(\mu_1/2) + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq & \; \mu_1 T/2 + \mu_1(B+1)/2 + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}.
\end{aligned}$$

In general, we thus have that

$$u_1(M_R, S', S^*) \leq \mu_1 T/2 + \mu_1(B+1)/2 + \max\left(0, \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right).$$

Therefore

$$\begin{aligned}
u_1(M, S', S^*) &= \mathbb{E}_R[u_1(M_R, S', S^*)] \\
&\leq \mu_1 T/2 + \mu_1(B+1)/2 + \mathbb{E}_R\left[\max\left(0, \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right)\right].
\end{aligned}$$

Notice that $\sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}$ is the regret of not playing arm 2 (i.e., R_2 in the proof of Lemma B.1.2). Since the mechanism M is (ρ, δ) low regret, with probability $1 - \rho$, this sum is at most δ (and in the worst case, it is bounded above by $T\mu_2$). We therefore have that:

$$\begin{aligned} u_1(M, S', S^*) &\leq \frac{\mu_1 T}{2} + \frac{\mu_1(B+1)}{2} + \delta + \rho T \mu_2 \\ &\leq \frac{\mu_1 T}{2} + O(\sqrt{T\delta}) \end{aligned}$$

From this and our earlier lower bound on $u_1(M, S^*, S^*)$, it follows that $u_1(M, S', S^*) - u_1(M, S^*, S^*) \leq O(\sqrt{T\delta})$, thus establishing that (S^*, S^*) is an $O(\sqrt{T\delta})$-Nash equilibrium for the arms.

Finally, to bound the revenue of the principal, note that if the arms both play according to S^* and $|c_{1,t} - c_{2,t}| \leq B$ for all t (so they do not defect), the principal gets a maximum of $T\theta = O(\sqrt{T\delta})$ revenue overall. Since (by Lemma B.1.2) this happens with probability at least $1 - \frac{4}{T}$ (and the total amount of revenue the principal is bounded above by T), it follows that the total expected revenue of the principal is at most $O(\sqrt{T\delta})$.

□

We now extend this proof to the K arm case, where K can be as large as $T^{1/3}/\log(T)$.

Theorem B.1.3. *Let mechanism M be a (ρ, δ)-low regret algorithm for the multi-armed bandit problem with K arms, where $K \leq T^{1/3}/\log(T)$, $\rho \leq T^{-2}$, and $\delta \geq \sqrt{T \log T}$. Then in the strategic multi-armed bandit problem under the tacit observational model, there exist distributions D_i and an $O(\sqrt{KT\delta})$-Nash Equilibrium for the arms where the principal gets at most $O(\sqrt{KT\delta})$ revenue.*

Proof of Theorem B.1.3. As in the proof of Theorem B.1.1, let μ_i denote the mean value of the ith arm's distribution D_i (supported on $[\sqrt{K\delta/T}, 1]$). Without loss of generality, further assume that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$. We will show that as long as $\mu_1 - \mu_2 \leq \frac{\mu_1}{K}$, there exists some $O(\sqrt{KT\delta})$-Nash equilibrium for the arms where the principal gets at most $O(\sqrt{KT\delta})$ revenue.

We begin by describing the equilibrium strategy S^* for the arms. Let $c_{i,t}$ denote the number of times arm i has been pulled up to time t. As before, set $B = 7\sqrt{KT\delta}$ and set $\theta = \sqrt{\frac{K\delta}{T}}$. The equilibrium strategy for arm i at time t is as follows:

1. If at any time $s \leq t$ in the past, there exists an arm j with $c_{j,s} - c_{i,s} \geq B$, defect and offer your full value $w_{i,t} = \mu_i$.

2. Compute the probability $p_{i,t}$, the probability that the principal will pull arm i conditioned on the history so far.

3. Offer $w_{i,t} = \theta(1 - p_{i,t})$.

We begin, as before, by showing that if all parties follow this strategy, then with high probability no one will ever defect.

Lemma B.1.4. *If all arms are using strategy S^*, then with probability $\left(1 - \frac{3}{T}\right)$, $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$.*

Proof. As before, assume that all arms are playing the strategy S^* with the modification that they never defect. This does not change the probability that $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$.

Define $R_{i,t} = \sum_{s=1}^{t} w_{i,s} - \sum_{s=1}^{t} w_{I_s,s}$ be the regret the principal experiences for not playing only arm i up until time t. We begin by showing that with probability at least $1 - \frac{2}{T}$, $R_{i,t}$ lies in $[-K\theta\sqrt{T \log T} - (K-1)\delta, \delta]$ for all $t \in [T]$ and $i \in [K]$.

To do this, first note that since the principal is using a (T^{-2}, δ)-low-regret algorithm, with probability at least $1 - T^{-2}$ the regrets $R_{i,t}$ are all upper bounded by δ at any fixed time t. Via the union bound, it follows that $R_{i,t} \leq \delta$ for all i and t with probability at least $1 - \frac{1}{T}$.

To lower bound $R_{i,t}$, we will first show that $\sum_{i=1}^{K} R_{i,t}$ is a submartingale in t. Note that, with probability $p_{j,t}$, $R_{i,t+1}$ will equal $R_{i,t} + \theta((1 - p_{j,t}) - (1 - p_{i,t}))$. We then have

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{K} R_{i,t+1} \middle| \sum_{i=1}^{K} R_{i,t}\right] &= \sum_{i=1}^{K} R_{i,t} + \sum_{i=1}^{K} p_{i,t} \sum_{j=1}^{K} \theta((1-p_{j,t}) - (1-p_{i,t})) \\
&= \sum_{i=1}^{K} R_{i,t} + \sum_{i=1}^{K} p_{i,t} \sum_{j=1}^{K} \theta(p_{i,t} - p_{j,t}) \\
&= \sum_{i=1}^{K} R_{i,t} + \theta \sum_{i=1}^{K} p_{i,t}(Kp_{i,t} - 1) \\
&= \sum_{i=1}^{K} R_{i,t} + \theta \left(K \sum_{i=1}^{K} p_{i,t}^2 - \sum_{i=1}^{K} p_{i,t} \right) \\
&\geq \sum_{i=1}^{K} R_{i,t}
\end{aligned}$$

where the last inequality follows by Cauchy-Schwartz. It follows that $\sum_{i=1}^{K} R_{i,t}$ forms a submartingale.

Moreover, note that (since $|p_i - p_j| \leq 1$) $|R_{i,t+1} - R_{i,t}| \leq \theta$. It follows that $\left|\sum_{i=1}^{K} R_{i,t+1} - \sum_{i=1}^{K} R_{i,t}\right| \leq K\theta$ and therefore by Azuma's inequality that, for any fixed $t \in [T]$,

$$\Pr\left[\sum_{i=1}^{K} R_{i,t} \leq -2K\theta\sqrt{T \log T}\right] \leq \frac{1}{T^2}.$$

With probability $1 - \frac{1}{T}$, this holds for all $t \in [T]$. Since (with probability $1 - \frac{1}{T}$) $R_{i,t} \leq \delta$, this implies that with probability $1 - \frac{2}{T}$, $R_{i,t} \in \left[-2K\theta\sqrt{T \log T} - (K-1)\delta, \delta\right]$.

We next proceed to bound the probability that $c_{i,t} - c_{j,t} > B$ for a i, j, and t. Define

$$S_t^{(i,j)} = \left(c_{i,t} - c_{j,t} + \frac{1}{\theta}(R_{i,t} - R_{j,t})\right).$$

We claim that $S_t^{(i,j)}$ is a martingale. To see this, we first claim that $R_{i,t+1} - R_{j,t+1} = R_{i,t} - R_{j,t} - \theta(p_{i,t} - p_{j,t})$. Note that, if arm k is pulled, then $R_{i,t+1} = R_{i,t} + \theta((1 - p_{i,t}) - (1 - p_{k,t})) = R_{i,t} + \theta(p_{k,t} - p_{i,t})$ and similarly, $R_{j,t+1} = R_{j,t} + \theta(p_{k,t} - p_{j,t})$. It follows that $R_{i,t+1} - R_{j,t+1} = R_{i,t} - R_{j,t} - \theta(p_{i,t} - p_{j,t})$.

Secondly, note that (for any arm k) $\mathbb{E}[c_{k,t+1} - c_{k,t}|p_t] = p_{k,t}$, and thus $\mathbb{E}[c_{i,t+1} - c_{j,t+1} - (c_{i,t} - c_{j,t})|p_t] = p_{i,t} - p_{j,t}$. It follows that

$$\begin{aligned}
\mathbb{E}[S_{t+1}^{(i,j)} - S_t^{(i,j)}|p_t] &= \mathbb{E}[(c_{i,t+1} - c_{j,t+1}) - (c_{i,t} - c_{j,t})|p_t] \\
&\quad + \frac{1}{\theta}\mathbb{E}[(R_{i,t+1} - R_{j,t+1}) - (R_{i,t} - R_{j,t})|p_t] \\
&= (p_{i,t} - p_{j,t}) - (p_{i,t} - p_{j,t}) \\
&= 0
\end{aligned}$$

and thus that $\mathbb{E}[S_{t+1}^{(i,j)}|S_t^{(i,j)}] = S_t^{(i,j)}$, and thus that $S_t^{(i,j)}$ is a martingale. Finally, note that $|S_{t+1}^{(i,j)} - S_t^{(i,j)}| \leq 2$, so by Azuma's inequality

$$\Pr\left[S_t^{(i,j)} \geq 4\sqrt{T\log(TK)}\right] \leq (TK)^{-2}$$

Taking the union bound, we find that with probability at least $1 - \frac{1}{T}$, $S^{(i,j)} \leq 4\sqrt{T\log(TK)}$ for all i, j, and t. Finally, since with probability at least $1 - \frac{2}{T}$ each $R_{i,t}$ lies in $\left[-2K\theta\sqrt{T\log T} - (K-1)\delta, \delta\right]$, with probability at least $1 - \frac{3}{T}$ we have that (for all i, j, and t)

$$
\begin{aligned}
c_{i,t} - c_{j,t} &= S_t^{(i,j)} - \frac{1}{\theta}(R_{i,t} - R_{j,t}) \\
&\leq 4\sqrt{T\log(TK)} + \frac{1}{\theta}\left|R_{i,t} - R_{j,t}\right| \\
&\leq 4\sqrt{T\log(TK)} + 2K\sqrt{T\log T} + \frac{K\delta}{\theta} \\
&\leq \frac{7K\delta}{\theta} \\
&= 7K\sqrt{T\delta} \\
&= B
\end{aligned}
$$

□

By Lemma B.1.4, we know that with probability $1 - \frac{3}{T}$, $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$. In this case, arm 1 never defect, and $c_{1,T} \geq T/K - B$. Therefore

$$
\begin{aligned}
u_1(M, S^*, S^*) &\geq \left(1 - \frac{3}{T}\right) \cdot (\mu_1 - \theta) \cdot (T/K - B) \\
&\geq \frac{\mu_1 T}{K}\left(1 - \frac{3}{T} - \frac{\theta}{\mu_1} - \frac{BK}{T}\right) \\
&= \frac{\mu_1 T}{K} - 3\mu_1/K - \frac{\theta T}{K} - B\mu_1 \\
&\geq \frac{\mu_1 T}{K} - O(\sqrt{KT\delta})
\end{aligned}
$$

Now we are going to show that $u_1(M, S', S^*) \leq \frac{\mu_1 T}{K} + O(\sqrt{KT\delta})$. Without loss of generality, we can assume S' is deterministic. Let M_R be the deterministic mechanism when M's randomness is fixed to some outcome R. Consider the situation when arm 1 is using strategy S', arm 2 is using strategy S^* and the principal is using mechanism M_R. There are two cases:

1. $c_{i,t} - c_{j,t} \leq B$ is true for all $t \in [T]$ and $i, j \in [K]$. In this case, we have

$$u_1(M_R, S', S^*) \leq c_{1,T} \cdot \mu_1 \leq \mu_1(T + (K-1)B)/K.$$

2. There exists some $t \in [T]$ and $i, j \in [K]$ such that $c_{i,t} - c_{j,t} > B$: Let $\tau_R + 1$ be the smallest t such that $c_{i,t} - c_{j,t} > B$ for some $i, j \in [K]$. We know that $c_{1,\tau_R} - c_{i,\tau_R} \leq B$ for all $i \in [K]$. Therefore we have

$$\begin{aligned}
u_1(M_R, S', S^*) &= \sum_{t=1}^{T} (\mu_1 - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
&= \sum_{t=1}^{T} (\mu_1 - w_{2,t}) \cdot \mathbb{1}_{I_t=1} + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
&\leq c_{1,\tau_R}\mu_1 + \mu_1 + (T - \tau_R - 1)\max(\mu_1 - \mu_2, 0) + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
&\leq \mu_1(\tau_R + B)/K + \mu_1 + (T - \tau_R - 1)(\mu_1/K) + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
&\leq \mu_1 T/K + \mu_1(B+1)(K-1)/K + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}.
\end{aligned}$$

In M_R, we also have

$$\begin{aligned}
\sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} &= \sum_{t=1}^{T} (w_{2,t} - w_{I_t,t}) - \sum_{t=1}^{T} (w_{2,t} - w_{I_t,t}) \cdot \mathbb{1}_{I_t \neq 1} \\
&\leq \sum_{t=1}^{T} (w_{2,t} - w_{I_t,t}) + \sum_{t=1}^{\tau_R} w_{I_t,t} \cdot \mathbb{1}_{I_t \neq 1} - \sum_{t=\tau_R+1}^{T} (\mu_2 - \mu_{I_t}) \cdot \mathbb{1}_{I_t \neq 1} \\
&\leq \sum_{t=1}^{T} (w_{2,t} - w_{I_t,t}) + T(\theta + B/T) + 0.
\end{aligned}$$

In general, we thus have that

$$u_1(M_R, S', S^*) \leq \mu_1 T/K + \mu_1(B+1)(K-1)/K + \max\left(0, \sum_{t=1}^{T}(w_{2,t} - w_{I_t,t}) + T\theta + B\right).$$

Therefore

$$\begin{aligned} u_1(M, S', S^*) &= \mathbb{E}_R[u_1(M_R, S', S^*)] \\ &\leq \mu_1 T/K + \mu_1(B+1)(K-1)/K \\ &\quad + \mathbb{E}_R\left[\max\left(0, \sum_{t=1}^{T}(w_{2,t} - w_{I_t,t}) + T\theta + B\right)\right]. \end{aligned}$$

Notice that $\sum_{t=1}^{T}(w_{2,t} - w_{I_t,t})$ is the regret of not playing arm 2. Since the mechanism M is (ρ, δ) low regret, with probability $1-\rho$, this sum is at most δ (and in the worst case, it is bounded above by $T\mu_2$). We therefore have that:

$$\begin{aligned} u_1(M, S', S^*) &\leq \mu_1 T/K + \mu_1(B+1)(K-1)/K + \delta + \rho T\mu_+ T\theta + B \\ &\leq \frac{\mu_1 T}{K} + O(\sqrt{KT\delta}). \end{aligned}$$

From this and our earlier lower bound on $u_1(M, S^*, S^*)$, it follows that $u_1(M, S', S^*) - u_1(M, S^*, S^*) \leq O(\sqrt{KT\delta})$, thus establishing that (S^*, S^*) is an $O(\sqrt{KT\delta})$-Nash equilibrium for the arms.

Finally, to bound the revenue of the principal, note that if the arms both play according to S^* and $|c_{i,t} - c_{j,t}| \leq B$ for all $t \in [T], i, j \in [K]$ (so they do not defect), the principal gets a maximum of $T\theta = O(\sqrt{KT\delta})$ revenue overall. Since (by Lemma B.1.2) this happens with probability at least $1 - \frac{3}{T}$ (and the total amount of revenue the principal is bounded above by T), it follows that the total expected revenue of the principal is at most $O(\sqrt{KT\delta})$. □

While the theorems above merely claim that a bad set of distributions for the arms exists, note that the proofs above show it is possible to collude in a wide range of instances - in particular, any set of distributions which satisfy $\mu_1 - \mu_2 \leq \mu_1/K$. A natural question is whether we can extend the above results to show that it is possible to collude in any set of distributions.

One issue with the collusion strategies in the above proofs is that if $\mu_1 - \mu_2 > \mu_1/K$, then arm 1 will have an incentive to defect in any collusive strategy that plays all the arms evenly (arm 1 can report a bit over μ_2 per round, and make $\mu_1 - \mu_2$ every round instead of μ_1 every K rounds). One solution to this is to design a collusive strategy that plays some arms more than others in equilibrium (for example, playing arm 1 90% of the time). We show how to modify our result for two arms to achieve an arbitrary market partition and thus work over a broad set of distributions.

Theorem B.1.5. *Let mechanism M be a (ρ, δ)-low regret algorithm for the multi-armed bandit problem with two arms, where $\rho \leq T^{-2}$ and $\delta \geq \sqrt{T \log T}$. Then, in the strategic multi-armed bandit problem under the tacit observational model, for* any *distributions D_1, D_2 of values for the arms (supported on $[\sqrt{\delta/T}, 1]$), there exists an $O(\sqrt{T\delta})$-Nash Equilibrium for the arms where a principal using mechanism M gets at most $O(\sqrt{T\delta})$ revenue.*

Unfortunately, it as not as easy to modify the proof of Theorem B.1.3 to prove the same result for K arms. It is an interesting open question whether there exist collusive strategies for K arms that can achieve an arbitrary partition of the market.

Proof. Let D_1 and D_2 be distributions with means μ_1 and μ_2 respectively, and both distributions supported on $[\sqrt{\delta/T}, 1]$. We now describe the equilibrium strategy S^* (the below description is for arm 1; S^* for arm 2 is symmetric):

1. Set parameters $B = 6\sqrt{T\delta}/\mu_2$ and $\theta = \sqrt{\frac{\delta}{T}}$.

2. Define $c_{1,t}$ to be the number times arm 1 is pulled in rounds $1, ..., t$. Similarly define $c_{2,t}$ to be the number times arm 2 is pulled in rounds $1, ..., t$.

3. For $t = 1, ..., T$.

   (a) If there exists a $t' \leq t-1$ such that $c_{1,t'}/\mu_1 < c_{2,t'}/\mu_2 - B$, set $w_{1,t} = v_{1,t}$.

   (b) If the condition in (a) is not true, let $p_{1,t}$ be the probability that the principal will pick arm 1 in this round conditioned on the history (assuming player 2 is also playing S^*), and let $p_{2,t} = 1 - p_{1,t}$. Then:

      i. If $c_{1,t-1}/\mu_1 < c_{2,t-1}/\mu_2$ and $p_{1,t}/\mu_1 < p_{2,t}/\mu_2$, set $w_{1,t} = \theta$.

      ii. Otherwise, set $w_{1,t} = 0$.

We will now show that (S^*, S^*) is an $O(\sqrt{T\delta})$-Nash equilibrium. To do this, for any deviating strategy S', we will both lower bound $u_1(M, S^*, S^*)$ and upper bound $u_1(M, S', S^*)$, hence bounding the net utility of deviation.

We begin by proving that $u_1(M, S^*, S^*) \geq \frac{\mu_1^2 T}{\mu_1+\mu_2} - O(\sqrt{T\delta})$. We need the following lemma.

Lemma B.1.6. *If both arms are using strategy S^*, then with probability $\left(1 - \frac{4}{T}\right)$, $|c_{1,t}/\mu_1 - c_{2,t}/\mu_2| \leq B$ for all $t \in [T]$.*

Proof. Assume that both arms are playing the strategy S^* with the modification that they never defect (i.e. condition (a) in the above strategy is removed). This does not change the probability that $|c_{1,t}/\mu_1 - c_{2,t}/\mu_2| \leq B$ for all $t \in [T]$.

Define $R_{1,t} = \sum_{s=1}^{t} w_{1,s} - \sum_{s=1}^{t} w_{I_s,s}$ be the regret the principal experiences for not playing only arm 1. Define $R_{2,t}$ similarly. We will begin by showing that with high probability, these regrets are bounded both above and below. In particular, we will show that with probability at least $1 - \frac{2}{T}$, $R_{i,t}$ lies in $\left[-\frac{\mu_1}{\mu_2}(2\theta\sqrt{T \log T} + \delta), \delta\right]$ for all $t \in [T]$ and $i \in \{1, 2\}$.

To do this, note that there are two cases where the regrets $R_{1,t}$ and $R_{2,t}$ can possibly change. The first is when $p_{1,t}/\mu_1 > p_{2,t}/\mu_2$ and $c_{1,t}/\mu_1 > c_{2,t}/\mu_2$. In this case, the arms offer $(w_{1,t}, w_{2,t}) = (0, \theta)$. With probability $p_{1,t}$ the principal chooses arm 1 and the regrets update to $(R_{1,t+1}, R_{2,t+1}) = (R_{1,t}, R_{2,t} + \theta)$, and with probability $p_{2,t}$ the principal chooses arm 2 and the regrets update to $(R_{1,t+1}, R_{2,t+1}) = (R_{1,t} - \theta, R_{2,t})$. It follows that $\mathbb{E}[R_{1,t+1}/\mu_2 + R_{2,t+1}/\mu_1 | R_{1,t}/\mu_2 + R_{2,t}/\mu_1] = R_{1,t}/\mu_2 + R_{2,t}/\mu_1 + (p_{1,t}/\mu_1 - p_{2,t}/\mu_2)\theta \geq R_{1,t}/\mu_2 + R_{2,t}/\mu_1$.

In the second case, $p_{1,t}/\mu_1 < p_{2,t}/\mu_2$ and $c_{2,t}/\mu_1 < c_{1,t}/\mu_2$, and a similar calculation shows again that $\mathbb{E}[R_{1,t+1}/\mu_2 + R_{2,t+1}/\mu_1 | R_{1,t}/\mu_2 + R_{2,t}/\mu_1] = R_{1,t}/\mu_2 + R_{2,t}/\mu_1 + (p_{2,t}/\mu_2 - p_{1,t}/\mu_1)\theta \geq R_{1,t} + R_{2,t}$. It follows that $R_{1,t}/\mu_2 + R_{2,t}/\mu_1$ forms a submartingale.

From the above analysis, it is also clear that $|(R_{1,t+1}/\mu_2 + R_{2,t+1}/\mu_1) - (R_{1,t}/\mu_2 + R_{2,t}/\mu_1)| \leq \theta/\mu_2$. It follows from Azuma's inequality that, for any fixed $t \in [T]$,

$$\Pr\left[R_{1,t}/\mu_2 + R_{2,t}/\mu_1 \leq -\frac{2\theta}{\mu_2}\sqrt{T \log T}\right] \leq \frac{1}{T^2}$$

Applying the union bound, with probability at least $1 - \frac{1}{T}$, $R_{1,t}/\mu_2 + R_{2,t}/\mu_1 \geq -\frac{2\theta}{\mu_2}\sqrt{T \log T}$ for all $t \in [T]$. Furthermore, since the principal is using a (T^{-2}, δ)-low-regret algorithm, it is also true that with probability at least $1 - T^{-2}$ (for any fixed t) both $R_{1,t}$ and $R_{2,t}$ are at most δ. Applying the union bound again, it is true that $R_{1,t} \leq \delta$ and $R_{2,t} \leq \delta$ for all t with probability at least $1 - \frac{1}{T}$. Finally, combining this with the earlier inequality (and applying union bound once more), with probability at least $1 - \frac{2}{T}$, $R_{i,t} \in \left[-\frac{\mu_1}{\mu_2}(2\theta\sqrt{T \log T} + \delta), \delta\right]$, as desired. For the remainder of the proof, condition on this being true.

We next proceed to bound the probability that (for a fixed t) $c_{1,t}/\mu_1 - c_{2,t}/\mu_2 \leq B$. Define the random variable $\tau - 1$ to be the largest value $s \leq t$ such that $c_{1,\tau}/\mu_1 - c_{2,\tau}/\mu_2 \leq 0$ – note that if $c_{1,t}/\mu_1 - c_{2,t}/\mu_2 \geq 0$, then $c_{1,s}/\mu_1 - c_{2,s}/\mu_2 \geq 0$ for all s

in the range $[\tau, t]$. Additionally let Δ_s denote the ± 1 random variable given by the difference $(c_{1,s}/\mu_1 - c_{2,s}/\mu_2) - (c_{1,s-1}/\mu_1 - c_{2,s-1}/\mu_2)$. We can then write

$$\begin{aligned} c_{1,t}/\mu_1 - c_{2,t}/\mu_2 &\leq \sum_{s=\tau+1}^{t} \Delta_s \\ &\leq \sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s}/\mu_1 > p_{2,s}/\mu_2} + \sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s}/\mu_1 \leq p_{2,s}/\mu_2} \end{aligned}$$

Here the first summand corresponds to times s where one of the arms offers θ (and hence the regrets change), and the second summand corresponds to times where both arms offer 0. Note that since $c_{1,s}/\mu_1 \geq c_{2,s}/\mu_2$ in this interval, the regret $R_{2,s}$ increases by θ whenever $\Delta_s = 1/\mu_1$ (i.e., arm 1 is chosen), and furthermore no choice of arm can decrease $R_{2,s}$ in this interval. Since we know that $R_{2,s}$ lies in the interval $\left[-\frac{\mu_1}{\mu_2}(2\theta\sqrt{T\log T} + \delta), \delta\right]$ for all s, this bounds the first sum by

$$\sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} > p_{2,s}} \leq \frac{\delta + \frac{\mu_1}{\mu_2}(2\theta\sqrt{T\log T} + \delta)}{\theta} \cdot (1/\mu_1) = \frac{1}{\mu_2}\left(\frac{2\delta}{\theta} + 2\sqrt{T\log T}\right)$$

On the other hand, when $p_{1,s}/\mu_1 \leq p_{2,s}/\mu_2$, then $\mathbb{E}[\Delta_s] = p_{1,s}/\mu_1 - p_{2,s}/\mu_2 \leq 0$. By Hoeffding's inequality, it then follows that with probability at least $1 - \frac{1}{T^2}$,

$$\sum_{s=\tau+1}^{t} \Delta_s \cdot \mathbb{1}_{p_{1,s} \leq p_{2,s}} \leq \frac{2}{\mu_2}\sqrt{T\log T}$$

Altogether, this shows that with probability at least $1 - \frac{1}{T^2}$,

$$c_{1,t} - c_{2,t} \leq \frac{1}{\mu_2}\left(\frac{2\delta}{\theta} + 4\sqrt{T\log T}\right) \leq 6\sqrt{T\delta}/\mu_2 = B$$

The above inequality therefore holds for all t with probability at least $1 - \frac{1}{T}$. Likewise, we can show that $c_{2,t}/\mu_2 - c_{1,t}/\mu_1 \leq B$ also holds for all t with probability

at least $1 - \frac{1}{T}$. Since we are conditioned on the regrets $R_{i,t}$ being bounded (which is true with probability at least $\frac{2}{T}$), it follows that $|c_{1,t}/\mu_1 - c_{2,t}/\mu_2| \leq B$ for all t with probability at least $1 - \frac{4}{T}$.

□

By Lemma B.1.2, we know that with probability $1 - \frac{4}{T}$, $|c_{1,t}/\mu_1 - c_{2,t}/\mu_2| \leq B$ throughout the mechanism. In this case, arm 1 never uses step (a), and $c_{1,T} \geq \frac{\mu_1}{\mu_1+\mu_2}T - \frac{\mu_1\mu_2}{\mu_1+\mu_2}B$. Therefore

$$\begin{aligned} u_1(M, S^*, S^*) &\geq \left(1 - \frac{4}{T}\right) \cdot (\mu_1 - \theta) \cdot \left(\frac{\mu_1}{\mu_1 + \mu_2}T - \frac{\mu_1\mu_2}{\mu_1 + \mu_2}B\right) \\ &\geq \frac{\mu_1^2 T}{\mu_1 + \mu_2} - O(\sqrt{T\delta}) \end{aligned}$$

Now we will show that $u_1(M, S', S^*) \leq \frac{\mu_1^2 T}{\mu_1+\mu_2} + O(\sqrt{T\delta})$. Without loss of generality, we can assume S' is deterministic. Let M_R be the deterministic mechanism when M's randomness is fixed to some outcome R. Consider the situation when arm 1 is using strategy S', arm 2 is using strategy S^* and the principal is using mechanism M_R. There are two cases:

1. $c_{1,t}/\mu_1 - c_{2,t}/\mu_2 \leq B$ is true for all $t \in [T]$. In this case, we have

$$u_1(M_R, S', S^*) \leq c_{1,T} \cdot \mu_1 \leq \frac{\mu_1}{\mu_1 + \mu_2}T + \frac{\mu_1\mu_2}{\mu_1 + \mu_2}B.$$

2. There exists some t such that $c_{1,t}/\mu_1 - c_{2,t}/\mu_2 > B$: Let $\tau_R + 1$ be the smallest t such that $c_{1,t}/\mu_1 - c_{2,t}/\mu_2 > B$. We know that $c_{1,\tau_R}/\mu_1 - c_{2,\tau_R}/\mu_2 \leq B$.

Therefore we have

$$
\begin{aligned}
& u_1(M_R, S', S^*) \\
= \; & \sum_{t=1}^{T} (\mu_1 - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
= \; & \sum_{t=1}^{T} (\mu_1 - w_{2,t}) \cdot \mathbb{1}_{I_t=1} + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq \; & c_{1,\tau_R}\mu_1 + \mu_1 + (T - \tau_R - 1)\max(\mu_1 - \mu_2, 0) + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq \; & \mu_1 \left(\frac{\mu_1}{\mu_1 + \mu_2}\tau_R + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} B \right) + \mu_1 + (T - \tau_R - 1)\frac{q_1^2}{\mu_1 + \mu_2} \\
& + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\
\leq \; & \frac{\mu_1^2}{\mu_1 + \mu_2} T + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} B + \mu_1 + \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}.
\end{aligned}
$$

In general, we thus have that

$$
u_1(M_R, S', S^*) \leq \frac{\mu_1^2}{\mu_1 + \mu_2} T + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} B + \mu_1 + \max\left(0, \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right).
$$

Therefore

$$
\begin{aligned}
u_1(M, S', S^*) &= \mathbb{E}_R[u_1(M_R, S', S^*)] \\
&\leq \frac{\mu_1^2}{\mu_1 + \mu_2} T + \frac{\mu_1\mu_2}{\mu_1 + \mu_2} B + \mu_1 + \mathbb{E}_R\left[\max\left(0, \sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right)\right].
\end{aligned}
$$

Notice that $\sum_{t=1}^{T} (w_{2,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1}$ is the regret of not playing arm 2 (i.e., R_2 in the proof of Lemma B.1.2). Since the mechanism M is (ρ, δ) low regret, with probability

$1-\rho$, this sum is at most δ (and in the worst case, it is bounded above by $T\mu_2$). We therefore have that:

$$\begin{aligned} u_1(M, S', S^*) &\leq \frac{\mu_1^2}{\mu_1+\mu_2}T + \frac{\mu_1\mu_2}{\mu_1+\mu_2}B + \mu_1 + \delta + \rho T\mu_2 \\ &\leq \frac{\mu_1^2}{\mu_1+\mu_2}T + O(\sqrt{T\delta}) \end{aligned}$$

From this and our earlier lower bound on $u_1(M, S^*, S^*)$, it follows that $u_1(M, S', S^*) - u_1(M, S^*, S^*) \leq O(\sqrt{T\delta})$, thus establishing that (S^*, S^*) is an $O(\sqrt{T\delta})$-Nash equilibrium for the arms.

Finally, to bound the revenue of the principal, note that if the arms both play according to S^* and $|c_{1,t}/\mu_1 - c_{2,t}/\mu_2| \leq B$ for all t (so they do not defect), the principal gets a maximum of $T\theta = O(\sqrt{T\delta})$ revenue overall. Since (by Lemma B.1.2) this happens with probability at least $1 - \frac{4}{T}$ (and the total amount of revenue the principal is bounded above by T), it follows that the total expected revenue of the principal is at most $O(\sqrt{T\delta})$.

□

B.1.2 Explicit Observational Model

In this section we show that in the explicit observational model, there is an approximate equilibrium for the arms that results in the principal receiving no revenue. Since arms can view other arms' reported values, it is easy to collude in the explicit model; simply defect and pass along the full amount as soon as you observe another arm passing along a positive amount.

Theorem B.1.7. *Let mechanism M be a δ-low regret algorithm for the multi-armed bandit problem. Then in the strategic multi-armed bandit problem under the explicit*

observational model, there exist distributions D_i and a $(\delta+1)$-Nash equilibrium for the arms where a principal using mechanism M receives zero revenue.

Proof. Consider the two-arm setting where D_1 and D_2 are both deterministic distributions supported entirely on $\{1\}$, so that $v_{i,t} = 1$ for all $i = 1, 2$ and $t \in [T]$. Consider the following strategy S^* for arm i:

1. Set $w_{i,t} = 0$ if at time $1, ..., t-1$, the other arm always reports 0 when pulled.
2. Set $w_{i,t} = 1$ otherwise.

We will show that (S^*, S^*) is a $(\delta+1)$-Nash Equilibrium. It suffices to show that arm 1 can get at most $\delta + 1$ more utility by deviating. Consider any deviating strategy S' for arm 1. By convexity, we can assume S' is deterministic (there is some best deterministic deviating strategy). Since mechanism M might be randomized, let R be the randomness used by M and define M_R to be the deterministic mechanism when M uses randomness R. Now, consider the case when arm 1 plays strategy S', arm 2 plays strategy S^* and the principal is usings mechanism M_R.

1. If arm 1 never reports any value larger than 0 when pulled, then S' behaves exactly the same as S^*. Therefore,

$$u_1(M_R, S', S^*) = u_1(M_R, S^*, S^*).$$

2. If arm 1 ever reports some value larger than 0 when pulled, let τ_R be the first time it does so. We know that S' behaves the same as S^* before τ_R. Therefore,

$$\begin{aligned} u_1(M_R, S', S^*) &\leq u_1(M_R, S^*, S^*) + \sum_{t=\tau_R}^{T} (v_{1,t} - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \\ &\leq u_1(M_R, S^*, S^*) + 1 + \sum_{t=\tau_R+1}^{T} (\max(w_{1,t}, w_{2,t}) - w_{1,t}) \cdot \mathbb{1}_{I_t=1} \end{aligned}$$

So in general, we have

$$u_1(M_R, S', S^*) \leq u_i(M_R, S^*, S^*) + 1 + \sum_{t=\tau_R+1}^{T} (\max(w_{1,t}, w_{2,t}) - w_{1,t}) \cdot \mathbb{1}_{I_t=1}.$$

Therefore

$$\begin{aligned} u_1(M, S', S^*) &= \mathbb{E}_R[u_1(M_R, S', S^*)] \\ &\leq \mathbb{E}_R[u_1(M_R, S^*, S^*)] + 1 + \mathbb{E}_R\left[\sum_{t=\tau_R+1}^{T} (\max(w_{1,t}, w_{2,t}) - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right] \\ &= u_1(M, S^*, S^*) + 1 + \mathbb{E}_R\left[\sum_{t=\tau_R+1}^{T} (\max(w_{1,t}, w_{2,t}) - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right]. \end{aligned}$$

Notice that this expectation is at most the regret of M in the classic multi-armed bandit setting when the adversary sets rewards equal to the values $w_{1,t}$ and $w_{2,t}$ passed on by the arms when they play (S', S^*). Therefore, by our low-regret guarantee on M, we have that

$$\mathbb{E}_R\left[\sum_{t=\tau_R+1}^{T} (\max(w_{1,t}, w_{2,t}) - w_{1,t}) \cdot \mathbb{1}_{I_t=1}\right] \leq \delta.$$

Thus

$$u_1(M, S', S^*) \leq u_1(M, S^*, S^*) + 1 + \delta$$

and this is a $(1 + \delta)$-approximate Nash equilibrium. Finally, it is easy to check that the principal receives zero revenue when both arms play according to this equilibrium strategy. □

B.2 Omitted Results and Proofs of Section 3.4

B.2.1 All Strategic Arms with Stochastic Values

Proof of Lemma 3.4.1. Note that the mechanism is naturally divided into three parts (in the same way the strategy above is divided into three parts): (1) the start, where each arm is played once and reports its mean, (2) the middle, where the principal plays the best arm and extracts the second-best arm's value (and plays each other arm once), and (3) the end, where the principal plays each arm some number of times, effectively paying them off for responding truthfully in step (1). To show the above strategy is dominant, we will proceed by backwards induction, showing that each part of the strategy is the best conditioned on an arbitrary history.

We start with step (3). It is easy to check that these rounds don't affect how many times the arm is played or not. It follows that it is strictly dominant to just report 0 (and receive your full value for the turn). Note that the reward the arm receives in expectation for this round is $(u + \log(w_i))\mu_i$; we will use this later.

For step (2), assume that $i = i^*$; otherwise, arm i is played only once, and the dominant strategy is to report 0 and receive expected reward μ_i. Depending on what happened in step (1), there are two cases; either $w' \leq \mu_i$, or $w' > \mu_i$. We will show that if $w' \leq \mu_i$, the arm should play w' for the next R rounds (not defecting) and report 0 for the bonus round. If $w' > \mu_i$, the arm should play 0 (defecting immediately).

Note that we can recast step (2) as follows: arm i starts by receiving a reward from his distribution D_i. For the next R turns, he can pay w' for the privilege of drawing a new reward from his distribution (ending the game immediately if he refuses to pay). If $w' \leq \mu_i$, then paying for a reward w' is positive in expectation, whereas if $w' > \mu_i$, then paying for a reward is negative in expectation. It follows that the dominant strategy is to continue to report w' if $w' \leq \mu_i$ (receiving a total expected reward of

$R(\mu_i - w') + \mu_i)$ and to immediately defect and report 0 if $w' > \mu_i$ (receiving a total expected reward of μ_i).

Finally, we analyze step (1). We will show that, regardless of the values reported by the other players, it is a dominant strategy for arm i to report its true mean μ_i. If arm i reports w_i, and $i \neq i^*$, then arm i will receive in expectation reward

$$G = (\mu_i - w_i) + \mu_i + \max(u + \log(w_i), 0)\mu_i$$

If $u + \log(w_i) > 0$, then this is maximized when $w_i = \mu_i$ and $G = (u + \log(\mu_i) + 1)\mu_i$ (note that by our construction of u, $u + \log(\mu_i) \geq 1$). On the other hand, if $u + \log(w_i) \leq 0$, then this is maximized when $w_i = 0$ and $G = 2\mu_i$. Since $u + \log(\mu_i) + 1 \geq 2$, the overall maximum occurs at $w_i = \mu_i$.

Similarly, when arm i reports w_i and $i = i^*$, then arm i receives in expectation reward

$$G' = (\mu_i - w_i) + \max(0, R(\mu_i - w')) + \mu_i + \max(u + \log(w_i), 0)\mu_i$$

which is similarly maximized at $w_i = \mu_i$. Finally, it follows that if $\mu_i \leq w'$, $G = G'$, so it is dominant to report $w_i = \mu_i$. On the other hand, if $\mu_i > w'$, then reporting $w_i = \mu_i$ will ensure $i = i^*$ and so once again it is dominant to report $w_i = \mu_i$. □

Proof of Lemma 3.4.3. Suppose there exists an truthful mechanism A guarantees $(\alpha\mu + (1 - \alpha)\mu')T$ revenue for any distributions. We will show this results in a contradiction.

We now consider $L > \exp(1/\alpha)$ inputs. The i-th input has $\mu = b_i = 1/2 + i/(2L)$ and $\mu' = 1/2$. Among these inputs, one arm (call it arm j^*) is always the arm with largest mean and another arm is always the arm with the second largest mean. Other arms have the same input distribution in all the inputs.

Consider all the arms are using their dominant strategies. For the i-th input, let $x_i T$ be the expected number of pulls by A on the arm k^* and $p_i T$ be the expected amount arm k^* gives to the principal. Because the mechanism is truthful, in the i-th distribution, arm k^* prefers its dominant strategy than the dominant strategy it uses in some j-th distribution ($i \neq j$). In other words, we have for $i \neq j$,

$$b_i x_i - p_i \geq b_i x_j - p_j.$$

We also have, for all i,

$$b_i x_i - p_i \geq 0.$$

By using these inequalities , we get for all i,

$$p_i \leq b_i x_i + \sum_{j=1}^{i-1} x_j (b_{j+1} - b_j).$$

On the other hand, A's revenue in the i-th distribution is at most $(p_i + (1 - x_i)\mu')T$. Therefore we have, for all i,

$$p_i + (1 - x_i)\mu' \geq \alpha \cdot b_i + (1 - \alpha)\mu'.$$

So we get

$$(1 - x_i)\mu' + b_i x_i + \sum_{j=1}^{i-1} x_j (b_{j+1} - b_j) \geq \alpha \cdot b_i + (1 - \alpha)\mu'.$$

It can be simplified as

$$x_i \geq \alpha + \sum_{j=1}^{i-1} x_j \frac{b_{j+1} - b_j}{b_i - \mu'} = \alpha + \frac{1}{i} \cdot \sum_{j=1}^{i-1} x_j.$$

By induction we get for all i,

$$x_i \geq \alpha \sum_{j=1}^{i} \frac{1}{i} > \alpha \ln(i).$$

Therefore we have

$$x_L > \alpha \ln(L) \geq 1.$$

Here we get a contradiction. □

B.2.2 Strategic and Non-strategic Arms with Stochastic Values

Proof of Lemma 3.4.4. Similarly as the proof of Lemma 3.4.1, the mechanism is divided into three parts: (1) the start, where each arm is played B times and reports its mean, (2) the middle, where the principal plays the best arm and extracts the second-best arm's value (and plays each other arm B times), and (3) the end, where the principal plays each arm some number of times, effectively paying them off for responding truthfully in step (1). To show the above strategy is dominant, we will proceed by backwards induction, showing that each part of the strategy is the best conditioned on an arbitrary history.

For step (3), similarly as the proof of Lemma 3.4.1, it is strictly dominant for the arm to report 0. The reward the arm receives in expectation for this step is $(u + \log(\bar{w}_i - M))\mu_i B$.

For step (2), assume that $i = i^*$; otherwise, arm i is played B times, and the dominant strategy is to report 0 and receive expected reward $\mu_i B$. Depending on what happened in step (1), there are two cases; either $w' - M \leq \mu_i$, or $w' - M > \mu_i$. Similarly as the proof of Lemma 3.4.1, we know that if $w' - M \leq \mu_i$, the arm should

play $w' - M$ for the next R rounds (not defecting) and report 0 for B bonus rounds. If $w' - M > \mu_i$, the arm should play 0 (defecting immediately).

For step (1), similar as the proof of Lemma 3.4.1, the expected reward of arm i is either

$$G = (\mu_i - \bar{w}_i)B + B\mu_i + \max(u + \log(\bar{w}_i - M), 0)B\mu_i$$

or

$$G' = \max(0, R(\mu_i - w' + M)) + (\mu_i - \bar{w}_i)B + B\mu_i + \max(u + \log(\bar{w}_i - M), 0)B\mu_i$$

Using the same argument as the proof of Lemma 3.4.1, we know arm i's dominant strategy is to make $\bar{w}_i = \mu_i + M$. □

Proof of Lemma 3.4.6. The only difference between the strategy in this lemma and the strategy in Lemma 3.4.4 is the first step, where instead of the arm reporting their mean every round (which they don't necessarily know), they instead report their value every round. It suffices to show that the expected difference in utility between running the above strategy and the strategy in Lemma 3.4.4 is at most $o(T)$.

To do this, let $\bar{w}_i' = \frac{1}{B}\sum_{t=1}^{B}(v_t + M)$ be the average value reported in the first phase by this new strategy, and let $\bar{w}_i = \mu_i + M$ be the optimal average to report. Let $\delta = \bar{w}_i' - \bar{w}$. From the formulas for net utility in the proof of Lemma 3.4.4, we note that reporting $\bar{w}_i'$ in the first phase instead of $\bar{w}_i$ results in at most $T\delta$ less utility overall. On the other hand, since $\mathbb{E}[v_t] = \mu_i$ for all t, by the Chernoff bound,

$$\Pr\left[|\delta| > 2\sqrt{\log T/B}\right] \leq 2\exp\left(\frac{1}{2}\left(2\sqrt{\frac{\log T}{B}}\right)^2 B\right) = \frac{2}{T^2}.$$

It follows that the expected difference in utility is at most

$$2\sqrt{\frac{\log T}{B}}T + \frac{2}{T^2}T = O(\epsilon^{-1/8}T^{5/8}) = o(T).$$

□

Proof of Corollary 3.4.5. Note that the proof of Lemma 3.4.4 works regardless of the values of B and M, so the strategy described in Lemma 3.4.4 is still a dominant strategy here. In an ϵ-Nash equilibrium, each player plays according to a strategy which gives them at least ϵ less than their payoff in the dominant equilibrium. We will show that if this is the case, then the principal gets at most $K\epsilon$ less than their payoff in the dominant equilibrium; since $K\epsilon = o(T)$, this proves the theorem.

Recall that $B = 2\epsilon^{1/4}T^{3/4}/\mu_{min}$ and define $\gamma = \epsilon^{1/3}/T^{1/3}$. We first claim that, similarly as in the proof of Lemma 3.4.4, if $i = i^*$ and $(1+\gamma)\mu_i \geq w'$, then if arm i is playing according to an ϵ-Nash equilibrium, it will not defect. This follows from the fact that modifying arm i's strategy to start repeatedly reporting w' as soon as arm i would have defected under the original strategy increases arm i's payoff by at least $B\mu_i - R\gamma\mu_i \geq 2\epsilon^{1/4}T^{3/4} - \epsilon^{1/3}T^{2/3} \geq \epsilon$ in expectation (where the additional $B\mu_i$ term comes via the payoff from the bonus rounds).

We next show that, in any ϵ-Nash equilibrium, each arm i reports an average value $\bar{w}_i$ between $\mu_i(1-\gamma) + M$ and $\mu_i(1+\gamma) + M$ with high probability.

To do this, we define

$$G_\mu(w) = ((\mu - (w + M)) + \mu + \max(u + \log w, 0)\mu) \cdot B.$$

Note that $G_\mu(w)$ upper bounds the expected reward an arm with mean μ which reports $w+M$ can get from all rounds except the R rounds in line 4 (but including the potential bonus rounds). Moreover, by the proof of Lemma 3.4.4, for $w = \mu$, $G_\mu(\mu)$ exactly equals the expected reward (in these rounds) of an arm following the dominant strategy. We'll first show that if $w < \mu(1-\gamma)$, then $G_\mu(\mu) - G_\mu(w) \geq \epsilon^{11/12}T^{1/12}$.

First, if $u+\log(w-M) < 0$, then $G_\mu(w) \leq 2B\mu - BM$, but by the proof of Lemma 3.4.4, $G_\mu(\mu) = B(u + \log\mu + 1)\mu - BM$. Since $u + \log\mu + 1 > 3$, $G_\mu(\mu) - G_\mu(w) \geq B\mu \geq \epsilon^{1/4}T^{3/4} \geq \epsilon^{11/12}T^{1/12}$. We can thus assume $\max(u + \log w, 0) = u + \log w$. Under this assumption

$$G_\mu(w) = ((\mu - w) + \mu + (u + \log w)\mu) \cdot B.$$

Then, if $w \leq \mu(1-\gamma)$,

$$\begin{aligned} G_\mu(\mu) - G_\mu(w) &= B(\mu\log\mu - \mu - \mu\log w + w) \\ &= B\mu(\log(1-\gamma) - \gamma) \\ &\geq B\mu\gamma^2 \\ &\geq 2\epsilon^{1/4}T^{3/4}\epsilon^{2/3}T^{-2/3} \\ &\geq \epsilon^{11/12}T^{1/12} \end{aligned}$$

Similarly, if $w > \mu(1+\gamma)$, we have that $G_\mu(\mu) - G_\mu(w) \geq \epsilon^{11/12}T^{1/12}$. Now, in expectation over w, $\mathbb{E}_w[G_\mu(\mu) - G_\mu(w)] \leq \epsilon$; otherwise, this player could increase their expected total reward by at least ϵ by switching to the dominant strategy (note that a player's expected reward from the R rounds in line 4 can only increase by switching to the dominant strategy). From Markov's inequality, it follows that

$$\Pr_{w_i}[w_i \in [\mu_i(1-\gamma), \mu_i(1+\gamma)]] \geq 1 - (\epsilon/T)^{1/12}.$$

Via the union bound, it follows that the probability that each w_i belongs to the interval $[\mu_i(1-\gamma)+M, \mu_i(1+\gamma)+M]$ is at least $1 - K(\epsilon/T)^{1/12} \geq 1 - o(1)$. Note that if this is the case, arm i^* will not defect, since $(1+\gamma)\mu_{i^*} \geq w_{i^*} \geq w'$. In addition, note that $w' \geq (1-\gamma)\mu' + M$ (since the two largest means are larger than μ', the two

largest reported values w will be at least $(1-\gamma)\mu' + M)$. It follows in this case that the principal receives at least $(1-\gamma)\mu' R = \mu' T - o(T)$. Since this occurs with probability $1 - o(1)$, it follows that the principal receives at least $\mu' T - o(T)$ in expectation, as desired. □

Proof of Theorem 3.4.7. Recall that $B = 2\epsilon^{1/4} T^{3/4}/\mu_{min}$ and $M = 8B^{-1/2} \ln(KT)$. We first show that with high probability non-strategic arms' reported values don't deviate too much from their means.

For each non-strategic arm i, by the Chernoff bound,

$$\Pr[|\bar{w}_i - \mu_i| \geq M/2] \leq 2\exp(-(M/2)^2 B/2) \leq 1/(KT)^8$$

By the union bound, with probability $1 - o(1/T)$, all non-strategic arms i satisfy $|\bar{w}_i - \mu_i| \leq M/2$. From now on, we will assume we are in the case when $|\bar{w}_i - \mu_i| < M/2$, for all i such that arm i is a non-strategic arm.

In the proof of Corollary 3.4.5, we showed that any strategic arm i playing according to an ϵ-Nash equilbrium, will report in Line 1 an average value $\bar{w}_i$ between $(1-\gamma)\mu_i + M$ and $(1+\gamma)\mu_i + M$ with high probability, where $\gamma = o(1)$. Note that this guarantee holds even in the presence of non-strategic arms, as we only use the fact that any strategy an arm plays in an ϵ-Nash equilibrium has an expected value of at least ϵ less than their dominant strategy's expected value. With this, we can consider two possible cases:

- **Case 1**: Arm i^* is a strategic arm. Then $w' \geq (1-\gamma)\mu_s + M$ and $w' \geq \mu_n - M/2$, and also $\mu_{i^*} = w_{i^*} - M \geq w' - M$. So, from only the third step of Mechanism

3, the principal will get reward at least

$$\begin{aligned}
& (w' - M)R = \max((1-\gamma)\mu_s, \mu_n - 3M/2)R \\
\geq\ & (1-\gamma)\max(\mu_s, \mu_n)R - 3MR/2 \\
\geq\ & (1-\gamma)\max(\mu_s, \mu_n)T - \max(\mu_s, \mu_n)(u+3)BK - 3MR/2 \\
=\ & \max(\mu_s, \mu_n)T - o(T).
\end{aligned}$$

- **Case 2**: Arm i^* is a non-strategic arm. We know that $\mu_{i^*} \geq w_{i^*} - M/2 \geq (w' - M) + M/2$. By using the Chernoff bound and union bound again, we know that arm i^* will defect in the line three with probability at most $o(1/T)$. We also know that $\mu_{i^*} \geq w_{i^*} - M/2 \geq (1-\gamma)\mu_s + M - M/2$ and $\mu_{i^*} \geq w_{i^*} - M/2 \geq u_n - M/2 - M/2$. It follows via the same argument as Case 1 that the principal will get reward at least $\max(u_s, u_n)T - o(T)$.

□

Appendix C

Appendix for Chapter 4

C.1 Analysis of the 1-dimensional case

We now analyze the policy of Kleinberg and Leighton for the one-dimensional case. Their keep a knowledge set $S_t = [a_t, a_t + \Delta_t]$ and choose price

$$p_t = a_t + 1/2^{2^{k_t}} \quad \text{where} \quad k_t = \lfloor 1 + \log_2 \log_2 \Delta_t^{-1} \rfloor$$

while $\Delta_t > 1/T$ after that, their policy prices at the lower end of the interval.

Clearly the total regret whenever $\Delta_t \leq 1/T$ is at most 1, so we only need to analyze the cases where $\Delta_t > 1/T$ and hence $k_t \leq O(\log \log T)$. To show a regret bound of $O(\log \log T)$ it is enough to argue that for every value of k, the total regret from timesteps where $k_t = k$ is $O(1)$.

We start by noting that if there is no sale then in the next period $\Delta_{t+1} = 1/2^{2^{k_t}}$ and therefore $k_{t+1} = k_t + 1$. Since k_t is monotone, there can be at most one no-sale for every value of k_t. The remaining periods where $k_t = k$ correspond to sales, where the loss is at most $\Delta_t \leq 1/2^{2^{k-1}}$, since by each sale Δ_t decreases by $1/2^{2^k}$, there are at most $2^{2^k}/2^{2^{k-1}} = 2^{2^{k-1}}$ sales. Since each of them incur loss $\Delta_t \leq 1/2^{2^{k-1}}$, the total

regret for sales with $k_t = k$ is at most 1. The total regret from no-sales is at most 1 since there is at most one no-sale.

Appendix D

Appendix for Chapter 5

D.1 More Details on the Coupling Argument

In this appendix we present examples of bracket transformations. Recall that our transformations took as input any "bad" bracket, where player i eventually meets player j, *and* player j will lose to some player k in the future if she advances past i (and k is the latest such player). The players benefit from manipulating these brackets. We transformed them into "good" brackets, where either player j is eliminated before even meeting player i, or where player j would be the champion conditioned on getting past i. The players have no incentive to manipulate these brackets.

We designed two injective transformations with disjoint images, σ_i and σ_j. σ_i was more straight-forward, but we include an example below anyway. σ_j was more complex. We include below an example showing that the complexity is necessary, and then an example of σ_j. All figures are at the end.

D.1.1 Example of the transformation $\sigma_i(B)$.

Recall that σ_i essentially swaps the sub-brackets rooted at i and k. See Section 5.3.2 for a formal description.

Consider the partial bracket B_1 shown in Figure D.1. Then, applying the transformation $\sigma_i(B_1)$ as described in our paper will yield the bracket B_1' shown in Figure D.2. Note that this mapping is injective: by examining $\sigma_i(B)$, we see exactly where j is eliminated, and conclude that this must be where i met j in the original B.

D.1.2 Counterexample to a naive $\sigma_j(B)$.

We could try using the same ideas in σ_i for σ_j: simply swap the subtrees rooted at k and j. Unfortunately, this mapping is not injective.

Consider the two brackets B_3, B_4 shown in Figure D.3. Then applying this naive transformation will map these brackets to the same bracket (see Figure D.4), showing that the mapping may not be injective. This motivates the need for the more involved transformation σ_j from Section 5.3.2.

Specifically, observe that in B_3, i meets j in round 2, so the depth-2 subtree rooted at k would get swapped with the depth-2 subtree rooted at j. In B_4, i meets j in round 1, so the single node i_1 would get swapped with the single node j. It is easy, but tedious, to complete this into a full tournament/bracket.

D.1.3 Example of the transformation $\sigma_j(B)$.

Essentially, the problem with the naive transformation is that it's hard to recover where i met j in the original B just from the naive $\sigma_j(B)$. This is because maybe on its path to j, i met many other competitors who also would have beaten j, in addition to the k we swap in from the mapping. Our more involved transformation fixes this by additionally swapping all such competitors out of the subtree below i, so we can again recover where i met j in the original B.

Consider the partial bracket B_2 shown in Figure D.5 and assume that in the tournament in case i_2 would beat j. Then, applying the transformation $\sigma_j(B_2)$ as described in our paper will yield the bracket B_2' shown in D.6.

Note that this mapping is injective! First, we can recover where i met j in the original B by looking at where i first encounters someone who would beat j in $\sigma_j(B)$. Once we learn this, we also know that in the original B, j actually advanced this far in the tournament to meet i, so we know exactly which subtrees we need to un-swap with subtrees of i.

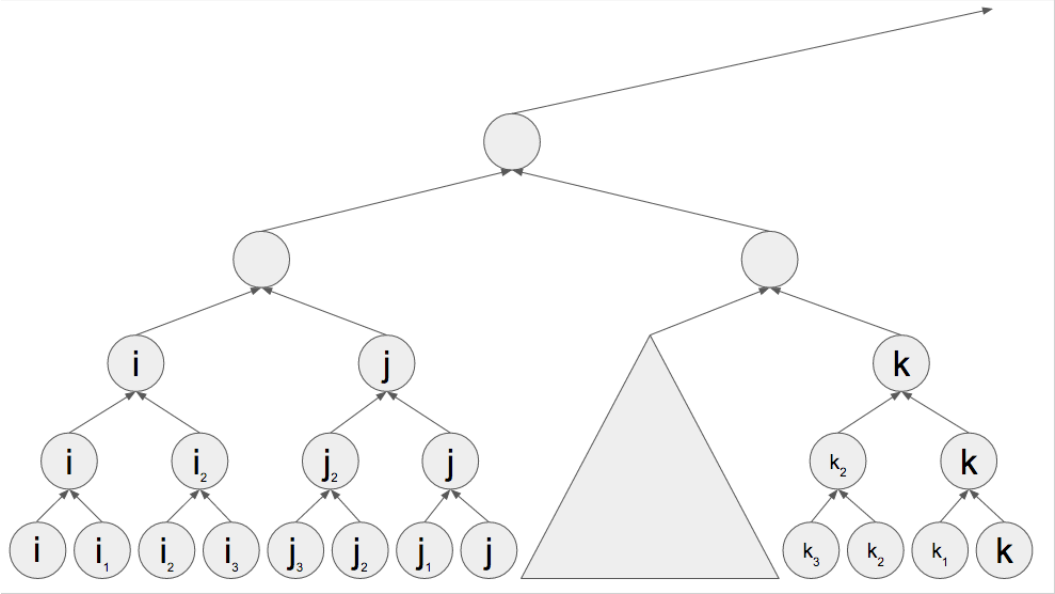


Figure D.1: A partial bracket B_1.

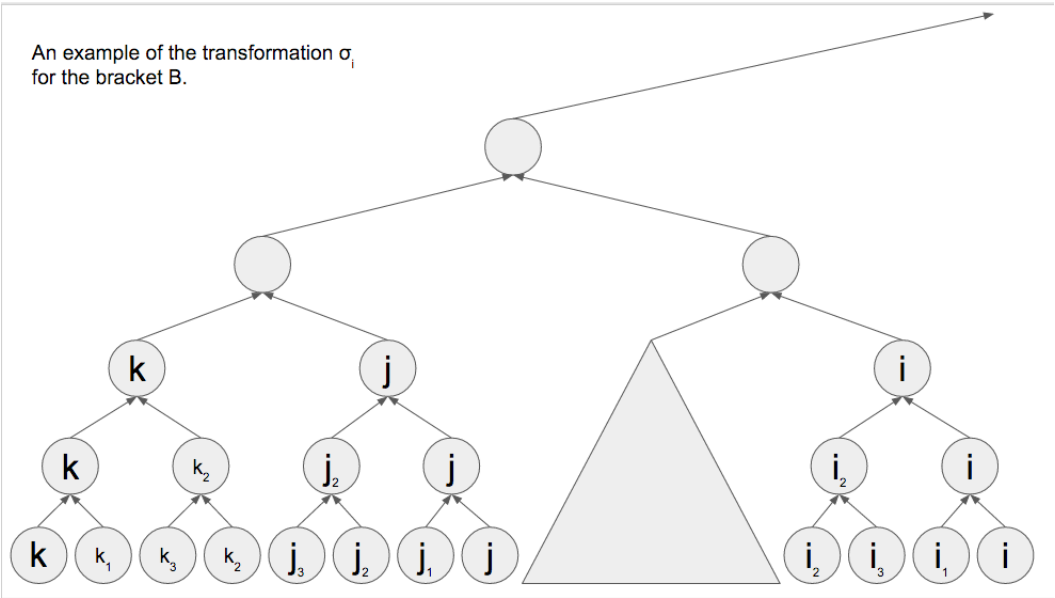


Figure D.2: $\sigma_i(B_1)$.

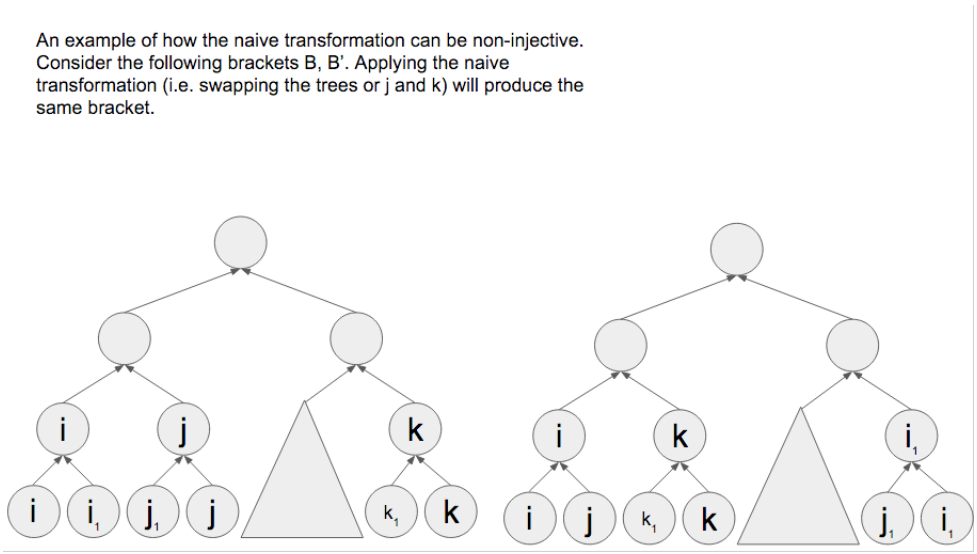


Figure D.3: Two partial brackets B_3, B_4.

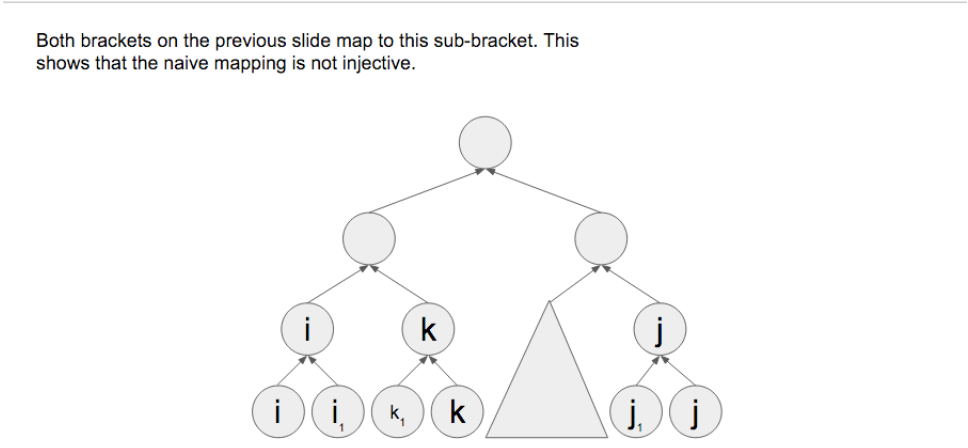


Figure D.4: Swapping the subtrees corresponding to j, k in both brackets above yields this bracket.

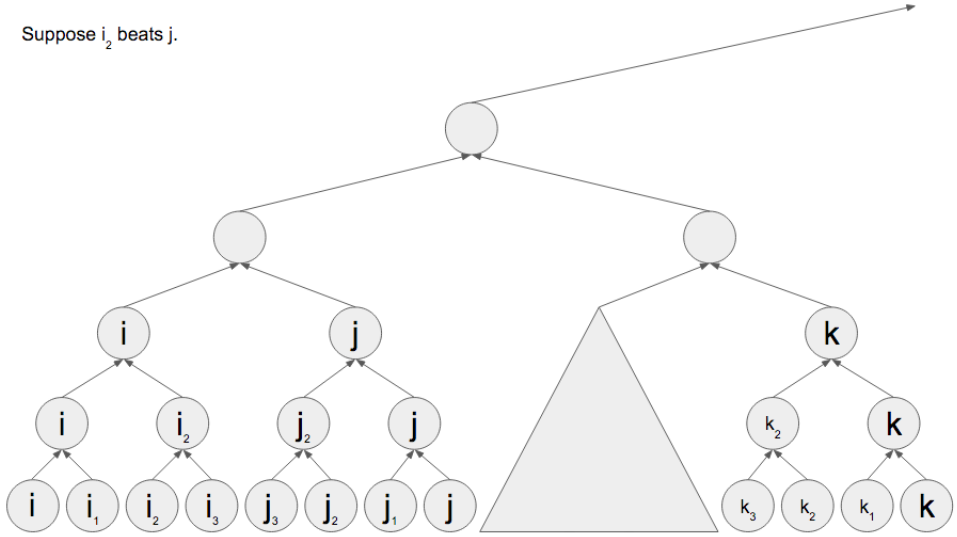


Figure D.5: A partial bracket B_2.

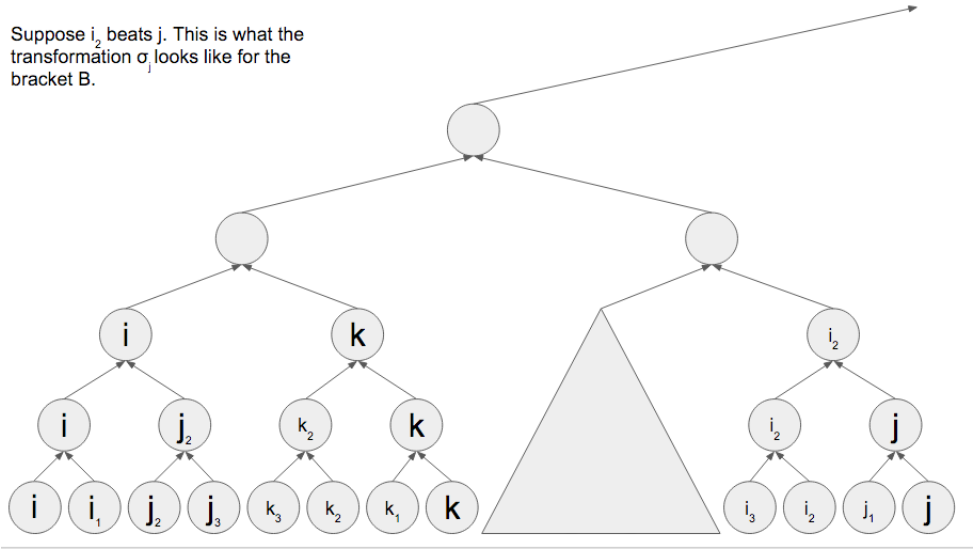


Figure D.6: $\sigma_j(B_2)$.

Appendix E

Appendix for Chapter 6

E.1 Probability and Information Theory Preliminaries

We briefly review some standard facts and definitions from information theory we will use throughout this paper. For a more detailed introduction, we refer the reader to [48].

Throughout this paper, we use log to refer to the base 2 logarithm and use ln to refer to the natural logarithm. If X is drawn from Bernoulli distribution B_p, we use $H(p) = -(p \log p + (1-p)(\log(1-p))$ to denote $H(X)$.

Fact E.1.1. *Let X_1, X_2, Y, Z be random variables, we have $I(X_1X_2; Y|Z) = I(X_1; Y|Z) + I(X_2; Y|X_1Z)$.*

Fact E.1.2. *Let X, Y, Z, W be random variables. If $I(Y; W|X, Z) = 0$, then $I(X; Y|Z) \geq I(X; Y|ZW)$.*

If X and Y are drawn from Bernoulli distribution B_p and B_q, we write $D(p\|q)$ as an abbreviation for $D(X\|Y)$.

Fact E.1.3. *Let X, Y, Z be random variables, we have $I(X;Y|Z) = \mathbb{E}_{x,z}[D((Y|X = x, Z = z)\|(Y|Z = z))]$.*

Fact E.1.4. *Let X, Y be random variables, $\sum_x \frac{|\Pr[X=x]-\Pr[Y=x]|^2}{2\max\{\Pr[X=x],\Pr[Y=x]\}} \leq \ln(2) \cdot D(X\|Y) \leq \sum_x \frac{|\Pr[X=x]-\Pr[Y=x]|^2}{\Pr[Y=x]}$.*

Proof. A proof of Fact E.1.4 can be found in [28]. □

We will also need the following quantitative version of the central limit theorem.

Lemma E.1.1 (Berry-Esseen Theorem)**.** *Let $Z_1, \cdots, Z_k$ be independent random variables and let $S = \sum_{i=1}^k Z_i$. Let $\mu = \mathbb{E}[S] = \sum_{i=1}^k \mathbb{E}[Z_i]$, $\sigma^2 = \mathrm{Var}[S] = \sum_{i=1}^k \mathrm{Var}[Z_i]$ and $\gamma = \sum_{i=1}^k \mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^3]$. Let Φ be the CDF of standard Gaussian. Then for all $t \in \mathbb{R}$,*

$$\left|\Pr\left[S < t\right] - \Phi\left(\frac{t-\mu}{\sigma}\right)\right| \leq \frac{\gamma}{\sigma^3}.$$

Finally, we will need the following estimates on the tails of the Gaussian distribution.

Lemma E.1.2. *Let $\Phi(t)$ be the CDF of standard Gaussian distribution then for $t > 0$,*

$$\frac{1}{\sqrt{2\pi}}\exp(-t^2/2)\left(\frac{1}{t} - \frac{1}{t^3}\right) \leq 1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}}\exp(-t^2/2)\frac{1}{t}.$$

Proof.

$$\begin{aligned}
1 - \Phi(t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty \exp(-x^2/2)dx \\
&= \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{1}{x}\cdot x\exp(-x^2/2)dx \\
&= \frac{1}{\sqrt{2\pi}}\left(\frac{\exp(-t^2/2)}{t} - \int_t^\infty \frac{1}{x^2}\exp(-x^2/2)dx\right) \qquad \text{(integration by parts)} \\
&= \frac{1}{\sqrt{2\pi}}\Big(\frac{\exp(-t^2/2)}{t} - \frac{\exp(-t^2/2)}{t^3} \\
&\quad + \int_t^\infty \frac{3}{x^4}\exp(-x^2/2)dx\Big). \qquad \text{(integration by parts again)}
\end{aligned}$$

From the last two expressions, we get the required upper and lower bounds. □

E.2 Missing proofs of Section 6.5

Proof of Lemma 6.5.1. Let p_e be the probability that $B = 0$ and $\mathcal{A}_{count}$ outputs $B = 1$ when provided with $r = \frac{2n\ln(\alpha^{-1})}{\|\mathbf{p}-\mathbf{q}\|_1^2}$ samples. By symmetry p_e is equal to the probability that we are in the case $B = 1$ and $\mathcal{A}_{count}$ outputs $B = 0$ when provided with r samples. It therefore suffices to show that p_e is at most α. When $B = 0$,

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=1}^{n} S_i\right] = r\sum_{i=1}^{n}(p_i - q_i).$$

By the Chernoff bound,

$$p_e \leq \Pr[Z \leq 0] \leq \exp\left(-\frac{nr}{2}\cdot\left(\frac{\sum_{i=1}^{n}(p_i - q_i)}{n}\right)^2\right) \leq \alpha.$$

The second part of the lemma follows from Corollary 6.4.4, along with the observation that $\|\mathbf{p}-\mathbf{q}\|_1^2 \geq \|\mathbf{p}-\mathbf{q}\|_2^2$. □

Proof of Lemma 6.5.2. Assume without loss of generality that $B = 0$, and let $\varepsilon = p_1 - q_1 = \|\mathbf{p}-\mathbf{q}\|_\infty$. Let E be the event that $\mathcal{A}_{max}$ makes an error and outputs $B = 1$ when given $r = \frac{8\ln 2n\alpha^{-1}}{\varepsilon^2}$ samples. We can upper bound the probability of error as

$$\Pr[E] \leq \Pr[E|S_1 > r\varepsilon/2] + \Pr[S_1 \leq r\varepsilon/2].$$

We will bound each term separately. Since $\mathbb{E}[S_1] = r(p_1 - q_1) = r\varepsilon$, by Hoeffding's inequality,

$$\Pr[S_1 \leq r\varepsilon/2] \leq \exp(-r\varepsilon^2/8) \leq \frac{\alpha}{2}.$$

Similarly, by Hoeffding's inequality and the union bound, $\Pr[E|S_1 > r\varepsilon/2] \leq \Pr[\exists i : S_i < -r\varepsilon/2] \leq n\exp(-r\varepsilon^2/8) \leq \frac{\alpha}{2}$. It follows that $\Pr[E] \leq \alpha$. The second part of

the lemma follows from Corollary 6.4.4, along with the observation that $\|\mathbf{p}-\mathbf{q}\|_2^2 \leq n\|\mathbf{p}-\mathbf{q}\|_\infty^2$. □

Proof of Lemma 6.5.3. Let k be an arbitrary integer between 1 and $n-1$. Let $\mathbf{p}, \mathbf{q}$ be any vectors satisfying the following constraints:

1. For all $i \in [n]$, $\frac{1}{4} < p_i, q_i < \frac{3}{4}$.
2. If $i \notin \{k, k+1\}$, $p_i = q_i$.
3. If $i \in \{k, k+1\}$, $q_i = p_i - \varepsilon$.

Note that $\|\mathbf{p}-\mathbf{q}\|_\infty = \varepsilon$. Therefore, by Lemma 6.5.2, $r_{min}(C, \mathcal{A}_{max}, 1-\frac{2}{n}) \leq \frac{16 \ln n}{\varepsilon^2}$, thus proving the first part of the lemma.

Now assume that $r \leq n/128\varepsilon^2$. We will show that with this many samples, $\mathcal{A}_{count}$ solves instance C with probability at most $3/4$, thus implying the second part of the lemma. Without loss of generality, assume that $B = 0$. Define the following random variables $U_{i,j}$:

1. $U_{i,j} = X_{i,j} - Y_{i,j}$ for $i = 1, ..., k-1, k+2, ..., n$ and $j = 1, .., r$.
2. $U_{i,j} = X_{i,j} - Y_{i,j} - \varepsilon$. $i = k, k+1$ and $j = 1, ..., r$.

It is straightforward to check that for all $i = 1, ..., n, j = 1, .., r$, $\mathbb{E}[U_{i,j}] = 0$, $\mathbb{E}[U_{i,j}^2] \geq 1/4$ and $\mathbb{E}[|U_{i,j}|^3] \leq 1$. Let Φ be the cdf of the standard normal distribution.

$$
\begin{aligned}
&\Pr[\mathcal{A}_{count} \text{ outputs } B = 1 \text{ (incorrectly)}] \\
&= \Pr[\sum_{i=1}^{n}\sum_{j=1}^{r}(X_{i,j} - Y_{i,j}) < 0] = \Pr[\sum_{i=1}^{n}\sum_{j=1}^{r} U_{i,j} < -2r\varepsilon] \\
&\geq \Phi\left(-2r\varepsilon \cdot \frac{1}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[U_{i,j}^2]}}\right) \\
&\quad - \frac{\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[|U_{i,j}|^3]}{(\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[U_{i,j}^2])^{-3/2}} \qquad \text{(By Berry-Esseen theorem (Lemma E.1.1))} \\
&\geq \Phi\left(-\sqrt{\frac{8r\varepsilon^2}{n}}\right) - \frac{8}{\sqrt{nr}} \geq \Phi(-1/4) - \frac{8}{\sqrt{nr}} \geq 1/4.
\end{aligned}
$$

□

Proof of Corallary 6.5.4. Let $i = \pi^{-1}(k)$ and $j = \pi^{-1}(k+1)$. The algorithm A' correctly places i in the set of the top k rows exactly when $\mathcal{A}_{count}$ correctly outputs that row i dominates row j. On the other hand, any two consecutive rows of $\mathbf{P}$ satisfy the constraints in the proof of Lemma 6.5.3. It follows that $r_{min}(S, A') \geq \Omega(\frac{n}{\varepsilon^2})$. □

Proof of Lemma 6.5.5. Consider the instance $C = (n, \mathbf{p}, \mathbf{q})$ where $p_i = \frac{1}{2} + \varepsilon$ and $q_i = \frac{1}{2}$, with $\varepsilon = \frac{1}{n^2}$. Since $\|\mathbf{p} - \mathbf{q}\|_1 = \frac{1}{n}$, by Lemma 6.5.1, $r_{min}(C, \mathcal{A}_{count}, 1 - \frac{1}{n}) \leq 2n^3 \ln n$.

Now assume $r = \frac{n^4}{2^{14} \ln n}$. We will now show that $\mathcal{A}_{max}$ solves $\textsc{Domination}(n, \mathbf{p}, \mathbf{q}, r)$ with probability strictly smaller than 4/5. Without loss of generality, assume that $B = 0$. Define random variables $S_i = \sum_{j=1}^{r}(X_{i,j} - Y_{i,j})$. Note that $S_1, \cdots, S_n$ are i.i.d random variables with $\mathbb{E}[S_i] = r\varepsilon$ and $\mathrm{Var}[S_i] = r(\frac{1}{2} - \varepsilon^2)$. Our algorithm $\mathcal{A}_{max}$ outputs $B = 1$ whenever $\inf_i S_i + \sup_i S_i < 0$. Let $\lambda > 0$ be a parameter whose value we will choose later. Note that:

$$
\begin{aligned}
&\Pr[\inf_i S_i + \sup_i S_i < 0] \\
&\geq \Pr[\inf_i S_i < -\lambda,\ \sup_i S_i < \lambda] \\
&\geq \Pr[\sup_i S_i < \lambda] - \Pr[\inf_i S_i \geq -\lambda,\ \sup_i S_i < \lambda] \\
&= \prod_{i=1}^{n} \Pr[S_i < \lambda]^n - \prod_{i=1}^{n} \Pr[-\lambda \leq S_i < \lambda]^n \\
&= \Pr[S_1 < \lambda]^n - \Pr[-\lambda \leq S_1 < \lambda]^n \\
&= \Pr[S_1 < \lambda]^n - (\Pr[S_1 < \lambda] - \Pr[S_1 < -\lambda])^n
\end{aligned}
$$

We will now apply the Berry-Esseen Theorem (Lemma E.1.1) with $Z_j = (X_{1,j} - Y_{1,j})$ to approximate the CDF of S_1. We have $\mu = \mathbb{E}[S_1] = r\varepsilon$, $\sigma^2 = \mathrm{Var}[S_1] = r(\frac{1}{2} - \varepsilon^2) \geq \frac{r}{4}$. and $\gamma = \sum_{j=1}^{r} \mathbb{E}[|Z_j - \varepsilon|^3] \leq 8r$. Therefore for all $t \in \mathbb{R}$,

$$
\left| \Pr[S_1 < t] - \Phi\left(\frac{t-\mu}{\sigma}\right) \right| \leq \frac{\gamma}{\sigma^3} \leq \frac{64}{\sqrt{r}} = \frac{2^{15}\sqrt{\ln n}}{n^2} \leq \frac{1}{n^{3/2}}
$$

when n is large enough. Let us choose $\lambda = \mu + \sigma\Phi^{-1}(1 - \frac{\ln 2}{n})$ and let $a = \frac{\lambda - \mu}{\sigma}$, $b = \frac{\lambda + \mu}{\sigma}$. Therefore $\Phi(a) = 1 - \frac{\ln 2}{n}$. When n is large enough, $a > 10$. By Fact E.1.2,

$$
\frac{1}{\sqrt{2\pi}} \exp(-a^2/2)\frac{1}{a} \geq \frac{\ln 2}{n} = 1 - \Phi(a) \geq \frac{1}{\sqrt{2\pi}} \exp(-a^2/2)\frac{1}{2a}.
$$

From the left hand side of the above inequality, we can conclude that $a \leq 2\sqrt{\ln n}$. Also,

$$
\begin{aligned}
&\Phi(-b) = 1 - \Phi(b) \\
&= \frac{\ln 2}{n} - (\Phi(b) - \Phi(a)) \\
&= \frac{\ln 2}{n} - \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-t^2/2\right) dt \\
&\geq \frac{\ln 2}{n} - \frac{1}{\sqrt{2\pi}} (b-a) \exp(-a^2/2) \\
&\geq \frac{\ln 2}{n} - \frac{2a(\ln 2)(b-a)}{n} && \left(\text{Since } \tfrac{1}{\sqrt{2\pi}} \exp(-a^2/2) \tfrac{1}{2a} \leq \tfrac{\ln 2}{n}\right) \\
&\geq \frac{\ln 2}{n} - \frac{4a\mu}{n\sigma} \\
&\geq \frac{\ln 2}{n} - \frac{16\varepsilon\sqrt{r \ln n}}{n} && \left(\mu = r\varepsilon, \sigma^2 \geq \tfrac{r}{4}, a \leq 2\sqrt{\ln n}\right) \\
&\geq \frac{\ln 2}{n} - 16 \frac{1}{n^2} \frac{1}{n} \sqrt{\frac{n^4}{2^{14} \ln n} \ln n} \\
&\geq \frac{\ln 2}{n} - \frac{1}{8n}
\end{aligned}
$$

Now we can bound the probability of error as follows:

$$
\begin{aligned}
&\Pr[\inf_i S_i + \sup_i S_i < 0] \\
&\geq \Pr[S_1 < \lambda]^n - (\Pr[S_1 < \lambda] - \Pr[S_1 < -\lambda])^n \\
&\geq \left(\Phi\left(\frac{\lambda - \mu}{\sigma}\right) - \frac{1}{n^{3/2}}\right)^n \\
&- \left(\Phi\left(\frac{\lambda - \mu}{\sigma}\right) - \Phi\left(\frac{-\lambda - \mu}{\sigma}\right) + 2 \cdot \frac{1}{n^{3/2}}\right)^n \\
&= \left(\Phi(a) - \frac{1}{n^{3/2}}\right)^n - \left(\Phi(a) - \Phi(-b) + \frac{2}{n^{3/2}}\right)^n \\
&\geq \left(1 - \frac{\ln 2}{n} - \frac{1}{n^{3/2}}\right)^n - \left(1 - \frac{2\ln 2}{n} + \frac{1}{8n} + \frac{2}{n^{3/2}}\right)^n \\
&\geq \exp(-\ln 2) - \exp(-2\ln 2 + 1/8) - 0.01 && \text{(when } n \text{ is large enough)} \\
&> \frac{1}{5}.
\end{aligned}
$$

□

E.3 Missing proofs of Section 6.6

Proof of Lemma 6.6.2. We will use Sanov's theorem (Lemma 6.6.1). Let $\Sigma = \{0,1\}^2$. Consider the set of distributions on Σ, $\mathcal{P}(\Sigma) = \{(p_{00}, p_{01}, p_{10}, p_{11}) : 0 \leq p_{00}, p_{01}, p_{10}, p_{11} \leq 1, p_{00} + p_{01} + p_{10} + p_{11} = 1\}$, and define $C \subset \mathcal{P}(\Sigma)$ as $C = \{(p_{00}, p_{01}, p_{10}, p_{11}) : p_{01} \geq p_{10}\}$. Clearly C is a closed convex set. Define $R = ((1-p)(1-q), (1-p)q, p(1-q), pq) \in \mathcal{P}(\Sigma)$; note that this is exactly the distribution of (X_i, Y_i) for each $i \in [k]$. Since $p > q$, $R \notin C$. Observe that $\sum_{i=1}^{r}(X_i - Y_i) \leq 0$ iff the empirical distribution generated by $(X_1, Y_1), \cdots, (X_k, Y_k)$, $\hat{P}_{((X_1,Y_1),\cdots,(X_k,Y_k))}$ belongs to C. We can assume that there is some $Q \in C$ such that $D(Q||R) < \infty$, otherwise the lemma is trivially true. Therefore by Lemma 6.6.1,

$$\Pr\left[\sum_{i=1}^{r}(X_i - Y_i) \leq 0\right] \leq \exp\left(-k(\ln 2)D(Q^*||R)\right)$$

where $Q^* = \text{argmin}_{Q \in C} D(Q||R)$ is unique. In addition, Q^* should lie on the boundary of C i.e. Q^* should satisfy $p_{01} = p_{10}$. So

$$D(Q^*||R) = \min_{0 \leq x,y \leq 1,\ x+2y \leq 1} D((1-x-2y, y, y, x)||R).$$

Let $f(x,y) = (\ln 2)D((1-x-2y, y, y, x||R)$. Since $D(Q||R)$ is convex as a function of Q, $f(x,y)$ is convex as well. We will show that there is always a point in the region $\{0 \leq x, y \leq 1,\ x + 2y \leq 1\}$ where the gradient of $f(x,y)$ is zero. Since f is convex,

this must be the minimizer of f. Note that

$$\begin{aligned}\frac{\partial f(x,y)}{\partial x} &= -1 - \ln(1-x-2y) + \ln((1-p)(1-q)) \\ &\quad + 1 + \ln x - \ln(pq) = 0 \\ \frac{\partial f(x,y)}{\partial y} &= -2 - 2\ln(1-x-2y) + \ln((1-p)(1-q)) \\ &\quad + 2 + 2\ln y - \ln(pq) = 0.\end{aligned}$$

Solving the above equations for x, y we get

$$x = \frac{pq}{\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)^2},$$

$$y = \frac{\sqrt{pq(1-p)(1-q)}}{\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right)^2}.$$

It is easy to check that $0 \le x, y \le 1$ and $x + 2y \le 1$. Substituting the values of x, y, we find that

$$D(Q^*||R) = -2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right).$$

□

Proof of Lemma 6.6.3. We can assume $0 < p, q < 1$, otherwise the required inequality follows from the fact that $-\ln(1-t) \ge t$ for $0 \le t < 1$. For example, when $p = 0$, the LHS simplifies to $-\log(1-q)$ and the RHS to $q/2$, and the inequality is satisfied. The other cases are similar. Hence, from now on, assume that $0 < p, q < 1$. Let $x = p(1-q)$ and $y = q(1-p)$. Thus $\mathbb{I}(p,q) = (x+y)(1 - H(x/x+y))$. We can also

the write the LHS of the inequality as:

$$\begin{aligned}
&-2\log\left(\sqrt{pq}+\sqrt{(1-p)(1-q)}\right)\\
&=-\log\left(pq+(1-p)(1-q)+2\sqrt{pq(1-p)(1-q)}\right)\\
&=-\log\left(1-x-y+2\sqrt{xy}\right)\\
&=-\log\left(1-(\sqrt{x}-\sqrt{y})^2\right)\\
&\geq(\sqrt{x}-\sqrt{y})^2/(\ln 2) && (-\log(1-t)\geq t/(\ln 2))\\
&=(x+y-2\sqrt{xy})/(\ln 2).
\end{aligned}$$

Now we need to show that

$$\frac{(x+y-2\sqrt{xy})}{\ln 2}\geq\frac{1}{2}(x+y)\left(1-H\left(\frac{x}{x+y}\right)\right).$$

We can scale x, y such that $x+y=1$, so let $x=\frac{1}{2}+z$ and $y=\frac{1}{2}-z$. Therefore it is enough to show that

$$1-\sqrt{1-4z^2}\geq\frac{\ln 2}{2}\left(1-H\left(\frac{1}{2}+z\right)\right).$$

We have $1-\sqrt{1-4z^2}\geq 2z^2$ and by Fact E.1.4,

$$1-H\left(\frac{1}{2}+z\right)=D\left(\frac{1}{2}+z\middle\|\frac{1}{2}\right)\leq\frac{4}{\ln 2}z^2.$$

Combining these two, we have the required inequality.

□

Proof of Lemma 6.6.6. Define $\mathbf{p}, \mathbf{q}$ to be

1. $p_1=\varepsilon$, $q_1=0$.

2. $p_i=q_i=1/2$ for $i=2,...,n$.

Note that $\mathbb{I}(p_1, q_1) = \varepsilon$, and $\mathbb{I}(p_2, q_2) = \cdots = \mathbb{I}(p_n, q_n) = 0$. Therefore, by Theorem 6.6.5, $\mathcal{A}_{coup}$ succeeds given $r = \frac{5184\sqrt{n}\log n}{\varepsilon}$ samples with probability at least $1 - 2/n$.

Now assume that $r \leq n/(16\varepsilon^2) \leq (n-1)/(8\varepsilon^2)$. We will now show that $\mathcal{A}_{count}$ solves DOMINATION$(n, \mathbf{p}, \mathbf{q}, r)$ with probability at most $3/4$. Without loss of generality assume that $B = 0$. Define the random variables $U_{i,j}$ as follows:

1. $U_{1,j} = X_{1,j} - Y_{1,j} - \varepsilon.\ j = 1, ..., r.$

2. $U_{i,j} = X_{i,j} - Y_{i,j}$ for $i = 2, ..., n, j = 1, .., r.$

It is straightforward to check that for all $i = 1, \ldots, n$ and $j = 1, \ldots, r$, $\mathbb{E}[U_{i,j}] = 0$ and $\mathbb{E}[|U_{i,j}|^3] \leq 1/2$. For all $i = 2, \ldots, n$ and $j = 1, \ldots, r$, we further have that $\mathbb{E}[U_{i,j}^2] = 1/2$. Let Φ be the cdf of the standard normal distribution.

$$
\begin{aligned}
&\Pr[\mathcal{A}_{count} \text{ outputs } B = 1 \text{ (incorrectly)}] \\
&= \Pr[\sum_{i=1}^{n}\sum_{j=1}^{r}(X_{i,j} - Y_{i,j}) < 0] = \Pr[\sum_{i=1}^{n}\sum_{j=1}^{r} U_{i,j} < -r \cdot \varepsilon] \\
&\geq \Phi\left(-r \cdot \varepsilon \cdot \frac{1}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[U_{i,j}^2]}}\right) \\
&\quad - \frac{\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[|U_{i,j}|^3]}{(\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbb{E}[U_{i,j}^2])^{-3/2}} \qquad \text{(By Berry-Esseen theorem (Lemma E.1.1))} \\
&\geq \Phi\left(-r \cdot \varepsilon/\sqrt{r(n-1)/2}\right) - \sqrt{\frac{2n^2}{(n-1)^3 r}} \geq \Phi(-1/4) \\
&\quad - \sqrt{\frac{2n^2}{(n-1)^3 r}} \\
&\geq 1/4.
\end{aligned}
$$

□

Proof of Lemma 6.6.7. Define $\mathbf{p}, \mathbf{q}$ as:

1. $p_1 = \varepsilon/100$, $q_1 = 0$.

2. $p_i = 1/2 + \varepsilon$, $q_i = 1/2$ $i = 2, ..., n$.

Note that $\mathbb{I}(p_1, q_1) = \varepsilon/100$ and $\mathbb{I}(p_2, q_2) = \cdots \mathbb{I}(p_n, q_n) = (1 - H(1/2 + \varepsilon))/2$. By Fact E.1.4, $\mathbb{I}(p_2, q_2) \leq \frac{4}{\ln(2)} \cdot (\varepsilon)^2 \leq \varepsilon/(100n)$. Thus $\varepsilon/100 \leq \sum_{i=1}^{n} \mathbb{I}(p_i, q_i) \leq \varepsilon/50$. Therefore, by Theorem 6.6.5, given at least $\frac{518400\sqrt{n}\ln n}{\varepsilon}$ samples, $\mathcal{A}_{coup}$ succeeds with probability at least $1 - 2/n$.

Now fix $r = \frac{1}{\varepsilon^2 2^{14} \ln n}$. We will now show that $\mathcal{A}_{max}$ solves $\textsc{Domination}(n, \mathbf{p}, \mathbf{q}, r)$ with probability at most $9/10$. Without loss of generality assume $B = 0$. Define random variable $S_i = \sum_{j=1}^{r}(X_{i,j} - Y_{i,j})$. S_1 is always non-negative. $S_2, \cdots, S_n$ are i.i.d random variables with $\mathbb{E}[S_i] = r\varepsilon$ and $\mathrm{Var}[S_i] = r(\frac{1}{2} - \varepsilon^2)$. Algorithm 10 outputs $B = 1$ when $\inf_i S_i + \sup_i S_i < 0$. Let $\lambda > 0$ be some parameter which we will choose later.

$$
\begin{aligned}
&\Pr[\inf_i S_i + \sup_i S_i < 0] \\
&\geq \Pr[\inf_i S_i < -\lambda,\ \sup_i S_i < \lambda] \\
&\geq \Pr[\sup_i S_i < \lambda] - \Pr[\inf_i S_i \geq -\lambda,\ \sup_i S_i < \lambda] \\
&= \prod_{i=1}^{n} \Pr[S_i < \lambda]^n - \prod_{i=1}^{n} \Pr[-\lambda \leq S_i < \lambda]^n \\
&= \Pr[S_1 < \lambda] \left(\Pr[S_2 < \lambda]^{n-1} - \Pr[-\lambda \leq S_2 < \lambda]^{n-1}\right) \\
&= \Pr[S_1 < \lambda] \\
&\quad \cdot \left(\Pr[S_2 < \lambda]^{n-1} - (\Pr[S_2 < \lambda] - \Pr[S_2 < -\lambda])^{n-1}\right)
\end{aligned}
$$

We will now apply Berry-Esseen Theorem (Lemma E.1.1) with $Z_j = (X_{2,j} - Y_{2,j})$ for $j = 1, \cdots, r$, to approximate the CDF of S_2. We have $\mu = \mathbb{E}[S_2] = r\varepsilon$, $\sigma^2 =$

$\mathrm{Var}[S_2] = r(\frac{1}{2} - \varepsilon^2) \geq \frac{r}{4}$. and $\gamma = \sum_{j=1}^{r} \mathbb{E}[|Z_j - \varepsilon|^3] \leq 8r$. Therefore for all $t \in \mathbb{R}$,

$$\left|\Pr[S_2 < t] - \Phi\left(\frac{t-\mu}{\sigma}\right)\right| \leq \frac{\gamma}{\sigma^3} \leq \frac{64}{\sqrt{r}} \leq \frac{64}{n^{3/2}}$$

when n is large enough. Let us choose $\lambda = \mu + \sigma\Phi^{-1}(1 - \frac{\ln 2}{n-1})$ and let $a = \frac{\lambda-\mu}{\sigma}$, $b = \frac{\lambda+\mu}{\sigma}$. Therefore $\Phi(a) = 1 - \frac{\ln 2}{n-1}$. When n is large enough, $a > 10$. By Fact E.1.2,

$$\frac{1}{\sqrt{2\pi}} \exp(-a^2/2)\frac{1}{a} \geq \frac{\ln 2}{n-1} = 1 - \Phi(a) \geq \frac{1}{\sqrt{2\pi}} \exp(-a^2/2)\frac{1}{2a}.$$

From the left hand side of the above inequality, we can conclude that $a \leq 2\sqrt{\ln(n-1)}$. Also,

$$\begin{aligned}
\Phi(-b) = 1 - \Phi(b) &= \frac{\ln 2}{n} - (\Phi(a) - \Phi(b)) \\
&= \frac{\ln 2}{n-1} - \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-t^2/2\right) dt \\
&\geq \frac{\ln 2}{n-1} - \frac{1}{\sqrt{2\pi}}(b-a)\exp(-a^2/2) \\
&\geq \frac{\ln 2}{n-1} - \frac{2a(\ln 2)(b-a)}{n-1} \\
&\geq \frac{\ln 2}{n-1} - \frac{4a\mu}{(n-1)\sigma} \\
&\geq \frac{\ln 2}{n-1} - \frac{16\varepsilon\sqrt{r\ln(n-1)}}{n-1} \quad (\mu = r\varepsilon, \sigma^2 \geq \tfrac{r}{4}, a \leq 2\sqrt{\ln(n-1)}) \\
&\geq \frac{\ln 2}{n-1} - 16\frac{\varepsilon}{n-1}\sqrt{\frac{1}{\varepsilon^2 2^{14}\ln n}\ln(n-1)} \\
&\geq \frac{\ln 2}{n-1} - \frac{1}{8(n-1)}
\end{aligned}$$

By Chernoff bound, we have $\Pr[S_1 < \lambda] \geq \Pr[S_1 \leq \mu] = 1 - e^{-r \cdot D(\varepsilon \| \varepsilon/100)} \geq 1 - e^{-\frac{2.5}{\varepsilon} \cdot \varepsilon} \geq 1/2$. Now we can bound the probability of error as follows:

$$
\begin{aligned}
&\Pr[\inf_i S_i + \sup_i S_i < 0] \\
&\geq \Pr[S_1 < \lambda] \\
&\quad \cdot \left(\Pr[S_2 < \lambda]^{n-1} - (\Pr[S_2 < \lambda] - \Pr[S_2 < -\lambda])^{n-1}\right) \\
&\geq \frac{1}{2}\left(\Phi\left(\frac{\lambda - \mu}{\sigma}\right) - \frac{64}{n^{3/2}}\right)^{n-1} \\
&\quad - \frac{1}{2}\left(\Phi\left(\frac{\lambda - \mu}{\sigma}\right) - \Phi\left(\frac{-\lambda - \mu}{\sigma}\right) + 2 \cdot \frac{64}{n^{3/2}}\right)^{n-1} \\
&= \frac{1}{2}\left(\Phi(a) - \frac{64}{n^{3/2}}\right)^{n-1} \\
&\quad - \frac{1}{2}\left(\Phi(a) - \Phi(-b) + \frac{128}{n^{3/2}}\right)^{n-1} \\
&\geq \frac{1}{2}\left(1 - \frac{\ln 2}{n-1} - \frac{64}{n^{3/2}}\right)^{n-1} \\
&\quad - \frac{1}{2}\left(1 - \frac{2\ln 2}{n-1} + \frac{1}{8(n-1)} + \frac{128}{n^{3/2}}\right)^{n-1} \\
&\geq \frac{1}{2}\left(\exp(-\ln 2) - \exp(-2\ln 2 + 1/8) - 0.01\right) \qquad \text{(when } n \text{ is large enough)} \\
&> \frac{1}{10}.
\end{aligned}
$$

□

E.4 Missing proofs of Section 6.7

Proof of Lemma 6.7.1. Pick $v \in [n]$, uniformly at random. Let $d^{in}(v)$ and $d^{out}(v)$ be the indegree and outdegree of vertex $v \in [n]$. Clearly $d^{in}(v) + d^{out}(v) = n - 1$. Also $v \in S$ iff $d^{in}(v) < k$. We can thus easily test if $v \in S$ by querying the $n - 1$ edges, $\{(i, v) : i \in [n] \setminus \{v\}\}$. Depending on whether $v \in S$, we now have two cases:

- **Case 1**: $v \in S$

For every i such that $(i, v) \in E$, we can conclude that $i \in S$. We can therefore remove these vertices and iterate. We have reduced the problem to a graph on $n - 1 - d^{in}(v) = d^{out}(v)$ vertices.

- **Case 2**: $v \notin S$

  For every i such that $(v, i) \in E$, we can conclude that $i \notin S$. We can therefore remove these vertices and iterate. We have reduced the problem to a graph on $n - 1 - d^{out}(v) = d^{in}(v)$ vertices.

Let n' be the number of vertices that remain after the above random process. Note that

$$\begin{aligned}
\mathbb{E}_v[n'] &= \Pr[v \in S] \cdot \mathbb{E}[d^{out}(v) | v \in S] \\
&\quad + \Pr[v \notin S] \cdot \mathbb{E}[d^{in}(v) | v \notin S] \\
&= \frac{k}{n}\left(n - k + \frac{k-1}{2}\right) + \frac{n-k}{n}\left(k + \frac{n-k-1}{2}\right) \\
&= \frac{n-1}{2} + \frac{k(n-k)}{2n} \leq \frac{3n}{4}.
\end{aligned}$$

By Markov's inequality, $\Pr[n' \geq 4n/5] \leq \frac{15}{16}$. We will repeatedly choose v at random until we find a v such that $n' < \frac{4n}{5}$. Once we find such a v, we can remove at least $n/5$ vertices from the graph and iterate the same procedure for the remaining graph. Let T_0 denote the random variable equal to the number of times we sample v. We have that $\Pr[T_0 \geq t] \leq (\frac{15}{16})^t$ and therefore

$$\mathbb{E}[T_0] = \sum_{t=1}^{\infty} \Pr[T_0 \geq t] \leq 15.$$

Similarly let T_i represent the number of times we must sample v in iteration i of this process; by the same logic, $\mathbb{E}[T_i] \leq 15$ for all i. If we let the random variable X

denote the number of edge queries the algorithm makes, then since the graph shrinks by a factor of $4/5$ at each iteration,

$$X = T_0 \cdot n + T_1 \cdot \left(\frac{4}{5}\right) n + T_2 \cdot \left(\frac{4}{5}\right)^2 n + \cdots$$
$$\mathbb{E}[X] \leq 15 \cdot \left(1 + \frac{4}{5} + \left(\frac{4}{5}\right)^2 + \ldots\right) \cdot n \leq 75n.$$

This completes the proof that $\mathbb{E}[X] = O(n)$, as required. We can similarly analyze the tail probability of X; note that:

$$\Pr[X > C\lambda n] \leq \Pr\left[\exists i : T_i > \frac{\lambda C}{9}\left(\frac{10}{9}\right)^i\right]$$

since $T_i \leq \frac{C\lambda}{9}\left(\frac{10}{9}\right)^i$ for every i implies that

$$X \leq \frac{C\lambda n}{9}\sum_{i=0}^{\infty}\left(\frac{4}{5}\right)^i\left(\frac{10}{9}\right)^i = \frac{C\lambda n}{9}\sum_{i=0}^{\infty}\left(\frac{8}{9}\right)^i \leq C\lambda n.$$

By the union bound,

$$\begin{aligned}
&\Pr\left[\exists i : T_i > \frac{C\lambda}{9}\left(\frac{10}{9}\right)^i\right] \\
&\leq \sum_{i=0}^{\infty}\Pr\left[T_i > \frac{C\lambda}{9}\left(\frac{10}{9}\right)^i\right] \\
&\leq \sum_{i=0}^{\infty}\exp\left(-\frac{C\lambda}{9}\ln\left(\frac{16}{15}\right)\left(\frac{10}{9}\right)^i\right) \\
&\leq \exp(-\lambda). && \text{(for sufficiently large } C\text{)}
\end{aligned}$$

□

Proof of Lemma 6.7.2. We have:

$$\begin{aligned}\frac{\partial \mathbb{I}(p,q)}{\partial p} &= (1-q)\log\left(\frac{2p(1-q)}{p(1-q)+(1-p)q}\right) \\ &\quad - q\log\left(\frac{2(1-p)q}{p(1-q)+(1-p)q}\right) \\ \frac{\partial \mathbb{I}(p,q)}{\partial q} &= (1-p)\log\left(\frac{2(1-p)q}{p(1-q)+(1-p)q}\right) \\ &\quad - p\log\left(\frac{2p(1-q)}{p(1-q)+(1-p)q}\right)\end{aligned}$$

When $p \geq q$,

$$\log\left(\frac{2p(1-q)}{p(1-q)+(1-p)q}\right) \geq 0,$$

$$\log\left(\frac{2(1-p)q}{p(1-q)+(1-p)q}\right) \leq 0.$$

Thus $\frac{\partial \mathbb{I}(p,q)}{\partial p} \geq 0$ and $\frac{\partial \mathbb{I}(p,q)}{\partial q} \leq 0$ when $p \geq q$. Thus increasing p or decreasing q cannot decrease $\mathbb{I}(p,q)$ when $p \geq q$. □

Bibliography

[1] Alekh Agarwal, Daniel J. Hsu, Satyen Kale, John Langford, Lihong Li, and Robert E. Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 1638–1646, 2014.

[2] N. Ailon. Active learning ranking from pairwise preferences with almost optimal query complexity. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2011.

[3] N Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM*, 55(5):23:1–23:27, 2008.

[4] Alon Altman and Robert Kleinberg. Nonmanipulable randomized tournament selections. 2010.

[5] Alon Altman, Ariel D. Procaccia, and Moshe Tennenholtz. Nonmanipulable selections from a tournament. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI'09, pages 27–32, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[6] Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Learning prices for repeated auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, pages 1169–1177, 2013.

[7] Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems*, pages 622–630, 2014.

[8] Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Repeated contextual auctions with strategic buyers. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 622–630, 2014.

[9] Masaki Aoyagi. Bid rotation and collusion in repeated auctions. *Journal of Economic Theory*, 112(1):79–105, 2003.

[10] Masaki Aoyagi. Efficient collusion in repeated auctions with communication. *Journal of Economic Theory*, 134(1):61–92, 2007.

[11] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In John Langford and Joelle Pineau, editors, *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1503–1510, New York, NY, USA, 2012. ACM.

[12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(6):121–164, 2012.

[13] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, 1950.

[14] Itai Ashlagi, Constantinos Daskalakis, and Nima Haghpanah. Sequential mechanisms with ex-post participation guarantees. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 213–214, 2016.

[15] Susan Athey and Kyle Bagwell. Optimal collusion with private information. *RAND Journal of Economics*, 32(3):428–65, 2001.

[16] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, January 2003.

[17] Moshe Babaioff, Robert D. Kleinberg, and Aleksandrs Slivkins. Truthful mechanisms with implicit payment computation. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pages 43–52, New York, NY, USA, 2010. ACM.

[18] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms: Extended abstract. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 79–88, New York, NY, USA, 2009. ACM.

[19] T. P. Ballinger and N. T. Wilcox. Decisions, error and heterogeneity. *The Economic Journal*, 107(443):1090–1105, 1997.

[20] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discrete & Computational Geometry*, 2(4):319–326, 1987.

[21] John J. Bartholdi, Craig A. Tovey, and Michael A. Trick. How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8):27 – 40, 1992.

[22] Hamsa Bastani and Mohsen Bayati. Online decision-making with high-dimensional covariates. *Working paper, Stanford University*, 2016.

[23] Dirk Bergemann and Juuso Vlimki. Learning and strategic pricing. *Econometrica*, 64(5):1125–49, 1996.

[24] Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

[25] R. Bradley and M. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[26] M. Braverman, J. Mao, and M. S. Weinberg. Parallel algorithms for se- lect and partition with noisy comparisons. In *Proccedings of the Annual Symposium on the Theory of Computing (STOC)*, 2016.

[27] M. Braverman and E. Mossel. Noisy sorting without resampling. In *Proceedings of the ACM-SIAM symposium on discrete algorithms (SODA)*, 2008.

[28] Mark Braverman and Jieming Mao. Simulating noisy channel interaction. In *Proceedings of the Conference on Innovations in Theoretical Computer Science*, 2015.

[29] Mark Braverman, Jieming Mao, Jon Schneider, and S Matthew Weinberg. Multi-armed bandit problems with strategic arms. *arXiv preprint arXiv:1706.09060*, 2017.

[30] Mark Braverman, Jieming Mao, Jon Schneider, and S Matthew Weinberg. Selling to a no-regret buyer. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 2018.

[31] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[32] Noam Brown and Tuomas Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, page eaao1733, 2017.

[33] Sebastian Bubeck, Tengyao Wang, and Nitin Viswanathan. Multiple identifications in multi-armed bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2013.

[34] Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

[35] Yang Cai and Constantinos Daskalakis. Learning multi-item auctions with (or without) samples. In *FOCS*, 2017.

[36] E. J. Candés. Modern statistical estimation via oracle inequalities. *Acta Numerica*, 15:257–325, 2006.

[37] Sylvain Chassang. Calibrated incentive contracts. *Econometrica*, 81(5):1935–1971, 2013.

[38] Sabyasachi Chatterjee, Adityanand Guntuboyina, and Bodhisattva Sen. On risk bounds in isotonic and other shape restricted regression problems. 43(4):1774–1800, 2014.

[39] Sabyasachi Chatterjee, Adityanand Guntuboyina, and Bodhisattva Sen. On matrix estimation under monotonicity constraints. arXiv preprint arXiv:1506.03430, 2015.

[40] Xi Chen, Sivakanth Gopi, Jieming Mao, and Jon Schneider. Competitive analysis of the top-k ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2017.

[41] Y. Chen and C. Suh. Spectral MLE: Top-K rank aggregation from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2015.

[42] Edward H. Clarke. Multipart Pricing of Public Goods. *Public Choice*, 11(1):17–33, 1971.

[43] Maxime C. Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, page 817, 2016.

[44] Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 817–817. ACM, 2016.

[45] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 243–252, New York, NY, USA, 2014. ACM.

[46] A.H. Copeland. A 'reasonable' social welfare function. *Seminar on Mathematics in Social Sciences*, 1951.

[47] S Cox. Tennis match fixing: Evidence of suspected match-fixing revealed, January 2016. http://www.bbc.com/sport/tennis/35319202.

[48] I. Csiszar and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[49] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[50] Constantinos Daskalakis and Vasilis Syrgkanis. Learning in auctions: Regret is hard, envy is easy. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 219–228, 2016.

[51] Constantinos Daskalakis and S. Matthew Weinberg. Symmetries and Optimal Multi-Dimensional Mechanism Design. In *the 13th ACM Conference on Electronic Commerce (EC)*, 2012.

[52] D. Davidson and J. Marschak. Experimental tests of a stochastic decision theory. *Measurement: Definitions and theories*, pages 233–269, 1959.

[53] Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. The sample complexity of auctions with side information. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 426–439, New York, NY, USA, 2016. ACM.

[54] Nikhil R. Devanur and Sham M. Kakade. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 99–106, New York, NY, USA, 2009. ACM.

[55] Nikhil R. Devanur, Yuval Peres, and Balasubramanian Sivan. Perfect bayesian equilibria in repeated sales. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 983–1002, Philadelphia, PA, USA, 2015. Society for Industrial and Applied Mathematics.

[56] Miroslav Dudík, Nika Haghtalab, Haipeng Luo, Robert E. Schapire, Vasilis Syrgkanis, and Jennifer Wortman Vaughan. Oracle-efficient learning and auction design. In *FOCS*, 2017.

[57] Bhaskar Dutta. Covering sets and a new condorcet choice correspondence. *Journal of Economic Theory*, 44(1):63 – 80, 1988.

[58] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the Tenth International World Wide Web Conference*, 2001.

[59] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

[60] B. Eriksson. Learning to top-k search using pairwise comparisons. In *Conference on Artificial Intelligence and Statistics*, 2013.

[61] Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *J. Comput. Syst. Sci.*, 66(4):614–656, 2003.

[62] P. C. Fishburn. Binary choice probabilities: on the varieties of stochastic transitivity. *Journal of Mathematical psychology*, 10(4):327–352, 1973.

[63] Peter C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

[64] Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 5–22, New York, NY, USA, 2014. ACM.

[65] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4):587–601, 1973.

[66] Allan Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

[67] Stephen Gillen, Christopher Jung, Michael Kearns, and Aaron Roth. Online learning with an unknown fairness metric. *arXiv preprint arXiv:1802.06936*, 2018.

[68] J.C. Gittins and D.M. Jones. A dynamic allocation index for the sequential design of experiments. In J. Gani, editor, *Progress in Statistics*, pages 241–266. North-Holland, Amsterdam, 1974.

[69] Yannai A. Gonczarowski and Noam Nisan. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 856–868, New York, NY, USA, 2017. ACM.

[70] Theodore Groves. Incentives in Teams. *Econometrica*, 41(4):617–631, 1973.

[71] A. Guntuboyina, D. Lieu, S. Chatterjee, and B. Sen. Spatial adaptation in trend filtering. arXiv preprint arXiv:1702.05113, 2017.

[72] Adityanand Guntuboyina and Bodhisattva Sen. Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields*, 2013. *To appear*, available at http://arxiv.org/abs/1305.1648.

[73] James Hannan. Approximation to bayes risk in repeated play. In *Contributions to the Theory of Games*, pages 3:97–139, 1957.

[74] R. Heckel, N. B. Shah, K. Ramchandran, and M. J. Wainwright. Active ranking from pairwise comparisons and when parametric assumptions dont help. arXiv preprint arXiv:1606.08842v2, 2016.

[75] Nicole Immorlica, Brendan Lucier, Emmanouil Pountourakis, and Samuel Taggart. Repeated sales with multiple strategic buyers. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 167–168. ACM, 2017.

[76] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil' ucb : An optimal exploration algorithm for multi-armed bandits. In *Proceedings of Conference on Learning Theory (COLT)*, 2014.

[77] K. Jamieson and R. Nowak. Active ranking using pairwise comparisons. In *Advances in Neural Information Processing Systems*, 2011.

[78] M. Jang, S. Kim, C. Suh, and S. Oh. Top-k ranking from pairwise comparisons: When spectral ranking is optimal. arXiv preprint arXiv:1603.04153, 2013.

[79] Adel Javanmard. Perishability of data: dynamic pricing under varying-coefficient models. *The Journal of Machine Learning Research*, 18(1):1714–1744, 2017.

[80] Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Working paper, University of Southern California*, 2016.

[81] Paul Johnson and Jacques Robert. Collusion in a model of repeated auctions. Cahiers de recherche, Universite de Montreal, Departement de sciences economiques, 1999.

[82] Sham M. Kakade, Ilan Lobel, and Hamid Nazerzadeh. Optimal dynamic mechanism design and the virtual-pivot mechanism. *Operations Research*, 61(4):837–854, 2013.

[83] Adam Kalai and Santosh Vempala. Geometric algorithms for online optimization. In *Journal of Computer and System Sciences*, pages 26–40, 2002.

[84] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci.*, 71(3):291–307, October 2005.

[85] P Kelso. Badminton pairs expelled from london 2012 olympics after 'match-fixing' scandal, August 2012. .

[86] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *Symposium on Theory of computing (STOC)*, 2007.

[87] Michael P. Kim, Warut Suksompong, and Virginia Vassilevska Williams. Who can win a single-elimination tournament? In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 516–522, 2016.

[88] Michael P. Kim and Virginia Vassilevska Williams. Fixing tournaments for kings, chokers, and more. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 561–567, 2015.

[89] Daniel A Klain and Gian-Carlo Rota. *Introduction to geometric probability*. Cambridge University Press, 1997.

[90] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 594–605. IEEE, 2003.

[91] Vladimir Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole dEté de Probabilités de Saint-Flour XXXVIII-2008*, volume 38. Springer Science & Business Media, 2011.

[92] Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988 – 1012, 2014.

[93] Jean-Jacques Laffont and David Martimort. *The Theory of Incentives: The Principal-Agent Model*. 2002.

[94] T.L Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.*, 6(1):4–22, March 1985.

[95] John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 817–824. Curran Associates, Inc., 2008.

[96] Jean-Francois Laslier. *Tournament solutions and majority voting*. Number 7. Springer Verlag, 1997.

[97] Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. *arXiv preprint arXiv:1804.03195. To appear in FOCS 2018.*, 2018.

[98] Siqi Liu and Christos-Alexandros Psomas. On the competition complexity of dynamic mechanism design. *CoRR*, abs/1709.07955, 2017.

[99] Ilan Lobel, Renato Paes Leme, and Adrian Vladu. Multidimensional binary search for contextual decision-making. *Operations Research*, 2017.

[100] T. Lu and C. Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2011.

[101] R. D. Luce. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.

[102] Randolph McAfee and John McMillan. Bidding rings. *American Economic Review*, 82(3):579–99, 1992.

[103] John McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42(9):654–655, 1956.

[104] D. H. McLaughlin and R. D. Luce. Stochastic transitivity and cancellation of preferences between bitter-sweet solutions. *Psychonomic Science*, 2(1–12):89–90, 1965.

[105] Peter McMullen. Inequalities between intrinsic volumes. *Monatshefte für Mathematik*, 111(1):47–53, 1991.

[106] Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Dynamic auctions with bank accounts. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 387–393, 2016.

[107] Vahab S. Mirrokni, Renato Paes Leme, Pingzhong Tang, and Song Zuo. Optimal dynamic mechanisms with ex-post IR via bank accounts. *CoRR*, abs/1605.08840, 2016.

[108] Jamie Morgenstern and Tim Roughgarden. The pseudo-dimension of near-optimal auctions. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 136–144, Cambridge, MA, USA, 2015. MIT Press.

[109] Jamie Morgenstern and Tim Roughgarden. Learning simple auctions. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1298–1318, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[110] H. Moulin. Choosing from a tournament. *Social Choice and Welfare*, 3(4):271–291, 1986.

[111] Hervé Moulin, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[112] Roger B. Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[113] Hamid Nazerzadeh, Amin Saberi, and Rakesh Vohra. Dynamic cost-per-action mechanisms and applications to online advertising. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 179–188, New York, NY, USA, 2008. ACM.

[114] S. Negahban, S. Oh, and D. Sha. Rank Centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.

[115] Denis Nekipelov, Vasilis Syrgkanis, and Eva Tardos. Econometrics for learning agents. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 1–18, New York, NY, USA, 2015. ACM.

[116] Christos Papadimitriou, George Pierrakos, Christos-Alexandros Psomas, and Aviad Rubinstein. On the complexity of dynamic mechanism design. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 1458–1475, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics.

[117] Marc Pauly. Can strategizing in round-robin subtournaments be avoided? *Social Choice and Welfare*, 43(1):29–46, 2014.

[118] B Phillips. The tennis triangle, July 2011. http://grantland.com/features/the-tennis-triangle/.

[119] Sheng Qiang and Mohsen Bayati. Dynamic pricing with demand covariates. *Available at SSRN 2765257*, 2016.

[120] A. Rajkumar and S. Agarwal. A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.

[121] Ronald L. Rivest and Emily Shen. An optimal single-winner preferential voting system based on game theory, 2010.

[122] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[123] Tim Roughgarden. The price of anarchy in games of incomplete information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 862–879, New York, NY, USA, 2012. ACM.

[124] Ariel Rubinstein. Equilibrium in Supergames with the Overtaking Criterion. *Journal of Economic Theory*, 21:1–9, 1979.

[125] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[126] Stephen H Schanuel. What is the length of a potato? *Lecture Notes in Mathematics, Springer*, pages 118–126, 1986.

[127] S Scherer. Italy breaks up soccer match-fixing network involving mafia, May 2015. http://www.bbc.com/news/world-europe-32793892.

[128] Jon Schneider, Ariel Schvartzman, and S Matthew Weinberg. Condorcet-consistent and approximately strategyproof tournament rules. *Proceedings of the 2017 ACM Conference on Innovations in Theoretical Computer Science*, 2017.

[129] T. Schwartz. Cyclic tournaments and cooperative majority voting: A solution. *Social Choice and Welfare*, 7(1):19–29, 1990.

[130] N. B. Shah, S. Balakrishnan, A. Guntuboyina, and M. J. Wainright. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Transactions on Information Theory*, 63(2):934–959, 2016.

[131] N. B. Shah, S. Balakrishnan, and M. J. Wainwright. Feeling the bern: Adaptive estimators for bernoulli probabilities of pairwise comparisons. arXiv preprint arXiv:1603.06881v1, 2016.

[132] N. B. Shah and M. Wainwright. Simple, robust and optimal ranking from pairwise comparisons. arXiv preprint arXiv:1512.08949, 2015.

[133] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.

[134] B Sinclair. 12 arrested in esports match fixing scandal - report, October 2015. http://www.gamesindustry.biz/articles/2015-10-19-12-arrested-in-esports-match-fixing-scandal-report.

[135] Andrzej Skrzypacz and Hugo Hopenhayn. Tacit collusion in repeated auctions. *Journal of Economic Theory*, 114(1):153–169, 2004.

[136] R Smyth. World cup: 25 stunning moments ... no3: West germany 1-0 austria in 1982, February 2014. http://www.theguardian.com/football/blog/2014/feb/25/world-cup-25-stunning-moments-no3-germany-austria-1982-rob-smyth.

[137] Isabelle Stanton and Virginia Vassilevska Williams. Rigging tournament brackets for weaker players. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 357–364, 2011.

[138] C. Suh, V. Tan, and R. Zhao. Adversarial top-K ranking. *IEEE Transactions on Information Theory (to appear)*, 2017. DOI 10.1109/TIT.2017.2659660.

[139] Vasilis Syrgkanis and Eva Tardos. Composable and efficient mechanisms. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 211–220, New York, NY, USA, 2013. ACM.

[140] L. L. Thurstone. A law of comparative judgement. *Psychological Reviews*, 34(4):273, 1927.

[141] A. Tversky. Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281–299, 1972.

[142] William Vickrey. Counterspeculations, Auctions, and Competitive Sealed Tenders. *Journal of Finance*, 16(1):8–37, 1961.

[143] Thuc Vu, Alon Altman, and Yoav Shoham. On the complexity of schedule control problems for knockout tournaments. In *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 1*, pages 225–232, 2009.

[144] F. Wauthier, M. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2013.

[145] Paul D Yoo, Maria H Kim, and Tony Jan. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 2, pages 835–841. IEEE, 2005.

[146] H. P. Young. Social choice scoring functions. *SIAM Journal on Applied Mathematics*, 28(4):824–838, 1975.

[147] Y. Zhou, X. Chen, and J. Li. Optimal PAC multiple arm identification with applications to crowdsourcing. In *Proceedings of International Conference on Machine Learning (ICML)*, 2014.