Dynamic Pricing with Limited Supply*

Moshe Babaioff† Shaddin Dughmi‡ Robert Kleinberg§ Aleksandrs Slivkins†

July 2011
Minor revision: February 2012

Abstract

We consider the problem of designing revenue maximizing online posted-price mechanisms when the seller has limited supply. A seller has k identical items for sale and is facing n potential buyers ("agents") that are arriving sequentially. Each agent is interested in buying one item. Each agent's value for an item is an independent sample from some fixed (but unknown) distribution with support $[0, 1]$. The seller offers a take-it-or-leave-it price to each arriving agent (possibly different for different agents), and aims to maximize his expected revenue.

We focus on mechanisms that do not use any information about the distribution; such mechanisms are called *detail-free* (or *prior-independent*). They are desirable because knowing the distribution is unrealistic in many practical scenarios. We study how the revenue of such mechanisms compares to the revenue of the optimal offline mechanism that knows the distribution ("offline benchmark").

We present a detail-free online posted-price mechanism whose revenue is at most $O((k \log n)^{2/3})$ less than the offline benchmark, for every distribution that is regular. In fact, this guarantee holds without *any* assumptions if the benchmark is relaxed to fixed-price mechanisms. Further, we prove a matching lower bound. The performance guarantee for the same mechanism can be improved to $O(\sqrt{k} \log n)$, with a distribution-dependent constant, if the ratio $\frac{k}{n}$ is sufficiently small. We show that, in the worst case over all demand distributions, this is essentially the best rate that can be obtained with a distribution-specific constant.

On a technical level, we exploit the connection to multi-armed bandits (MAB). While dynamic pricing with unlimited supply can easily be seen as an MAB problem, the intuition behind MAB approaches breaks when applied to the setting with limited supply. Our high-level conceptual contribution is that even the limited supply setting can be fruitfully treated as a bandit problem.

Keywords: mechanism design; revenue maximization; posted price; multi-armed bandits; regret.

1 Introduction

Consider an airline that is interested in selling k tickets for a given flight. The seller is interested in maximizing her revenue from selling these tickets, and is offering the tickets on a website such as Expedia. Potential buyers ("agents") arrive one after another, each with the goal of purchasing a ticket if the price is smaller than the agent's valuation. The seller expects n such agents to arrive. Whenever an agent arrives the seller presents to him a take-it-or-leave-it price, and the agent makes a purchasing decision according to that price.

*A preliminary version of this paper, titled "Detail-free, Posted-Price Mechanisms for Limited Supply Online Auctions", has appeared in the *Workshop on Bayesian Mechanism Design* at *ACM EC 2011*. That version did not include the results in Section 6.

†Microsoft Research Silicon Valley, Mountain View CA, USA. Email: {moshe,slivkins} @ microsoft.com.

‡Microsoft Research, Redmond WA, USA. Email: shaddin@microsoft.com.

§Department of Computer Science, Cornell University, Ithaca NY, USA. Email: rdk@cs.cornell.edu.

The seller can update the price taking into account the observed history and the number of remaining items and agents.

We adopt a Bayesian view that the valuations of the buyers are IID samples from a fixed distribution, called *demand distribution*. A standard assumption in a Bayesian setting is that the demand distribution is known to the seller, who can design a specific mechanism tailored to this knowledge. (For example, the Myerson optimal auction for one item sets a reserve price that is a function of the distribution). However, in some settings this assumption is very strong, and should be avoided if possible. For example, when the seller enters a new market, she might not know the demand distribution, and learning it through market research might be costly. Likewise, when the market has experienced a significant recent change, the new demand function might not be easily derived from the old data.

Ideally we would like to design mechanisms that perform well for any demand distribution, and yet do not rely on knowing it. Such mechanisms are called *detail-free*,[1] in the sense that the specification of the mechanism does not depend on the details of the "environment", in the spirit of Wilson's Doctrine [39]. Learning about the demand distribution is an integral part of the problem that a detail-free mechanism faces. The performance of such mechanisms is compared to a benchmark that *does* depend on the specific demand distribution, as in [30, 27, 12, 23] and many other papers.

In this paper we take this approach and design detail-free, online posted-price mechanisms with revenue that is close to the revenue of the optimal offline mechanism (that can depend on the demand distribution and is not restricted to be posted price). Our main results are for any demand distribution that is regular, or any demand distribution that satisfies the stronger condition of "monotone hazard rate". Both conditions are mild and standard, and even the stronger one is satisfied by most common distributions, such as the normal, uniform, and exponential distributions.

Posted price mechanisms are commonly used in practice, and are appealing for several reasons. First, an agent only needs to evaluate her offer rather than compute her private value exactly. Human agents tend to find the former task much easier than the latter. Second, agents do not reveal their entire private information to the seller: rather, they only reveal whether their private value is larger than the posted price. Third, posted-price mechanisms are truthful (in dominant strategies) and moreover also group strategy-proof (a notion of collusion resistance when side payments are not allowed). Further, detail-free posted-price mechanisms are particularly useful in practice as the seller is not required to estimate the demand distribution in advance. Similar arguments can be found in prior work, e.g. [20].

Our model. We consider the following limited supply auction model, which we term *dynamic pricing with limited supply*. A seller has k items she can sell to a set of n agents (potential buyers), aiming to maximize her expected revenue. The agents arrive sequentially to the market and the seller interacts with each agent before observing future agents (in an online manner). We make the simplifying assumption that each agent interacts with the seller only once, and the timing of the interaction cannot be influenced by the agent. (This assumption is also made in other papers that consider our problem for special supply amounts [30, 7, 12].) Each agent i ($1 \leq i \leq n$) is interested in buying one item, and has a private value v_i for an item. The private values are independently drawn from the same *demand distribution* F. The demand distribution F is *unknown* to the seller, but it is known that F has support in $[0, 1]$.[2]

Whenever agent i arrives to the market the seller offers him a price p_i for an item. The agent buys the item if and only if $v_i \geq p_i$, and in case she buys the item she pays p_i (so the mechanism is incentive-compatible). The seller never learns the exact value of v_i, she only observes the agent's binary decision to buy the item or not. The seller selects prices p_i using an online algorithm, that we henceforth call

[1] An alternative term used to describe these mechanisms is *prior-independent*.

[2] Assuming that $\texttt{support}(F) \subset [0, 1]$ is w.l.o.g. (by normalizing) as long as the seller knows an upper bound on the support.

pricing strategy. We are interested in designing pricing strategies with high revenue compared to a natural benchmark, with minimal assumptions on the demand distribution.

Our main benchmark is the maximal expected revenue of an offline mechanism that is allowed to use the demand distribution; henceforth, we will call it *offline benchmark*. This is a very strong benchmark, as it has the following advantages over our mechanism: it is allowed to use the demand distribution, it is not constrained to posted prices and is not constrained to run online. It is realized by a well-known Myerson Auction [35] (which *does* rely on knowing the demand distribution).

High-level discussion. Absent the supply constraint, our problem fits into the *multi-armed bandit* (MAB) framework [18]: in each round, an algorithm chooses among a fixed set of alternatives ("arms") and observes a payoff, and the objective is to maximize the total payoff over a given time horizon. Our setting corresponds to (prior-free) MAB with *stochastic payoffs* [31]: in each round, the payoff is an independent sample from some unknown distribution that depends on the chosen "arm" (price). This connection is exploited in [30, 15] for the special case of unlimited supply ($k = n$). The authors use a standard algorithm for MAB with stochastic payoffs, called `UCB1` [4]. Specifically, they focus on the prices $\{i\delta : i \in \mathbb{N}\}$, for some parameter δ, and run `UCB1` with these prices as "arms". The analysis relies on the regret bound from [4].

However, neither the analysis nor the intuition behind `UCB1` and similar MAB algorithms is directly applicable for the setting with limited supply. Informally, the goal of an MAB algorithm would be to converge to a price p that maximizes the expected per-round revenue $R(p) \triangleq p(1 - F(p))$. This is, in general, a wrong approach if the supply is limited: indeed, selling at a price that maximizes $R(\cdot)$ may quickly exhaust the inventory, in which case a higher price would be more profitable.

Our high-level conceptual contribution is showing that even the limited supply setting can be fruitfully treated as a bandit problem. The MAB perspective here is that we focus on the trade-off between *exploration* (acquiring new information) and *exploitation* (taking advantage of the information available so far). In particular, we recover an essential feature of `UCB1` that it does not separate exploration and exploitation, and instead explores arms (prices) according to a schedule that unceasingly adapts to the observed payoffs. This feature results, both for `UCB1` and for our algorithm, in a much more efficient exploration of suboptimal arms: very suboptimal arms are chosen very rarely even while they are being "explored".

We use an "index-based" algorithm where each arm is deterministically assigned a numerical score ("index") based on the past history, and in each round an arm with a maximal index is chosen; the index of an arm depends on the past history of this arm (and not on other arms). One key idea is that we define the index of an arm according to the estimated expected total payoff from this arm given the known constraints, rather than according to its estimated expected payoff in a single round. This idea leads to an algorithm that is simple and (we believe) very natural. However, while the algorithm is simple its analysis is not: some new ideas are needed, as the elegant tricks from prior work do not apply (see Section 4 for further discussion).

Contributions. In all results below, we consider the dynamic pricing problem with limited supply: n agents and $k \leq n$ items. We present pricing strategies with expected revenue that is close to the offline benchmark, for large families of natural distributions. All our pricing strategies are deterministic and (trivially) run in polynomial time. Our main result follows.

Theorem 1.1. *There exists a detail-free pricing strategy such that for any regular demand distribution its expected revenue is at least the offline benchmark minus* $O((k \log n)^{2/3})$.

We emphasize that Theorem 1.1 holds for a pricing strategy that does *not* know the demand distribution. The resulting mechanism is incentive-compatible as it is a posted price mechanism. The specific bound $O((k \log n)^{2/3})$ is most informative when $k \gg \log n$, so that the dependence on n is insignificant; the focus here is to optimize the power of k. (Note that any non-trivial bound must be below k.)

The proof of Theorem 1.1 consists of two stages. The first stage (immediate from Yan [40]) is to observe that for any regular demand distribution the expected revenue of the best fixed-price strategy[3] is close to the offline benchmark. Henceforth, the expected revenue of the best fixed-price strategy will be called the *fixed-price benchmark*. The second stage, which is our main technical contribution, is to show that our pricing strategy achieves expected revenue that is close to the fixed-price benchmark. Surprisingly, this holds without *any* assumptions on the demand distribution.

Theorem 1.2. *There exists a detail-free pricing strategy whose expected revenue is at least the fixed-price benchmark minus $O((k \log n)^{2/3})$. This result holds for every demand distribution. Moreover, this result is the best possible up to a factor of $O(\log n)$.*

As discussed above, we recover the MAB technique from [4] for the unlimited supply setting. The corresponding contribution to the literature on MAB may be of independent interest.

If the demand distribution is regular and moreover the ratio $\frac{k}{n}$ is sufficiently small then the guarantee in Theorem 1.1 can be improved to $O(\sqrt{k} \log n)$, with a distribution-specific constant.

Theorem 1.3. *There exists a detail-free pricing strategy whose expected revenue, for any regular demand distribution F, is at least the offline benchmark minus $O(c_F \sqrt{k} \log n)$ whenever $\frac{k}{n} \leq s_F$, where c_F and s_F are positive constants that depend on F. For monotone hazard rate distributions one can take $s_F = \frac{1}{4}$.*

The bound in Theorem 1.3 is achieved using the pricing strategy from Theorem 1.1 with a different parameter. Varying this parameter, we obtain a family of strategies that improve over the bound in Theorem 1.1 in the "nice" setting of Theorem 1.3, and moreover have non-trivial additive guarantees for arbitrary demand distributions. However, we cannot match both theorems with the same parameter.

Note that the rate-$\sqrt{k}$ dependence on k in Theorem 1.3 contains a distribution-dependent constant c_F (which can be arbitrarily large, depending on F), and thus is not directly comparable to the rate-$k^{2/3}$ dependence in Theorem 1.2. The distinction (and a significant gap) between bounds with and without distribution-dependent constants is not uncommon in the literature on sequential decision problems, e.g. in [4, 30, 29].[4]

In fact, we show that the $c_F \sqrt{k}$ dependence on k is essentially the best possible.[5] We focus on the fixed-price benchmark (which is a weaker benchmark, so it gives to a stronger lower bound). Following the literature, we define *regret* as the fixed-price benchmark minus the expected revenue of our pricing strategy.

Theorem 1.4. *For any $\gamma < \frac{1}{2}$, no detail-free pricing strategy can achieve regret $O(c_F\, k^{\gamma})$ for all demand distributions F and arbitrarily large k, n, where the constant c_F can depend on F.*

The bounds in Theorem 1.1 and Theorem 1.2 are uninformative when $k = O(\log^2 n)$. We next provide another detail-free, online posted-price mechanism that gives meaningful bounds – not depending on n – in the case that k is very small (but bigger than some constant).

Theorem 1.5. *There exists a detail-free pricing strategy such that for any MHR demand distribution its expected revenue is at least the offline benchmark minus $O(k^{3/4} \operatorname{poly} \log(k))$.*

[3] A fixed-price strategy is a pricing strategy that offers the same price to all agents, as long as it has items to sell. The "best" fixed-price strategy is one with the maximal expected revenue for a given demand distribution.

[4] For a particularly pronounced example, for the K-armed bandit problem with stochastic payoffs the best possible rates for regret with and without a distribution dependent constant are respectively $O(c_F \log n)$ and $O(\sqrt{Kn})$ [4, 5, 3].

[5] However, the lower bound in Theorem 1.4 does not match the upper bound in Theorem 1.3 since the latter assumes regularity.

2 Related Work

Dynamic pricing. Dynamic pricing problems and, more generally, revenue management problems, have a rich literature in Operations Research. A proper survey of this literature is beyond our scope; see [12] for an overview. The main focus is on parameterized demand distributions, with priors on the parameters.

The study of dynamic pricing with *unknown* demand distribution (without priors) has been initiated in [15, 30]. Several special cases of our setting have been studied in [30, 7, 12], detailed below.

First, Kleinberg and Leighton [30] consider the unlimited supply case (building on the earlier work [15]). Among other results, they study IID valuations, i.e. our setting with $k = n$. They provide upper bounds on regret of order $O(n^{2/3})$ and $O(c_F \sqrt{n})$. [6] The latter bound is akin to Theorem 1.3 in that it assumes a version of regularity, and depends on a distribution-specific constant c_F. Further, they prove matching lower bounds which, in particular, imply Theorem 1.4 for the special case of unlimited supply. [7]

On the other extreme, Babaioff et al. [7] consider the case that the seller has only one item to sell ($k = 1$). They provide a super-constant multiplicative lower bound for unrestricted demand distribution (with respect to the online optimal mechanism), and a constant-factor approximation assuming MHR. Note that we also use MHR to derive bounds that apply to the case of a very small k.

Besbes and Zeevi [12] consider a continuous-time version which (when specialized to discrete time) is essentially equivalent to our setting with $k = \Omega(n)$. They prove a number of upper bounds on regret with respect to the fixed-price benchmark, with guarantees that are inferior to ours. The key distinction is that their pricing strategies separate exploration and exploitation. Assuming that the demand distribution $F(\cdot)$ and its inverse $F^{-1}(\cdot)$ are Lipschitz-continuous, they achieve regret $O(n^{3/4})$. They improve it to $O(n^{2/3})$ if furthermore the demand distributions are parameterized, and to $O(\sqrt{n})$ if this is a single-parameter parametrization. Both results rely on knowing the parametrization: the mechanisms continuously update the estimates of the parameter(s) and revise the current price according to these estimates. The upper bounds in [12] should be contrasted with our $O(k^{2/3})$ upper bound that applies to an arbitrary k and makes no assumptions on the demand distribution, and the $O(c_F \sqrt{k})$ improvement for MHR demand distributions.

Also, [12] contains an $\Omega(\sqrt{n})$ lower bound for their notion of regret. Essentially, this lower bound compares the best pricing strategy for a given demand distribution to the best (distribution-dependent) pricing strategy for a fictitious environment where in every round the mechanism sells a fractional amount of good. In particular, this lower bound does not have any immediate implications on regret with respect to either of the two benchmarks that we use in this paper.

Online mechanisms. The study of online mechanisms was initiated by Lavi and Nisan [32], who unlike us consider the case that each agent is interested in multiple items, and provide a logarithmic multiplicative approximation. Below we survey only the most relevant papers in this line of work, in addition to the special cases of our setting that we have already discussed.

Several papers [10, 15, 30, 14] consider online mechanisms with unlimited supply and adversarial valuations (as opposed to limited supply and IID valuations in our setting). The mechanism in the initial paper [10] requires the agents to submit bids and so is not posted-price. The subsequent work [15, 30, 14] provides various improvements. In particular, Blum et al. [15] (among other results) design a simple *posted-price* mechanism which achieves multiplicative approximation $1 + \epsilon$, for any $\epsilon > 0$, with an additive term that depends on ϵ. [8] Blum and Hartline [14] use a more elaborate posted-price mechanism to improve the additive term. Kleinberg and Leighton [30] show that the simple mechanism in [15] achieves regret $O(n^{2/3})$;

[6] Throughout this section, we omit the $\log$ factors in regret bounds.

[7] The construction in [30] that proves Theorem 1.4(a) for the unlimited supply case is contained in the proof of a theorem on *adversarial* valuations, but the construction itself only uses IID valuations.

[8] This result considers valuations in the range $[1, H]$, and the additive term also depends on H.

moreover, they provide a nearly matching lower bound of $\Omega(n^{2/3})$.

Papers [26, 21] study online mechanisms for limited supply and IID valuations (same as us), but their mechanisms are not posted-price. Hajiaghayi et al. [26] consider an online auction model where players arrive and depart online, and may misreport the time period during which they participate in the auction. This makes designing strategy-proof mechanisms more challenging, and as a result their mechanisms achieve a constant multiplicative approximation rather than additive regret. Devanur and Hartline [21] study several variants of the limited-supply mechanism design problem: supply is known or unknown, online or offline. Most related to our paper is their mechanism for limited, known, online supply. This mechanism is based on random sampling and achieves constant (multiplicative) approximation, but is not posted-price. Our mechanism is posted-price and achieves low (additive) regret.

Other work. Absent the supply constraint, our problem (and a number of related formulations) fit into the *multi-armed bandit* (MAB) framework.[9] MAB has a rich literature in Statistics, Operations Research, Computer Science and Economics. A proper discussion of this literature is beyond the scope of this paper; a reader can refer to [18, 11] for background. Most relevant to our specific setting is the work on (prior-free) MAB with stochastic payoffs, e.g. [31, 4], and MAB with Lipschitz-continuous stochastic payoffs, e.g. [2, 28, 6, 29, 17]. The posted-price mechanisms in [15, 30, 14] described above are based on a well-known MAB algorithm [5] for adversarial payoffs. The connection between online learning and online mechanisms has been explored in a number of other papers, including [36, 22, 9, 8].

Recently, [20, 19, 40] studied the problem of designing an offline, sequential posted-price mechanisms in Bayesian settings, where the distributions of valuations are not necessarily identical, yet are known to the seller. Chawla et al. [20] provide constant multiplicative approximations. Yan [40] obtains a multiplicative bound that is optimal for large k, and Chakraborty et al. [19] obtain a PTAS for all k.

3 Preliminaries

Throughout, we assume that agents' valuations are drawn independently from a distribution F with support in $[0, 1]$, called *demand distribution*. We use $p \in [0, 1]$ to denote a price. We let $F(p)$ denote the c.d.f, and $S(p) = 1 - F(p)$ denote the *survival rate* at price p. Let $R(p) = p\,S(p)$ denote the *revenue function*: the expected single-round revenue at price p given that there is still at least one item left. The demand distribution F is called *regular* if $F(\cdot)$ is twice differentiable and the revenue function $R(\cdot)$ is concave: $R''(\cdot) \leq 0$. We call F *strictly regular* if furthermore $R''(\cdot) < 0$. Then $R(p)$ is increasing for $p \leq p_{\text{r}}$ and decreasing for $p \geq p_{\text{r}}$, where p_{r} is the unique maximizer, known as the *Myerson reserve price*. Moreover, the survival rate $S(\cdot)$ is strictly decreasing, so the inverse S^{-1} is well-defined. We say F is a *Monotone Hazard Rate (MHR)* distribution if $F(\cdot)$ is twice differentiable and the hazard rate $H(p) \triangleq F'(p)/S(p)$ is non-decreasing. All MHR distributions are regular.

A *fixed-price strategy* with n agents, k items and price p, denoted $\mathcal{A}^n_k(p)$, is a pricing strategy that makes a fixed offer price p to every agent so long as fewer than k items have been sold, and stops afterwards (equivalently, from that point always sets the price to ∞). Note that for the unlimited supply case $\mathcal{A}^n_n(p)$ sells $n\,S(p)$ items in expectation.

A pricing strategy is called *detail-free* if it does not use the knowledge of the demand distribution. We are interested in designing detail-free pricing strategies with good performance for *every* demand distribution in some (large) family of distributions. We compare our mechanisms to two benchmarks that depend on

[9] To void a possible confution, we note that the supply constraint in our setting may appear similar to the budget constraint in line of work on *budgeted MAB* (see [16, 25] for details and further references). However, the "budget" in budgeted MAB is essentially the duration of the experimentation phase (n), rather than the number of rounds with positive reward (k).

the demand distribution: the maximal expected revenue of an offline mechanism (the *offline benchmark*), and the maximal expected revenue of a fixed price mechanism (the *fixed-price benchmark*). An offline mechanism that maximizes expected revenue was given in the seminal paper of Myerson [35]; it is not an online posted price mechanism.

Let $\texttt{Rev}(\mathcal{A})$ be the total expected revenue achieved by mechanism $\mathcal{A}$. We define the *regret* of $\mathcal{A}$ with respect to the fixed-price benchmark as follows: $\texttt{Regret}(\mathcal{A}) \triangleq \max_p \texttt{Rev}[\mathcal{A}_k^n(p)] - \texttt{Rev}(\mathcal{A})$. Thus, regret is the additive loss in expected revenue compared to the best fixed-price mechanism. (Note that the regret of $\mathcal{A}$ could, in principle, be a negative number, since the fixed-price benchmark is not generally the Bayesian optimal pricing strategy for distribution F.)

Benchmarks Comparison. We observe that for regular demand distributions, the fixed-price benchmark is close to the offline benchmark. This result is immediate from Yan [40]; we provide a self-contained proof in Appendix A.

Lemma 3.1 (Yan [40])**.** *For each regular demand distribution there exists a fixed-price strategy whose expected revenue is at least the offline benchmark minus $O(\sqrt{k})$.*

Lemma 3.1 implies that any pricing strategy with regret $O(R)$, $R = \Omega(\sqrt{k})$ with respect to the fixed-price benchmark has the same asymptotic regret $O(R)$ with respect to the offline benchmark, as long as the demand distribution is regular, and in particular if it is MHR. Therefore, the rest of the paper can focus on the fixed-price benchmark. In particular, our main result, Theorem 1.1 for regular distributions, follows from Theorem 1.2 that addresses the fixed-price benchmark.

Furthermore, the expected revenue of a fixed-price mechanism has an easy characterization:

Claim 3.2. *Let $\mathcal{A}$ be the fixed-price mechanism with price p. Let $\nu(p) = p \min(k, n\,S(p)))$. Then*

$$\nu(p) - O(p\sqrt{k \log k}) \leq \texttt{Rev}(\mathcal{A}) \leq \nu(p). \tag{1}$$

It follows that for a strictly regular demand distribution the bound in Lemma 3.1 is satisfied for the fixed price $p^* = \operatorname{argmax}_p \nu(p) = \max(p_r,\, S^{-1}(\frac{k}{n}))$, where $p_r = \operatorname{argmax}_p p\,S(p)$ is the Myerson reserve price.

Proof. Let us focus on the first inequality in (1) (the second one is obvious). Let X_t be the indicator variable of sale in round t. Denote $X = \sum_{t=1}^{n} X_t$ and let $\mu = \mathbb{E}[X]$. Then by Chernoff Bounds (Theorem 4.7(a)) with probability at least $1 - \frac{1}{k}$ it holds that $X \geq \mu - O(\sqrt{\mu \log k})$, in which case

$$\text{\#sales} = \min(k, X) \geq \min(k, \mu - O(\sqrt{\mu \log k})) \geq \min(k, \mu) - O(\sqrt{k \log k}),$$

which implies the claim since $\mu = n\,S(p)$. □

4 The main technical result: the upper bound in Theorem 1.2

This section is devoted to the main technical result (the upper bound in Theorem 1.2) which asserts that there exists a detail-free pricing strategy whose regret with respect to the fixed-price benchmark is at most $O(k \log n)^{2/3}$. This result is very general, as it makes no assumptions on the demand distribution.

As discussed in Section 1, we design an algorithm that carefully optimizes the trade-off between exploration and exploitation. We use an *index-based* algorithm in which each arm is assigned a numerical score, called *index*, so that in each round an arm with the highest index is picked. The index of an arm depends only on the past history of this arm. In prior work on index-based bandit algorithms the index of an arm was

defined according to estimated expected payoff from this arm in a single round. Instead, we define the index according to estimated expected *total payoff* from this arm given the constraints.

We apply the above idea to UCB1. The index in UCB1 is, essentially, the best available Upper Confidence Bound (UCB) on the expected single-round payoff from a given arm. Accordingly, we define a new index, so that the index of a given price corresponds to a UCB on the expected total payoff from this price (i.e., from a fixed-price strategy with this price), given the number of agents and the inventory size. Such index takes into account both the average payoff from this arm ("exploitation") and the number of samples for this arm ("exploration"), as well as the supply constraint. In particular we recover the appealing property of UCB1 that it does not separate "exploration" and "exploitation", and instead explores arms (prices) according to a schedule that unceasingly adapts to the observed payoffs.

There are several steps to make this approach more precise. First, while it is tempting to use the current values for the number of agents and the inventory size to define the index, we adopt a non-obvious (but more elegant) design choice to use the original values, i.e. the n and the k. Second, since the exact expected total payoff for a given price is hard to quantify, we will instead use a natural approximation thereof provided by $\nu(p)$ in Claim 3.2. In other words, our index will be a UCB on $\nu(p)$. Third, in specifying the UCB we will use non-standard estimator from [29] to better handle prices with very low survival rate.

The main technical hurdle in the analysis is to "charge" each suboptimal price for each time that it is chosen, in a way that the total regret is bounded by the sum of these charges and this sum can be usefully bounded from above. The analysis of UCB1 accomplishes this via simple (but very elegant) tricks which, unfortunately, fail in the limited supply setting.

An additional difficulty comes from the probabilistic nature of the analysis. While we adopt a well-known trick – we define some high-probability events and assume that these events hold deterministically in the rest of the analysis – choosing an appropriate collection of events is, in our case, non-trivial. Proving that these events indeed hold with high probability relies on some non-standard tail bounds from prior work.

4.1 Our pricing strategy

Let us define our pricing strategy, called CappedUCB. The pricing strategy is initialized with a set $\mathcal{P}$ of "active prices". In each round t, some price $p \in \mathcal{P}$ is chosen. Namely, for each price $p \in \mathcal{P}$ we define a numerical score, called *index*, and we pick a price with the highest index, breaking ties arbitrarily. Once k items are sold, CappedUCB sets the price to ∞ and never sells any additional item.

Recall from Claim 3.2 that the expected revenue from the fixed-price strategy $\mathcal{A}^n_k(p)$ is approximated by $\nu(p) \triangleq p \cdot \min(k,\, n\, S(p))$. In each round t, we define the *index* $I_t(p)$ as a UCB on $\nu(p)$:

$$I_t(p) \triangleq p \cdot \min(k,\, n\, S^{\texttt{UB}}_t(p)).$$

Here $S^{\texttt{UB}}_t(p)$ is a UCB on the survival rate $S(p)$, as defined below.

For each $p \in \mathcal{P}$, let $N_t(p)$ be the number of rounds before t in which price p has been chosen, and let $k_t(p)$ be the number of items sold in these rounds. Then $\widehat{S}_t(p) \triangleq k_t(p)/N_t(p)$ is the current average survival rate. To avoid division by zero, we define $\widehat{S}_t(p)$ to be equal to 1 when $N_t(p) = 0$. We will define $S^{\texttt{UB}}_t(p) = \widehat{S}_t(p) + r_t(p)$, where $r_t(p)$ is a *confidence radius*: some number such that

$$|S(p) - \widehat{S}_t(p)| \leq r_t(p) \quad (\forall\, p \in \mathcal{P}, t \leq n). \tag{2}$$

holds with high probability, namely with probability at least $1 - n^{-2}$.

We need to define a suitable confidence radius $r_t(p)$, which we want to be as small as possible subject to (2). Note that $r_t(p)$ must be defined in terms of quantities that are observable at time t, such as $N_t(p)$ and $\widehat{S}_t(p)$. A standard confidence radius used in the literature is (essentially) $r_t(p) = \sqrt{\frac{\Theta(\log n)}{N_t(p)+1}}$.

Instead, we use a more elaborate confidence radius from [29]:

$$r_t(p) \triangleq \frac{\alpha}{N_t(p)+1} + \sqrt{\frac{\alpha\, \widehat{S}_t(p)}{N_t(p)+1}}, \quad \text{for some } \alpha = \Theta(\log n). \tag{3}$$

The confidence radius in (3) performs as well as the standard one in the worst case: $r_t(p) \leq \sqrt{\frac{O(\log n)}{N_t(p)+1}}$, and much better for very small survival rates: $r_t(p) \leq \frac{O(\log n)}{N_t(p)+1}$; see Appendix 4.3 for a self-contained proof.

To recap, we have

$$I_t(p) \triangleq p \cdot \min(k,\, n\,(\widehat{S}_t(p) + r_t(p))), \text{ where } r_t(p) \text{ is from (3).} \tag{4}$$

Finally, the active prices are given by

$$\mathcal{P} = \{\delta(1+\delta)^i \in [0,1] :\ i \in \mathbb{N}\}, \text{ where } \delta \in (0,1) \text{ is a parameter.} \tag{5}$$

This completes the specification of `CappedUCB`. See Mechanism 1 for the pseudocode.

Mechanism 1 Pricing strategy `CappedUCB` for n agents and k items

Parameter: $\delta \in (0,1)$

1: $\mathcal{P} \leftarrow \{\delta(1+\delta)^i \in [0,1] :\ i \in \mathbb{N}\}$ {"active prices"}

2: While there is at least one item left, in each round t
pick any price $p \in \operatorname{argmax}_{p \in \mathcal{P}} I_t(p)$, where $I_t(p)$ is the "index" given by (4).

3: For all remaining agents, set price $p = \infty$.

4.2 Analysis of the pricing strategy

Our goal is to bound from above the *regret* of `CappedUCB`, which is the difference between the optimal expected revenue of a fixed-price strategy and the expected revenue of `CappedUCB`. We prove that `CappedUCB` achieves regret $O(k \log n)^{2/3}$ for a suitable choice of parameter δ in (5).

Lemma 4.1. `CappedUCB` *with parameter* $\delta = k^{-1/3}\,(\log n)^{2/3}$ *achieves regret* $O(k \log n)^{2/3}$.

Since the bound in Lemma 4.1 is trivial for $k < \log^2 n$, we will assume that $k \geq \log^2 n$ from now on.

Note that `CappedUCB` "exits" (sets the price to ∞) after it sells k items. For a thought experiment, consider a version of this pricing strategy that does not "exit" and continues running as if it has unlimited supply of items; let us call this version `CappedUCB`$'$. Then the realized revenue of `CappedUCB` is exactly equal to the realized revenue obtained by `CappedUCB`$'$ from selling the first k items. Thus from here on we focus on analyzing the latter.

We will use the following notation. Let X_t be the indicator variable of the random event that `CappedUCB`$'$ makes a sale in round t. Note that X_t is a 0-1 random variable with expectation $S(p_t)$, where p_t depends on $X_1, \ldots, X_{t-1}$. Let $X \triangleq \sum_{t=1}^{n} X_t$ be the total number of sales if the inventory were unlimited. Note that $\mathbb{E}[X] = S \triangleq \sum_{t=1}^{n} S(p_t)$. Going back to our original algorithm, let $\widehat{\texttt{Rev}}$ denote the realized revenue of `CappedUCB` (revenue that is realized in a given execution). Then

$$\widehat{\texttt{Rev}} = \textstyle\sum_{t=1}^{N} p_t\, X_t, \qquad \text{where } N = \max\{N \leq n :\ \sum_{t=1}^{N} X_t \leq k\}. \tag{6}$$

High-probability events. We tame the randomness inherent in the sales X_t by setting up three high-probability events, as described below. In the rest of the analysis, we will argue deterministically under the assumption that these three events hold. It suffices because the expected loss in revenue from the low-probability failure events will be negligible. The three events are summarized in the following claim:

Claim 4.2. *With probability at least $1 - n^{-2}$ holds, for each round t and each price $p \in \mathcal{P}$:*

$$|S(p) - \widehat{S}_t(p)| \leq r_t(p) \leq 3 \left(\tfrac{\alpha}{N_t(p)+1} + \sqrt{\tfrac{\alpha\, S_t(p)}{N_t(p)+1}} \right), \tag{7}$$

$$|X - S| < O(\sqrt{S \log n} + \log n), \tag{8}$$

$$|\textstyle\sum_{t=1}^{n} p_t (X_t - S(p_t))| < O(\sqrt{S \log n} + \log n). \tag{9}$$

The probability bounds on the three events in Claim 4.2 are derived via appropriate concentration inequalities, some of which are non-standard; see Section 4.3 for further discussion. In the first event, the left inequality asserts that $r_t(p)$ is a confidence radius, and the right inequality gives the performance guarantee for it. The other two events focus on `CappedUCB'`, and bound the deviation of the total number of sales (X) and the realized revenue ($\sum_{t=1}^{n} p_t X_t$) from their respective expectations; importantly, these bound are in terms of $\sqrt{S}$ rather than $\sqrt{n}$.

In the rest of the analysis we will assume that the three events in Claim 4.2 hold deterministically.

Single-round analysis. Let us analyze what happens in a particular round t of the pricing strategy. Let p_t be the price chosen in round t. Let $p^*_{\text{act}} \in \operatorname{argmax}_{p \in \mathcal{P}} \nu(p)$ be the best active price according to $\nu(\cdot)$, and let $\nu^*_{\text{act}} \triangleq \nu(p^*_{\text{act}})$. Let $\Delta(p) \triangleq \max(0, \frac{1}{n}\nu^*_{\text{act}} - p\, S(p))$ be our notion of "badness" of price p, compared to the optimal approximate revenue ν^*. We will use this notation throughout the analysis, and eventually we will bound regret in terms of $\sum_{p \in \mathcal{P}} \Delta(p)\, N(p)$, where $N(p)$ is the total number of times price p is chosen.

Claim 4.3. *For each price $p \in \mathcal{P}$ it holds that*

$$N(p)\, \Delta(p) \leq O(\log n) \left(1 + \tfrac{k}{n}\, \tfrac{1}{\Delta(p)}\right). \tag{10}$$

Proof. By definition (2) of the confidence radius, for each price $p \in \mathcal{P}$ and each round t we have

$$\nu(p) \leq I_t(p) \leq p \cdot \min\left(k,\, n\, \left(S(p) + 2\, r_t(p)\right)\right). \tag{11}$$

Let us use this to connect each choice p_t with ν^*_{act}:

$$\begin{cases} I_t(p_t) \geq I_t(p^*_{\text{act}}) \geq \nu(p^*_{\text{act}}) \triangleq \nu^*_{\text{act}} \\ I_t(p_t) \leq p_t \cdot \min\left(k,\, n\, \left(S(p_t) + 2\, r_t(p_t)\right)\right). \end{cases}$$

Combining these two inequalities, we obtain the key inequality:

$$\tfrac{1}{n}\nu^*_{\text{act}} \leq p_t \cdot \min\left(\tfrac{k}{n},\, S(p_t) + 2\, r_t(p_t)\right). \tag{12}$$

There are several consequences for p_t and $\Delta(p_t)$:

$$\begin{cases} p_t & \geq & \frac{1}{k}\nu^*_{\text{act}} \\ \Delta(p_t) & \leq & 2\, p_t\, r_t(p_t) \\ \Delta(p_t) > 0 & \Rightarrow & S(p_t) < \frac{k}{n} \end{cases}. \tag{13}$$

The first two lines in (13) follow immediately from (12). To obtain the third line, note that $\Delta(p_t) > 0$ implies $p_t\, k \geq \nu^*_{\text{act}} > n\, p_t\, S(p_t)$, which in turn implies $S(p_t) < \frac{k}{n}$.

Note that we have not yet used the definition (3) of the confidence radius. For each price $p = p_t$, let t be the last round in which this price has been selected by the pricing strategy. Note that $N(p)$ (the total number of times price p is chosen) is equal to $N_t(p) + 1$. Then using the second line in (13) to bound $\Delta(p)$, Eq. (7) to bound the confidence radius $r_t(p)$, and the third line in (13) to bound the survival rate, we obtain:

$$\Delta(p) \leq O(p) \times \max\left(\tfrac{\log n}{N(p)},\ \sqrt{\tfrac{k}{n}\,\tfrac{\log n}{N(p)}}\right).$$

Rearranging the terms, we can bound $N(p)$ in terms of $\Delta(p)$ and obtain (10). □

Analyzing the total revenue. A key step is the following claim that allows us to consider $\sum_{t=1}^n p_t\, S(p_t)$ instead of the realized revenue $\widehat{\texttt{Rev}}$, effectively ignoring the capacity constraint. This is where we use the high-probability events (8) and (9). For brevity, let us denote $\beta(S) = O(\sqrt{S \log n} + \log n)$.

Claim 4.4. $\widehat{\texttt{Rev}} \geq \min(\nu^*_{act},\ \sum_{t=1}^n p_t\, S(p_t)) - \beta(k)$.

Proof. Recall that $p_t \geq \frac{1}{k}\nu^*_{\text{act}}$ by (13). It follows that $\widehat{\texttt{Rev}} \geq \nu^*_{\text{act}}$ whenever $\sum_{t=1}^n X_t > k$. Therefore, if $\widehat{\texttt{Rev}} < \nu^*_{\text{act}}$ then $\sum_{t=1}^n X_t \leq k$ and so $\widehat{\texttt{Rev}} = \sum_{t=1}^n p_t\, X_t$. Thus, by (9) it holds that

$$\widehat{\texttt{Rev}} \geq \min\left(\nu^*_{\text{act}},\ \textstyle\sum_{t=1}^n p_t\, X_t\right) \geq \min\left(\nu^*_{\text{act}},\ \textstyle\sum_{t=1}^n p_t\, S(p_t) - \beta(S)\right).$$

So the claim holds when $S \leq k$. On the other hand, if $S > k$ then by (8) it holds that

$$\begin{aligned} X &\geq S - \beta(S) \geq k - \beta(k) \\ \widehat{\texttt{Rev}} &\geq \min(k, X)\,(\tfrac{1}{k}\,\nu^*_{\text{act}}) \geq \nu^*_{\text{act}} - \beta(k). \end{aligned}$$

□

In light of Claim 4.4, we can now focus on $\sum_{t=1}^n p_t\, S(p_t)$.

$$\begin{aligned} \textstyle\sum_{t=1}^n p_t\, S(p_t) &\geq \textstyle\sum_{t=1}^n \frac{1}{n}\,\nu^*_{\text{act}} - \Delta(p_t) \\ &= \nu^*_{\text{act}} - \textstyle\sum_{t=1}^n \Delta(p_t) \\ &= \nu^*_{\text{act}} - \textstyle\sum_{p\in\mathcal{P}} \Delta(p)\, N(p). \end{aligned} \tag{14}$$

Fix a parameter $\epsilon > 0$ to be specified later, and denote

$$\begin{cases} \mathcal{P}_{\text{sel}} & \triangleq \{p \in \mathcal{P} :\ N(p) \geq 1\} \\ \mathcal{P}_\epsilon & \triangleq \{p \in \mathcal{P}_{\text{sel}} : \Delta(p) \geq \epsilon\} \end{cases}$$

to be, respectively, be the set of prices that have been selected at least once and the set of prices of badness at least ϵ that have been selected at least once. Plugging (10) into (14), we obtain

$$\begin{aligned} \textstyle\sum_{p\in\mathcal{P}} \Delta(p)\, N(p) &\leq \textstyle\sum_{p\in\mathcal{P}_{\text{sel}}\setminus\mathcal{P}_\epsilon} \Delta(p)\, N(p) + \sum_{p\in\mathcal{P}_\epsilon} \Delta(p)\, N(p) \\ &\leq \epsilon n + O(\log n) \textstyle\sum_{p\in\mathcal{P}_\epsilon} \left(1 + \frac{k}{n}\frac{1}{\Delta(p)}\right) \\ &\leq \epsilon n + O(\log n)\left(|\mathcal{P}_\epsilon| + \frac{k}{n}\textstyle\sum_{p\in\mathcal{P}_\epsilon} \frac{1}{\Delta(p)}\right). \end{aligned} \tag{15}$$

Combining (14), (15) and Claim 4.4 yields a claim that summarizes our findings so far.

Claim 4.5. *For any set $\mathcal{P}$ of active prices and any parameter $\epsilon > 0$ it holds that*

$$\nu^*_{act} - \mathbb{E}[\widehat{\texttt{Rev}}] \leq \epsilon n + O(\log n) \left(|\mathcal{P}_\epsilon| + \tfrac{k}{n} \sum_{p \in \mathcal{P}_\epsilon} \tfrac{1}{\Delta(p)} \right) + \beta(k).$$

Interestingly, this claim holds for any set of active prices. The following claim, however, takes advantage of the fact that the active prices are given by (5).

Claim 4.6. *$\nu^*_{act} \geq \nu^* - \delta k$, where $\nu^* \triangleq \max_p \nu(p)$.*

Proof. Let $p^* \in \operatorname{argmax}_p \nu(p)$ denote the best fixed price with respect to $\nu(\cdot)$, ties broken arbitrarily. If $p^* \leq \delta$ then $\nu^* \leq \delta k$. Else, letting $p_0 = \max\{p \in \mathcal{P} : p \leq p^*\}$ we have $p_0/p \geq \frac{1}{1+\delta} \geq 1 - \delta$, and so

$$\nu^*_{\text{act}} \geq \nu(p_0) \geq \tfrac{p_0}{p^*}\, \nu(p^*) \geq \nu^*(1-\delta) \geq \nu^* - \delta k. \qquad \square$$

It follows that for any $\epsilon > 0$ and $\delta \in (0,1)$ we have:

$$\texttt{Regret} \leq O(\log n) \left(|\mathcal{P}_\epsilon| + \tfrac{k}{n} \sum_{p \in \mathcal{P}_\epsilon} \tfrac{1}{\Delta(p)} \right) + \epsilon n + \delta k + \beta(k). \tag{16}$$

The rest is a standard computation. Plugging in $\Delta(p) \geq \epsilon$ for each $p \in \mathcal{P}_\epsilon$ in (16), we obtain:

$$\texttt{Regret} \leq O(|\mathcal{P}_\epsilon| \log n) \left(1 + \tfrac{1}{\epsilon}\, \tfrac{k}{n}\right) + \epsilon n + \delta k + \beta(k).$$

Note that $|\mathcal{P}| \leq \frac{1}{\delta} \log n$. To simplify the computation, we will assume that $\delta \geq \frac{1}{n}$ and $\epsilon = \delta\, \frac{k}{n}$. Then

$$\texttt{Regret} \leq O\left(\delta k + \tfrac{1}{\delta^2} (\log n)^2 + \sqrt{k \log n} \right). \tag{17}$$

Finally, it remains to pick δ to minimize the right-hand side of (17). Let us simply take δ such that the first two summands are equal: $\delta = k^{-1/3} (\log n)^{2/3}$. Then the two summands are equal to $O(k \log n)^{2/3}$. This completes the proof of Lemma 4.1.

4.3 Concentration inequalities and the proof of Claim 4.2

We use an elementary concentration inequality known as *Chernoff Bounds*, in a formulation from [34].

Theorem 4.7 (Chernoff Bounds)**.** *Consider n i.i.d. random variables $X_1 \ldots X_n$ with values in $[0,1]$. Let $X = \frac{1}{n} \sum_{i=1}^n X_i$ be their average, and let $\mu = \mathbb{E}[X]$. Then:*

(a) $\Pr[|X - \mu| > \delta\mu] < 2\, e^{-\mu n \delta^2/3}$ *for any $\delta \in (0,1)$.*

(b) $\Pr[X > a] < 2^{-an}$ *for any $a > 6\mu$.*

Further, we use a non-standard corollary from [29] [10] which provides us with a sharper (i.e., smaller) confidence radius when μ is small; we include the proof for the sake of completeness.

Theorem 4.8 ([29])**.** *Consider n i.i.d. random variables $X_1 \ldots X_n$ on $[0,1]$. Let X be their average, and let $\mu = \mathbb{E}[X]$. Then for any $\alpha > 0$, letting $r(\alpha, x) = \frac{\alpha}{n} + \sqrt{\frac{\alpha x}{n}}$, we have:*

$$\Pr\left[\, |X - \mu| < r(\alpha, X) < 3\, r(\alpha, \mu) \,\right] > 1 - e^{-\Omega(\alpha)},$$

[10] This is Lemma 4.9 in the full (arXiv) version of [29].

Proof. First, suppose $\mu \geq \frac{\alpha}{6n}$. Apply Theorem 4.7(a) with $\delta = \frac{1}{2}\sqrt{\frac{\alpha}{6\mu n}}$. Thus with probability at least $1 - e^{-\Omega(\alpha)}$ we have $|X - \mu| < \delta\mu \leq \mu/2$. Plugging in the δ,

$$|X - \mu| < \tfrac{1}{2}\sqrt{\tfrac{\alpha\mu}{n}} \leq \sqrt{\tfrac{\alpha X}{n}} \leq r(\alpha, X) < 1.5\, r(\alpha, \mu).$$

Now suppose $\mu < \frac{\alpha}{6n}$. Then using Theorem 4.7(b) with $a = \frac{\alpha}{n}$, we obtain that with probability at least $1 - 2^{-\Omega(\alpha)}$ we have $X < \frac{\alpha}{n}$, and therefore $|X - \mu| < \frac{\alpha}{n} < r(\alpha, X)$ and

$$|X - \mu| < \tfrac{\alpha}{n} < r(\alpha, X) < (1 + \sqrt{2})\, \tfrac{\alpha}{n} < 3\, r(\alpha, \mu). \quad \square$$

Proof of (7) in Claim 4.2. For each price $p \in \mathcal{P}$ let $\{Z_{i,p}\}_{i \leq n}$ be a family of independent 0-1 random variables with expectation $S(p)$. Without loss of generality, let us pretend that the i-th time that price p is selected by the pricing strategy, sale happens if and only if $Z_{i,p} = 1$. Then by Lemma 4.8 after the i-th play of price p the bound (7) holds with probability at least $1 - n^{-4}$. Taking the Union Bound over all choices of i and all choices of p, we obtain that (7) holds with probability at least $1 - n^{-2}$ as long as $|\mathcal{P}| \leq n$ (which is the case for us). $\square$

Sharper Azuma-Hoeffding inequality. We use a concentration inequality on the sum of n random variables $X_t \in \{0, 1\}$ such that each variable X_t is a random coin toss with probability M_t that depends on the previous variables $X_1, \ldots, X_{t-1}$. We are interested in bounding the deviation $|X - M|$, where $X = \sum_t X_t$ and $M = \sum_t M_t$. The well-known Azuma-Hoeffding inequality states that with high probability we have $|X - M| \leq O(\sqrt{n \log n})$. However, we need a sharper high-probability bound: $|X - M| \leq O(\sqrt{M \log n})$. Moreover, we need an extension of such bound which considers deviation $|\sum_{t=1}^{n} \alpha_t (X_t - M_t)|$, where each multiplier $\alpha_t \in [0, 1]$ is determined by $X_1, \ldots, X_{t-1}$.

We use the following concentration inequality from the literature.

Theorem 4.9 (Theorem 3.15 in [33])**.** *Let $Z_1, \ldots, Z_n$ be random variables which take values in $[-1, 1]$. Let $Z = \sum_{t=1}^{n} Z_t$, $\mu = \mathbb{E}[Z]$. Let $V = \sum_{t=1}^{n} \mathit{Var}(Z_t | Z_1, \ldots, Z_{t-1})$. Then for any $a > 0, v > 0$ we have*

$$\Pr\left[(|Z - \mu| \geq a) \wedge (V \leq v)\right] \leq e^{-\Omega(\frac{a^2}{v+a})}.$$

We use the above bound to bound the deviation for $|\sum_{t=1}^{n} \alpha_t (X_t - M_t)|$.

Theorem 4.10. *Let $X_1, \ldots, X_n$ be 0-1 random variables. Let $M = \sum_{t=1}^{n} \mathbb{E}[X_t | X_1, \ldots, X_{t-1}]$. For each t, let $\alpha_t \in [0, 1]$ be the multiplier determined by $X_1, \ldots, X_{t-1}$. Then for any $b \geq 1$ the event*

$$|\textstyle\sum_{t=1}^{n} \alpha_t (X_t - M_t)| \leq b(\sqrt{M \log n} + \log n).$$

holds with probability at least $1 - n^{-\Omega(b)}$.

Proof. Let $Z_t = X_t - y_t$, where $y_t \in [0, 1]$ is a function of $X_1, \ldots, X_{t-1}$, and let $Z = \sum_{t=1}^{n} Z_t$.

We claim that

$$\Pr\left[|\textstyle\sum_{t=1}^{n} \alpha_t (Z_t - \mathbb{E}[Z_t])| \leq b(\sqrt{M \log n} + \log n)\right] \geq 1 - n^{-\Omega(b)}, \quad \text{for any } b \geq 1. \tag{18}$$

To prove (18), let $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$ be the σ-algebra generated by $X_1, \ldots, X_t$, and let $M_t = \mathbb{E}[X_t | X_1, \ldots, X_{t-1}]$. Then conditional on $\mathcal{F}_{t-1}$, Z_t is a random variable with expectation $M_t - y_t$ and two possible values, $-\alpha_t\, y_t$ and $\alpha_t\, (1 - y_t)$, where α_t and y_t are constants. It follows that $\mathrm{Var}(Z_t | \mathcal{F}_{t-1}) = \alpha_t^2 (M_t - M_t^2) \leq M_t$, and therefore $V \triangleq \sum_{t=1}^{n} \mathrm{Var}(Z_t | \mathcal{F}_{t-1}) \leq M$.

Taking Theorem 4.9 with $a = b(\sqrt{v \log n} + \log n)$, we have that for any $b \geq 1$ the event

$$(|Z - E[Z]| \geq b(\sqrt{v \log n} + \log n)) \wedge (V \leq v).$$

holds with probability at most $n^{-\Omega(b)}$. Finally, we take the Union Bound over (say) all integer v between $\log n$ and n, noting that $V \leq M$. This completes the proof of (18).

Finally, to prove the theorem take (18) with $y_t = M_t$ and note that $Z_t = X_t - M_t$ and so $\mathbb{E}[Z_t] = 0$. □

Proof of (8) ***and*** (9) ***in Claim 4.2.*** Recall that for each t, X_t is a 0-1 random variable with expectation $S(p_t)$, where p_t depends on $X_1, \ldots, X_{t-1}$. Using Lemma 4.10 with $\alpha_t \equiv 1$ we obtain (8). Using Lemma 4.10 with $\alpha_t = p_t$ we obtain (9). □

5 The $O(\sqrt{k} \log n)$ regret bound (Theorem 1.3)

We show that the pricing strategy from Section 4 (with a different parameter) satisfies an improved regret bound, $O(\sqrt{k} \log n)$, if the demand distribution is regular and moreover the ratio $\frac{k}{n}$ is sufficiently small. The regret bound depends on a distribution-specific constant.

Theorem 5.1. *For any regular demand distribution F there exist positive constants s_F and c_F such that* `CappedUCB` *with parameter $\delta = k^{-1/2} \log(n)$ achieves regret $O(c_F \sqrt{k} \log n)$ whenever $\frac{k}{n} \leq s_F$. For monotone hazard rate distributions we can take $s_F = \frac{1}{4}$.*

Proof. Let $g(s) \triangleq s\, S^{-1}(s)$ be a function from $[S(1), 1]$ to $[0, 1]$ that maps a survival rate to the corresponding revenue. Regularity implies $g''(\cdot) \leq 0$. Since $g'(0) > 0$, we can pick a constant $s_F > 0$ such that $C \triangleq g'(s_F) > 0$. For monotone hazard rate distributions we can take $s_F = \frac{1}{4}$ because for any maximizer s of $g(\cdot)$ it holds that $s \geq \frac{1}{e}$ (see Claim B.2). Now, for any $\frac{k}{n} \leq s_F$ we have that $g'(\frac{k}{n}) \geq C$. We will use this to obtain a lower bound on $\Delta(p)$; any such lower bound is absent in the analysis in Section 4. This improvement results in savings in (16), which in turn implies the claimed regret bound.

We will use the notation from Section 4.2, particularly the "badness" $\Delta(p)$ and the set $\mathcal{P}_\epsilon$ of arms of badness $\geq \epsilon$ that have been selected at least once. Note that by regularity $g'(s) \geq C$ for any $s \in (0, \frac{k}{n})$. Let $p^* = S^{-1}(\frac{k}{n})$ and $p \in \mathcal{P}_\epsilon$. By the third line in (13) it holds that $S(p) < \frac{k}{n}$ and then $p > p^*$.

First, we claim that $S(p) < \frac{p^*}{p} \frac{k}{n}$. Indeed, this is because $p\, S(p) = g(S(p)) < g(\frac{k}{n}) = p^* \frac{k}{n}$.

Second, we bound $\Delta(p)$ from below:

$$\begin{aligned}
\tfrac{1}{n} \nu^*_{\text{act}} &\geq (1-\delta)\, \tfrac{\nu^*}{n} \geq (1-\delta)\, g(\tfrac{k}{n}) \\
\Delta(p) &\geq (1-\delta)\, g(\tfrac{k}{n}) - g(S(p)) \\
&\geq [g(\tfrac{k}{n}) - g(S(p))] - \delta\, g(\tfrac{k}{n}) \\
&\geq C(\tfrac{k}{n} - S(p)) - \delta \tfrac{k}{n}\, p^* \\
&\geq C\, \tfrac{k}{n}\, (1 - \tfrac{p^*}{p}) - \delta \tfrac{k}{n}\, p^* \\
&\geq C\, \tfrac{k}{n}\, (1 - \tfrac{p^*}{p}(1 + \tfrac{\delta}{C})).
\end{aligned}$$

Since $\mathcal{P}$ is given by (5), it holds that $\mathcal{P}_\epsilon \subset \{p^* \alpha\, (1+\delta)^i : i \in \mathbb{N}\}$ for some $\alpha \geq 1$. Define

$$\mathcal{P}' \triangleq \{p \in \mathcal{P}_\epsilon : p = p^* \alpha\, (1+\delta)^i \text{ with } i \geq \tfrac{2}{C}\}.$$

Then for any $p \in \mathcal{P}'$ it holds that $p/p^* = \alpha(1+\delta)^i \geq 1 + i\delta$ and therefore

$$\Delta(p) \geq C\, \tfrac{k}{n}\, (1 - \tfrac{1+\delta/C}{1+i\delta}) \geq \tfrac{C}{2}\, \tfrac{k}{n}\, \tfrac{i\delta}{1+i\delta}.$$

Therefore, noting that $|\mathcal{P}'| \leq |\mathcal{P}| \leq O(\frac{1}{\delta}\,\log\frac{1}{\delta})$, we have

$$\begin{aligned}
&\tfrac{k}{n}\sum_{p\in\mathcal{P}'}\tfrac{1}{\Delta(p)} \leq \tfrac{2}{C}\sum_{p\in\mathcal{P}'}(1+\tfrac{1}{i\delta}) \leq \tfrac{2}{C}\,(|\mathcal{P}'|+\tfrac{1}{\delta}\,\log|\mathcal{P}'|) \leq O(\tfrac{1}{C}\tfrac{1}{\delta}\,\log\tfrac{1}{\delta})\\
&\textstyle\sum_{p\in\mathcal{P}_\epsilon\setminus\mathcal{P}'}\tfrac{1}{\Delta(p)} \leq \tfrac{1}{\epsilon}\,|\mathcal{P}\setminus\mathcal{P}'| \leq \tfrac{1}{\epsilon}\,(\tfrac{2}{C}+1).
\end{aligned}$$

Plugging this into (16) with $\epsilon = \delta\,\frac{k}{n}$, we obtain:

$$\begin{aligned}
\tfrac{k}{n}\textstyle\sum_{p\in\mathcal{P}_\epsilon}\tfrac{1}{\Delta(p)} &\leq O(\tfrac{1}{\delta}\,\log\tfrac{1}{\delta})(1+\tfrac{1}{C})\\
\texttt{Regret} &\leq O(\delta k + \tfrac{1}{\delta}(1+\tfrac{1}{C})(\log n)^2 + \sqrt{k\log n}) \qquad (19)\\
&\leq O(c_F\,\sqrt{k}\log n),\ \text{where } c_F = 1+1/C.
\end{aligned}$$

The regret bound (19) improves over the corresponding bound (17) in Section 4. We obtain the final bound by plugging $\delta = k^{-1/2}\,\log n$. □

It is desirable to achieve the bounds in Theorem 1.2 and Theorem 5.1 using the same pricing strategy. Unfortunately, the choice of parameter δ in Theorem 5.1 results in a trivial $O(k)$ regret guarantee for arbitrary demand distributions (as per Equation (17)). However, varying δ and using Equations (17) and (19) we obtain a family of pricing strategies that improve over the bound in Theorem 1.2 for the "nice" setting in Theorem 5.1, and moreover have non-trivial regret bounds for arbitrary demand distributions.

Theorem 5.2. *For each $\gamma \in [\frac{1}{3}, \frac{1}{2}]$, consider pricing strategy* `CappedUCB` *with parameter $\delta = \widetilde{O}(k^{-\gamma})$. This pricing strategy achieves regret $\widetilde{O}(k^{1-\gamma})(1+1/g'(\frac{k}{n}))$ if the demand distribution is regular and $g'(\frac{k}{n}) > 0$, and regret $\widetilde{O}(k^{2\gamma})$ for arbitrary demand distributions.*

6 Lower Bounds

We prove two lower bounds on regret over all demand distributions which match the upper bounds in Theorem 1.2 and Theorem 1.3, respectively. (Note that the latter upper bound is specific to regular distributions.) Throughout this section, *regret* is with respect to the fixed-price benchmark.

Theorem 6.1. *Consider the dynamic pricing problem with limited supply: with n agents and $k \leq n$ items.*

(a) No detail-free pricing strategy can achieve regret $o(k^{2/3})$ for arbitrarily large k, n.

(b) For any $\gamma < \frac{1}{2}$, no detail-free pricing strategy can achieve regret $O(c_F\,k^{\gamma})$ for all demand distributions F and arbitrarily large k, n, where the constant c_F can depend on F.

Our proof is a black-box reduction to the unlimited supply case ($k = n$). The unlimited supply case of Theorem 6.1 is proved in [30] (see Footnote 7 on page 5).

Proof. Suppose that some pricing strategy $\mathcal{A}$ violates part (a). Then there is a sequence $\{k_i, n_i\}_{i\in\mathbb{N}}$, where $k_i \leq n_i$ and $\{k_i\}_{i\in\mathbb{N}}$ is strictly increasing, such that $\mathcal{A}$ achieves regret $o(k^{2/3})$ for all problem instances with n_i agents and k_i items, for each $i \in \mathbb{N}$. To obtain a contradiction, let us use $\mathcal{A}$ to solve the unlimited supply problem with regret $o(n^{2/3})$. Specifically, we will solve problem instances with $k_i/4$ agents, for each i.

Fix $i \in \mathbb{N}$ and let $k = k_i$ and $n = n_i$. Consider a problem instance $\mathcal{I}$ with unlimited supply and $k/4$ agents and survival rate $S(\cdot)$. Let $\mathcal{I}'$ be an artificial problem instance with unlimited supply and n agents, so that the first $k/4$ agents in $\mathcal{I}'$ correspond to $\mathcal{I}$. Form an artificial problem instance $\mathcal{J}$ with k items and n agents as follows: in each round, $\mathcal{A}$ outputs a price, then with probability $k/2n$ this price is offered to the next agent in $\mathcal{I}'$, and with the remaining probability there is no interaction with agents in $\mathcal{I}'$ and no sale.

Since the demand distribution for $\mathcal{J}$ is a mixture of the "no sale" event which happens with probability $1 - \frac{k}{2n}$ and the original demand distribution for $\mathcal{I}$, the survival rate for J is given by $S_{\mathcal{J}}(p) = \frac{k}{2n} S(p)$.

Running $\mathcal{A}$ on problem instance $\mathcal{J}$ induces a pricing strategy $\mathcal{A}'$ on the original problem instance $\mathcal{I}$.[11] In the rest of the proof we show that $\mathcal{A}'$ achieves regret $o(k^{2/3})$ on $\mathcal{I}$.

Let $\texttt{Rev}_{\mathcal{J}}(\mathcal{A})$ and $\widehat{\texttt{Rev}}_{\mathcal{J}}(\mathcal{A})$ be, respectively, the expected revenue and the realized revenue of $\mathcal{A}$ on problem instance $\mathcal{J}$. Let $r = \operatorname{argmax}_p pS(p)$ be the Myerson reserve price, and let $\mathcal{A}_r$ be the fixed-price strategy with price r. By our assumption, we have that $\texttt{Rev}_{\mathcal{J}}(\mathcal{A}) \geq \texttt{Rev}_{\mathcal{J}}(\mathcal{A}_r) - o(k^{2/3})$. We need to deduce that $\texttt{Rev}_{\mathcal{I}}(\mathcal{A}') \geq \texttt{Rev}_{\mathcal{I}}(\mathcal{A}_r) - o(k^{2/3})$.

Let N be the number of rounds in $\mathcal{J}$ in which $\mathcal{A}$ interacts with the agents in $\mathcal{I}'$. With high probability $\frac{k}{4} < N < k$. Let us condition on N and the event $\mathcal{E}_N \triangleq \{k/4 < N < k\}$:

$$\mathbb{E}[\,\widehat{\texttt{Rev}}_{\mathcal{J}}(\mathcal{A}_r) \,|\, N, \mathcal{E}_N\,] = NrS(r)$$
$$\mathbb{E}[\,\widehat{\texttt{Rev}}_{\mathcal{J}}(\mathcal{A}) - \widehat{\texttt{Rev}}_{\mathcal{I}}(\mathcal{A}') \,|\, N, \mathcal{E}_N\,] \leq (N - \tfrac{k}{4})\, rS(r).$$

Since $\mathbb{E}[N] = \frac{k}{2}$, it follows that

$$\begin{aligned}
\texttt{Rev}_{\mathcal{I}}(\mathcal{A}') &\geq \texttt{Rev}_{\mathcal{J}}(\mathcal{A}) - \tfrac{k}{4}\, rS(r) - o(1) \\
&\geq \texttt{Rev}_{\mathcal{J}}(\mathcal{A}_r) - \tfrac{k}{4}\, rS(r) - o(k^{2/3}) \\
&= \tfrac{k}{4}\, rS(r) - o(k^{2/3}) \\
&= \texttt{Rev}_{\mathcal{I}}(\mathcal{A}_r) - o(k^{2/3}),
\end{aligned}$$

as required. The reduction for part (b) proceeds similarly. □

7 Selling very few items: proof of Theorem 1.5

In this section we target a case when very few items are available for sale (roughly, $k < O(\log^2 n)$), so that the bound in Theorem 1.1 becomes trivial. We provide a different pricing strategy whose regret does not depend on n, under the mild assumption of monotone hazard rate.

We rely on the characterization in Claim 3.2: we look for the price $p^* = \max(p_{\mathrm{r}},\, S^{-1}(\frac{k}{n}))$, where $p_{\mathrm{r}} = \operatorname{argmax}_p p\, S(p)$ is the Myerson reserve price. The pricing strategy proceeds as follows (see Mechanism 2 on page 17). It considers prices $p_\ell = (1-\delta)^\ell$, $\ell \in \mathbb{N}$ sequentially in the descending order. For each ℓ, it offers the price p_ℓ to a fixed number of agents. The loop stops once the pricing strategy detects that, essentially, the "best" p_ℓ has been reached: either $S(p_\ell)$ is close to $\frac{k}{n}$, or we are near a maximum of $p\, S(p)$. Parameters are chosen so as to minimize regret.

Theorem 7.1. *For some parameters ϵ and δ, Mechanism 2 achieves regret $O\left(k^{3/4} \operatorname{poly}\log(k)\right)$ with respect to the offline benchmark, for any demand distribution that satisfies the monotone hazard rate condition.*

The rest of this section is devoted to proving Theorem 7.1 for parameters $\epsilon = k^{-1/4}$ and $\delta = (\frac{1}{k}\, \log k)^{1/4}$. We will assume that the demand distribution is MHR, without further notice. We derive Theorem 7.1 from the following multiplicative bound; it appears difficult to prove the additive version directly.

Lemma 7.2. *Assume $p^* \geq \epsilon$. Set $\delta = \sqrt[4]{\frac{1}{k}\, \log k \log \frac{1}{\epsilon} \log\log \frac{1}{\epsilon}}$. Then the expected revenue of Mechanism 2 is at least $1 - O(\delta)$ fraction of the offline benchmark.*

Proof of Theorem 7.1. If $p^* \leq \epsilon$ then the expected loss in revenue is at most ϵk. Else by Lemma 7.2 the expected loss in revenue is at most $O(\delta k)$, where δ is from Lemma 7.2. In both cases the additive regret compared to the offline benchmark is at most $\max(\epsilon k, O(k\delta))$. Finally, pick $\epsilon = k^{-1/4}$. □

[11] If $\mathcal{A}$ stops before it iterates through all agents in $\mathcal{I}$, the remaining agents in $\mathcal{I}$ are offered a price of ∞.

Mechanism 2 Descending prices

Parameter: Approximation parameters $\delta, \epsilon \in [0,1]$

Let $\alpha = \left(\frac{k}{n}\right)^{1-\delta}$, $\gamma = \min(\alpha, 1/e)$.
$\ell \leftarrow 0,\ \ell_{\max} \leftarrow 0,\ R_{\max} \leftarrow 0$.
repeat
  $\ell \leftarrow \ell + 1,\ p_\ell \leftarrow (1+\delta)^{-\ell}$
  Offer price p_ℓ to $m = \lceil \delta \frac{n}{\log_{1+\delta}(1/\epsilon)} \rceil$ agents.
  Let S_ℓ be the fraction of them who accept.
  Let $R_\ell = p_\ell S_\ell$ be the average per agent revenue.
  If $S_\ell \geq (1+\delta)^{-1}\gamma$ and $R_\ell \geq R_{\max}$,
    then $R_{\max} \leftarrow R_\ell,\ \ell_{\max} \leftarrow \ell$
until $p_\ell \leq \epsilon$ or $S_\ell \geq (1+\delta)\alpha$ or $R_\ell \leq (1+\delta)^{-2} R_{\max}$
Offer price $\widetilde{p} = p_\ell$ so long as unsold items remain.

7.1 Proof of Lemma 7.2

We use a multiplicative bound in which fixed-price strategies for limited supply are compared to those for unlimited supply (which in turn can be compared to the offline benchmark using Claim A.2).

Lemma 7.3. *Assume the demand distribution is regular. Let $p' \leq p$ be two prices such that $p \geq S^{-1}(k/n)$. Let $n' \leq n$. Then* $\texttt{Rev}(\mathcal{A}_k^{n'}(p')) \geq \frac{n'}{n}\frac{p'}{p}\left(1 - \frac{1}{\sqrt{2\pi k}}\right)\texttt{Rev}(\mathcal{A}_n^n(p))$.

The proof uses a technique from [40], see Appendix A. Also, we take advantage of several properties of MHR distributions, detailed in Appendix B.

We say the exploration phase is δ-*approximate* if

$$S(p_\ell) \geq \gamma \ \Rightarrow\ \tfrac{1}{1+\delta} \leq S_\ell / S(p_\ell) \leq 1+\delta.$$

Claim 7.4. *The exploration phase is δ-approximate with probability at least* $1 - 2\left(\log_{1+\delta}\frac{1}{\epsilon}\right) e^{-\delta^2\gamma m/4}$.

Proof. This follows directly by applying Chernoff bounds (both the upper and lower tail form) to the event that some S_ℓ violates the condition, then applying the union bound over all choices of ℓ. □

Claim 7.5. *When the exploration phase is δ-approximate, we have* $(1-7\delta)S^{-1}\left(\frac{k}{n}\right) \leq \widetilde{p} \leq p^*$.

Proof. It is easy to see that none of the stopping conditions of the exploration phase can be triggered until the price goes below p^*. Therefore $\widetilde{p} \leq p^*$. For the other inequality observe that, by Claim B.3 it holds that $S^{-1}(\alpha) \geq (1-\delta)\, S^{-1}(\frac{k}{n})$. Therefore it suffices to show that $\widetilde{p} \geq (1-6\delta)\, S^{-1}(\alpha)$.

Assume for a contradiction that the stopping conditions are not triggered in some phase ℓ such that $p_{\ell+1} < (1+\delta)^{-6}\, S^{-1}(\alpha)$. Therefore, at round ℓ we have

$$p_\ell = (1+\delta)p_{\ell+1} < (1+\delta)^{-5}\, S^{-1}(\alpha) \tag{20}$$

Examining the stopping conditions, and using our assumption above, we deduce that:

$$S_\ell < (1+\delta)\alpha \tag{21}$$

$$R_{\max}/R_\ell < (1+\delta)^2, \tag{22}$$

Combining (20) and (21), we get

$$R_\ell = p_\ell S_\ell < (1+\delta)^{-4}\alpha S^{-1}(\alpha) \tag{23}$$

Note that, since we chose round ℓ such that $p_\ell \ll S^{-1}(\alpha)$, the pricing strategy already encountered some round $t < \ell$ such that p_t is"close" to $S^{-1}(\alpha)$ – in particular

$$(1+\delta)^{-1}S^{-1}(\alpha) \le p_t \le S^{-1}(\alpha) \tag{24}$$

and therefore also $S(p_t) \ge \alpha$. Since we assume the exploration phase is δ-approximate, the estimated survival rate at round t satisfies $S_t \ge (1+\delta)^{-1}S(p_t) \ge (1+\delta)^{-1}\alpha$. Combining this with (24), we get that the estimated revenue R_t at round t satisfies

$$R_t = p_t S_t \ge (1+\delta)^{-2}\alpha S^{-1}(\alpha) \tag{25}$$

The value of $R_{\max}$ in round ℓ is at least R_t. Combining (25) with (23), this shows that at round ℓ we have $\frac{R_{\max}}{R_\ell} > (1+\delta)^2$, contradicting (22). □

Claim 7.6. *When the exploration phase is δ-approximate, we have $R(\widetilde{p}) \ge (1-7\delta)R(p^*)$.*

Proof. By Claim 7.5, we are done when $p^* = S^{-1}\left(\frac{k}{n}\right)$. Therefore, assume $p^* = p_{\mathrm{r}}$, the Myerson reserve price. It is easy to see that $R(p_{\ell+1}) \ge \frac{1}{1+\delta}\, R(p_\ell)$ for each ℓ. Let t be the first integer such that $p_t \le p^* = p_{\mathrm{r}}$. Note that $(1+\delta)^{-1}p^* \le p_t \le p^*$. Claim 7.5 says that $\widetilde{p} \le p^* = p_{\mathrm{r}}$, therefore $\widetilde{\ell} \ge t$ and by Claim B.2 $S(p_t) \ge S(p_{\mathrm{r}}) \ge 1/e \ge \gamma$. It suffices to show that a stopping condition must be triggered before $R(p_\ell)$ gets too small.

Assume for a contradiction that the stopping condition is not triggered by phase $\ell \ge t$, for some ℓ such that $R(p_{\ell+1}) < (1-7\delta)R(p^*)$. Since R decreases slowly as described above, it follows that $t < \ell$. Moreover, since we assumed the exploration phase is δ-approximate, $S_t \ge \frac{1}{1+\delta}\, S(p_t) \ge \frac{1}{1+\delta}\, \gamma$. Therefore, during phase ℓ we have $R_{max} \ge R_t = S_t p_t \ge (\frac{1}{1+\delta})^2\, R(p^*)$. Since no stopping condition is triggered for phase ℓ, it must be that $R_\ell \ge (\frac{1}{1+\delta})^2\, R_{max} \ge (\frac{1}{1+\delta})^4\, R(p^*)$. Moreover $R(p_{\ell+1}) \ge \frac{1}{1+\delta}\, R(p_\ell) \ge (\frac{1}{1+\delta})^2\, R_\ell \ge (\frac{1}{1+\delta})^6 R(p^*)$, a contradiction. □

We can now complete the proof of Lemma 7.2.

We condition on the exploration phase being δ-approximate. Let n' and k' be the number of players and items left after the exploration phase, respectively. In the exploitation phase, we attain expected revenue $\mathtt{Rev}(A^{n'}_{k'}(\widetilde{p}))$. Moreover, in the exploration phase we attained revenue at least $(k-k')\widetilde{p}$, since we only used prices greater than or equal to $\widetilde{p}$. Therefore, the total expected revenue of our pricing strategy is at least $\mathtt{Rev}(\mathcal{A}^{n'}_{k'}(\widetilde{p})) + (k'-k)\widetilde{p}$. It is easy to see that this is at least $\mathtt{Rev}(\mathcal{A}^{n'}_{k}(\widetilde{p}))$.

It remains to bound the expected revenue of $\mathcal{A}^{n'}_{k}(\widetilde{p})$. Observe that $\frac{n'}{n} \ge 1-\delta$. For brevity, denote $\beta \triangleq (1 - \frac{1}{\sqrt{2\pi k}})$.

There are two cases. In the first case, $p^* = S^{-1}(\frac{k}{n})$. Lemma 7.3 and Claim 7.5 imply that

$$\mathtt{Rev}(\mathcal{A}^{n'}_{k}(\widetilde{p})) \ge \beta\, \frac{n'}{n}\frac{\widetilde{p}}{p^*}\, \mathtt{Rev}(\mathcal{A}^{n}_{n}(p^*)) \ge \beta\,(1-8\delta)\, \mathtt{Rev}(\mathcal{A}^{n}_{n}(p^*)).$$

The second case is $p^* = p_{\mathrm{r}}$. By Claim 7.6 and unimodality of R, we have that

$$\mathtt{Rev}(\mathcal{A}^{n}_{n}(\max(S^{-1}(\tfrac{k}{n}), \widetilde{p}))) \ge \mathtt{Rev}(\mathcal{A}^{n}_{n}(\widetilde{p})) \ge (1-7\delta)\, \mathtt{Rev}(A^{n}_{n}(p^*)).$$

Moreover, using Lemma 7.3, Claim 7.5, and the equation above we show that

$$\texttt{Rev}(\mathcal{A}_k^{n'}(\widetilde{p})) \geq \beta\ (1-8\delta)\ \texttt{Rev}(\mathcal{A}_n^n(\max(S^{-1}(\tfrac{k}{n}),\widetilde{p}))) \geq \beta\ (1-15\delta)\ \texttt{Rev}(\mathcal{A}_n^n(p^*).$$

By Lemma A.2, Mechanism 2 achieves, in expectation, at least the following fraction of the expected revenue of the offline benchmark:

$$\beta\ (1-O(\delta))\ \left(1-2\ \log_{1+\delta}(\tfrac{1}{\epsilon})\exp(-\tfrac{1}{4}\,\delta^2\gamma m)\right).$$

Now, plug δ into Lemma 7.2, and m as defined in the pricing strategy. Note that $m=\Theta(\frac{\delta^2 n}{\log 1/\epsilon})$. We obtain the final bound replacing γ by the lesser quantity $\frac{k}{n}$, and using the fact that $\log_{1+\delta}(x)=\Theta(\frac{1}{\delta}\,\log x)$.

8 Conclusions and open questions

We consider dynamic pricing with limited supply and achieve near-optimal performance using an index-based bandit-style algorithm. A key idea in designing this algorithm is that we define the index of an arm (price) according to the estimated expected *total payoff* from this arm given the known constraints.

It is worth noting that a good index-based algorithm did not *have* to exist in our setting. Indeed, many bandit algorithms in the literature are not index-based, e.g. `EXP3` [5] and "zooming algorithm" [29] and their respective variants. The fact that Gittins algorithm [24] and `UCB1` [4] achieve (near-)optimal performance with index-based algorithms was widely seen as an impressive contribution.

While in this paper we apply the above key idea to a specific index-based algorithm (`UCB1`), it can be seen as an (informal) general reduction for index-based algorithms for dynamic pricing, from unlimited supply to limited supply. This reduction may help with more general dynamic pricing settings (more on that below), and moreover it can be extended to other bandit-style settings where the "best arm" is *not* an arm with the best expected per-round payoff. In particular, an ongoing project [1] uses this reduction in the context of adaptive crowd-selection in crowdsourcing.

It is an interesting open question whether a reduction such as above can be made more formal, and which algorithms and which settings it can be applied to. An ambitions conjecture for our setting is that there is a simple black-box reduction from unlimited supply to limited supply that applies to arbitrary "reasonable" algorithms. In the full generality this conjecture appears problematic; in particular, some reasonable bandit algorithms such as `EXP3` are hard-coded to spend a prohibitively large amount of time on exploration.

This paper gives rise to a number of more concrete open questions. The most immediate ones concern extending our upper and lower bounds for, respectively, more general and more specific classes of demand functions. First, it is desirable to extend Theorem 1.1 to possibly irregular distributions, i.e. obtain non-trivial regret bounds with respect to the offline benchmark. Second, one wonders whether the optimal $O(c_F\,\sqrt{k})$ regret rate from Theorem 1.3 can be extended to all regular demand distributions. Third, it is open whether our lower bounds can be strengthened to regular demand distributions.

Further, it is desirable to extend dynamic pricing with limited supply beyond IID valuations. A recent result in this direction is [13], where the demand distribution can change exactly once, at some point in time that is unknown to the mechanism. Some of the natural specific targets for further work are slowly changing valuations and adversarial valuations. One promising approach for slowly changing valuations is to apply the reduction from this paper to index-based algorithms for the corresponding bandit setting [38, 37].

Acknowledgements. We are grateful to Jason Hartline, Qiqi Yan and Assaf Zeevi for their comments.

References

[1] Ittai Abraham, Omar Alonso, Vasilis Kandylas, and Aleksandrs Slivkins. Adaptive algorithms for crowdsourcing [preliminary title]. Ongoing project, 2011-2012.

[2] Rajeev Agrawal. The continuum-armed bandit problem. *SIAM J. Control and Optimization*, 33(6):1926–1951, 1995.

[3] J.Y. Audibert and S. Bubeck. Regret Bounds and Minimax Policies under Partial Monitoring. *J. of Machine Learning Research (JMLR)*, 11:2785–2836, 2010. A preliminary version has been published in *COLT 2009*.

[4] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002. Preliminary version in *15th ICML*, 1998.

[5] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. Preliminary version in *36th IEEE FOCS*, 1995.

[6] Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved Rates for the Stochastic Continuum-Armed Bandit Problem. In *20th Conf. on Learning Theory (COLT)*, pages 454–468, 2007.

[7] Moshe Babaioff, Liad Blumrosen, Shaddin Dughmi, and Yaron Singer. Posting prices with unknown distributions. In *Symp. on Innovations in CS*, 2011.

[8] Moshe Babaioff, Robert Kleinberg, and Aleksandrs Slivkins. Truthful mechanisms with implicit payment computation. In *11th ACM Conf. on Electronic Commerce (EC)*, pages 43–52, 2010.

[9] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. In *10th ACM Conf. on Electronic Commerce (EC)*, pages 79–88, 2009.

[10] Z. Bar-Yossef, K. Hildrum, and F. Wu. Incentive-compatible online auctions for digital goods. In *13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2002.

[11] Dirk Bergemann and Juuso Välimäki. Bandit Problems. In Steven Durlauf and Larry Blume, editors, *The New Palgrave Dictionary of Economics, 2nd ed.* Macmillan Press, 2006.

[12] Omar Besbes and Assaf Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57:1407–1420, 2009.

[13] Omar Besbes and Assaf Zeevi. On the minimax complexity of pricing in a changing environment. *Operations Reseach*, 59:66–79, 2011.

[14] Avrim Blum and Jason Hartline. Near-optimal online auctions. In *16th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, 2005.

[15] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. In *14th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 202–204, 2003.

[16] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure Exploration in Multi-Armed Bandit Problems. In *20th Intl. Conf. on Algorithmic Learning Theory (ALT)*, 2009.

[17] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvari. Online Optimization in X-Armed Bandits. In *Advances in Neural Information Processing Systems (NIPS)*, pages 201–208, 2008.

[18] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge Univ. Press, 2006.

[19] Tanmoy Chakraborty, Eyal Even-Dar, Sudipto Guha, Yishay Mansour, and S. Muthukrishnan. Approximation schemes for sequential posted pricing in multi-unit auctions. In *Workshop on Internet & Network Economics (WINE)*, pages 158–169, 2010.

[20] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *ACM Symp. on Theory of Computing (STOC)*, pages 311–320, 2010.

[21] Nikhil Devanur and Jason Hartline. Limited and online supply and the bayesian foundations of prior-free mechanism design. In *ACM Conf. on Electronic Commerce (EC)*, 2009.

[22] Nikhil Devanur and Sham M. Kakade. The price of truthfulness for pay-per-click auctions. In *10th ACM Conf. on Electronic Commerce (EC)*, pages 99–106, 2009.

[23] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. In *ACM Conf. on Electronic Commerce (EC)*, pages 129–138, 2010.

[24] J. C. Gittins. Bandit processes and dynamic allocation indices (with discussion). *J. Roy. Statist. Soc. Ser. B*, 41:148–177, 1979.

[25] Ashish Goel, Sanjeev Khanna, and Brad Null. The Ratio Index for Budgeted Learning, with Applications. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 18–27, 2009.

[26] Mohammad T. Hajiaghayi, Robert Kleinberg, and David C. Parkes. Adaptive limited-supply online auctions. In *Proc. ACM Conf. on Electronic Commerce*, pages 71–80, 2004.

[27] J.D. Hartline and T. Roughgarden. Optimal mechanism design and money burning. In *ACM Symp. on Theory of Computing (STOC)*, 2008.

[28] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *18th Advances in Neural Information Processing Systems (NIPS)*, 2004.

[29] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-Armed Bandits in Metric Spaces. In *40th ACM Symp. on Theory of Computing (STOC)*, pages 681–690, 2008.

[30] Robert D. Kleinberg and Frank T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2003.

[31] T.L. Lai and Herbert Robbins. Asymptotically efficient Adaptive Allocation Rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[32] Ron Lavi and Noam Nisan. Competitive analysis of incentive compatible on-line auctions. In *ACM Conference on Electronic Commerce*, pages 233–241, 2000.

[33] Colin McDiarmid. Concentration. In M. Habib. C. McDiarmid. J. Ramirez and B. Reed, editors, *Probabilistic Methods for Discrete Mathematics*, pages 195–248. Springer-Verlag, Berlin, 1998.

[34] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[35] R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[36] Hamid Nazerzadeh, Amin Saberi, and Rakesh Vohra. Dynamic cost-per-action mechanisms and applications to online advertising. In *17th WWW*, 2008.

[37] Aleksandrs Slivkins. Contextual Bandits with Similarity Information. In *24th Conf. on Learning Theory (COLT)*, 2011.

[38] Aleksandrs Slivkins and Eli Upfal. Adapting to a Changing Environment: the Brownian Restless Bandits. In *21st Conf. on Learning Theory (COLT)*, pages 343–354, 2008.

[39] Robert B. Wilson. Efficient and competitive rationing. *Econometrica*, 57:1–40, 1989.

[40] Qiqi Yan. Mechanism design via correlation gap. In *22nd ACM-SIAM Symp. on Discrete Algorithms (SODA)*, 2011.

Appendix A: Benchmark comparison

We start with a self-contained proof of a slightly weaker version of Lemma 3.1 (which suffices for the purposes of this paper).

Lemma A.1 (Yan [40])**.** *For each regular demand distribution there exists a fixed-price strategy whose expected revenue is at least the offline benchmark minus* $O(\sqrt{k \log k})$.

Recall that $A_k^n(p)$ denotes the fixed-price strategy with k items, n agents, and fixed price p. Let M_k^n denote the optimal (expected revenue maximizing) offline auction with n-players and k-items. As in Claim 3.2, let $p^* = \max(p_{\mathrm{r}},\, S^{-1}(\frac{k}{n}))$, where $p_{\mathrm{r}} = \operatorname{argmax}_p p\, S(p)$ is the Myerson reserve price.

Claim A.2. *If the demand distribution is regular then* $\texttt{Rev}(A_n^n(p^*)) \geq \texttt{Rev}(M_k^n)$.

Proof of Claim A.2. Let q_i be the probability that M_k^n sells to agent i. By symmetry, $q_i = q_j$ for all players i and j, so we simply denote this probability by q. Let $p = S^{-1}(q)$ be the single price we would need to offer a agent in order to sell to him with probability q. Since R is a concave function of the selling probability, Jensen's inequality implies that $R(p)$ is an upper bound on the revenue collected by the Myerson auction from a single agent. Equivalently: $nR(p) \geq \texttt{Rev}(M_k^n)$.

Now, observe that the expected number of items sold by M_k^n is nq. Since M_k^n never sells more than k items, it must be that $q \leq \frac{k}{n}$. Therefore, $p \geq S^{-1}(\frac{k}{n})$. By definition of p^*, we deduce that there are two cases: (1) $p^* = p_{\mathrm{r}}$, or (2) $p_{\mathrm{r}} \leq p^* = S^{-1}(\frac{k}{n}) \leq p$. In case (1) it is clear that $R(p^*) \geq R(p)$. In case (2) we get that $R(p^*) \geq R(p)$ since $R(x)$ is decreasing for $x \geq p_{\mathrm{r}}$. Then

$$\texttt{Rev}(A_n^n(p^*)) = nR(p^*) \geq nR(p) \geq \texttt{Rev}(M_k^n). \square$$

Lemma A.1 follows from Claim A.2 and Claim 3.2 because for $p = p^*$ we have $S(p) \leq \frac{k}{n}$, and so

$$\nu(p) = p\, \min(k, n\, S(p))) = np^*\, S(p^*) = \texttt{Rev}(A_n^n(p^*)) \geq \texttt{Rev}(M_k^n).$$

Multiplicative bounds. Further, we derive a multiplicative bound in which fixed-price strategies for limited supply are compared to those for unlimited supply. We use this bound to prove Lemma 7.3.

Claim A.3. *For any regular demand distribution and any* $p \geq S^{-1}(\frac{k}{n})$ *it holds that*

$$\texttt{Rev}(\mathcal{A}_k^n(p)) \geq \left(1 - \tfrac{1}{\sqrt{2\pi k}}\right) \texttt{Rev}(\mathcal{A}_n^n(p)).$$

Proof. The proof uses a technique from [40]. As a thought experiment, consider an environment where agent valuations are *correlated* as follows: The joint distribution of agent valuations can be sampled by choosing a set S' of k players uniformly at random, then for each agent in S' sampling from the conditional distribution $F(x)|_{x \geq S^{-1}(k/n)}$, and for each agent not in S' sampling from the conditional distribution $F(x)|_{x < S^{-1}(k/n)}$. Observe that each agent's valuation is distributed according to F, yet at any point exactly k players have value exceeding $S^{-1}(k/n)$.

Let T' be the set of players in this correlated environment whose valuation exceeds p. The probability of a particular agent being included in T' is $S(p)$, and $\mathbb{E}[|T'|] = nS(p)$. Since $p \geq S^{-1}(k/n)$, it is clear that $T' \subseteq S'$ and therefore $0 \leq |T'| \leq k$.

Now consider our original environment where each agent's valuation is drawn i.i.d from F. Let T be the set of players in this environment whose valuations exceed p. The probability of a agent being included in T is $S(p)$ – the same as the probability of being included in T'. However, each agent is included in T

independently with probability $S(p)$. As a result, some of the players in T do not win an item – this happens when $|T| > k$. We can write the revenue of $\mathcal{A}_k^n(p)$ in this i.i.d environment as follows.

$$\texttt{Rev}(\mathcal{A}_k^n(p))] = p\,\mathbb{E}[\,\min(|T|,k)\,] \tag{26}$$

Now, observe that $r(Y) = \min(|Y|,k)$ is the rank function of the k-uniform matroid. Moreover, it was shown in [40] that the correlation gap of this function is $\beta \triangleq \left(1 - \frac{1}{\sqrt{2\pi k}}\right)$. Therefore, since each agent is included in T independently, we know by the definition of the correlation gap and the fact that T and T' have the same marginals that

$$\mathbb{E}[r(T)] \geq \beta\,\mathbb{E}[r(T')]. \tag{27}$$

Recall that T' is always bounded between 0 and k, therefore $r(T') = |T'|$. Combining (26) and (27), we get

$$\texttt{Rev}(A_k^n(p)) = p\,\mathbb{E}[\,\min(|T|,k)\,] \geq \beta\,p\,\mathbb{E}[|T'|] = \beta\,pnS(p)) = \beta\,\texttt{Rev}(\mathcal{A}_n^n(p))). \quad \square$$

Corollary (Lemma 7.3)**.** *Assume the demand distribution is regular. Let $p \leq p'$ be two prices such that $p \geq S^{-1}(k/n)$. Let $n' \leq n$. Then $\texttt{Rev}(\mathcal{A}_k^{n'}(p')) \geq \frac{n'}{n}\frac{p'}{p}\left(1 - \frac{1}{\sqrt{2\pi k}}\right)\texttt{Rev}(\mathcal{A}_n^n(p))$.*

Proof. Observe that $\mathcal{A}_k^n(p')$ sells at least as many items as $\mathcal{A}_k^n(p)$ for every realization of the bids, but at price p' instead of p. Therefore $\texttt{Rev}(A_k^n(p')) \geq \frac{p'}{p}\texttt{Rev}(A_k^n(p))$. Combining with Claim A.3 we get that

$$\texttt{Rev}(A_k^n(p')) \geq \frac{p'}{p}\left(1 - \frac{1}{\sqrt{2\pi k}}\right)\texttt{Rev}(\mathcal{A}_n^n(p)).$$

Next, a simple (omitted) argument shows that the revenue $\texttt{Rev}(A_k^n(p))$ of a fixed price auction exhibits diminishing marginal returns in the number n of players. Therefore, $\texttt{Rev}(A_k^{n'}(p)) \geq \frac{n'}{n}\texttt{Rev}(A_k^n(p))$. $\square$

Let us note in passing that Claim A.3 and Claim 3.2 imply a stronger, multiplicative version of Lemma 3.1, which is also immediate from [40].

Lemma A.4 (Yan [40])**.** *Assume that the demand distribution is regular. Then there exists a fixed-price strategy whose expected revenue approximates the offline benchmark up to a factor $1 - \frac{1}{\sqrt{2\pi k}}$.*

Appendix B: Monotone Hazard Rate distributions

Let us state and prove several properties of Monotone Hazard Rate (MHR) distributions which we use in Section 5 and Section 7. Throughout, for a distribution F we use $F(x)$ to denote the c.d.f, $S(x) = 1 - F(x)$ to denote the survival rate, and $f(x)$ to denote the p.d.f.

We begin with a simple known characterization of MHR distributions.

Fact B.1. *A distribution is MHR if and only if $S(\cdot)$ is log-concave (i.e. $\log S(x)$ is a concave function of x).*

Next, we bound the survival probability at the Myerson reserve price.

Claim B.2. *Let F be an MHR distribution with support on $[0,\infty]$, and let $S(x) = 1 - F(x)$. Let $r \in \operatorname{argmax} R(\cdot)$ where $R(x) = x\,S(x)$. Then $S(r) \geq 1/e$.*

Proof. We have $R'(r) = S(r) + rS'(r) = 0$. Moreover, by Fact B.1 we deduce that

$$\frac{\log S(r)}{r} \geq \frac{d}{dx}\log(S(x))|_r = \frac{S'(r)}{S(r)}$$

Combining with the previous equality, we have $\frac{-1}{r} \leq \frac{\log(S(r))}{r}$ which is equivalent to $S(r) \geq \frac{1}{e}$. $\square$

We now use log-concavity to bound the sensitivity of the inverse of the survival function.

Claim B.3. *Let F be an MHR distribution with support on $[0,\infty]$, and let $\alpha, \beta \in [0,1]$ with $\beta \geq \alpha$. Then*

$$S^{-1}(\beta) \geq \frac{\log(\beta)}{\log(\alpha)} S^{-1}(\alpha)$$

Proof. By Fact B.1, $\log(S(x))$ is a concave, decreasing function of x with $\log(S(0)) = 0$ and $\lim_{x\to\infty} \log(S(x)) = -\infty$. Let $a = S^{-1}(\alpha)$ and $b = S^{-1}(\beta)$. By Jensen's inequality, this gives for every $a, b \in [0,\infty]$ with $b \leq a$

$$\frac{\log(S(b))}{\log(S(a))} \leq \frac{b}{a}$$

Let $a = S^{-1}(\alpha)$ and $b = S^{-1}(\beta)$. Plugging into the above inequality completes the proof. □