

# The Value of Knowing a Demand Curve: Bounds on Regret for On-line Posted-Price Auctions

Robert Kleinberg <sup>\*</sup>      Tom Leighton <sup>†</sup>

April 17, 2007

## Abstract

We consider the revenue-maximization problem for a seller with an unlimited supply of identical goods, interacting sequentially with a population of  $n$  buyers through an on-line posted-price auction mechanism, a paradigm which is frequently available to vendors selling goods over the Internet. For each buyer, the seller names a price between 0 and 1; the buyer decides whether or not to buy the item at the specified price, based on her privately-held valuation. The price offered is allowed to vary as the auction proceeds, as the seller gains information from interactions with the earlier buyers.

The *additive regret* of a pricing strategy is defined to be the difference between the strategy's expected revenue and the revenue derived from the optimal fixed-price strategy. In the case where buyers' valuations are independent samples from a fixed probability distribution (usually specified by a demand curve), one can interpret the regret as specifying how much the seller should be willing to pay for knowledge of the demand curve from which buyers' valuations are sampled.

The answer to the problem depends on what assumptions one makes about the buyers' valuations. We consider three such assumptions: that the valuations are all equal to some unknown number  $p$ , that they are independent samples from an unknown probability distribution, or that they are chosen by an oblivious adversary. In each case, we derive upper and lower bounds on regret which match within a factor of  $\log n$ ; the bounds match up to a constant factor in the case of identical valuations.

---

<sup>\*</sup>Department of Mathematics, MIT, Cambridge MA 02139, and Akamai Technologies, 8 Cambridge Center, Cambridge, MA 02142. Email: rdk@math.mit.edu. Supported by a Fannie and John Hertz Foundation Fellowship.

<sup>†</sup>Department of Mathematics, MIT, Cambridge MA 02139, and Akamai Technologies, 8 Cambridge Center, Cambridge, MA 02142. Email: ftl@math.mit.edu.

# 1 Introduction

The rising popularity of Internet commerce has spurred much recent research on market mechanisms which were either unavailable or impractical in traditional markets, because of the amount of communication or computation required. We consider one such mechanism, the on-line posted-price auction, in which a seller with an unlimited supply of identical goods interacts sequentially with a population of  $n$  buyers. For each buyer, the seller names a price between 0 and 1; the buyer then decides whether or not to buy the item at the specified price, based on her privately-held valuation for the good. This transaction model is dictated by the following considerations:

- Following earlier authors [5], [7], [8], [13], we are interested in auction mechanisms which are *strategyproof*, meaning that buyers weakly maximize their utility by truthfully revealing their preferences. As shown in [5], this requirement in the on-line auction setting is equivalent to requiring that the seller charge buyer  $i$  a price which depends only on the valuations of previous buyers.
- Given that the price offered to buyer  $i$  does not depend on any input from that buyer, it is natural for the seller to announce this price before the buyer reveals any preference information. In fact, for reasons of trust, the buyers may not want to reveal their preferences before an offer price is quoted [6].
- For privacy reasons, the buyers generally do not wish to reveal any preference information *after* the price is quoted either, apart from their decision whether or not to purchase the good. Also, buyers are thus spared the effort of precisely determining their valuation, since the mechanism only requires them to determine whether it is greater or less than the quoted price.

The seller's pricing strategy will tend to converge to optimality over time, as she gains information about how the buyers' valuations are distributed. A natural question which arises is: what is the cost of not knowing the distribution of the buyers' valuations in advance? In other words, assume our seller pursues a pricing strategy  $\mathcal{S}$  which maximizes her expected revenue  $\rho(\mathcal{S})$ . As is customary in competitive analysis of auctions, we compare  $\rho(\mathcal{S})$  with the revenue  $\rho(\mathcal{S}^{\text{opt}})$  obtained by a seller who knows the buyers' valuations in advance but is constrained to charge the same price to all buyers ([5],[6],[7],[8]). While previous authors have analyzed auctions in terms of their competitive ratio (the ratio between  $\rho(\mathcal{S})$  and  $\rho(\mathcal{S}^{\text{opt}})$ ), we instead analyze the *additive regret*, i.e. the difference  $\rho(\mathcal{S}) - \rho(\mathcal{S}^{\text{opt}})$ . This is a natural parameter to study for two reasons. First, it roughly corresponds to the amount the seller should be willing to pay to gain knowledge of the buyers' valuations, e.g. by doing market research. Second, it was shown by Blum et al in [6] that there are randomized pricing strategies achieving competitive ratio  $1 + \varepsilon$  for any  $\varepsilon > 0$ ; thus it is

natural to start investigating the lower-order terms, i.e. the  $o(1)$  term in the ratio  $\rho(\mathcal{S})/\rho(\mathcal{S}^{\text{opt}})$  for the optimal pricing strategy  $\mathcal{S}$ .

One can envision several variants of this problem, depending on what assumptions are made about the buyers' valuations. We will study three valuation models.

**Identical:** All buyers' valuations are equal to a single price  $p \in [0, 1]$ . This price is unknown to the seller.

**Random:** Buyers' valuations are independent random samples from a fixed probability distribution on  $[0, 1]$ . The probability distribution is not known to the seller.

**Worst-case:** The model makes no assumptions about the buyers' valuations. They are chosen by an adversary who is oblivious to the algorithm's random choices.

Our results are summarized in the following three theorems. In all of them, the term "pricing strategy" refers to a randomized on-line algorithm for choosing offer prices, unless noted otherwise.

**Theorem 1.1.** *Assuming identical valuations, there is a deterministic pricing strategy achieving regret  $O(\log \log n)$ . No pricing strategy can achieve regret  $o(\log \log n)$ .*

**Theorem 1.2.** *Assuming random valuations, there is a pricing strategy achieving regret  $O(\sqrt{n} \log n)$ , under the hypothesis that the function*

$$f(x) = x \cdot \Pr(\text{buyer's valuation is } \geq x)$$

*has a unique global maximum  $x^*$  in the interior of  $[0, 1]$ , and that  $f''(x^*) < 0$ . No pricing strategy can achieve regret  $o(\sqrt{n})$ , even under the same hypothesis on the distribution of valuations.*

**Theorem 1.3.** *Assuming worst-case valuations, there is a pricing strategy achieving regret  $O((n^{2/3}(\log n)^{1/3}))$ . No pricing strategy can achieve regret  $o(n^{2/3})$ .*

The lower bound in the random-valuation model is the most difficult of the results stated above, and we believe it is this paper's main contribution. No such bound was known previously, and our proof introduces novel techniques which we believe may be applicable elsewhere. Moreover, our lower bound does not rely on constructing a contrived demand curve to defeat a given pricing strategy. Rather, we will show that for any family  $\mathcal{D}$  of demand curves satisfying some reasonably generic axioms, and for any randomized pricing strategy, the probability of achieving expected regret  $o(\sqrt{n})$  when the demand curve is chosen randomly from  $\mathcal{D}$  is zero. Note the order of quantification here, which differs from  $\sqrt{n}$  lower bounds which have appeared in the literature on the closely-related multi-armed bandit problem. In those theorems it was shown that, given foreknowledge of  $n$ , one could construct a random sequence of payoffs forcing

any strategy to have expected regret  $\Omega(\sqrt{n})$ . In our theorem, the demand curve is chosen randomly without foreknowledge of  $n$  or of the pricing strategy, and it is *still* the case that the probability of the strategy achieving regret  $o(\sqrt{n})$  is zero.

## 1.1 Related work

There has been much recent activity applying notions from the theory of algorithms to the analysis of auction mechanisms. While much of this work focuses on combinatorial auctions — a subject not touched on here — there has also been a considerable amount of work on auction mechanisms for selling identical individual items, the setting considered in this paper. In [7], [8], the authors consider mechanisms for off-line auctions, i.e. those in which all buyers reveal their valuations before any goods are sold. The authors characterize mechanisms which are truthful (a term synonymous with “strategyproof”, defined above), and show that no such mechanism can be constant-competitive with respect to the optimal single-price auction, assuming worst-case valuations. In contrast, they present several randomized off-line auction mechanisms which are truthful and constant-competitive with respect to the optimal auction which is constrained to set a single price *and to sell at least two copies of the good*.

On-line auctions were considered in [5], [6], in the posted-price setting considered here as well as the setting where buyers reveal their valuations but are charged a price which depends only on the information revealed by prior buyers. In the latter paper, techniques from machine learning theory are applied to yield a  $(1 + \varepsilon)$ -competitive on-line mechanism (for any  $\varepsilon > 0$ ) under the hypothesis that the optimal single-price auction achieves revenue  $\Omega(h \log h \log \log h)$ , where  $[1, h]$  is the interval to which the buyers’ valuations belong. In Section 4.1, we use their algorithm (with a very minor technical modification) to achieve expected regret  $O(n^{2/3}(\log n)^{1/3})$  assuming worst-case valuations.

An interesting hybrid of the off-line and on-line settings is considered by Hartline in [9]. In that paper, the mechanism interacts with the set of buyers in two rounds, with prices in the second round influenced by the preferences revealed by buyers in the first round. Assuming the set of buyers participating in the first round is a uniform random subset of the pool of  $n$  buyers, the paper exhibits a posted-price mechanism which is 4-competitive against the optimal single-price auction.

On-line multi-unit auctions (in which buyers may bid for multiple copies of the item) are considered in [4], which presents a randomized algorithm achieving competitive ratio  $O(\log B)$  where  $B$  is the ratio between the highest and lowest per-unit prices offered. This result is sharpened in [10], where the optimal competitive ratio (as a function of  $B$ ) is determined exactly.

The preceding papers have all adopted the worst-case model for buyers’ valuations, as is customary in the computer science literature. The traditional approach in the economics literature (e.g. [12]) is to assume that buyers’ valuations are i.i.d. samples from a *known* probability distribution. Our random-valuations model occupies a middle ground between these two sets of assumptions, in that

the i.i.d. hypothesis is preserved but the probability distribution (i.e. the demand curve) is unknown to the seller. The same set of hypotheses is made by Ilya Segal in [13], a paper which considers strategyproof off-line multi-unit auction mechanisms. (A multi-unit auction is one in which buyers may purchase more than one copy of the good.) Segal compares the expected regret of the optimal strategyproof off-line mechanism with that of the optimal on-line posted-price mechanism (which he calls the “optimal experimentation mechanism”) under three assumptions on the space  $\mathcal{D}$  of possible demand curves:

- $\mathcal{D}$  is a finite set (“Hypothesis testing”);
- $\mathcal{D}$  is parametrized by a finite-dimensional Euclidean space (“Parametric estimation”);
- $\mathcal{D}$  is arbitrary (“Non-parametric estimation”).

In this terminology, our paper is concerned with bounding the expected regret of the optimal experimentation mechanism in the non-parametric case. Segal explicitly refrains from addressing this case, writing, “The optimal experimentation mechanism would be very difficult to characterize in [the non-parametric] setting. Intuitively, it appears that its convergence rate may be slower [than that of the optimal off-line mechanism] because the early purchases at prices that are far from  $p^*$  will prove useless for fine-tuning the price around  $p^*$ .” This intuition is confirmed by the lower bound we prove in Section 3.2.

Our work is also closely tied to the literature on the so-called “multi-armed bandit problem,” in which a gambler in a casino with  $K$  slot machines must decide which machine to play in a sequence of  $n$  trials, basing his decisions on the payoffs observed in prior trials. As in our auction problem, the regret is defined as the difference between the gambler’s expected net payoff and the net payoff obtained from the best single action (slot machine) over the sequence of  $n$  trials.

In their pioneering work on the multi-armed bandit problem, Lai and Robbins [11] assumed that for each action, the payoffs on each of the  $n$  trials are i.i.d. random variables, but the distribution of the payoff variable varies from one action to another. Under this hypothesis, they exhibited an algorithm achieving expected regret  $O(\log n)$  as  $n \rightarrow \infty$ , and proved that this is the optimal regret, up to constant factors. Auer et al [2] sharpened this analysis to obtain explicit regret bounds which hold for finite  $n$ .

If we view each potential offer price in  $[0, 1]$  as a slot machine with random payoff, then our on-line posted-price auction problem (in the random-valuations model) becomes a special case of the “continuum-armed bandit problem”, i.e. the variant of the multi-armed bandit problem in which there is an uncountable set of slot machines indexed by a real parameter  $t$ , with the expected reward depending continuously on  $t$ . This problem is considered by Agrawal in [1], who describes an algorithm achieving regret  $O(n^{3/4+\epsilon})$  in the case that the expected reward is a differentiable function of  $t$ . (The paper also gives a regret bound under a weaker continuity hypothesis on the expected reward, but it is more

difficult to state.) Our upper bound of  $O(\sqrt{n \log n})$  in the random-valuations model is better than the one obtained by applying Agrawal’s algorithm, as might be expected because our auction problem is a highly-specific special case of the continuum-armed bandit problem. However, our lower bound of  $O(\sqrt{n})$  for the random-valuations model directly implies the same lower bound for Agrawal’s continuum-armed bandit problem. This answers a question left open at the end of [1], when Agrawal writes, “We do not have any tighter bounds on the learning loss other than those available for the finite case,” referring to the  $\Omega(\log n)$  lower bound proved by Lai and Robbins. Our paper is thus the first demonstration of an exponential separation between the expected regret when the set of possible actions is finite and when it is infinite.

In [3], the multi-armed bandit problem is studied from an adversarial perspective, parallel to our worst-case valuation model. The authors present an algorithm achieving expected regret  $O(\sqrt{nK \log K})$  and a nearly-matching lower bound of  $\Omega(\sqrt{nK})$  for this problem. Their algorithm forms the basis for the on-line posted-price auction algorithms in [6] and in Section 4.1 of this paper, and our lower-bound proof in the worst-case model (Theorem 4.3) is an adaptation of their lower-bound proof.

We would like to elucidate the difference between the lower bound for regret appearing in [3] and the lower bound presented here in the random-valuations model (Theorem 3.9). In both constructions, the rewards are i.i.d. random variables whose distribution is also random, and the theorem establishes that for any strategy the expected regret is  $\Omega(\sqrt{n})$ , where the expectation is over the random choices of both the distribution and of the rewards themselves. However, in our random-valuations lower bound we define regret as the difference in expected payoff between the on-line pricing strategy and the optimal strategy which has foreknowledge of the demand curve *but not of the individual buyers’ valuations*. This definition of regret accords with the definition adopted in papers studying the multi-armed bandit problem with i.i.d. random rewards (e.g. [1], [11]) and differs from the definition of regret adopted in [3], namely the difference in expected payoff between the on-line strategy and the *ex post* optimal single action. Because we choose to measure regret relative to the *ex ante* rather than the *ex post* optimal strategy, a subtler analysis is required. In [3], the lower bound comes from constructing a counterexample in which the reward distribution is so close to uniform that it is information-theoretically impossible, in  $n$  trials, to learn which of the possible reward distributions is generating the payoffs. More precisely, there is no sequence of  $n$  experiments such that an observer, after seeing the outcomes of the  $n$  experiments and being asked to guess which reward distribution is generating the payoffs, could outperform a random guesser by more than a constant factor. In our random-valuations model, it may be possible in some cases to gain rather precise information on which reward distribution is generating the payoffs in the course of  $n$  trials, but the process of gaining such knowledge will generally require offering prices which are too far from the optimal price, thereby incurring a cost of  $\Omega(\sqrt{n})$  in the process of learning this information. (This is a nice illustration of the trade-off between exploration and exploitation.) The difference between the two lower bounds is

most clearly illustrated by considering the case in which the algorithm is allowed to engage in an initial  $n$  trials using “play money,” and is then judged in terms of the expected regret incurred in the course of the subsequent  $n$  trials. The example which furnishes the lower bound in [3] has the property that no strategy could take advantage of the  $n$  free trials to achieve  $o(\sqrt{n})$  expected regret on the following  $n$  trials. In contrast, the example furnishing our  $\sqrt{n}$  lower bound in Section 3.2 has the property that there exist pricing strategies which can take advantage of the  $n$  free trials and achieve  $O(1)$  expected regret on the following  $n$  trials, yet it is impossible to achieve  $o(\sqrt{n})$  regret on the first  $n$  trials.

## 2 Identical valuations

### 2.1 Upper bound

When all buyers have the same valuation  $p \in [0, 1]$ , the situation is completely different from the scenarios considered above, because there is no randomness in the buyer’s response. Every response gives the seller perfect information about a lower or upper bound on  $p$ , depending on whether the buyer’s response was to accept or to reject the price offered.

A pricing strategy  $\mathcal{S}$  which achieves regret  $O(\log \log n)$  may be described as follows. The strategy keeps track of a *feasible interval*  $[a, b]$ , initialized to  $[0, 1]$ , and a *precision parameter*  $\varepsilon$ , initialized to  $1/2$ . In a given phase of the algorithm, the seller offers the prices  $a, a + \varepsilon, a + 2\varepsilon, \dots$  until one of them is rejected. If  $a + k\varepsilon$  was the last offer accepted in this phase, then  $[a + k\varepsilon, a + (k+1)\varepsilon]$  becomes the new feasible interval, and the new precision parameter is  $\varepsilon^2$ . This process continues until the length of the feasible interval is less than  $1/n$ ; then the seller offers a price of  $a$  to all remaining buyers.

**Theorem 2.1.** *Strategy  $\mathcal{S}$  achieves regret  $O(\log \log n)$ .*

*Proof.* The number of phases is equal to the number of iterations of repeated squaring necessary to get from  $1/2$  to  $1/n$ , i.e.  $O(\log \log n)$ . Let  $p$  denote the valuation shared by all buyers. The seller accrues regret for two reasons:

- Items are sold at a price  $q < p$ , accruing regret  $p - q$ .
- Buyers decline items, accruing regret  $p$ .

At most one item is declined per phase, incurring  $p < 1$  units of regret, so the declined offers contribute  $O(\log \log n)$  to the total regret.

In each phase except the first and the last, the length  $b - a$  of the feasible interval is  $\sqrt{\varepsilon}$  (i.e. it is the value of  $\varepsilon$  from the previous phase), and the set of offer prices carves up the feasible interval into subintervals of length  $\varepsilon$ . There are  $1/\sqrt{\varepsilon}$  such subintervals, so there are at most  $1/\sqrt{\varepsilon}$  offers made during this phase. Each time one of them is accepted, this contributes at most  $b - a = \sqrt{\varepsilon}$  to the total regret. Thus, the total regret contribution from accepted offers in this phase is less than or equal to  $(1/\sqrt{\varepsilon}) \cdot \sqrt{\varepsilon} = 1$ . There are  $O(\log \log n)$

phases, so the total regret contribution from accepted offers in these phases is also  $O(\log \log n)$ .

In the final phase, the length of the feasible interval is less than  $1/n$ , and each offer is accepted. There are at most  $n$  such offers, so they contribute at most 1 to the total regret.  $\square$

## 2.2 Lower bound

**Theorem 2.2.** *If  $\mathcal{S}$  is any randomized pricing strategy, and  $p$  is randomly sampled from the uniform distribution on  $[0, 1]$ , the expected regret of  $\mathcal{S}$  when the buyers' valuations are  $p$  is  $\Omega(\log \log n)$ .*

*Proof.* It suffices to prove the lower bound for a deterministic pricing strategy  $\mathcal{S}$ , since any randomized pricing strategy is a probability distribution over deterministic ones. At any stage of the game, let  $a$  denote the highest price that has yet been accepted, and  $b$  the lowest price that has yet been declined; thus  $p \in [a, b]$ . As before, we will refer to this interval as the *feasible interval*. It is counterproductive to offer a price less than  $a$  or greater than  $b$ , so we may assume that the pricing strategy works as follows: it offers an ascending sequence of prices until one of them is declined; it then limits its search to the new feasible interval, offering an ascending sequence of prices in this interval until one of them is declined, and so forth.

Divide the pool of buyers into phases (starting with phase 0) as follows: phase  $k$  begins immediately after the end of phase  $k - 1$ , and ends after an additional  $2^{2^k} - 1$  buyers, or after the first rejected offer following phase  $k - 1$ , whichever comes earlier. The number of phases is  $\Omega(\log \log n)$ , so it suffices to prove that the expected regret in each phase is  $\Omega(1)$ .

**Claim 2.3.** *Let  $\mathcal{I}_k$  denote the set of possible feasible intervals at the start of phase  $k$ . The cardinality of  $\mathcal{I}_k$  is at most  $2^{2^k}$ .*

*Proof.* The proof is by induction on  $k$ . The base case  $k = 0$  is trivial. Now assume the claim is true for a particular value of  $k$ , and let  $I_k = [a_k, b_k]$  be the feasible interval at the start of phase  $k$ . Let  $x_1 \leq x_2 \leq \dots \leq x_j$  denote the ascending sequence of prices that  $\mathcal{S}$  will offer during phase  $k$  if all offers are accepted. (Here  $j = 2^{2^k} - 1$ .) Then the feasible interval at the start of phase  $k + 1$  will be one of the subintervals  $[a_k, x_1], [x_1, x_2], [x_2, x_3], \dots, [x_{j-1}, x_j], [x_j, b_k]$ . There are at most  $j = 2^{2^k}$  such subintervals, and at most  $2^{2^k}$  possible choices for  $I_k$  (by the induction hypothesis), hence there are at most  $2^{2^{k+1}}$  elements of  $\mathcal{I}_{k+1}$ .  $\square$

**Claim 2.4.** *Let  $|I|$  denote the length of an interval  $I$ . With probability at least  $3/4$ ,  $|I_k| \geq \frac{1}{4} \cdot 2^{-2^k}$ .*

*Proof.* The expectation of  $1/|I_k|$  may be computed as follows:

$$\mathbf{E}(1/|I_k|) = \sum_{I \in \mathcal{I}_k} \Pr(p \in I)(1/|I|) = \sum_{I \in \mathcal{I}_k} |I|/|I| = \sum_{I \in \mathcal{I}_k} 1 \leq 2^{2^k},$$



where the last inequality follows from Claim 2.3. Now use Markov's Inequality:

$$\Pr(|I_k| < \frac{1}{4} \cdot 2^{-2^k}) = \Pr(1/|I_k| > 4 \cdot 2^{2^k}) < 1/4.$$

□

**Claim 2.5.** *The expected regret in phase  $k$  is at least  $\frac{1}{64}$ .*

*Proof.* Let  $\mathcal{E}_k$  denote the event that  $p \geq 1/4$  and  $|I_k| \geq \frac{1}{4} \cdot 2^{-2^k}$ . This is the intersection of two events, each having probability  $\geq 3/4$ , so  $\Pr(\mathcal{E}_k) \geq 1/2$ . It suffices to show that the expected regret in phase  $k$ , conditional on  $\mathcal{E}_k$ , is at least  $1/32$ . So from now on, assume that  $p \geq 1/4$  and  $|I_k| \geq \frac{1}{4} \cdot 2^{-2^k}$ .

Let  $m$  denote the midpoint of  $I_k$ . As before, let  $j = 2^{2^k} - 1$  and let  $x_1 \leq x_2 \leq \dots \leq x_j$  denote the ascending sequence of prices which  $\mathcal{S}$  would offer in phase  $k$  if no offers were rejected. We distinguish two cases:

**Case 1:**  $x_j \geq m$ . With probability at least  $1/2$ ,  $p < m$  and the phase ends in a rejected offer, incurring a regret of  $p$ , which is at least  $1/4$ . Thus the expected regret in this case is at least  $1/8$ .

**Case 2:**  $x_j < m$ . The event  $\{p > m\}$  occurs with probability  $1/2$ , and conditional on this event the expectation of  $p - m$  is  $|I_k|/4 \geq 2^{-2^k}/16$ . Thus with probability at least  $1/2$ , there will be  $2^{2^k} - 1$  accepted offers, each contributing  $2^{-2^k}/16$  to the expected regret, for a total of  $(2^{2^k} - 1)(2^{-2^k})/16 \geq 1/32$ .

□

Thus there are  $\Omega(\log \log n)$  phases, each contributing  $\Omega(1)$  to the expected regret of  $\mathcal{S}$ , which establishes the theorem.

□

## 3 Random valuations

### 3.1 Preliminaries

In this section we will consider the case each buyer's valuation  $v$  is an independent random sample from a fixed but unknown probability distribution on  $[0, 1]$ . It is customary to describe this probability distribution in terms of its *demand curve*

$$D(x) = \Pr(v \geq x).$$

Given foreknowledge of the demand curve, but not of the individual buyers' valuations, it is easy to see what the optimal pricing strategy would be. The expected revenue obtained from setting price  $x$  is  $xD(x)$ . Since buyers' valuations are independent and the demand curve is known, the individual buyers'

responses provide no useful information about future buyers' valuations. The best strategy is thus to compute

$$x^* = \arg \max_{x \in [0,1]} xD(x)$$

and to offer this price to every buyer. We denote this strategy by  $\mathcal{S}^*$ , and its expected revenue by  $\rho(\mathcal{S}^*)$ . Clearly, for any on-line pricing strategy  $\mathcal{S}$ , we have

$$\rho(\mathcal{S}) \leq \rho(\mathcal{S}^*) \leq \rho(\mathcal{S}^{\text{opt}}),$$

and it may be argued that in the context of random valuations it makes the most sense to compare  $\rho(\mathcal{S})$  with  $\rho(\mathcal{S}^*)$  rather than  $\rho(\mathcal{S}^{\text{opt}})$ . We address this issue by proving a lower bound on  $\rho(\mathcal{S}^*) - \rho(\mathcal{S})$  and an upper bound on  $\rho(\mathcal{S}^{\text{opt}}) - \rho(\mathcal{S})$ .

A deterministic pricing strategy can be specified by a sequence of rooted planar binary trees  $T_1, T_2, \dots$ , where the  $n$ -th tree specifies the decision tree to be applied by the seller when interacting with a population of  $n$  buyers. (Thus  $T_n$  is a complete binary tree of depth  $n$ .) We will use  $a$  to denote a generic internal node of such a decision tree, and  $\ell$  to denote a generic leaf. The relation  $a \prec b$  will denote that  $b$  is a descendant of  $a$ ; here  $b$  may be a leaf or another internal node. If  $e$  is an edge of  $T$ , we will also use  $a \prec e$  (resp.  $e \prec a$ ) to denote that  $e$  is below (resp. above)  $a$  in  $T$ , i.e. at least one endpoint of  $e$  is a descendant (resp. ancestor) of  $a$ . The left subtree rooted at  $a$  will be denoted by  $T_l(a)$ , the right subtree by  $T_r(a)$ . Note that  $T_l(a)$  (resp.  $T_r(a)$ ) includes the edge leading from  $a$  to its left (resp. right) child.

The internal nodes of the tree are labeled with numbers  $x_a \in [0, 1]$  denoting the price offered by the seller at node  $a$ , and random variables  $v_a \in [0, 1]$  denoting the valuation of the buyer with whom the seller interacts at that node. The buyer's choice is represented by a random variable

$$\chi_a = \begin{cases} 1 & \text{if } v_a \geq x_a \\ 0 & \text{if } v_a < x_a \end{cases}.$$

In other words,  $\chi_a$  is 1 if the buyer accepts the price offered, 0 otherwise.

The tree  $T_n$  specifies a pricing strategy as follows. The seller starts at the root  $r$  of the tree and offers the first buyer price  $x_r$ . The seller moves from this node to its left or right child depending on whether the buyer declines or accepts the offer, and repeats this process until reaching a leaf which represents the outcome of the auction.

A strategy as defined above is called a *non-uniform deterministic pricing strategy*. A uniform deterministic pricing strategy is one in which there is a single infinite tree  $T$  whose first  $n$  levels comprise  $T_n$  for each  $n$ . (This corresponds to a pricing strategy which is not informed of the value of  $n$  at the outset of the auction.) A randomized pricing strategy is a probability distribution over deterministic pricing strategies.

As mentioned above, the outcome of the auction may be represented by a leaf  $\ell \in T_n$ , i.e. the unique leaf such that for all ancestors  $a \prec \ell$ ,  $\ell \in T_r(a) \Leftrightarrow \chi_a = 1$ . A probability distribution on the buyers' valuations  $v_a$  induces a probability

distribution on outcomes  $\ell$ . We will use  $p_D(\ell)$  to denote the probability assigned to  $\ell$  under the valuation distribution represented by demand curve  $D$ . For an internal node  $a$ ,  $p_D(a)$  denotes the probability that the outcome leaf is a descendant of  $a$ . We define  $p_D(e)$  similarly for edges  $e \in T$ .

## 3.2 Lower bound

### 3.2.1 A family of random demand curves

The demand curves  $D$  appearing in our lower bound will be random samples from a space  $\mathcal{D}$  of possible demand curves. In this section we single out a particular random demand-curve model, and we enumerate the properties which will be relevant in establishing the lower bound. The choice of a particular random demand-curve model is done here for ease of exposition, and not because of a lack of generality in the lower bound itself. In Section 3.2.6 we will indicate that Theorem 3.9 applies to much broader classes  $\mathcal{D}$  of demand curves. In particular we believe that it encompasses random demand-curve models which are realistic enough to be of interest in actual economics and e-commerce applications.

For now, however,  $\mathcal{D}$  denotes the one-parameter family of demand curves  $\{D_t : 0.3 \leq t \leq 0.4\}$  defined as follows. Let

$$\tilde{D}_t(x) = \max \left\{ 1 - 2x, \frac{2t - x}{7t^2}, \frac{1 - x}{2} \right\}.$$

In other words, the graph of  $\tilde{D}_t$  consists of three line segments: the middle segment is tangent to the curve  $xy = 1/7$  at the point  $(t, 1/7t)$ , while the left and right segments belong to lines which lie below that curve and are independent of  $t$ . Now we obtain  $D_t$  by smoothing  $\tilde{D}_t$ . Specifically, let  $b(x)$  be a non-negative, even  $C^\infty$  function supported on the interval  $[-0.01, 0.01]$  and satisfying  $\int_{-0.01}^{0.01} b(x) dx = 1$ . Define  $D_t$  by convolving  $\tilde{D}_t$  with  $b$ , i.e.

$$D_t(x) = \int_{-\infty}^{\infty} \tilde{D}_t(y) b(x - y) dy.$$

We will equip  $\mathcal{D} = \{D_t : 0.3 \leq t \leq 0.4\}$  with a probability measure by specifying that  $t$  is uniformly distributed in  $[0.3, 0.4]$ .

Let  $x_t^* = \arg \max_{x \in [0, 1]} x D_t(x)$ . It is an exercise to compute that  $x_t^* = t$ . (With  $\tilde{D}_t$  in place of  $D_t$  this would be trivial. Now  $D_t(x) = \tilde{D}_t(x)$  unless  $x$  is within 0.01 of one of the two points where  $\tilde{D}_t$  is discontinuous, and these two points are far from maximizing  $x \tilde{D}_t(x)$ , so  $x D_t(x)$  is also maximized at  $x = t$ .)

The specifics of the construction of  $\mathcal{D}$  are not important, except insofar as they enable us to prove the properties specified in the following lemma.

**Lemma 3.1.** *There exist constants  $\alpha, \beta > 0$  and  $\gamma < \infty$  such that for all  $D = D_{t_0} \in \mathcal{D}$  and  $x \in [0, 1]$ :*

1.  $\frac{d}{dt}(x_t^*)|_{t=t_0} > \alpha$ ;

2.  $x^*D(x^*) - xD(x) > \beta(x^* - x)^2$ ;
3.  $|\dot{D}(x)/D(x)| < \gamma|x^* - x|$  and  $|\dot{D}(x)/(1 - D(x))| < \gamma|x^* - x|$ ;
4.  $|D^{(k)}(x)/D(x)| < \gamma$  and  $|D^{(k)}(x)/(1 - D(x))| < \gamma$ , for  $k = 2, 3, 4$ .

Here  $x^*$  denotes  $x_{t_0}^*$ ,  $D^{(k)}(x)$  denotes the  $k$ -th  $t$ -derivative of  $D_t(x)$  at  $t = t_0$ , and  $\dot{D}(x)$  denotes  $D^{(1)}(x)$ .

The proof of the lemma is elementary but tedious, so it is deferred to Appendix A.

### 3.2.2 High-level description of the proof

The proof of the lower bound on regret is based on the following intuition. If there is uncertainty about the demand curve, then no single price can achieve a low expected regret for all demand curves. The family of demand curves exhibited above is parametrized by a single parameter  $t$ , and we will see that if the uncertainty about  $t$  is on the order of  $\varepsilon$  then the regret per buyer is  $\Omega(\varepsilon^2)$ . (This statement will be made precise in Lemma 3.7 below.) So to avoid accumulating  $\Omega(\sqrt{n})$  regret on the last  $\Omega(n)$  buyers, the pricing strategy must ensure that it reduces the uncertainty to  $O(n^{-1/4})$  during its interactions with the initial  $O(n)$  buyers. However — and this is the crux of the proof — we will show that offering prices far from  $x^*$  is much more informative than offering prices near  $x^*$ , so there is a quantifiable cost to reducing the uncertainty in  $t$ . In particular, reducing the uncertainty to  $O(n^{-1/4})$  costs  $\Omega(\sqrt{n})$  in terms of expected regret.

To make these ideas precise, we will introduce a notion of “knowledge” which quantifies the seller’s ability to distinguish the actual demand curve from nearby ones based on the information obtained from past transactions, and a notion of “conditional regret” whose expectation is a lower bound on the pricing strategy’s expected regret. We will show that the ratio of conditional regret to knowledge is bounded below, so that the strategy cannot accumulate  $\Omega(\sqrt{n})$  knowledge without accumulating  $\Omega(\sqrt{n})$  regret. Finally, we will show that when the expected knowledge is less than a small constant multiple of  $\sqrt{n}$ , there is so much uncertainty about the true demand curve that the expected regret is  $\Omega(\sqrt{n})$  with high probability (taken over the probability measure on demand curves).

### 3.2.3 Definition of knowledge

In the following definitions,  $\log$  denotes the natural logarithm function.  $T$  denotes a finite planar binary tree, labeled with a pricing strategy as explained in Section 3.1. When  $f$  is a function defined on leaves of  $T$ , we will use the notation  $\mathbf{E}_D f$  to denote the expectation of  $f$  with respect to the probability distribution  $p_D$  on leaves, i.e.

$$\mathbf{E}_D f = \sum_{\ell \in T} p_D(\ell) f(\ell).$$

For a given demand curve  $D = D_{t_0}$ , we define the *infinitesimal relative entropy* of a leaf  $\ell \in T$  by

$$IRE_D(\ell) = \frac{d}{dt}(-\log p_{D_t}(\ell))|_{t=t_0},$$

and we define the *knowledge* of  $\ell$  as the square of the infinitesimal relative entropy:

$$K_D(\ell) = IRE_D(\ell)^2.$$

Those familiar with information theory may recognize  $IRE_D(\ell)$  as the  $t$ -derivative of  $\ell$ 's contribution to the weighted sum defining the relative entropy  $RE(D||D_t)$ , and  $K_D(\ell)$  as a random variable whose expected value is a generalization of the notion of *Fisher information*.

An important feature of  $IRE_D(\ell)$  is that it may be expressed as a sum of terms coming from the edges of  $T$  leading from the root to  $\ell$ . For an edge  $e = (a, b) \in T$ , let

$$\begin{aligned} ire_D(e) &= \begin{cases} \frac{d}{dt}(\log D(x_a)) & \text{if } e \in T_r(a) \\ \frac{d}{dt}(\log(1 - D(x_a))) & \text{if } e \in T_l(a) \end{cases} \\ &= \begin{cases} \dot{D}(x_a)/D(x_a) & \text{if } e \in T_r(a) \\ -\dot{D}(x_a)/(1 - D(x_a)) & \text{if } e \in T_l(a) \end{cases} \end{aligned}$$

Then

$$IRE_D(\ell) = \sum_{e \prec \ell} ire_D(e).$$

### 3.2.4 Definition of conditional regret

For a given  $D$ , the conditional regret  $R_D(\ell)$  may be informally defined as follows. At the end of the auction, if the demand curve  $D$  were revealed to the seller and then she were required to repeat the same sequence of offered prices  $\{x_a : a \prec \ell\}$  to a new, independent random population of buyers whose valuations are distributed according to  $D$ , then  $R_D(\ell)$  is the expected regret incurred by the seller during this second round of selling. Formally,  $R_D(\ell)$  is defined as follows. Let

$$r_D(x) = x^*D(x^*) - xD(x),$$

where  $x^* = \arg \max_{x \in [0,1]} \{xD(x)\}$  as always. Note that if two different sellers offer prices  $x^*, x$ , respectively, to a buyer whose valuation is distributed according to  $D$ , then  $r_D(x)$  is the difference in their expected revenues. Now let

$$R_D(\ell) = \sum_{a \prec \ell} r_D(x_a).$$

Although  $R_D(\ell)$  is *not* equal to the seller's actual regret conditional on outcome  $\ell$ , it is a useful invariant because  $\mathbf{E}_D R_D(\ell)$  is equal to the actual expected regret of  $\mathcal{S}$  relative to  $\mathcal{S}^*$ . (It is also therefore a lower bound on the expected regret

of  $\mathcal{S}$  relative to  $\mathcal{S}^{\text{opt}}$ .) This fact is far from obvious, because the distribution of the actual buyers' valuations, conditioned on their responses to the prices they were offered, is very different from the distribution of  $n$  new independent buyers. In general the expected revenue of  $\mathcal{S}$  or  $\mathcal{S}^*$  on the hypothetical independent population of  $n$  buyers will not equal the expected revenue obtained from the actual population of  $n$  buyers, conditioned on those buyers' responses. Yet the expected difference between the two random variables, i.e. the regret, is the same for both populations of buyers. This fact is proved in the following lemma.

**Lemma 3.2.** *Let  $S$  be a strategy with decision tree  $T$ , and let  $S^*$  be the fixed-price strategy which offers  $x^*$  to each buyer. If the buyers' valuations are independent random samples from the distribution specified by  $D$ , then the expected revenue of  $S^*$  exceeds that of  $S$  by exactly  $\mathbf{E}_D R_D(\ell)$ .*

*Proof.* Let

$$\chi_a^* = \begin{cases} 1 & \text{if } v_a \geq x^* \\ 0 & \text{if } v_a < x^* \end{cases} .$$

At a given point of the sample space, let  $\ell$  denote the outcome leaf, and let  $a_1, a_2, \dots, a_n$  be the ancestors of  $\ell$ . Then the revenue of  $S^*$  is  $\sum_{i=1}^n \chi_{a_i}^* x^*$ , and the revenue of  $S$  is  $\sum_{i=1}^n \chi_{a_i} x_{a_i}$ . It follows that the expected difference between the two is

$$\begin{aligned} \sum_{a \in T} p_D(a) [\mathbf{E}_D(\chi_a^* x^*) - \mathbf{E}_D(\chi_a x_a)] &= \sum_{a \in T} p_D(a) [x^* D(x^*) - x_a D(x_a)] \\ &= \sum_{a \in T} \sum_{\ell > a} p_D(\ell) r_D(x_a) \\ &= \sum_{\ell \in T} p_D(\ell) \left( \sum_{a < \ell} r_D(x_a) \right) \\ &= \mathbf{E}_D R_D(\ell). \end{aligned}$$

□

### 3.2.5 Proof of the lower bound

In stating the upcoming lemmas, we will introduce constants  $c_1, c_2, \dots$ . When we introduce such a constant we are implicitly asserting that there exists a constant  $0 < c_i < \infty$  depending only on the demand curve family  $\mathcal{D}$ , and satisfying the property specified in the statement of the corresponding lemma.

We begin with a series of lemmas which establish that  $\mathbf{E}_D K_D$  is bounded above by a constant multiple of  $\mathbf{E}_D R_D$ . Assume for now that  $D$  is fixed, so  $x^*$  is also fixed, and put

$$h_a = x_a - x^*.$$

**Lemma 3.3.**  $\mathbf{E}_D R_D(\ell) \geq c_1 \sum_{a \in T} p_D(a) h_a^2$ .

*Proof.* Recall from Lemma 3.1 that

$$(x^* + h)D(x^* + h) < x^*D(x^*) - \beta h^2,$$

hence

$$r_D(x_a) = x^*D(x^*) - x_aD(x_a) > \beta h_a^2.$$

Now we see that

$$\begin{aligned} \mathbf{E}_D R_D(\ell) &= \sum_{\ell \in T} p_D(\ell) \left( \sum_{a \prec \ell} r_D(x_a) \right) \\ &= \sum_{a \in T} \left( \sum_{\ell \succ a} p_D(\ell) \right) r_D(x_a) \\ &= \sum_{a \in T} p_D(a) r_D(x_a) \\ &> \beta \sum_{a \in T} p_D(a) h_a^2. \end{aligned}$$

so the lemma holds with  $c_1 = \beta$ .  $\square$

**Lemma 3.4.**  $\mathbf{E}_D K_D(\ell) \leq c_2 \sum_{a \in T} p_D(a) h_a^2$ .

*Proof.* As in the preceding lemma, the idea is to rewrite the sum over leaves as a sum over internal nodes and then bound the sum term-by-term. (In this case, actually it is a sum over internal edges of  $T$ .) A complication arises from the fact that the natural expression for  $\mathbf{E}_D K_D(\ell)$  involves summing over pairs of ancestors of a leaf; however, we will see that all of the cross-terms cancel, leaving us with a manageable expression.

$$\begin{aligned} \mathbf{E}_D K_D(\ell) &= \sum_{\ell} p_D(\ell) I R E_D(\ell)^2 \\ &= \sum_{\ell} p_D(\ell) \left( \sum_{e \prec \ell} ire_D(e) \right)^2 \\ &= \sum_{\ell} p_D(\ell) \left[ \sum_{e \prec \ell} ire_D(e)^2 + 2 \sum_{e \prec e' \prec \ell} ire_D(e) ire_D(e') \right] \\ &= \left[ \sum_e \sum_{\ell \succ e} p_D(\ell) ire_D(e)^2 \right] + 2 \left[ \sum_e \sum_{e' \succ e} ire_D(e) \left( \sum_{\ell \succ e'} p_D(\ell) ire_D(e') \right) \right] \\ &= \left[ \sum_e p_D(e) ire_D(e)^2 \right] + 2 \left[ \sum_e ire_D(e) \left( \sum_{e' \succ e} p_D(e') ire_D(e') \right) \right]. \quad (1) \end{aligned}$$

For any  $e \in T$ , the sum  $\sum_{e' \succ e} p_D(e') ire_D(e')$  vanishes because the terms may be grouped into pairs  $p_D(e') ire_D(e') + p_D(e'') ire_D(e'')$  where  $e', e''$  are the edges

joining a node  $a \in T$  to its right and left children, respectively, and we have

$$\begin{aligned} & p_D(e')ire_D(e') + p_D(e'')ire_D(e'') \\ &= p_D(a) \left[ D(x_a) \left( \frac{\dot{D}(x_a)}{D(x_a)} \right) + (1 - D(x_a)) \left( -\frac{\dot{D}(x_a)}{1 - D(x_a)} \right) \right] = 0. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{E}_D K_D(\ell) &= \sum_{e \in T} p_D(e)ire_D(e)^2 \\ &= \sum_a p_D(a) \left[ D(x_a) \left( \frac{\dot{D}(x_a)}{D(x_a)} \right)^2 + (1 - D(x_a)) \left( -\frac{\dot{D}(x_a)}{1 - D(x_a)} \right)^2 \right] \\ &\leq \sum_a p_D(a) \left[ \left( \frac{\dot{D}(x_a)}{D(x_a)} \right)^2 + \left( -\frac{\dot{D}(x_a)}{1 - D(x_a)} \right)^2 \right] \\ &< \sum_a p_D(a) (\gamma^2 h_a^2 + \gamma^2 h_a^2), \end{aligned}$$

so the lemma holds with  $c_2 = 2\gamma^2$ .  $\square$

**Corollary 3.5.**  $\mathbf{E}_D K_D(\ell) \leq c_3 \mathbf{E}_D R_D(\ell)$ .

The relevance of Corollary 3.5 is that it means that when  $\mathbf{E}_D R_D$  is small, then  $p_{D_t}(\ell)$  cannot shrink very rapidly as a function of  $t$ , for most leaves  $\ell$ . This is made precise by the following Lemma. Here and throughout the rest of this section,  $D$  refers to a demand curve  $D_{t_0} \in \mathcal{D}$ .

**Lemma 3.6.** *For all sufficiently large  $n$ , if  $\mathbf{E}_D R_D < \sqrt{n}$  then there exists a set  $S$  of leaves such that  $p_D(S) \geq 1/2$ , and  $p_{D_t}(\ell) > c_4 p_D(\ell)$  for all  $\ell \in S$  and all  $t \in [t_0, t_0 + n^{-1/4}]$ .*

The proof is quite elaborate, so we have deferred it to Appendix B.

We will also need a lemma establishing the growth rate of  $R_{D_t}(\ell)$  for a fixed leaf  $\ell$ , as  $t$  varies.

**Lemma 3.7.**  $R_D(\ell) + R_{D_t}(\ell) > c_5(t - t_0)^2 n$  for all leaves  $\ell \in T_n$  and for all  $D_t \in \mathcal{D}$ .

*Proof.* We know that

$$\begin{aligned} R_D(\ell) &= \sum_{a \prec \ell} r_D(x_a) \\ R_{D_t}(\ell) &= \sum_{a \prec \ell} r_{D_t}(x_a) \end{aligned}$$

so it suffices to prove that  $r_D(x) + r_{D_t}(x) > c_4(t - t_0)^2$  for all  $x \in [0, 1]$ . Assume without loss of generality that  $t - t_0 > 0$ . (Otherwise, we may reverse the roles



of  $D$  and  $D_t$ .) Let  $x^*$  and  $x_t^*$  denote the optimal prices for  $D, D_t$ , respectively. (Recall that  $D = D_{t_0}$ .) Note that  $x_t^* - x^* > \alpha(t - t_0)$ , by property 1 of Lemma 3.1.

Let  $h = x - x^*$ ,  $h_t = x - x_t^*$ , and note that  $|h| + |h_t| > \alpha(t - t_0)$ . Now

$$\begin{aligned} r_D(x) &> c_1|h|^2 \\ r_{D_t}(x) &> c_1|h_t|^2 \\ r_D(x) + r_{D_t}(x) &> c_1(|h|^2 + |h_t|^2) \\ &\geq \frac{1}{2}c_1(|h| + |h_t|)^2 \\ &> \frac{1}{2}c_1\alpha^2 t^2, \end{aligned}$$

so the lemma holds with  $c_5 = \frac{1}{2}c_1\alpha^2$ .  $\square$

We now exploit Lemmas 3.6 and 3.7 to prove that if  $\mathbf{E}_D R_D$  is less than some small constant multiple of  $\sqrt{n}$  when  $D = D_{t_0}$ , then  $\mathbf{E}_{D_t} R_{D_t} = \Omega(\sqrt{n})$  on a large fraction of the interval  $[t_0, t_0 + n^{-1/4}]$ . The idea behind the proof is that Lemma 3.6 tells us there is a large set  $S$  of leaves whose measure does not vary by more than a constant factor as we move  $t$  across this interval, while Lemma 3.7 tells us that the regret contribution from leaves in  $S$  is  $\Omega(\sqrt{n})$  for a large fraction of the  $t$ -values in this interval. In the following proposition,  $c(M)$  denotes the function  $\min\{1, \frac{1}{2}c_4c_5(1 + c_4)^{-1}M^{-2}\}$ .

**Proposition 3.8.** *For all  $M$  and all sufficiently large  $n$ , if  $\mathbf{E}_D R_D < c(M)\sqrt{n}$ , then  $\mathbf{E}_{D_t} R_{D_t} > c(M)\sqrt{n}$  for all  $t \in [t_0 + (1/M)n^{-1/4}, t_0 + n^{-1/4}]$ .*

*Proof.* If  $\mathbf{E}_D R_D < c(M)\sqrt{n}$ , we may apply Lemma 3.6 to produce a set  $S$  of leaves such that  $p_D(S) \geq 1/2$  and  $p_{D_t}(\ell) > c_4 p_D(\ell)$  for all  $\ell \in S$  and all  $t \in [t_0, t_0 + n^{-1/4}]$ . Now,

$$\mathbf{E}_D R_D \geq \sum_{\ell \in S} p_D(\ell) R_D(\ell)$$

and, for all  $t \in [t_0 + (1/M)n^{-1/4}, t_0 + n^{-1/4}]$ ,

$$\begin{aligned} \mathbf{E}_{D_t} R_{D_t} &\geq \sum_{\ell \in S} p_{D_t}(\ell) R_{D_t}(\ell) \\ &\geq c_4 \sum_{\ell \in S} p_D(\ell) R_{D_t}(\ell) \\ &> c_4 \sum_{\ell \in S} p_D(\ell) (c_5(t - t_0)^2 n - R_D(\ell)) \\ &> c_4 p_D(S) c_5 \sqrt{n} / M^2 - c_4 \mathbf{E}_D R_D \\ &> (c_4 c_5 / 2 M^2) \sqrt{n} - c_4 c(M) \sqrt{n} \\ &\geq c(M) \sqrt{n} \end{aligned}$$

where the fourth line is derived from the third by applying the inequality  $R_{D_t}(\ell) > c_4(t - t_0)^2 n - R_D(\ell)$  coming from Lemma 3.7.  $\square$

**Theorem 3.9.** *Let  $\mathcal{S}$  be any randomized non-uniform strategy, and let  $\mathcal{R}_D(\mathcal{S}, n)$  denote the expected ex ante regret of  $\mathcal{S}$  on a population of  $n$  buyers whose valuations are independent random samples from the probability distribution specified by the demand curve  $D$ . Then*

$$\Pr_{D \leftarrow \mathcal{D}} \left( \limsup_{n \rightarrow \infty} \frac{\mathcal{R}_D(\mathcal{S}, n)}{\sqrt{n}} > 0 \right) = 1.$$

*In other words, if  $D$  is drawn at random from  $\mathcal{D}$ , then almost surely  $\mathcal{R}_D(\mathcal{S}, n)$  is not  $o(\sqrt{n})$ .*

*Proof.* It suffices to prove the theorem for a deterministic strategy  $\mathcal{S}$ , since any randomized strategy is a probability distribution over such strategies. Now assume, to the contrary, that

$$\Pr_{D \leftarrow \mathcal{D}} \left( \limsup_{n \rightarrow \infty} \frac{\mathcal{R}_D(\mathcal{S}, n)}{\sqrt{n}} = 0 \right) > 0. \quad (2)$$

and choose  $M$  large enough that the left side of (2) is greater than  $1/M$ . Recall from Lemma 3.3 that  $\mathbf{E}_D R_D = \mathcal{R}_D(\mathcal{S}, n)$ . We know that for every  $D = D_{t_0} \in \mathcal{D}$  such that  $\mathbf{E}_D R_D < c(M)\sqrt{n}$ ,

$$\mathbf{E}_{D_t} R_{D_t} > c(M)\sqrt{n} \quad \forall t \in [t_0 + (1/M)n^{-1/4}, t_0 + n^{-1/4}]. \quad (3)$$

Now choose  $N$  large enough that the set

$$X_N = \left\{ D \in \mathcal{D} : \sup_{n > N} \frac{\mathcal{R}_D(\mathcal{S}, n)}{\sqrt{n}} < c(M) \right\}$$

has measure greater than  $1/M$ . Replacing  $X_N$  if necessary with a proper subset still having measure greater than  $1/M$ , we may assume that  $\{t : D_t \in X_N\}$  is disjoint from  $[0.4 - \varepsilon, 0.4]$  for some  $\varepsilon > 0$ . Choosing  $n$  large enough that  $n > N$  and  $n^{-1/4} < \varepsilon$ , equation (3) ensures that the sets

$$X_N^k = \{D_s : s = t + (k/M)n^{-1/4}, D_t \in X_N\}$$

are disjoint for  $k = 0, 1, \dots, M-1$ . But each of the sets  $X_N^k$ , being a translate of  $X_N$ , has measure greater than  $1/M$ . Thus their total measure is greater than 1, contradicting the fact that  $\mathcal{D}$  has measure 1.  $\square$

### 3.2.6 General demand-curve models

The methods of the preceding section extend to much more general families of demand curves. Here we will merely sketch the ideas underlying the extension. Suppose that  $\mathcal{D}$  is a compact subset of the space  $C^4([0, 1])$  of functions on  $[0, 1]$  with continuous fourth derivative, and that the demand curves  $D \in \mathcal{D}$  satisfy the following two additional hypotheses:

- **(Unique global max)** The function  $f(x) = xD(x)$  has a unique global maximum  $x^* \in [0, 1]$ , and it lies in the interior of the interval.

- **(Non-degeneracy)** The second derivative of  $f$  is strictly negative at  $x^*$ .

Suppose  $\mathcal{D}$  is also endowed with a probability measure, denoted  $\mu$ . The proof of the lower bound relied heavily on the notion of being able to make a “one-parameter family of perturbations” to a demand curve. This notion may be encapsulated using a *flow*  $\phi(D, t)$  mapping an open set  $U \subseteq \mathcal{D} \times \mathbb{R}$  into  $\mathcal{D}$ , such that  $(D \times \{0\}) \cap U$  has measure 1, and  $\phi(D, 0) = D$  when defined. We will use the shorthand  $D_t$  for  $\phi(D, t)$ . The flow must satisfy the following properties:

- **(Additivity)**  $\phi(D, s + t) = \phi(\phi(D, s), t)$ .
- **(Measure-preservation)** If  $X \subseteq \mathcal{D}$  and  $\phi(D, t)$  is defined for all  $D \in X$ , then  $\mu(\phi(X, t)) = \mu(X)$ .
- **(Smoothness)** The function  $g(t, x) = D_t(x)$  is a  $C^4$  function of  $t$  and  $x$ .
- **(Profit-preservation)** If  $x_t^*$  denotes the point at which the function  $x D_t(x)$  achieves its global maximum, then  $x_t^* D_t(x_t^*) = x_0^* D_0(x_0^*)$  for all  $t$  such that  $D_t$  is defined.
- **(Non-degeneracy)**  $\frac{d}{dt}(x_t^*) \neq 0$ .
- **(Rate dampening at 0 and 1)** For  $k = 1, 2, 3, 4$ , the functions  $\left| \frac{D^{(k)}}{D} \right|$  and  $\left| \frac{D^{(k)}}{1-D} \right|$  are uniformly bounded above, where  $D^{(k)}$  denotes the  $k$ -th derivative of  $D$  with respect to  $t$ .

Provided that these axioms are satisfied, it is possible to establish all of the properties specified in Lemma 3.1. Property 1 follows from compactness of  $\mathcal{D}$  and non-degeneracy of  $\phi$ , property 2 follows from the compactness of  $\mathcal{D}$  together with the non-degeneracy and “unique global max” axioms for  $\mathcal{D}$ , and property 4 is the rate-dampening axiom. Property 3 is the subtlest: it follows from the smoothness, profit-preservation, and rate-dampening properties of  $\phi$ . The key observation is that profit-preservation implies that

$$x^* D_t(x^*) \leq x_t^* D_t(x_t^*) = x^* D(x^*),$$

so that  $x^* D_t(x^*)$ , as a function of  $t$ , is maximized at  $t = 0$ . This, coupled with smoothness of  $\phi$ , proves that  $\dot{D}(x^*) = 0$ . Another application of smoothness yields the desired bounds.

The final steps of Theorem 1 used the translation-invariance of Lebesgue measure on the interval  $[0.3, 0.4]$  to produce  $M$  sets whose disjointness yielded the desired contradiction. This argument generalizes, with the flow  $\phi$  playing the role of the group of translations. It is for this reason that we require  $\phi$  to satisfy the additivity and measure-preservation axioms.

### 3.3 Upper bound

The upper bound on regret in the random-valuation model is based on applying techniques from the literature on the multi-armed bandit problem, specifically [2]. To do so, we discretize the set of possible actions by limiting the seller

to strategies which only offer prices belonging to the set  $\{1/K, 2/K, \dots, 1 - 1/K, 1\}$ , for suitably-chosen  $K$ . (It will turn out that  $K = \theta((n/\log n)^{1/4})$  is the best choice.)

We are now in a setting where the seller must choose one of  $K$  possible actions on each of  $n$  trials, where each action yields a reward which is a random variable taking values in  $[0, 1]$ , whose distribution depends on the action chosen, but the rewards for a given action are i.i.d. across the  $n$  trials. This is the scenario studied in [2]. They define  $\mu^i$  to be the expected reward of action  $i$ ,  $\mu^* = \max\{\mu_1, \dots, \mu_K\}$ , and

$$\Delta_i = \mu^* - \mu_i. \quad (4)$$

Having made these definitions, the following theorem is proven.

**Theorem 3.10** ([2], Theorem 1.). *There exists a strategy UCB1 such that, for all  $K > 1$ , if UCB1 is run on a set of  $K$  actions having arbitrary reward distributions  $P_1, \dots, P_K$  with support in  $[0, 1]$ , then its expected regret after any number  $n$  of plays is at most*

$$\left[ 8 \sum_{i: \mu_i < \mu^*} \left( \frac{\log n}{\Delta_i} \right) \right] + \left( 1 + \frac{\pi^2}{3} \right) \left( \sum_{j=1}^K \Delta_j \right).$$

To be precise, the model in [2] assumes that the reward variables for different actions are independent, an assumption which does not hold in our scenario. However, this assumption is not used in their proof of Theorem 3.10, so we may still apply the theorem.

To apply this theorem, we need to know something about the values of  $\Delta_1, \dots, \Delta_K$  in the special case of interest to us. When the buyer's valuation is  $v$ , the payoff of action  $i/K$  is

$$X_i = \begin{cases} i/K & \text{if } v \geq i/K \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Hence

$$\mu_i = \mathbf{E}(X_i) = (i/K)D(i/K). \quad (6)$$

Recall that we are making the following hypothesis on the demand curve  $D$ : the function  $f(x) = xD(x)$  has a unique global maximum at  $x^* \in (0, 1)$ , and  $f''(x^*)$  is defined and strictly negative. This hypothesis is useful because it enables us to establish the following lemma, which translates directly into bounds on  $\Delta_i$ .

**Lemma 3.11.** *There exist constants  $C_1, C_2$  such that  $C_1(x^* - x)^2 < f(x^*) - f(x) < C_2(x^* - x)^2$  for all  $x \in [0, 1]$ .*

*Proof.* The existence and strict negativity of  $f''(x^*)$  guarantee that there are constants  $A_1, A_2, \varepsilon > 0$  such that  $A_1(x^* - x)^2 < f(x^*) - f(x) < A_2(x^* - x)^2$  for all  $x \in (x^* - \varepsilon, x^* + \varepsilon)$ . The compactness of  $X = \{x \in [0, 1] : |x^* - x| \geq \varepsilon\}$ , together with the fact that  $f(x^*) - f(x)$  is strictly positive for all  $x \in X$ , guarantees that

there are constants  $B_1, B_2$  such that  $B_1(x^* - x)^2 < f(x^*) - f(x) < B_2(x^* - x)^2$  for all  $x \in X$ . Now put  $C_1 = \min\{A_1, B_1\}$  and  $C_2 = \max\{A_2, B_2\}$  to obtain the lemma.  $\square$

**Corollary 3.12.**  $\Delta_i \geq C_1(x^* - i/K)^2$  for all  $i$ . If  $\tilde{\Delta}_0 \leq \tilde{\Delta}_1 \leq \dots \leq \tilde{\Delta}_{K-1}$  are the elements of the set  $\{\Delta_1, \dots, \Delta_k\}$  sorted in ascending order, then  $\tilde{\Delta}_j \geq C_1(j/2K)^2$ .

*Proof.* The inequality  $\Delta_i \geq C_1(x^* - i/K)^2$  is a restatement of the lemma using the formulae for  $\Delta_i, \mu_i$  given above in (4),(6). The lower bound on  $\tilde{\Delta}_j$  follows upon observing that at most  $j$  elements of the set  $\{1/K, 2/K, \dots, 1\}$  lie within a distance  $j/2K$  of  $x^*$ .  $\square$

**Corollary 3.13.**  $\mu^* > x^*D(x^*) - C_2/K^2$ .

*Proof.* At least one of the numbers  $\{1/K, 2/K, \dots, 1\}$  lies within  $1/K$  of  $x^*$ ; now apply the upper bound on  $f(x^*) - f(x)$  stated in the lemma.  $\square$

Putting all of this together, we have derived the following upper bound.

**Theorem 3.14.** *Assuming that the function  $f(x) = xD(x)$  has a unique global maximum  $x^* \in (0, 1)$ , and that  $f''(x^*)$  is defined and strictly negative, the strategy UCB1 with  $K = \lceil (n/\log n)^{1/4} \rceil$  achieves expected regret  $O(\sqrt{n \log n})$ .*

*Proof.* Consider the following four strategies:

- UCB1, the strategy defined in [2].
- $\mathcal{S}^{\text{opt}}$ , the optimal fixed-price strategy.
- $\mathcal{S}^*$ , the fixed-price strategy which offers  $x^*$  to every buyer.
- $\mathcal{S}_K^*$ , the fixed-price strategy which offers  $i^*/K$  to every buyer, where  $i^*/K$  is the element of  $\{1/K, 2/K, \dots, 1\}$  closest to  $x^*$ .

As usual, we will use  $\rho(\cdot)$  to denote the expected revenue obtained by a strategy. We will prove a  $O(\sqrt{n \log n})$  upper bound on each of  $\rho(\mathcal{S}_K^*) - \rho(\text{UCB1})$ ,  $\rho(\mathcal{S}^*) - \rho(\mathcal{S}_K^*)$ , and  $\rho(\mathcal{S}^{\text{opt}}) - \rho(\mathcal{S}^*)$ , from which the theorem follows immediately.

We first show, using 3.10, that  $\rho(\mathcal{S}_K^*) - \rho(\text{UCB1}) = O(\sqrt{n \log n})$ . By Corollary 3.12,

$$\begin{aligned} \sum_{i: \mu_i < \mu^*} \left( \frac{1}{\Delta_i} \right) &< \sum_{i=1}^K C_1^{-1} (i/2K)^{-2} \\ &< \frac{4K^2}{C_1} \sum_{i=1}^{\infty} i^{-2} \\ &= \left( \frac{2\pi^2}{3C_1} \right) K^2 \\ &= O((n/\log n)^{1/2}) \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^K \Delta_j &< K \\ &= O((n/\log n)^{1/4}). \end{aligned}$$

Plugging these estimates into the regret bound in Theorem 3.10, we see that the regret of UCB1 relative to  $\mathcal{S}_K^*$  is  $O((n \log n)^{1/2})$ , as claimed.

Next we bound the difference  $\rho(\mathcal{S}^*) - \rho(\mathcal{S}_K^*)$ . The expected revenues of  $\mathcal{S}^*$  and  $\mathcal{S}_K^*$  are  $nx^*D(x^*)$  and  $n\mu^*$ , respectively. Applying Corollary 3.13, the regret of  $\mathcal{S}_K^*$  relative to  $\mathcal{S}^*$  is bounded above by

$$\frac{C_2 n}{K^2} \leq \frac{C_2 n}{(n/\log n)^{1/2}} = O((n \log n)^{1/2}).$$

Finally, we must bound  $\rho(\mathcal{S}^{\text{opt}}) - \rho(\mathcal{S}^*)$ . For any  $x \in [0, 1]$ , let  $\rho(x)$  denote the revenue obtained by the fixed-price strategy which offers price  $x$ , and let  $x^{\text{opt}} = \arg \max_{x \in [0, 1]} \rho(x)$ . We begin by observing that for all  $x < x^{\text{opt}}$ ,

$$\rho(x) \geq \rho(x^{\text{opt}}) - n(x^{\text{opt}} - x).$$

This is simply because every buyer that accepts price  $x^{\text{opt}}$  would also accept  $x$ , and the amount of revenue lost by setting the lower price is  $x^{\text{opt}} - x$  per buyer. Now

$$\begin{aligned} \int_0^1 \Pr(\rho(x) - \rho(x^*) > \lambda) dx &\geq \int_0^1 \Pr(\rho(x^{\text{opt}}) - \rho(x^*) > 2\lambda \text{ and } x^{\text{opt}} - x < \lambda/n) dx \\ &= \lambda/n \Pr(\rho(x^{\text{opt}}) - \rho(x^*) > 2\lambda), \end{aligned}$$

so a bound on  $\Pr(\rho(x) - \rho(x^*) > \lambda)$  for fixed  $x$  translates into a bound on  $\Pr(\rho(x^{\text{opt}}) - \rho(x^*) > \lambda)$ . But for fixed  $x$ , the probability in question is the probability that a sum of  $n$  i.i.d. random variables, each supported in  $[-1, 1]$  and with negative expectation, exceeds  $\lambda$ . The Chernoff-Hoeffding bound tells us that

$$\Pr(\rho(x) - \rho(x^*) > \lambda) < e^{-\lambda^2/2n},$$

so

$$\Pr(\rho(x^{\text{opt}}) - \rho(x^*) > 2\lambda) < \min\left\{1, \frac{n}{\lambda} e^{-\lambda^2/2n}\right\}.$$

Finally,

$$\begin{aligned} \mathbf{E}(\rho(x^{\text{opt}}) - \rho(x^*)) &< \int_0^\infty \Pr(\rho(x^{\text{opt}}) - \rho(x^*) > y) dy \\ &< \int_0^\infty \min\left\{1, \frac{2n}{y} e^{-y^2/2n}\right\} dy \\ &< \int_0^{\sqrt{4n \log n}} dy + \frac{2n}{\sqrt{4n \log n}} \int_{\sqrt{4n \log n}}^\infty e^{-y^2/2n} dy \\ &= O(\sqrt{n \log n}). \end{aligned}$$

□

**Remark 3.15.** If the seller does not have foreknowledge of  $n$ , it is still possible to achieve regret  $O(\sqrt{n \log n})$  by maintaining an estimate  $n_{\text{est}}$  of  $n$ , initialized to 1. When  $2^k$  buyers have been seen, the seller sets  $n_{\text{est}}$  to  $2^{k+1}$  and reinitializes UCB1 using this new value of  $n_{\text{est}}$ .

## 4 Worst-case valuations

In the worst-case valuation model, we assume that the buyers' valuations are chosen by an adversary who has knowledge of  $n$  and of the pricing strategy, but is oblivious to the algorithm's random choices. A similar adversarial model for the multi-armed bandit problem was considered by Auer et al in [3]. In that paper the authors present an algorithm **Exp3** achieving regret  $O(\sqrt{nK \log K})$ , where  $K$  is the number of possible actions, and they exhibit a lower bound of  $\Omega(\sqrt{nK})$  on regret. The algorithm, which is based on the weighted-majority learning algorithm of Littlestone and Warmuth, was applied in the setting of on-line auctions by Blum et al [6], who normalize the buyers' valuations to lie in an interval  $[1, h]$  and then prove the following theorem:

**Theorem 4.1** ([6], Theorem 5.). *There exists a pricing strategy **Exp3** and a constant  $c(\varepsilon)$  such that for all valuation sequences, if the optimal fixed-price revenue  $\rho(\mathcal{S}^{\text{opt}})$  satisfies  $\rho(\mathcal{S}^{\text{opt}}) > c(\varepsilon)h \log h \log \log h$ , then **Exp3** is  $(1 + \varepsilon)$ -competitive relative to  $\rho(\mathcal{S}^{\text{opt}})$ .*

Our upper and lower bounds for regret in the worst-case valuation model are based on the techniques employed in these two papers. The upper bound (Theorem 4.2) is virtually a restatement of Blum et al's theorem, though the change in emphasis from competitive ratio to additive regret necessitates a minor change in technical details. Our worst-case lower bound (Theorem 4.3) is influenced by Auer et al's proof of the corresponding lower bound for the adversarial multi-armed bandit problem in [3]. While it is possible to prove our result entirely using the techniques from their paper, we will instead present a proof using the techniques developed in Section 3.2, partly in the interest of making the paper more self-contained and partly to illustrate the power of those techniques.

### 4.1 Upper bound

Following [6], as well as the technique used in Section 3.3 above, we specify a finite set of offer prices  $X = \{1/K, 2/K, \dots, 1\}$  and constrain the seller to select prices from this set only. This reduces the posted-price auction problem to an instance of the multi-armed bandit problem, to which the algorithm **Exp3** of [3] may be applied. Denote this pricing strategy by  $\mathcal{S}$ . The relevant theorem about **Exp3** is the following.

**Theorem 4.2** ([3], Corollary 4.2.). *If one runs the algorithm **Exp3** with a set of  $K$  actions, over  $n$  steps, with the rewards for each action in each step belonging*

to  $[0, 1]$ , then the expected regret of **Exp3** relative to the best fixed action is at most  $2\sqrt{e-1}\sqrt{nK\log K}$ .

Thus, if  $\mathcal{S}_X^{\text{opt}}$  denotes the fixed-price strategy which chooses the best offer price  $i^*/K$  from  $X$ , and  $\mathcal{S}^{\text{opt}}$  denotes the fixed-price strategy which chooses the best offer price  $x^*$  from  $[0, 1]$ , we have the following inequalities:

$$\begin{aligned}\rho(\mathcal{S}_K^{\text{opt}}) - \rho(\mathcal{S}) &< 2\sqrt{e-1}\sqrt{nK\log K} \\ \rho(\mathcal{S}^*) - \rho(\mathcal{S}_K^{\text{opt}}) &< n(1/K) = n/K\end{aligned}$$

where the second inequality follows from the fact that  $\mathcal{S}_K^{\text{opt}}$  is no worse than the strategy which offers  $\frac{1}{K}\lfloor Kx^* \rfloor$  to each buyer.

If we pick  $K = \lceil n/\log n \rceil^{1/3}$ , then both  $\sqrt{nK\log K}$  and  $n/K$  are  $O(n^{2/3}(\log n)^{1/3})$ . We have thus expressed the regret of **Exp3** as a sum of two terms, each of which is  $O(n^{2/3}(\log n)^{1/3})$ , establishing the upper bound asserted in Theorem 1.3.

Readers familiar with [6] will recognize that the only difference between this argument and their Theorem 5 is that they choose the prices in  $X$  to form a geometric progress (so as to optimize the competitive ratio) while we choose them to form an arithmetic progression (so as to optimize the additive regret).

## 4.2 Lower bound

In [3], the authors present a lower bound of  $\sqrt{nK}$  for the multi-armed bandit problem with payoffs selected by an oblivious adversary. Ironically, the power of the adversary in this lower bound comes not from adapting to the on-line algorithm  $A$ , but from adapting to the number of trials  $n$ . In fact, the authors define a model of random payoffs (depending on  $n$  but not the algorithm) such that the expected regret of *any* algorithm on a random sample from this distribution is  $\Omega(\sqrt{nK})$ . The idea is select one of the  $K$  actions uniformly at random and designate it as the “good” action. For all other actions, the payoff in each round is a uniform random sample from  $\{0, 1\}$ , but for the good action the payoff is a biased sample from  $\{0, 1\}$ , which is 1 with probability  $1/2 + \varepsilon$ , where  $\varepsilon = \theta(\sqrt{K/n})$ . A strategy which knows the good action will achieve expected payoff  $(1/2 + \varepsilon)n = 1/2 + \theta(\sqrt{nK})$ . It can be shown, for information-theoretic reasons, that no strategy can learn the good action rapidly and reliably enough to play it more than  $n/K + \theta(\varepsilon\sqrt{n^3/K})$  times in expectation, from which the lower bound on regret follows.

A similar counterexample can be constructed in the context of our on-line posted-price auction problem, i.e. a random distribution of buyers’ valuations (depending on  $n$  but not the algorithm) such that the expected regret of any algorithm on a random sample from this distribution is  $\Omega(n^{2/3})$ . The idea is roughly the same as above: one randomly chooses a subinterval of  $[0, 1]$  of length  $1/K$  to be the interval of “good prices”, and chooses the distribution of buyers’ valuations so that the expected revenue per buyer is a constant independent of the offer price outside the interval of good prices, and is  $\varepsilon$  higher than this constant inside the interval of good prices. As above, there is a trade-off between



choosing  $\varepsilon$  too large (which makes it too easy for strategies to learn which prices belong to the good interval) or too small (which leads to a negligible difference in revenue between the best strategy and all others), and the optimal trade-off is achieved when  $\varepsilon = \theta(\sqrt{K/n})$ . However, in our setting there is an additional constraint that  $\varepsilon \leq 1/K$ , since the expected payoff can grow by no more than  $1/K$  on an interval of length  $1/K$ . This leads to the values  $K = \theta(n^{1/3}), \varepsilon = \theta(n^{-1/3})$  and yields the stated lower bound of  $\Omega(n^{2/3})$ .

There are two complications that come up along the way. One is that the seller’s algorithm has a continuum of alternatives at every step, rather than a finite set of  $K$  alternatives as in the example from [3]. This can be dealt with by restricting the buyers’ valuations to lie in a finite set  $V = \{v_1, v_2, \dots, v_K\}$ . Then there is no incentive for the seller to offer a price which lies outside of  $V$ , so we may assume the seller is constrained to offer prices in  $V$  and prove lower bounds for this restricted class of strategies.

The second complication that arises is that the adversary in [3] was more powerful: he could specify the reward for each action independently, whereas our adversary can only set a valuation  $v$ , and this  $v$  determines the rewards for all actions simultaneously. While this entails choosing a more complicated reward distribution, the complication only makes the computations messier without introducing any new ideas into the proof.

**Theorem 4.3.** *For any given  $n$ , there exists a finite family  $\mathcal{P} = \{p_j^n\}_{j=1}^K$  of probability distributions on  $[0, 1]$ , such that if  $p_j^n$  is chosen uniformly at random from  $\mathcal{P}$  and then buyers’ valuations are sampled independently at random according to  $p_j^n$ , no pricing strategy can achieve expected regret  $o(n^{2/3})$ , where the expectation is over both the random choice of  $D$  and the randomly-sampled valuations.*

*Proof sketch.* For simplicity, assume  $n = 8K^3$ , and put  $\varepsilon = 1/2K$ . The valuations will be independent random samples from the set  $V = \{\frac{1}{2}, 12 + \varepsilon, 12 + 2\varepsilon, \dots, 1 - \varepsilon, 1\}$ . A “baseline probability distribution”  $p_{\text{base}}$  on  $V$  is defined so that

$$p_{\text{base}}(\{v \geq 1 - i\varepsilon\}) = \frac{1}{2}(1 - i\varepsilon)^{-1}.$$

A random sample from  $p_j^n \in \mathcal{P}$  is generated by sampling  $v \in V$  at random from the distribution  $p_{\text{base}}$ , and then adding  $\varepsilon$  to it with probability  $1/10$  if and only if  $v = 1 - j\varepsilon$ .

For any random variable  $X$  depending on a sequence of samples from  $V$ , we’ll write  $\mathbf{E}_{\text{base}}(X)$ ,  $\mathbf{E}_j(X)$  to denote the expectation of  $X$  with respect to the distributions  $p_{\text{base}}, p_j^n$ , respectively. We let  $r_t$  denote the Boolean random variable which is 1 if and only if the  $t$ -th buyer accepted the price offered. As in [3], let  $\mathbf{r}^t$  denote the vector  $(r_1, \dots, r_t)$ , and let  $\mathbf{r} = \mathbf{r}^T$ . Assume that the seller’s pricing strategy  $\mathcal{S}$  only offers prices in  $V$  — since it is counterproductive to offer a price outside  $V$  when the buyers’ valuations always belong to  $V$  — and let  $N_i$  denote the random variable specifying the number of times price  $1 - i\varepsilon$  is offered.

**Lemma 4.4.** *Let  $f : \{0, 1\}^T \rightarrow [0, M]$  be any function defined on sequences  $\mathbf{r}$ . Then for any action  $i$ ,*

$$\mathbf{E}_i[f(\mathbf{r})] \leq \mathbf{E}_{\text{base}}[f(\mathbf{r})] + \frac{1}{3}\varepsilon M \sqrt{\mathbf{E}_{\text{base}}[N_i]}.$$

The proof is exactly the same as the proof of Lemma B.1 in [3], except for the estimate of the relative entropy  $\text{KL}(p_{\text{base}}\{r_i|\mathbf{r}^{t-1}\}||p_i^n\{r_i|\mathbf{r}^{t-1}\})$ . The term  $\text{KL}(\frac{1}{2}||\frac{1}{2} + \varepsilon)$  must be replaced by

$$\text{KL}\left(\frac{1}{2v}||\frac{9}{10}\left(\frac{1}{2v}\right) + \frac{1}{10}\left(\frac{1}{2(v-\varepsilon)}\right)\right)$$

for some  $v \in \{3/4, 3/4 + \varepsilon, \dots, 1\}$ . A tedious computation verifies that this is bounded above by  $0.36\varepsilon^2$ , and one finishes up as in their proof of Lemma B.1.

For a deterministic on-line pricing strategy  $\mathcal{S}$ , the random variable  $N_i$  is a function of  $\mathbf{r}$ , so we may apply the above lemma to conclude that

$$\mathbf{E}_i[N_i] \leq \mathbf{E}_{\text{base}}[N_i] + \frac{1}{3}\varepsilon n \sqrt{\mathbf{E}_{\text{base}}[N_i]},$$

hence

$$\frac{1}{K} \sum_{i=1}^K \mathbf{E}_i[N_i] = \frac{1}{K} \left\{ \sum_{i=1}^K \mathbf{E}_{\text{base}}[N_i] + \frac{1}{3}\varepsilon n \sum_{i=1}^K \sqrt{\mathbf{E}_{\text{base}}[N_i]} \right\} \leq n/K + \frac{1}{3}\varepsilon n \sqrt{n/K} \quad (7)$$

In other words, the expected number of times  $\mathcal{S}$  chooses the “good” offer price is at most  $(1/3)\varepsilon n \sqrt{n/K}$ . Recalling that  $K = \frac{1}{2}n^{1/3}$  and  $\varepsilon = 2n^{-1/3}$ , we see that  $\mathcal{S}$  makes the right choice at most  $(2\sqrt{2}/3)n$  times in expectation, so it makes the wrong choice  $\Omega(n)$  times in expectation. Each time it does so, it incurs an expected regret of  $\varepsilon/10 = \Omega(n^{-1/3})$ . Thus the total expected regret of  $\mathcal{S}$  is  $\Omega(n^{2/3})$ , as claimed. □

## 5 Acknowledgements

We would like to thank Jason Hartline, Jon Kleinberg, and Leah Brooks for helpful discussions relating to this work. We would also like to thank Denny Denker and Sean Moriarity for helpful discussions that led us to investigate these issues.

## References

- [1] R. AGRAWAL. The continuum-armed bandit problem. *SIAM J. Control and Optimization*, 33:1926-1951, 1995.

- [2] P. AUER, N. CESA-BIANCHI, AND P. FISCHER. Finite-time analysis of the multi-armed bandit problem. *Machine Learning*, 47:235-256, 2002.
- [3] P. AUER, N. CESA-BIANCHI, Y. FREUND, AND R. SCHAPIRE. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science*, 1995.
- [4] A. BAGCHI, A. CHAUDHARY, R. GARG, M. GOODRICH, AND V. KUMAR. Seller-focused algorithms for online auctioning. In *Proc. 7th International Workshop on Algorithms and Data Structures (WADS 2001)*, vol. 2125. Springer Verlag LNCS, 2001.
- [5] Z. BAR-YOSSEF, K. HILDRUM, AND F. WU. Incentive-compatible online auctions for digital goods. In *Proc. 13th Symp. on Discrete Alg.* 964-970, 2002.
- [6] A. BLUM, V. KUMAR, A. RUDRA, AND F. WU. Online learning in online auctions. In *Proc. 14th Symp. on Discrete Alg.* 202-204, 2003.
- [7] A. FIAT, A. GOLDBERG, J. HARTLINE, AND A. WRIGHT. Competitive generalized auctions. In *Proc. 34th ACM Symposium on the Theory of Computing*. ACM Press, New York, 2002.
- [8] A. GOLDBERG, J. HARTLINE, AND A. WRIGHT. Competitive auctions and digital goods. In *Proc. 12th Symp. on Discrete Alg.*, 735-744, 2001.
- [9] J. HARTLINE. Dynamic posted price mechanisms. Manuscript, 2001.
- [10] R. LAVI AND N. NISAN. Competitive analysis of incentive compatible online auctions. In *Proceedings of the 2nd ACM Conference on Electronic Commerce (EC-00)*. 233-241, 2000.
- [11] T. L. LAI AND H. ROBBINS. Asymptotically efficient adaptive allocations rules. *Adv. in Appl. Math.*, 6:4-22, 1985.
- [12] R. MYERSON. Optimal auction design. *Mathematics of Operations Research*, 6:58-73, 1981.
- [13] I. SEGAL. Optimal pricing mechanisms with unknown demand. Manuscript, 2002.

## A Proof of Lemma 3.1

In this section we restate and prove Lemma 3.1.

**Lemma A.1.** *There exist constants  $\alpha, \beta > 0$  and  $\gamma < \infty$  such that for all  $D = D_{t_0} \in D$  and  $x \in [0, 1]$ :*

1.  $\frac{d}{dt}(x_t^*)|_{t=t_0} > \alpha$ ;

2.  $x^*D(x^*) - xD(x) > \beta(x^* - x)^2$ ;
3.  $|\dot{D}(x)/D(x)| < \gamma|x^* - x|$  and  $|\dot{D}(x)/(1 - D(x))| < \gamma|x^* - x|$ ;
4.  $|D^{(k)}(x)/D(x)| < \gamma$  and  $|D^{(k)}(x)/(1 - D(x))| < \gamma$ , for  $k = 2, 3, 4$ .

Here  $x^*$  denotes  $x_{t_0}^*$ ,  $D^{(k)}(x)$  denotes the  $k$ -th  $t$ -derivative of  $D_t(x)$  at  $t = t_0$ , and  $\dot{D}(x)$  denotes  $D^{(1)}(x)$ .

*Proof.* We begin with some useful observations about the relation between  $\tilde{D}_t$  and  $D_t$ . The function  $\tilde{D}_t$  is piecewise-linear, and linear functions are preserved under convolution with an even function whose integral is 1. Recall that the bump function  $b$  is an even function supported in  $[-0.01, 0.01]$  and satisfying  $\int_{-0.01}^{0.01} b(x)dx = 1$ ; hence  $D_t(x) = \tilde{D}_t(x)$  unless  $x$  is within 0.01 of one of the two points where the derivative of  $\tilde{D}_t$  is discontinuous. The  $x$ -coordinates of these two points are given by

$$\begin{aligned} x_0 &= \frac{7t^2 - 2t}{14t^2 - 1} \\ x_1 &= \frac{7t^2 - 4t}{7t^2 - 2}. \end{aligned}$$

For  $t$  in the range  $[0.3, 0.4]$  this means that  $x_0 \in (0.115, 0.259)$ ,  $x_1 \in (0.416, 0.546)$ . Recalling that  $b$  is a  $C^\infty$  function, we find that  $t \mapsto \tilde{D}_t$  is a continuous mapping from  $[0.3, 0.4]$  to  $C^\infty([0, 1])$ . Hence  $\{\tilde{D}_t : 0.3 \leq t \leq 0.4\}$  is a compact subset of  $C^\infty([0, 1])$ , and consequently for  $1 \leq k < \infty$ , the  $k$ -th derivative of  $\tilde{D}_t$  is bounded uniformly in  $t$ .

We now proceed to prove each of the properties stated in the Lemma.

1. First we verify that  $x_t^* = t$ , as stated in Section 3.2. If  $x$  lies in the interval  $I_t = [x_0 + 0.01, x_1 - 0.01]$  where  $D_t(x) = 2/7t - x/7t^2$ , then  $x D_t(x) = 2x/7t - x^2/7t^2 = \frac{1}{7}[1 - (1 - x/t)^2]$ , which is uniquely maximized when  $x = t$  and  $x D_t(x) = 1/7$ . Note that the estimates given above for  $x_0$  and  $x_1$  ensure that  $[x_0 + 0.01, x_1 - 0.01]$  always contains  $[0.3, 0.4]$ , so  $t$  always lies in this interval. If  $x$  lies in the interval where  $D_t(x) = 1 - 2x$  or  $D_t(x) = (1 - x)/2$ , then  $x D_t(x)$  is equal to  $2(1/16 - (x - 1/4)^2)$  or  $(1/2)(1/4 - (x - 1/2)^2)$ , and in either case  $x D_t(x)$  can not exceed  $1/8$ . It is straightforward but tedious to verify that  $x D_t(x)$  is bounded away from  $1/7$  when  $|x - x_0| \leq 0.01$  or  $|x - x_1| \leq 0.01$ ; this confirms that  $x_t^* = t$  is the unique global maximum of the function  $x D_t(x)$ . Having verified this fact, it follows immediately that  $d/dt(x_t^*)|_{t=t_0} = 1$ .

2. On the interval  $I_t$  where  $D_t(x) = 2/7t - x/7t^2$ , we have

$$\begin{aligned} x D_t(x) &= 1/7 - \frac{1}{7t^2}(t - x)^2 = x_t^* D_t(x_t^*) - \frac{1}{7t^2}(x_t^* - x)^2 \\ x_t^* D_t(x_t^*) - x D_t(x) &= \frac{1}{7t^2}(x_t^* - x)^2. \end{aligned} \tag{8}$$

We have seen that for  $t \in [0.3, 0.4]$ ,  $xD_t(x)$  attains its maximum value of  $1/7$  at a point  $x_t^* \in I_t$  and is strictly less than  $1/7$  at all other points of  $[0, 1]$ . By compactness it follows that there exist  $\varepsilon, \delta > 0$  such that

$$|x - x_t^*| > \delta \implies x_t^* D_t(x_t^*) - x D_t(x) > \varepsilon, \quad (9)$$

for all  $x \in [0, 1], t \in [0.3, 0.4]$ . Combining (8), which holds when  $x \in I_t$ , with (9), which holds when  $x \notin (x_t^* - \delta, x_t^* + \delta)$ , we obtain

$$x_t^* D_t(x_t^*) - x D_t(x) \geq \min\{1/7t^2, \varepsilon/\delta^2\}(x - x_t^*)^2$$

for all  $x \in [0, 1]$ .

3. If  $x < 0.1$  or  $x > 0.6$ , then  $D_t(x)$  is independent of  $t$ , so  $\dot{D}(x) = 0$ , which establishes the desired inequality. If  $x \in [0.1, 0.6]$ , then  $D(x)$  and  $1 - D(x)$  are both bounded below by  $0.2$ , so it remains to verify that  $\sup\{|\dot{D}_t(x)/(x_t^* - x)|\} < \infty$ . The function  $|\dot{D}_t(x)|$  is a continuous function of  $t$  and  $x$ , so by compactness it is bounded above by a constant. It follows that for any constant  $\varepsilon > 0$ ,  $\sup\{|\dot{D}_t(x)/(x_t^* - x)| : \varepsilon < |x_t^* - x|\} < \infty$ . Choose  $\varepsilon$  small enough that  $[x_t^* - \varepsilon, x_t^* + \varepsilon]$  is contained in the interval  $I_t$  where  $D_t(x) = 2/7t - x/7t^2$  for all  $t \in [0.3, 0.4]$ . Then for  $|x_t^* - x| \leq \varepsilon$ ,

$$\dot{D}_t(x) = -2/7t^2 + 2x/7t^3 = 2(x - t)/7t^3 = -\frac{2}{7t^3}(x_t^* - x),$$

so  $\sup\{|\dot{D}_t(x)/(x_t^* - x)|\} < \infty$  as claimed.

4. As before, if  $x \notin [0.1, 0.6]$  then  $D^{(k)}(x) = 0$  so there is nothing to prove. If  $x \in [0.1, 0.6]$  then  $D(x)$  and  $1 - D(x)$  are both bounded below by  $0.2$ , and  $|D^{(k)}(x)|$  is uniformly bounded above, by compactness.

□

## B Proof of Lemma 3.6

In this section we restate and prove Lemma 3.6.

**Lemma B.1.** *For all sufficiently large  $n$ , if  $\mathbf{E}_D R_D < \sqrt{n}$  then there exists a set  $S$  of leaves such that  $p_D(S) \geq 1/2$ , and  $p_{D_t}(\ell) > c_4 p_D(\ell)$  for all  $\ell \in S$  and all  $t \in [t_0, t_0 + n^{-1/4}]$ .*

*Proof.* It suffices to prove that there exists a set  $S$  of leaves such that  $p_D(S) \geq 1/2$  and  $|\log(p_{D_t}(\ell)/p_D(\ell))|$  is bounded above by a constant for  $\ell \in S$ . Let  $F(t, \ell) = \log(p_{D_t}(\ell))$ . By Taylor's Theorem, we have

$$\begin{aligned} F(t, \ell) - F(t_0, \ell) &= F'(t_0, \ell)(t - t_0) + \frac{1}{2}F''(t_0, \ell)(t - t_0)^2 + \\ &\quad \frac{1}{6}F'''(t_0, \ell)(t - t_0)^3 + \frac{1}{24}F''''(t_1, \ell)(t - t_0)^4, \end{aligned}$$

for some  $t_1 \in [t_0, t]$ . (Here  $F', F'', F''', F''''$  refer to the  $t$ -derivatives of  $F$ . Throughout this section, we will adopt the same notational convention when referring to the  $t$ -derivatives of other functions, in contrast to the “dot” notation used in other sections of this paper.) This means that

$$\left| \log \left( \frac{p_{D_t}(\ell)}{p_D(\ell)} \right) \right| \leq |F'(t_0, \ell)|n^{-1/4} + \frac{1}{2}|F''(t_0, \ell)|n^{-1/2} + \frac{1}{6}|F'''(t_0, \ell)|n^{-3/4} + \frac{1}{24}|F''''(t_1, \ell)|n^{-1}. \quad (10)$$

We will prove that, when  $\ell$  is randomly sampled according to  $p_D$ , the expected value of each term on the right side of (10) is bounded above by a constant. By Markov's Inequality, it will follow that right side is bounded above by a constant for a set  $S$  of leaves satisfying  $p_D(S) \geq 1/2$ , thus finishing the proof of the Lemma.

Unfortunately, bounding the expected value of the right side of (10) requires a separate computation for each of the four terms. For the first term, we observe that  $|F'(t_0, \ell)|^2$  is precisely  $K_D(\ell)$ , so  $\mathbf{E}_D(|F'(t_0, \ell)|^2) \leq c_3\sqrt{n}$  by Corollary 3.5. It follow, using the Cauchy-Schwarz Inequality, that  $\mathbf{E}_D(|F'(t_0, \ell)|n^{-1/4}) \leq \sqrt{c_3}$ .

To bound the remaining three terms, let  $a_0, a_1, \dots, a_n = \ell$  be the nodes on the path in  $T$  from the root  $a_0$  down to the leaf  $\ell$ . Let

$$q(a_i) = \begin{cases} D(x_{a_i}) & \text{if } \chi(a_i) = 1 \\ 1 - D(x_{a_i}) & \text{if } \chi(a_i) = 0. \end{cases}$$

We have

$$p_D(\ell) = \prod_{i=0}^{n-1} q(a_i),$$

so

$$F(t_0, \ell) = \sum_{i=0}^{n-1} \log q(a_i) \quad (11)$$

$$F'(t_0, \ell) = \sum_{i=0}^{n-1} \frac{q'(a_i)}{q(a_i)} \quad (12)$$

$$F''(t_0, \ell) = \sum_{i=0}^{n-1} \frac{q''(a_i)}{q(a_i)} - \left( \frac{q'(a_i)}{q(a_i)} \right)^2 \quad (13)$$

$$F'''(t_0, \ell) = \sum_{i=0}^{n-1} \frac{q'''(a_i)}{q(a_i)} - 3 \left( \frac{q'(a_i)q''(a_i)}{q(a_i)^2} \right) + 2 \left( \frac{q'(a_i)}{q(a_i)} \right)^3. \quad (14)$$

To prove that  $\mathbf{E}_D(|F''(t_0, \ell)|) = O(\sqrt{n})$ , we use the fact that the random variable  $F''(t_0, \ell)$  is a sum of two random variables  $\sum_{i=0}^{n-1} \frac{q''(a_i)}{q(a_i)}$  and  $-\sum_{i=0}^{n-1} \left( \frac{q'(a_i)}{q(a_i)} \right)^2$ . We bound the expected absolute value of each of these two terms separately.

For the second term, we use the fact that  $|q'(a_i)/q(a_i)| = O(h_{a_i})$ , which is property 3 from Lemma 3.1. Thus

$$\mathbf{E} \left( \left| \sum_{i=0}^{n-1} \left( \frac{q'(a_i)}{q(a_i)} \right)^2 \right| \right) = O \left( \sum_{a \in T} p_D(a) h_a^2 \right),$$

and the right side is  $O(\sqrt{n})$  using Lemma 3.3 and our hypothesis that  $\mathbf{E}_D R_D \leq \sqrt{n}$ . To bound the first term,  $\sum_{i=0}^{n-1} \frac{q''(a_i)}{q(a_i)}$ , we start by observing that, conditional on the value of  $a_i$ , the random variable  $\frac{q''(a_i)}{q(a_i)}$  has mean zero and variance  $O(1)$ . The bound on the conditional variance follows from property 4 in Lemma 3.1. The mean-zero assertion follows from the computation

$$\mathbf{E}_D \left( \frac{q''(a_i)}{q(a_i)} \middle| a_i \right) = D(x_{a_i}) \left( \frac{D''(x_{a_i})}{D(x_{a_i})} \right) + (1 - D(x_{a_i})) \left( \frac{-D''(x_{a_i})}{1 - D(x_{a_i})} \right) = 0.$$

This means that the random variables  $q''(a_i)/q(a_i)$  form a martingale difference sequence, hence

$$\mathbf{E}_D \left[ \left( \sum_{i=0}^{n-1} \frac{q''(a_i)}{q(a_i)} \right)^2 \right] = \sum_{i=0}^{n-1} \mathbf{E}_D \left[ \left( \frac{q''(a_i)}{q(a_i)} \right)^2 \right] = O(n).$$

The bound  $\mathbf{E}_D \left( \left| \sum_{i=0}^{n-1} \frac{q''(a_i)}{q(a_i)} \right| \right) = O(\sqrt{n})$  follows using the Cauchy-Schwarz Inequality, as before.

We turn now to proving that  $\mathbf{E}_D(|F'''(t_0, \ell)|) = O(n^{3/4})$ . As before, the first step is to use (14) to express  $F'''(t_0, \ell)$  as a sum of three terms

$$\begin{aligned} X &= \sum_{t=0}^{n-1} \frac{q'''(a_t)}{q(a_t)} \\ Y &= -3 \sum_{t=0}^{n-1} \frac{q'(a_t)q''(a_t)}{q(a_t)} \\ Z &= 2 \sum_{t=0}^{n-1} \left( \frac{q'(a_t)}{q(a_t)} \right)^2, \end{aligned}$$

and then to bound the expected absolute value of each of these terms separately. Exactly as above, one proves that the random variables  $q'''(a_i)/q(a_i)$  form a martingale difference sequence and have bounded variance, and consequently  $\mathbf{E}_D(|X|) = O(\sqrt{n})$ . Recalling that  $|q'(a_i)/q(a_i)| = O(h_{a_i})$  and  $|q''(a_i)/q(a_i)| = O(1)$  (properties 3 and 4 from Lemma 3.1, respectively) we find that

$$\begin{aligned} \frac{1}{3} \mathbf{E}_D(|Y|) &\leq \mathbf{E}_D \left( \sum_{i=0}^{n-1} \left| \frac{q'(a_i)q''(a_i)}{q(a_i)} \right| \right) \\ &\leq \mathbf{E}_D \left( \sum_{i=0}^{n-1} \left( \frac{q'(a_i)}{q(a_i)} \right)^2 \right)^{1/2} \mathbf{E}_D \left( \sum_{i=0}^{n-1} \left( \frac{q''(a_i)}{q(a_i)} \right)^2 \right)^{1/2} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{E}_D \left( \sum_{i=0}^{n-1} O(h_{a_i}^2) \right)^{1/2} \mathbf{E}_D \left( \sum_{i=0}^{n-1} O(1) \right)^{1/2} \\
&= \mathbf{E}_D \left( \sum_{a \in T} p_D(a) h_a^2 \right)^{1/2} \cdot O(\sqrt{n}) \\
&= O(n^{3/4}),
\end{aligned}$$

where the last line follows from Lemma 3.3. Finally, we have

$$\begin{aligned}
\frac{1}{2} \mathbf{E}_D(|Z|) &\leq \mathbf{E}_D \left( \sum_{i=0}^{n-1} \left| \frac{q'(a_i)}{q(a_i)} \right|^3 \right) \\
&= \mathbf{E}_D \left( \sum_{i=0}^{n-1} O(h_{a_i}^3) \right) \\
&= \mathbf{E}_D \left( \sum_{a \in T} p_D(a) h_a^3 \right) \\
&\leq \mathbf{E}_D \left( \sum_{a \in T} p_D(a) h_a^2 \right) \\
&= O(\sqrt{n}).
\end{aligned}$$

Combining the estimates for  $\mathbf{E}_D(|X|)$ ,  $\mathbf{E}_D(|Y|)$ ,  $\mathbf{E}_D(|Z|)$ , we obtain the bound  $\mathbf{E}_D(|F'''(t_0, \ell)|) = O(n^{3/4})$  as desired.

Finally, to prove  $|F''''(t_1, \ell)| = O(n)$ , we use the formula

$$\begin{aligned}
F''''(t_1, \ell) &= \sum_{i=0}^{n-1} \frac{q''''(a_i)}{q(a_i)} - 4 \left( \frac{q'(a_i)}{q(a_i)} \right) \left( \frac{q'''(a_i)}{q(a_i)} \right) - 3 \left( \frac{q''(a_i)}{q(a_i)} \right)^2 + \\
&\quad 12 \left( \frac{q'(a_i)}{q(a_i)} \right)^2 \left( \frac{q''(a_i)}{q(a_i)} \right) - 6 \left( \frac{q'(a_i)}{q(a_i)} \right)^4. \tag{15}
\end{aligned}$$

Each of the random variables  $q^{(k)}(a_i)/q(a_i)$  for  $k = 1, 2, 3, 4$  is  $O(1)$ , hence each summand on the right side of (15) is  $O(1)$ . Summing all  $n$  terms, we obtain  $|F''''(t_1, \ell)| = O(n)$  as desired.  $\square$