Online learning in repeated auctions.

Jonathan Weed JWEED@MIT.EDU
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307, USA

Vianney Perchet VIANNEY.PERCHET@NORMALESUP.ORG
Laboratoire de Statistique
ENSAE - CREST
Malakoff, France

Philippe Rigollet RIGOLLET@MATH.MIT.EDU
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307, USA

Abstract

Motivated by online advertising auctions, we consider repeated Vickrey auctions where goods of unknown value are sold sequentially and bidders only learn (potentially noisy) information about a good's value once it is purchased. We adopt an online learning approach with bandit feedback to model this problem and derive bidding strategies for two models: stochastic and adversarial. In the stochastic model, the observed values of the goods are random variables centered around the true value of the good. In this case, logarithmic regret is achievable when competing against well behaved adversaries. In the adversarial model, the goods need not be identical. Comparing our performance against that of the best fixed bid in hindsight, we show that sublinear regret is also achievable in this case. For both the stochastic and adversarial models, we prove matching minimax lower bounds showing our strategies to be optimal up to lower-order terms. To our knowledge, this is the first complete set of strategies for bidders participating in auctions of this type.

Keywords: Second price auctions, Vickrey auctions, Repeated auctions, Bandit problems

1. Introduction

Online advertising has been a driving force behind most of the recent work on online learning, particularly in the realm of bandit problems. During the first half of 2015 alone, internet advertising generated $27.5 billion in revenue, according to the Interactive Advertising Bureau. A large fraction of advertising space is sold on platforms known as *ad exchanges* such as Google's DoubleClick and AppNexus, which facilitate transactions between the owner of advertising space and advertisers. These transactions occur within a fraction of a second using auctions (Muthukrishnan, 2009), thus placing the actors squarely within the framework of game-theoretic auctions with a single item and multiple bidders. In this context, we refer to the advertising space as the *good*, its owner as the *seller* and the advertisers as *bidders*, respectively. From the seller's perspective, this is a well understood problem in mechanism design: the Vickrey (a.k.a. second price) auction is optimal in the sense that each bidder bidding their private value constitutes an equilibrium. Because of this property, the Vickrey auction is said to be *truthful*.

The seller may also maximize her revenue while maintaining truthfulness of the auction by optimizing a *reserve price* below which no transaction occur. For example, when the bidders' values are drawn independently from known distributions, the optimal reserve price may be computed in closed form (Myerson, 1981; Riley and Samuelson, 1981). The independence assumption was questioned already by Myerson (Myerson, 1981) and it was shown later (Crémer and McLean, 1988) that when the assumption is violated, the seller can take advantage of the situation to extract more revenue at the cost of a more complicated auction mechanism. In particular, this mechanism allows bidders to be charged even if they do not win the auction, which is arguably undesirable.

In short, the Vickrey auction is a reasonable compromise between simplicity and optimality, which likely explains its prevalence on ad exchanges. Nevertheless, it suffers from a major limitation: it relies on knowledge of the bidders' value distributions, which are unlikely to be known to the seller in practice (Wilson, 1987). This limitation has driven a recent line of work on approximately optimal auctions (Roughgarden et al., 2012; Hartline and Roughgarden, 2009; Fu et al., 2013) that are robust to misspecification of these distributions. In recent years the ubiquitous collection of data has presented new opportunities, insofar as unknown quantities, such as the bidders' value distributions or relevant functionals, may potentially be learned from past observations. This new paradigm was perhaps initiated by Kleinberg and Leighton (2003) and Blum et al. (2003) and has been investigated in several recent papers (Cesa-Bianchi et al., 2013; Chawla et al., 2014; Fu et al., 2014; Ostrovsky and Schwarz, 2011; Cole and Roughgarden, 2014; Amin et al., 2015; Kanoria and Nazerzadeh, 2014; Dhangwatnotai et al., 2015; Blum et al., 2015; Mohri and Medina, 2014; Amin et al., 2014). One of the take-home messages of this literature is that a few observations are sufficient to maximize the seller's revenue in the Vickrey auction.

Like the bulk of the literature on auctions, the aforementioned work adopts the seller's perspective. It focuses on designing mechanisms to maximize the seller's revenue. In this work, we instead take the perspective of a bidder engaged in repeated Vickrey auctions. Practical questions about online advertising have motivated many proposed bidding strategies in the literature, starting with an empirical study of Kitts and Leblanc (2004). However, prior work has largely focused on the case where the bidder has a fixed budget and is trying to maximize clicks under this budget. Notable recent examples include Gummadi et al. (2011) and Balseiro et al. (2015), where optimal strategies are derived when all parameters of the problem are known. More recently, McAfee (2011) raised the question of learning unknown parameters and studying the auction problem from an online learning perspective. This idea was successfully applied to the fixed budget case mentioned above (Amin et al., 2012; Badanidiyuru et al., 2013; Tran-Thanh et al., 2014).

Our work departs from these lines of research by considering a different notion of performance based on the net (expected) revenue generated by an ad rather than the number of purchased ads. In turn, our work ignores budget constraints, but we acknowledge that combining the two lines of research could be a fruitful research direction. In this framework, we identify and analyze several strategies that can be employed by a bidder in order to maximize his reward while simultaneously learning the value of a good sold repeatedly. This paradigm can be expressed as a learning problem with partial feedback, or *bandit problem* (Bubeck and Cesa-Bianchi, 2012).

Repeated auctions have been studied in the bandit framework, primarily in the context of *truthful* bandits (Devanur and Kakade, 2009; Babaioff et al., 2010, 2009). However, this line of literature also takes the seller's point of view and aims at designing an auction mechanism rather than designing an optimal bidding strategy under the constraint of a simple mechanism such as the Vickrey auction.

More generally, the problem we describe falls into the category of *partial monitoring games*, in which the learner receives only limited feedback about the loss associated with a given action. By analyzing the feedback structure of such games, it is possible to develop essentially optimal algorithms for many games in this class (Bartók et al., 2014). However, the performance guarantees of these algorithms degrade drastically as the number of actions increases. This renders these results unusable in our context, where the bidder's number of moves at each stage is essentially unbounded.

Our contribution. We present optimal strategies for a revenue-maximizing bidder engaged in a repeated auction game, under both stochastic and adversarial assumptions. In addition, this paper makes several technical contributions to the partial monitoring literature. We present a novel analysis of a UCB-type algorithm for the stochastic setting, showing that such a strategy can work here even though the payoff and information structures are very different from those of the bandit case. In the adversarial setting, we overcome a key challenge: in sharp contrast to other partial monitoring games, where it is possible to discretize the strategy space and play the game on only a finite number of "arms," discretization *cannot* achieve sublinear regret in our setup. (See Appendix B.) Our algorithm is notable for nevertheless achieving optimal regret in this setting.

2. Sequential Vickrey auctions

We restrict our attention to bounded values and bids in the interval $[0, 1]$.

Let us first recall the mechanism of a Vickrey auction for a single good. Each bidder $k \in [K+1] := \{1, \ldots, K+1\}$ submits a written bid $b[k] \in [0, 1]$. The highest bidder $k^\star \in \operatorname{argmax}_k b[k]$ wins the auction and pays the second highest bid $m^\star = \max_{k \neq k^\star} b[k]$. In case of ties, the winner is chosen uniformly at random among the highest bidders.

Each bidder $k \in [K + 1]$ has a private but unknown individual value $v[k] \in [0, 1]$, which represents the utility of the good. Note that this value is independent of the auction itself and is only measured by the bidder once the good is delivered to him. For example, in the case of advertising space, this value may be measured by the expected profit generated from this ad or the probability that it generates a click (McAfee, 2011). The reward of the winner is given by his *net utility* $v[k^\star] - m^\star$, while the reward of a losing bidder is 0.

Perhaps the most salient feature of the Vickrey auction is that it is optimal for bidder k to be *truthful*, that is to bid $b[k] = v[k]$ (assuming that the bidder knows this value). Here optimality is understood in the equilibrium sense: any other bid $b[k] \neq v[k]$, even random, could never lead to a strict improvement in expected utility and might lead to a net loss for that bidder. An implicit crucial assumption for the implementability of this bidding strategy is that each bidder must know his own value, a hypothesis that is not necessarily met in online repeated auctions. Nevertheless, a bidder may *learn* the value $v[k]$ from past observations. Like bandit problems, this problem exhibits an exploration-exploitation tradeoff: Higher bids increase the number of observations and thus give the bidder a more accurate estimate of the value $v[k]$ (exploration) while bids closer to the best estimate of the value at time t are more likely to be optimal in the sense described above (exploitation). We will see that auctions when viewed as bandit problems possess an idiosyncratic information feedback structure: information is collected only for higher bids, but these should be avoided due to the phenomenon known as the winner's curse (Wilson, 1969).

We consider a set of $T \geq 2$ goods $t \in [T] := \{1, \ldots, T\}$ that are sold sequentially in a Vickrey auction. Using a slight abuse of terminology, we will also call the auction at which good t is sold auction t. We take the point of view of bidder 1, hereafter referred to as *the bidder*, and denote respectively by $v_t, b_t, m_t \in [0, 1]$ the unknown private value of the bidder for the t^{th} good, his bid, and the maximum bid of all other bidders for this good. Without loss of generality[1], we assume that bids are never equal. At time $t \geq 2$, the bidder is aware of the outcomes of past auctions[2] $\{(b_s, m_s),\ s \in [t-1]\}$ as well as a (potentially noisy) measurement of the values of goods $[t-1]$ *at times when the bidder won the auction*. Our goal is to construct bidding strategies that mitigate potential losses (overbidding) and opportunity cost (underbidding) for the bidder.

We consider two generating processes for the sequence of values $\{v_t\}_t$: *stochastic* and *adversarial*. The stochastic setup is the most benign one: consecutive values $\{v_t\}_t$ are independent and identically distributed (i.i.d.) random variables in the unit interval $[0, 1]$. On the other side of the stationarity spectrum is the adversarial setup, where the sequence $\{v_t\}_t$ may be any sequence in $[0, 1]$. This framework has become quite standard in the online learning literature (Cesa-Bianchi and Lugosi, 2006; Bubeck and Cesa-Bianchi, 2012) where a game-theoretic setup prevails and arbitrary dependencies between rounds occur.

3. The stochastic setup

Recall that consecutive values $\{v_t\}_t$ are independent and identically distributed (i.i.d.) random variables in the unit interval $[0, 1]$. This assumption is appropriate when the goods to be sold are identical but for small independent variations or when the values v_t represent noisy realizations of some underlying value. Let $v = \mathbb{E}[v_t]$ denote the common expected value of these random variables. We also assume that the value v_t is independent of m_t, though we allow m_t to depend on v and on v_s for $s \leq t-1$. It is easy to see that the expected net utility of the bidder at time t, $\mathbb{E}(v_t - m_t)\mathbb{1}\{b_t > m_t\}$, is maximized at $b_t \equiv v$. Therefore, a constant bid equal to v is optimal among all sequences of deterministic bids. This implies that the Vickrey auction is truthful in expectation. Since v is unknown, the bidder may not be able to achieve the best net utility over t rounds, so his performance is measured by his (cumulative) *pseudo-regret*[3] $\bar{R}_T$ defined by

$$\bar{R}_T = \max_{b \in [0,1]} \sum_{t=1}^{T} \mathbb{E}(v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T} \mathbb{E}(v_t - m_t)\mathbb{1}\{b_t > m_t\}\,, \tag{3.1}$$

where the expectations are taken with respect to the randomness in v_t and possibly in m_t, if the other bidders are playing randomly. Regret and pseudo-regret as measures of performance are studied primarily in the bandit literature but rarely in the context of auctions. One benefit of adopting regret as a measure of performance in an auction is that regret automatically takes opportunity cost into account. Indeed, a net utility of zero can be obtained trivially at any round by bidding zero, but if the other bidders tend to bid below the value of the good, the regret will still scale linearly in T.

[1] This can been achieved at an arbitrarily small cost by slightly perturbing original bids randomly.

[2] The bidder knows m_t for auctions that he won since it is the paid price, and we assume that the winning bid m_t at times when he lost is made available publicly after each auction in order to incentivize higher future bids.

[3] The benchmark in the (true) regret is the *random bid* that maximizes $b \mapsto \sum_{t=1}^{T}(v_t - m_t)\mathbb{1}\{b > m_t\}$. This quantity is more difficult to control and yields worse bounds, as detailed in Section 4.

Algorithm 1: UCBID

```
input: $b_1 = 1, \omega = 1, \bar{v} = v_1$
for $t = 2, \dots, T$ do
    Bid $b_t = \min\left(\bar{v} + \sqrt{\frac{3 \log t}{2\omega}}, 1\right)$
    Observe $m_t$
    if $b_t > m_t$ (win auction) then
        Observe $v_t$
        $\bar{v} \leftarrow (\omega\bar{v} + v_t)/(\omega + 1)\,, \quad \omega \leftarrow \omega + 1$
    end
end
```

We introduce a bidding strategy called UCBID because it is inspired by the UCB algorithm (Lai and Robbins, 1985; Auer et al., 2002) but tailored to the auction setup under investigation (see Algorithm 1). For the first auction, it prescribes to place the bid $b_1 = 1$ and thus win the auction. At auction $t+1$, $t \geq 1$, this strategy prescribes to place the bid b_{t+1} defined by

$$b_{t+1} = \min\left(\overline{v}_{\omega_t} + \sqrt{\frac{3 \log t}{2\omega_t}}, 1\right),$$

where ω_t is the number of auctions won up to stage t and $\overline{v}_{\omega_t} = \sum_{s=1}^{\omega_t} v_{\tau_s}/\omega_t$ with τ_s being the stage of the s^{th} won auction. Unlike the UCB strategy, which computes such estimates for each possible action, the UCBID strategy uses the specific feedback structure of the auction problem to handle an infinite number of actions simultaneously.

Interestingly, the UCBID strategy does not require the knowledge of past bids of other bidders $\{m_1, \dots, m_{t-1}\}$. This feature is particularly attractive in the setup of ad exchanges, where the process takes place so fast that it may be useful for the platform to not communicate the cost of an auction to bidders until the end of the day, for example.

While the implementation of the UCBID strategy does not require the knowledge of $\{m_t\}_t$, its performance is affected by other bids that are larger but close to the optimal bid v. This is not surprising as such bids force the bidder to overpay in order to collect information about the unknown v. However, sub-linear regret of order $\sqrt{T}$ is achievable regardless of the sequence $\{m_t\}_t$. We prove two results that show that this strategy automatically *adapts* to more favorable sequences $\{m_t\}_t$. Both proofs are deferred to Appendix A.

3.1. Pseudo-regret bounds

Theorem 1 *Consider the stochastic setup where the values $v_1, \dots, v_T \in [0,1]$ are independent such that $\mathbb{E}[v_i] = v$. For any sequence $m_1, \dots, m_T \in [0,1]$ such that m_t is independent of v_t, the* UCBID *strategy yields pseudo-regret bounded as follows:*

$$\bar{R}_T \leq 3 + \frac{12 \log T}{\Delta} \wedge 2\sqrt{6T \log T}\,,$$

where $x \wedge y = \min(x,y)$ and $\Delta \in [0,1]$ is the largest number such that no bid m_t is the interval $(v, v+\Delta)$.

We note that the dependence on the parameter Δ in Theorem 1 cannot be improved. This follows from a reduction to a bandit problem: if $m_t = 1/2$ for all values of m_t, then the auction problem is equivalent to a two-armed stochastic bandit game, where the arms have expectation 0 and $-\Delta = v - 1/2$. (Bidding below $1/2$ corresponds to pulling the first arm, and bidding above corresponds to pulling the second.) It is known (Bubeck et al., 2013) that for any policy, there is a choice of Δ and sequence of outcomes such that the regret is of order $\log(T)/\Delta$.

Theorem 1 shows an interesting phenomenon: While UCB type strategies are usually very sensitive to the assumption that the rewards are stochastic, this strategy is actually robust to *any* sequence $\{m_t\}_t$ that may be generated by other bidders, including malicious ones, as long as m_t is independent of the stochastic value v_t at each time step t. Indeed, in this hybrid setup, where the v_t's are random but the m_t's may not be, the UCBID strategy exhibits a sublinear regret that can even be logarithmic in the favorable case where no bid m_t is the interval $(v, v + \Delta)$ for some $\Delta > 0$. It turns out that this condition can be softened and can be well captured by a simple margin condition under the assumption that the m_t's are also stochastic.

3.2. Margin condition

Assume in the rest of this section that $m_1, \ldots, m_T \overset{\text{iid}}{\sim} \mu$ for some unknown probability measure μ. Borrowing terminology from binary classification (Mammen and Tsybakov, 1999; Tsybakov, 2006), we define the *margin condition* as follows.

Definition 2 *A probability measure μ on $[0, 1]$ satisfies the margin condition with parameter $\alpha > 0$ around $v \in (0, 1)$ if there exists a constant $C_\mu > 0$ such that*

$$\mu\{(v, v + u]\} \leq C_\mu u^\alpha \quad \forall\, u > 0\,.$$

The parameter α is an indication of the difficulty of the problem—the larger the α, the easier the problem. Under the margin condition, we can interpolate between between the two bounds for the regret—$O(\log T)$ and $O(\sqrt{T \log T})$—that arise in Theorem 1.

Theorem 3 *Fix $T \geq 2$ and consider the stochastic setup where the values $v_1, \ldots, v_T \in [0, 1]$ are independent such that $\mathbb{E}[v_i] = v$. For any random sequence $m_1, \ldots, m_T \overset{\text{iid}}{\sim} \mu$, where μ on $[0, 1]$ satisfies the margin condition with parameter $\alpha > 0$ around $v \in (0, 1)$, the* UCBID *strategy yields pseudo-regret bounded as follows:*

$$\bar{R}_T \leq \begin{cases} c_1 T^{\frac{1-\alpha}{2}} \log^{\frac{1+\alpha}{2}}(T) & \text{if } \alpha < 1 \\ c_2 \log^2(T) & \text{if } \alpha = 1 \\ c_3 \log(T) & \text{if } \alpha > 1 \end{cases}$$

where c_1, c_2 and c_3 are positive constants that depend on α, v and C_μ.

As we can see from Theorem 3, the margin parameter α allows us to interpolate between $O(\log T)$ and $O(\sqrt{T})$ regret bounds. Since UCBID does not require the knowledge of α, we say that it is *adaptive to the margin parameter* α.

In fact, the above result holds, with the exact same proof, under a weaker assumption. Denote by μ_t the law of m_t conditional on the past history $\{b_s, v_s, m_s\}_{s \leq t-1}$. Then the conclusions of Theorem 6 remain true if all μ_t satisfy the margin condition with respect to the same parameters α and C_μ.

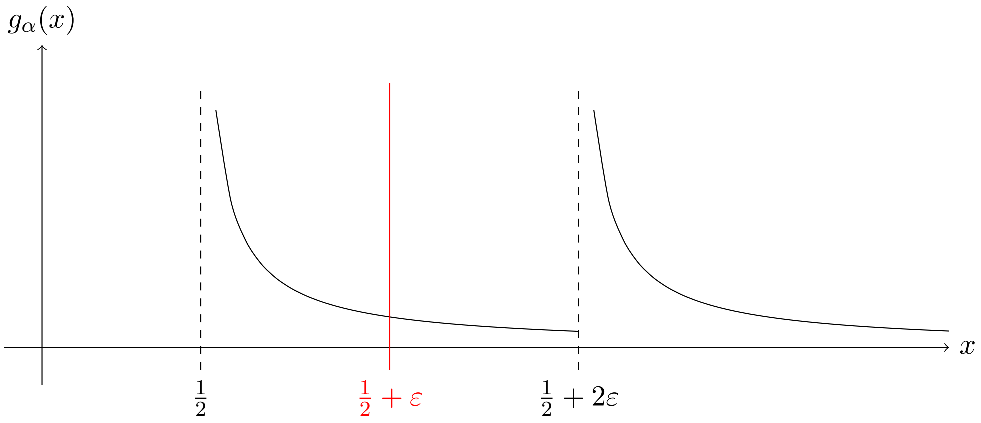

Figure 1: Representation of the density $g_{.95}$ of bids m_t.

3.3. Lower bound

We now show that the family of rates obtained in Theorem 3—indexed by α—is optimal up to logarithmic terms. As we shall see, the upper bound is tight already in the case where the bids $\{m_t\}_t$ are i.i.d., independent of $\{v_s\}_{s\le t-1}$. The proof that the construction below gives the claimed lower bound appears in Appendix B

We first consider the case where $\alpha \in (0,1)$. For any α in this interval, let μ_α denote the distribution on $[0,1]$ with density g_α with respect to the Lebesgue measure, where g_α is defined by

$$g_\alpha(x) = C_\alpha\Big[\big(x-\frac{1}{2}\big)^{\alpha-1}\mathbb{1}\big\{x\in(1/2,1/2+2\varepsilon]\big\} + \big(x-\frac{1}{2}-2\varepsilon\big)^{\alpha-1}\mathbb{1}\big\{x\in(1/2+2\varepsilon,1]\big\}\Big]\,,$$

where C_α is an appropriate normalizing constant. See Figure 1 for a representation of this density. Observe that μ_α satisfies the margin condition with parameter $\alpha>0$ around both $1/2$ and $1/2+2\varepsilon$.

For $\alpha \ge 1$, define the distribution μ_α to be the point mass at $1/2+\varepsilon$. This distribution also satisfies the margin condition with parameter α around both values.

Let ν denote the joint distribution of (v_t, m_t) and denote by $\bar R_T(\nu)$ the pseudo-regret associated to a strategy when the expectation in (3.1) is taken with respect to ν.

Theorem 4 *Fix $\alpha>0$. Let $\nu = \mathsf{Bern}(1/2)\otimes\mu_\alpha$ and $\nu' = \mathsf{Bern}(1/2+2\varepsilon)\otimes\mu_\alpha$, where $\varepsilon = \frac{1}{2}T^{-1/2}$ if $\alpha<1$ and $\varepsilon=\frac{1}{8}$ if $\alpha\ge 1$. Then, for any strategy, it holds*

$$\bar R_T(\nu)\vee \bar R_T(\nu') \ge \begin{cases} C_\alpha T^{\frac{1-\alpha}{2}} & \text{if } \alpha<1\\ C_\alpha \log T & \text{if } \alpha \ge 1\end{cases}$$

4. The adversarial setup

In this section, unlike the stochastic case, we make no assumptions on the sequences $\{v_t\}_t$ and $\{m_t\}_t$, even allowing the seller and other bidders to coordinate their plays according to a non-stationary process. As in the stochastic case, we compare the performance of a sequence $\{b_t\}_t$ of bids generated by a data-driven strategy to the *best fixed bid* in hindsight. As a consequence the (cumulative) *regret* R_T of the bidder for not knowing his own sequence of values is defined as

$$R_T = \max_{b\in[0,1]} \sum_{t=1}^{T} (v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T} (v_t - m_t)\mathbb{1}\{b_t > m_t\}\,. \tag{4.2}$$

As in the stochastic case, we will also consider the pseudo-regret $\bar{R}_T$, defined in (3.1), which is easier to handle and will serve as an illustration of the techniques used in our proofs.

Clearly, $\bar{R}_T \le \mathbb{E}[R_T]$ and it is well known that $\bar{R}_T = \mathbb{E}[R_T]$ when the adversary is oblivious (Cesa-Bianchi and Lugosi, 2006; Bubeck and Cesa-Bianchi, 2012), that is, when it generates its sequence of moves independently of the past actions of the bidder. In the sequel, we study both oblivious and non-oblivious (a.k.a. adaptive) adversaries.

For notational convenience, assume hereafter that $m_t \in (0,1]$ and that $v_t \in [0,1]$. Precluding $m_t = 0$ has no effect on the problem if we replace $m_t = 0$ by an arbitrarily small value.

4.1. Oblivious adversaries

Algorithm 2: EXPTREE

```
input: $\eta \in (0,1)$, $\mathcal{L} = \{(0,1]\}$, $w_{(0,1]} = 1$, $p_{(0,1]} = 1\,.$
for $t = 1, \dots, T$ do
    Select $\ell \in \mathcal{L}$ with probability $p_\ell$ and $b \sim \mathsf{Unif}(\ell)$
    Bid $b_t = b$
    Observe $m_t \in \bar{\ell} = (x, y] \in \mathcal{L}$
    $\bar{\ell}_l \leftarrow (x, m_t]$, $\bar{\ell}_r \leftarrow (m_t, y]$
    $\mathcal{L} \leftarrow (\mathcal{L} \setminus \bar{\ell}) \cup \bar{\ell}_l \cup \bar{\ell}_r$
    $\omega_{\bar{\ell}_l} \leftarrow \omega_{\bar{\ell}}, \omega_{\bar{\ell}_r} \leftarrow \omega_{\bar{\ell}}$
    for $\ell \in \mathcal{L}$ do
        $\hat{g}(\ell) \leftarrow \big(1 - \frac{1-(v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\big) \mathbb{1}\{m_t \preceq \ell\}$
        $\omega_\ell \leftarrow \omega_\ell \exp(\eta \hat{g}(\ell))$
        $p_\ell \leftarrow \frac{|\ell| \omega_\ell}{\sum_{\kappa \in \mathcal{L}} |\kappa| \omega_\kappa}$
    end
end
```

One popular strategy for adversarial partial-information problems of this kind is the celebrated EXP3 algorithm (Auer et al., 2002/03). However, EXP3 and similar approaches are tailored to problems with a *fixed* number of actions. In the auction setup, by contrast, the number of actions is *a priori* unbounded, and even the number of actions up to equivalence grows with T.

Moreover, since the payoff associated with a bid is sharply discontinuous, discretizing the interval and playing a standard bandit strategy on a finite set of arms fails to achieve sublinear regret, even when the discretization is exponentially fine. (A proof of this fact appears in Appendix B.)

Standard tools are therefore unusable in this regime. In Algorithm 2, we present a novel strategy for bandit games of this type that allows the number of actions to grow over time.

The algorithm maintains a sequence of nested partitions $\mathcal{L}_t, t \geq 1$ of $(0,1]$ into t intervals of the form $(x,y]$ for $0 \leq x < y \leq 1$. We set $\mathcal{L}_1 = \{(0,1]\}$ and the refinement of the partition $\mathcal{L}_t$ is done as follows. Let $\bar{\ell} = (x,y] \in \mathcal{L}_t$ be the unique interval in $\mathcal{L}_t$ such that $m_t \in \bar{\ell}$. Then $\bar{\ell}$ is *split* into two subintervals $\bar{\ell}_l = (x, m_t]$ and $\bar{\ell}_r = (m_t, y]$: $\mathcal{L}_{t+1} = (\mathcal{L}_t \setminus \bar{\ell}) \cup \bar{\ell}_l \cup \bar{\ell}_r$. This procedure is illustrated in Figure 2.

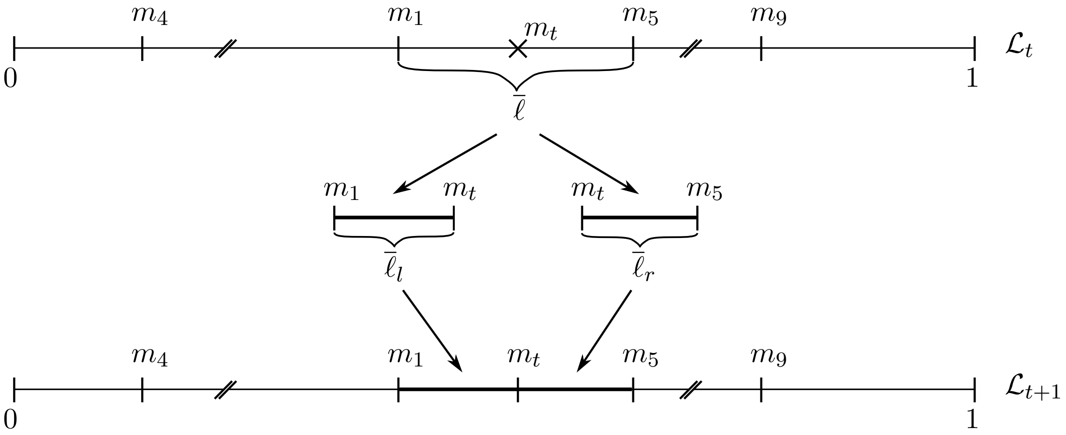

Figure 2: Illustration of the splitting procedure for constructing $\mathcal{L}_{t+1}$ from $\mathcal{L}_t$

Each element $\ell \in \mathcal{L}_t$ is assigned a probability $p_{\ell,t}$ defined in (4.5) below and such that $p_{\ell,t} > 0$ and $\sum_{\ell \in \mathcal{L}_t} p_{\ell,t} = 1$. At round t, the EXPTREE strategy prescribes to bid randomly as follows. First draw $\ell \in \mathcal{L}_t$ with probability $p_{\ell,t}$ and then draw a bid $b_t \sim \mathsf{Unif}(\ell)$ uniformly over the interval ℓ. We denote the resulting distribution of b_t by $\mathcal{B}_t$ and by $\mathbb{P}_{\mathcal{B}_t}$ the associated probability. Note that $\mathcal{B}_t$ is a mixture of uniform distributions that can be computed explicitly given $p_{\ell,t}, \ell \in \mathcal{L}_t$:

$$\mathbb{P}_{\mathcal{B}_t}(A) = \sum_{\ell \in \mathcal{L}_t} p_{\ell,t} |A \cap \ell| \,, \qquad \forall A \subseteq (0,1) \text{ measurable,} \tag{4.3}$$

where here and in what follows, $|A|$ denotes the Lebesgue measure of $A \subset [0,1]$. It remains only to specify the distribution $p_{\ell,t}, \ell \in \mathcal{L}_t$. Intuitively, we hope to construct this distribution based on the intervals' past performance, but since the player only observes the value v_t when $b_t > m_t$, we cannot evaluate the gain $g(b,t)$ of an arbitrary bid b at round t. Instead, we compute an unbiased estimate $\hat{g}(b,t)$ of $g(b,t)$ by

$$\hat{g}(b,t) = \Big(1 - \frac{1-(v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\Big) \mathbb{1}\{b > m_t\} \,.$$

It is not hard to check that $\mathbb{E}_{b_t \sim \mathcal{B}_t}[\hat{g}(b,t)] = g(b,t)$. Moreover, this estimate is constant on each interval $\ell \in \mathcal{L}_{t+1}$ and depends only on whether $m_t \preceq \ell$ (i.e., $m_t \leq x$ for all $x \in \ell$) or $m_t \succeq \ell$. As a result, overloading the notation, we define the following estimate for the gain of a bid in the interval ℓ:

$$\hat{g}(\ell,t) = \Big(1 - \frac{1-(v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\Big) \mathbb{1}\{m_t \preceq \ell\} \,. \tag{4.4}$$

With this estimate, we can compute $p_{\ell,t+1}, \ell \in \mathcal{L}_{t+1}$ using exponential weights:

$$p_{\ell,t+1} = \frac{|\ell| w_{\ell,t+1}}{\sum_{\kappa \in \mathcal{L}_{t+1}} |\kappa| w_{\kappa,t+1}}, \quad w_{\ell,t+1} = \exp\left(\eta \sum_{s=1}^{t} \hat{g}(\ell, s)\right), \ell \in \mathcal{L}_{t+1} \tag{4.5}$$

for some tuning parameter $\eta > 0$ to be chosen carefully. The reweighing by the length $|\ell|$ of the interval ℓ in (4.5) is one the main novelties of our algorithm.

The performance of the EXPTREE algorithm depends to some extent on the behavior of the adversary, in the following way. After T auctions, the best fixed bid in hindsight will lie in some interval $\ell^\circ \in \mathcal{L}_T$ of width $\Delta^\circ = |\ell^\circ|$. (If more than one interval contains an optimal bid, ℓ° can denote the widest such interval.) We obtain the following theorem.

Theorem 5 *Let $v_1, \ldots, v_T \in [0,1]$ and $m_1, \ldots, m_T \in [0,1]$ be arbitrary sequences, and let Δ° be defined as above. The strategy* EXPTREE *run with parameter $\eta = \sqrt{\log(1/\Delta^\circ)/4T} \wedge 1$ achieves the pseudo-regret bound*

$$\bar{R}_T \leq 4\sqrt{T \log(1/\Delta^\circ)}. \tag{4.6}$$

The proof appears in Appendix A.

Note that choosing a value of η appears to require knowledge of Δ° and T in advance. However, the so-called "generic doubling trick" allows the bidder to learn these values adaptively at the price of a constant factor (Hazan and Kale, 2010). (This change also requires replacing the width of the interval containing ℓ° by the width of the narrowest interval in $\mathcal{L}_T$.) Since this technique is standard, we relegate the details to Appendix A.

It is tempting to assert that a $O(\sqrt{T \log(1/\Delta^\circ)})$ rate of convergence could also be achieved by constructing a Δ°-discretization of the interval and optimizing over this finite set of actions with a standard bandit algorithm. This line of reasoning is incorrect for two reasons. First, the logarithmic dependence on the number of actions $(1/\Delta^\circ)$ is generally only achievable in full-information settings and not in those with partial feedback. Moreover, even with full information, a fixed discretization is doomed to incur linear regret, as shown in Appendix B.2.1.

4.2. Adaptive adversaries—Regret bound in high probability

Theorem 5 establishes an upper bound on the pseudo-regret against any adversary. Moreover, when the adversary is oblivious, the same bound holds for the expected regret. When the adversary is adaptive, however, achieving a bound on the expected regret requires a slightly modified algorithm, Algorithm 3. Actually, this algorithm achieves regret bound not only in expectation but also with high probability. We henceforth consider a shifted version of the auction described above where the reward associated to bid b at time t is given by

$$g(b,t) = (v_t - m_t)\mathbb{1}\{b > m_t\} + m_t\,.$$

Shifting the reward of the game in this way does not affect the regret, but it has the convenient effect that the bidder's net utility at each round is positive.

Algorithm 3 differs from Algorithm 2 chiefly in the method of calculating the estimated gain in (4.4). In place of $\hat{g}(\ell, t)$, EXPTREE.P employs a *biased* estimate $\tilde{g}(\ell, t)$ defined by

$$\tilde{g}(\ell, t) = \frac{v_t \mathbb{1}\{b_t > m_t\} + \beta}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \preceq \ell\} + \frac{m_t \mathbb{1}\{b_t \leq m_t\} + \beta}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \succeq \ell\}\,. \tag{4.7}$$

Algorithm 3: EXPTREE.P

```
input: $\eta \in (0, 1/8)$, $\gamma \in (0, 1/4)$, $\beta \in (0, 1)$, $\mathcal{L} = \{(0, 1]\}$, $w_{(0,1]} = 1$, $p_{(0,1]} = 1$.
for $t = 1, \dots, T$ do
    Select $\ell \in \mathcal{L}$ with probability $p_\ell$ and $b \sim \mathsf{Unif}(\ell)$
    Bid $b_t = \begin{cases} 1 & \text{with probability } \gamma \\ 0 & \text{with probability } \gamma \\ b & \text{with probability } 1 - 2\gamma \end{cases}$
    Observe $m_t \in \bar{\ell} = (x, y]$
    $\bar{\ell}_l = (x, m_t]$, $\bar{\ell}_r = (m_t, y]$
    $w_{\bar{\ell}_l} \leftarrow w_{\bar{\ell}}$, $w_{\bar{\ell}_r} \leftarrow w_{\bar{\ell}}$
    $\mathcal{L} \leftarrow (\mathcal{L} \setminus \bar{\ell}) \cup \bar{\ell}_l \cup \bar{\ell}_r$
    for $\ell \in \mathcal{L}$ do
        if $b_t > m_t$ then
            Observe $v_t$
            $\tilde{g}(\ell) \leftarrow \frac{v_t + \beta}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \preceq \ell\} + \frac{\beta}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \succeq \ell\}$
        else
            $\tilde{g}(\ell) \leftarrow \frac{\beta}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \preceq \ell\} + \frac{m_t + \beta}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{m_t \succeq \ell\}$
        end
        $w_\ell \leftarrow w_\ell \exp(\eta \tilde{g}(\ell))$,    $p_\ell \leftarrow \frac{|\ell| w_{\ell,t}}{\sum_{\ell \in \mathcal{L}} |\ell| w_\ell}$
    end
end
```

The following theorem holds.

Theorem 6 *Let $v_1, \dots, v_T \in [0, 1]$ and $m_1, \dots, m_T \in [0, 1]$ be arbitrary sequences. Let $\ell^\circ \in \mathcal{L}_T$ denote the narrowest interval in the finest partition $\mathcal{L}_T$ and let $\Delta^\circ = |\ell^\circ|$ denote its width. The strategy* EXPTREE.P *run with parameters*

$$\eta = \sqrt{\frac{\log(1/\Delta^\circ)}{8T}} \wedge \frac{1}{8}, \quad \gamma = 2\eta, \quad \textit{and } \beta = \sqrt{\frac{\log T}{2T}}$$

yields

$$R_T \leq 2\sqrt{8T \log(1/\Delta^\circ)} + 3\sqrt{2T \log T} \log(1/\delta)\,,$$

with probability at least $1 - \delta$. Moreover,

$$\mathbb{E}[R_T] \leq 2\sqrt{8T \log(1/\Delta^\circ)} + 3\sqrt{2T \log T}\,.$$

The proof of Theorem 5 appears in Appendix A.

4.3. Lower bound

The dependence on Δ° in Theorems 5 and 6 is unfortunate, since the resulting bounds become vacuous when Δ° is exponentially small. However, it turns out that this dependence is unavoidable.

We prove in this section a lower bound on the pseudo-regret $\bar{R}_T$. Since $\bar{R}_T \leq \mathbb{E}R_T$, this bound will also hold for expected regret.

We begin with a lemma establishing that the rate $\sqrt{T}$ is optimal, using standard information theoretic techniques for lower bounds (see, e.g., (Tsybakov, 2009)). The proof is standard and is deferred to Appendix B.

Lemma 7 *Fix an $m \in [1/4, 3/4]$. There exist a pair of adversaries U and L such that $m_t = m$ for all t and the sequence $v_1, \dots, v_T$ is i.i.d. conditional on the choice of adversary and such that*

$$\max_{A \in \{U,L\}} \max_{b \in [0,1]} \mathbb{E}_A \Big[\sum_{t=1}^T g(b,t) - \sum_{t=1}^T g(b_t,t) \Big] \geq \frac{1}{32}\sqrt{T}.$$

Moreover, under adversary U any bet $b > m$ is optimal, and under adversary L any bet $b < m$ is optimal.

We are now in a position to prove a tight minimax lower bound.

Theorem 8 *For any strategy and any value of $\Delta^\circ \in (0, 1/4)$, there exists sequences $v_1, \dots, v_T \in [0,1]$ and $m_1, \dots, m_T \in [0,1]$ such that Δ° is the smallest positive gap between the adversary's bids and*

$$\max_{b \in [0,1]} \mathbb{E} \sum_{t=1}^T g(b,t) - \mathbb{E} \sum_{t=1}^T g(b_t,t) \geq \frac{1}{32}\sqrt{T \lfloor \log_2(1/2\Delta^\circ) \rfloor}.$$

The proof appears in Appendix B.

5. Conclusion and open questions

Building on established strategies for the bandit problem, we propose a first set of strategies tailored to online learning in repeated auctions. Depending on the model, stochastic or adversarial, we obtain several regret bounds ranging from $O(\log T)$ to $O(\sqrt{T})$ and exhibit a reasonable family of models where regret bounds $\tilde{O}(T^{\beta/2})$ are achievable for all $\beta \in (0,1)$.

In both setups, several questions are beyond the scope of this paper and are left open.

1. As illustrated by the bulk of the recent research on online auctions, budget constraints are inherent to how online auctions are managed (see Tran-Thanh et al., 2014, and references therein). Devising strategies under such constraints that lead to good bounds on the pseudo-regret (3.1) is perhaps the most intriguing line of research following this work.

2. What is the effect of covariates on this problem? In practice, potentially relevant information about the value of the good is available before bidding (Mohri and Medina, 2014) and incorporating such covariates can allow for a better model. This question falls into the realm of *contextual bandits* that has been studied both in the stochastic and the adversarial framework (Wang et al., 2003; Kakade et al., 2008; Bubeck and Cesa-Bianchi, 2012; Perchet and Rigollet, 2013; Slivkins, 2014).

3. In the adversarial case, our benchmark is the best fixed bid in hindsight. While this is rather standard in the online learning literature, recent developments have allowed for more complicated benchmarks, namely sophisticated but fixed strategies (Han et al., 2013). Such developments are available only for the full information case, however.

4. Our results indicate that when facing well behaved bidders, better regret bounds are achievable in the stochastic case. Similar results are of interest in the adversarial case too (Hazan and Kale, 2010; Rakhlin and Sridharan, 2013; Foster et al., 2015). Here too, unfortunately, existing results are limited to the full information case.

5. The proof of Theorem 6 involves a union bound which leads to a $O(\sqrt{T\log(T)}\log(1/\delta))$ regret upper bound. The result is a gap of order $\sqrt{\log(T)}$ between the upper and lower bound. Is this term really present?

Acknowledgments

Jonathan Weed is supported in part by NSF Graduate Research Fellowship DGE-1122374 and NSF grants DMS-1317308 and CAREER-DMS-1053987. Vianney Perchet is supported in part by ANR grant ANR-13-JS01-0004-01, PGMO grant Nougat, and CNRS grant PEPS Parasol. Philippe Rigollet is supported in part by NSF grants DMS-1317308 and CAREER-DMS-1053987. The authors wish to thank the anonymous reviewers for their helpful suggestions.

References

Kareem Amin, Michael Kearns, Peter Key, and Anton Schwaighofer. Budget optimization for sponsored search: Censored learning in MDPs. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, August 14-18, 2012*, pages 54–63, 2012.

Kareem Amin, Afshin Rostamizadeh, and Umar Syed. Repeated contextual auctions with strategic buyers. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 622–630. Curran Associates, Inc., 2014. URL http://papers.nips.cc/paper/5589-repeated-contextual-auctions-with-strategic-buyers.pdf.

Kareem Amin, Rachel Cummings, Lili Dworkin, Michael Kearns, and Aaron Roth. Online learning and profit maximization from revealed preferences. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 770–776, 2015.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, 2002.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77 (electronic), 2002/03. ISSN 0097-5397. doi: 10.1137/S0097539701398375. URL http://dx.doi.org/10.1137/S0097539701398375.

Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms: Extended abstract. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 79–88, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-458-4. doi: 10.1145/1566374.1566386. URL http://doi.acm.org/10.1145/1566374.1566386.

Moshe Babaioff, Robert D. Kleinberg, and Aleksandrs Slivkins. Truthful mechanisms with implicit payment computation. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pages 43–52, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-822-3. doi: 10.1145/1807342.1807349. URL http://doi.acm.org/10.1145/1807342.1807349.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. In *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, FOCS '13, pages 207–216, Washington, DC, USA, 2013. IEEE Computer Society. ISBN 978-0-7695-5135-7. doi: 10.1109/FOCS.2013.30. URL http://dx.doi.org/10.1109/FOCS.2013.30.

Santiago R. Balseiro, Omar Besbes, and Gabriel Y. Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Manage. Sci.*, 61(4):864–884, April 2015. ISSN 0025-1909. doi: 10.1287/mnsc.2014.2022. URL http://dx.doi.org/10.1287/mnsc.2014.2022.

Gábor Bartók, Dean P. Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. Partial monitoring—classification, regret bounds, and algorithms. *Math. Oper. Res.*, 39(4):967–997, 2014. ISSN 0364-765X. doi: 10.1287/moor.2014.0663. URL http://dx.doi.org/10.1287/moor.2014.0663.

Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 202–204, 2003.

Avrim Blum, Yishay Mansour, and Jamie Morgenstern. Learning valuation distributions from partial observation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 798–804, 2015.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet. Bounded regret in stochastic multi-armed bandits. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *COLT 2013 - The 26th Conference on Learning Theory, Princeton, NJ, June 12-14, 2013*, volume 30 of *JMLR W&CP*, pages 122–134, 2013.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-84108-5; 0-521-84108-9. doi: 10.1017/CBO9780511546921. URL http://dx.doi.org/10.1017/CBO9780511546921.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Regret minimization for reserve prices in second-price auctions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13, pages 1190–1204. SIAM, 2013.

Shuchi Chawla, Jason Hartline, and Denis Nekipelov. Mechanism design for data science. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 711–712, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2565-3. doi: 10.1145/2600057.2602881. URL http://doi.acm.org/10.1145/2600057.2602881.

Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 243–252, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2710-7. doi: 10.1145/2591796.2591867. URL http://doi.acm.org/10.1145/2591796.2591867.

Jacques Crémer and Richard P. McLean. Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, 56(6):pp. 1247–1257, 1988. ISSN 00129682. URL http://www.jstor.org/stable/1913096.

Nikhil R. Devanur and Sham M. Kakade. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 99–106, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-458-4. doi: 10.1145/1566374.1566388. URL http://doi.acm.org/10.1145/1566374.1566388.

Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. *Games Econom. Behav.*, 91:318–333, 2015. ISSN 0899-8256. doi: 10.1016/j.geb.2014.03.011. URL http://dx.doi.org/10.1016/j.geb.2014.03.011.

Dylan Foster, Alexander Rakhlin, and Karthik Sridharan. Adaptive online learning. In *NIPS*, 2015.

Hu Fu, Jason Hartline, and Darrell Hoy. Prior-independent auctions for risk-averse agents. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 471–488, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1962-1. doi: 10.1145/2482540.2482551. URL http://doi.acm.org/10.1145/2482540.2482551.

Hu Fu, Nima Haghpanah, Jason Hartline, and Robert Kleinberg. Optimal auctions for correlated buyers with sampling. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 23–36, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2565-3. doi: 10.1145/2600057.2602895. URL http://doi.acm.org/10.1145/2600057.2602895.

Ramki Gummadi, Peter Key, and Alexandre Proutiere. Repeated auctions with budget constraints: optimal bidding strategies and equilibria. In *Proc. of Informs Applied Probability conf.*, 2011.

Wei Han, Alexander Rakhlin, and Karthik Sridharan. Competing with strategies. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *COLT 2013 - The 26th Conference on Learning Theory, Princeton, NJ, June 12-14, 2013*, volume 30 of *JMLR W&CP*, pages 966–992, 2013.

Jason D. Hartline and Tim Roughgarden. Simple versus optimal mechanisms. *SIGecom Exch.*, 8 (1):5:1–5:3, July 2009. ISSN 1551-9031. doi: 10.1145/1598780.1598785. URL http://doi.acm.org/10.1145/1598780.1598785.

Elad Hazan and Satyen Kale. Extracting certainty from uncertainty: regret bounded by variation in costs. *Mach. Learn.*, 80(2-3):165–188, 2010. ISSN 0885-6125. doi: 10.1007/s10994-010-5175-x. URL http://dx.doi.org/10.1007/s10994-010-5175-x.

Sham M. Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Efficient bandit algorithms for online multiclass prediction. In William W. Cohen, Andrew McCallum, and Sam T. Roweis, editors, *ICML*, volume 307 of *ACM International Conference Proceeding Series*, pages 440–447. ACM, 2008. ISBN 978-1-60558-205-4.

Yash Kanoria and Hamid Nazerzadeh. Dynamic reserve prices for repeated auctions: Learning from bids. In Tie-Yan Liu, Qi Qi, and Yinyu Ye, editors, *Web and Internet Economics*, volume 8877 of *Lecture Notes in Computer Science*, pages 232–232. Springer International Publishing, 2014. ISBN 978-3-319-13128-3. doi: 10.1007/978-3-319-13129-0_17. URL http://dx.doi.org/10.1007/978-3-319-13129-0_17.

Brendan Kitts and Benjamin Leblanc. Optimal bidding on keyword auctions. *Electronic Markets*, 14(3):186–201, 2004. doi: 10.1080/1019678042000245119. URL http://www.tandfonline.com/doi/abs/10.1080/1019678042000245119.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations* *of Computer Science*, FOCS '03, pages 594–, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-2040-5. URL http://dl.acm.org/citation.cfm?id=946243.946352.

T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

E. Mammen and A. B. Tsybakov. Smooth discrimination analysis. *Ann. Statist.*, 27(6):1808–1829, 1999. ISSN 0090-5364.

R.Preston McAfee. The design of advertising exchanges. *Review of Industrial Organization*, 39(3): 169–185, 2011. ISSN 0889-938X. doi: 10.1007/s11151-011-9300-1. URL http://dx.doi.org/10.1007/s11151-011-9300-1.

Mehryar Mohri and Andres Muñoz Medina. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 262–270, 2014.

S. Muthukrishnan. Ad exchanges: Research issues. In Stefano Leonardi, editor, *Internet and Network Economics*, volume 5929 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-10840-2. doi: 10.1007/978-3-642-10841-9_1. URL http://dx.doi.org/10.1007/978-3-642-10841-9_1.

Roger B. Myerson. Optimal auction design. *Math. Oper. Res.*, 6(1):58–73, 1981. ISSN 0364-765X. doi: 10.1287/moor.6.1.58. URL http://dx.doi.org/10.1287/moor.6.1.58.

Michael Ostrovsky and Michael Schwarz. Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, pages 59–60, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0261-6. doi: 10.1145/1993574.1993585. URL http://doi.acm.org/10.1145/1993574.1993585.

Vianney Perchet and Philippe Rigollet. The multi-armed bandit problem with covariates. *Ann. Statist.*, 41(2):693–721, 2013.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *COLT 2013 - The 26th Conference on Learning Theory, Princeton, NJ, June 12-14, 2013*, volume 30 of *JMLR W&CP*, pages 993–1019, 2013.

John Riley and William F Samuelson. Optimal auctions. *American Economic Review*, 71(3): 381–92, 1981. URL http://EconPapers.repec.org/RePEc:aea:aecrev:v:71:y:1981:i:3:p:381-92.

Tim Roughgarden, Inbal Talgam-Cohen, and Qiqi Yan. Supply-limiting mechanisms. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 844–861, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1415-2. doi: 10.1145/2229012.2229077. URL http://doi.acm.org/10.1145/2229012.2229077.

Aleksandrs Slivkins. Contextual bandits with similarity information. *J. Mach. Learn. Res.*, 15 (1):2533–2568, January 2014. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2627435.2670330.

Long Tran-Thanh, Lampros C. Stavrogiannis, Victor Naroditskiy, Valentin Robu, Nicholas R. Jennings, and Peter Key. Efficient regret bounds for online bid optimisation in budget-limited sponsored search auctions. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, UAI 2014, Quebec City, Quebec, Canada, July 23-27, 2014*, pages 809–818, 2014.

Alexandre Tsybakov. Statistique appliquée. Lecture Notes, 2006.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Chih-Chun Wang, S.R. Kulkarni, and H.V. Poor. Bandit problems with arbitrary side observations. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 3, pages 2948–2953 Vol.3, Dec 2003. doi: 10.1109/CDC.2003.1273074.

Robert Wilson. Game-theoretic analyses of trading processes. In Truman Fassett Bewley, editor, *Advances in Economic Theory*, pages 33–70. Cambridge University Press, 1987. ISBN 9781139052054. URL http://dx.doi.org/10.1017/CCOL0521340446.002. Cambridge Books Online.

Robert B. Wilson. Competitive bidding with disparate information. *Management Science*, 15(7): 446–448, 1969. ISSN 00251909, 15265501. URL http://www.jstor.org/stable/2628640.

Appendix A. Proofs of upper bounds

A.1. The stochastic case

A.1.1. PROOF OF THEOREM 1

Since v_t is independent of (m_t, b_t) and $\mathbb{E}[v_t] = v$, we have

$$\begin{aligned}\bar{R}_T &= \max_{b\in[0,1]} \mathbb{E}\sum_{t=1}^{T}(v - m_t)\mathbb{1}\{b > m_t\} - \mathbb{E}\sum_{t=1}^{T}(v - m_t)\mathbb{1}\{b_t > m_t\}\\ &= \mathbb{E}\sum_{t=1}^{T}(v - m_t)\mathbb{1}\{v > m_t\} - \mathbb{E}\sum_{t=1}^{T}(v - m_t)\mathbb{1}\{b_t > m_t\}\end{aligned}$$

where in the second equality we used the fact that the supremum is attained at $b = v$ because $(v - m_t)\mathbb{1}\{b > m_t\} \le (v - m_t)_+ = (v - m_t)\mathbb{1}\{v > m_t\}$, where $x_+ = \max(x, 0)$.

Next, decomposing the regret on the the events $\{b_t < m_t\}$ and $\{b_t > m_t\}$, on which the bidder lost and won auction t respectively, we get

$$(v - m_t)\mathbb{1}\{v > m_t\} \le (v - m_t)\mathbb{1}\{v > m_t > b_t\} + (v - m_t)_+\mathbb{1}\{b_t > m_t\}\,.$$

This yields

$$\begin{aligned}\bar{R}_T &\le \mathbb{E}\sum_{t=1}^{T}(v - m_t)\mathbb{1}\{v > m_t > b_t\} + \mathbb{E}\sum_{t=1}^{T}(m_t - v)_+\mathbb{1}\{b_t > m_t\}\\ &\le \sum_{t=1}^{T}\mathbb{P}\{b_t < v\} + \mathbb{E}\sum_{t=1}^{T}(m_t - v)\mathbb{1}\{v < m_t < b_t\}\,.\end{aligned}$$

To control the first sum, using a union bound and Hoeffding's inequality, we get

$$\mathbb{P}\{b_t < v\} \le \sum_{s=1}^{t}\mathbb{P}\big\{\bar{v}_s - v < -\sqrt{\frac{3\log t}{2s}}\big\} \le t^{-2}\,.$$

so that

$$\bar{R}_T \le \frac{\pi^2}{6} + \mathbb{E}\sum_{t=1}^{T}(m_t - v)\mathbb{1}\{v < m_t < b_t\}\,. \tag{A.8}$$

Denote by ω_t the value of ω during the tth round. To control the second sum in (A.8), observe that, since $b_t > m_t$ implies that the bidder won auction t, we have $\omega_{t+1} = \omega_t + 1$. Denote by $\mathcal{W} = \{t \in [T] \,:\, b_t > m_t\}$ the set of auctions that the bidder has won. If $m_t \ge v + \Delta$, we have

$$\begin{aligned}S &:= \mathbb{E}\sum_{t=1}^{T}(m_t - v)\mathbb{1}\{v < m_t < b_t\}\\ &\le \mathbb{E}\sum_{t\in\mathcal{W}}(m_t - v)\mathbb{1}\big\{\Delta < m_t - v < \bar{v}_{\omega_t} - v + \sqrt{\frac{3\log t}{2\omega_t}}\big\}\\ &\le \sum_{t=1}^{T}\int_0^\infty \mathbb{P}\big\{\bar{v}_t + \sqrt{\frac{3(\log T)}{2t}} - v > u + \Delta\big\}du\,.\end{aligned}$$

Using Hoeffding's inequality, we get

$$\mathbb{P}\big\{\bar{v}_t - v > \sqrt{\frac{3\log T}{2t}} + u\big\} \le T^{-3}e^{-u^2/2}\,.$$

It yields, on the one hand, that for any $t \in [T]$,

$$\begin{aligned}\int_0^\infty \mathbb{P}\big\{\bar{v}_t + \sqrt{\frac{3\log T}{2t}} - v > u + \Delta\big\}du &\le \sqrt{\frac{6\log T}{t}} + \int_0^\infty \mathbb{P}\big\{\bar{v}_t - v > \sqrt{\frac{3\log T}{2t}} + u\big\}du \\ &\le \sqrt{\frac{6\log T}{t}} + T^{-3}\sqrt{\frac{\pi}{2}}\end{aligned}$$

On the other hand, if $t > t_\Delta := 6(\log T)/\Delta^2$, we have

$$\int_0^\infty \mathbb{P}\big\{\bar{v}_t + \sqrt{\frac{3\log T}{2t}} - v > u + \Delta\big\}du \le \int_0^\infty \mathbb{P}\big\{\bar{v}_t - v > \sqrt{\frac{3\log T}{2t}} + u\big\}du \le T^{-3}\sqrt{\frac{\pi}{2}}\,.$$

It yields

$$S \le \sum_{t=1}^{t_\Delta \wedge T} \sqrt{\frac{6\log T}{t}} + \sqrt{\frac{\pi}{2}} \le \frac{12\log T}{\Delta} \wedge 2\sqrt{6T\log T} + \sqrt{\frac{\pi}{2}}\,.$$

■

A.1.2. PROOF OF THEOREM 3

We will prove the following bound:

$$\bar{R}_T \le \begin{cases} C_\mu\Big(\frac{12}{1-\alpha}T^{\frac{1-\alpha}{2}}\log^{\frac{1+\alpha}{2}}T + 1\Big) & \text{if } \alpha < 1 \\ 6C_\mu\Big(\log(T) + 1\Big)^2 & \text{if } \alpha = 1 \\ 6\log(T)\Big(1 + 2\frac{C_\mu}{\alpha\wedge 2-1}\Big) + \frac{4C_\mu}{\alpha\wedge 2-1} + 1 & \text{if } \alpha > 1 \end{cases}$$

Recall from the proof of Theorem 1 that

$$\begin{aligned} S &:= \mathbb{E}\sum_{t=1}^T (m_t - v)\mathbb{1}\{v < m_t < b_t\} \\ &\le \mathbb{E}\sum_{t\in\mathcal{W}} (m_t - v)\mathbb{1}\big\{0 < m_t - v < \bar{v}_{w_t} - v + \sqrt{\frac{3\log t}{2w_t}}\big\} \\ &\le \mathbb{E}\sum_{t\in\mathcal{W}} \Big(\bar{v}_{w_t} - v + \sqrt{\frac{3\log T}{2w_t}}\Big)\mathbb{1}\big\{0 < m_t - v < \bar{v}_{w_t} - v + \sqrt{\frac{3\log T}{2w_t}}\big\} \wedge 1 \\ &\le \mathbb{E}\sum_{t=1}^T \Big(\bar{v}_t - v + \sqrt{\frac{3\log T}{2t}}\Big)\mathbb{1}\big\{0 < m_t - v < \bar{v}_t - v + \sqrt{\frac{3\log T}{2t}}\big\} \wedge 1\,, \end{aligned}$$

where we used the fact that bids always belong to $[0,1]$. Using the margin condition, we get that for $\alpha \geq 0$

$$S \leq C_\mu \mathbb{E} \sum_{t=1}^{T} \Big(\bar{v}_t - v + \sqrt{\frac{3\log T}{2t}} \Big)_+^{1+\alpha} \wedge 1 \,.$$

Hoeffding's inequality yields that $\mathbb{P}\{\bar{v}_t - v \geq \varepsilon\} \leq e^{-2t\varepsilon^2}$, thus we get that

$$\begin{aligned} \mathbb{E}\Big(\bar{v}_t - v + \sqrt{\frac{3\log T}{2t}} \Big)_+^{1+\alpha} &\leq (1+\alpha) \int_{-\sqrt{\frac{3\log T}{2t}}}^{\infty} \Big(\varepsilon + \sqrt{\frac{3\log T}{2t}} \Big)^{\alpha} e^{-2t\varepsilon^2} d\varepsilon \\ &\leq \frac{1+\alpha}{t^{\frac{1+\alpha}{2}}} \int_{-\sqrt{\frac{3\log T}{2}}}^{\infty} \Big(s + \sqrt{\frac{3\log T}{2}} \Big)^{\alpha} e^{-2s^2} ds \\ &\leq \Big(\frac{6\log T}{t} \Big)^{\frac{1+\alpha}{2}} + \frac{1+\alpha}{2t^{\frac{1+\alpha}{2}}} \int_{\sqrt{6\log T}}^{\infty} u^{\alpha} e^{-u^2/2} du \,. \end{aligned}$$

As a consequence, if $\alpha \leq 1$, we obtain

$$S \leq C_\mu \sum_{t=1}^{T} \Big(\frac{6\log T}{t} \Big)^{\frac{1+\alpha}{2}} + \frac{C_\mu}{T^2} \,.$$

and for $\alpha < 1$, this yields that

$$S \leq C_\mu \Big(\frac{12}{1-\alpha} T^{\frac{1-\alpha}{2}} \log^{\frac{1+\alpha}{2}} T + 1 \Big) \,,$$

while, for $\alpha = 1$, we get

$$S \leq 6 C_\mu \Big(\log(T) + 1 \Big)^2 \,.$$

When $\alpha \leq 2$, it holds that

$$\int_{\sqrt{6\log T}}^{\infty} u^{\alpha} e^{-u^2/2} du \leq \int_{\sqrt{6\log T}}^{\infty} u^{2} e^{-u^2/2} du \leq 2$$

hence

$$\begin{aligned} S &\leq 6\log T + 1 + C_\mu \Big(\sum_{t=\lceil 6\log T \rceil + 1}^{T} \Big(\frac{6\log T}{t} \Big)^{\frac{1+\alpha}{2}} + \frac{1+\alpha}{t^{\frac{1+\alpha}{2}}} \Big) \\ &\leq 6\log(T) \Big(1 + 2\frac{C_\mu}{\alpha - 1} \Big) + \frac{4C_\mu}{\alpha - 1} + 1 \,. \end{aligned}$$

For bigger values of α, we shall use the fact that if the margin condition is satisfied for $\alpha \geq 2$, then it is also satisfied for the value $\alpha = 2$. As a consequence, plugging the value $\alpha = 2$ in the above equation, we obtain that

$$S \leq 6\log(T) \Big(1 + 2C_\mu \Big) + 4C_\mu + 1 \,.$$

■

A.2. The adversarial case

A.2.1. PROOF OF THEOREM 5

When $\sqrt{\log(1/\Delta^\circ)/4T} > 1$, the claimed bound is vacuous, so we assume in what follows that $\eta = \sqrt{\log(1/\Delta^\circ)/4T}$.

Define $W_t = \sum_{\kappa \in \mathcal{L}_t} |\kappa| w_{\kappa,t}$. By extending the definition (4.5) of $p_{\ell,t}$ to all $\ell \in \mathcal{L}_{t+1}$, we can write

$$\log \frac{W_{t+1}}{W_t} = \log \sum_{\ell \in \mathcal{L}_{t+1}} \frac{|\ell| w_{\ell,t} \exp(\eta \hat{g}(\ell,t))}{W_t} = \log \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \exp(\eta \hat{g}(\ell,t)).$$

By construction, $\hat{g}(\ell,t) \le 1$. Since $e^x \le 1 + x + x^2$ for $x \le 1$ and $\eta \le 1$, this implies

$$\begin{aligned} \log \frac{W_{t+1}}{W_t} &\le \log \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t}(1 + \eta \hat{g}(\ell,t) + \eta^2 \hat{g}(\ell,t)^2) \qquad (A.9) \\ &= \log \big(1 + \eta \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t) + \eta^2 \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t)^2\big) \\ &\le \eta \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t) + \eta^2 \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t)^2 . \end{aligned}$$

It follows from (4.4) and the fact that $\sum_{\ell \in \mathcal{L}_{t+1}, m_t \preceq \ell} p_{\ell,t} = \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)$ that

$$\begin{aligned} \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t) &= \sum_{\substack{\ell \in \mathcal{L}_{t+1} \\ m_t \preceq \ell}} p_{\ell,t} \Big(1 - \frac{1 - (v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\Big) \\ &= \mathbb{P}_{\mathcal{B}_t}(b_t > m_t) - \Big(1 - (v_t - m_t)\Big) \mathbb{1}\{b_t > m_t\} \\ &= g(b_t,t) + \mathbb{P}_{\mathcal{B}_t}(b_t > m_t) - \mathbb{1}\{b_t > m_t\} . \end{aligned}$$

As a consequence, we obtain

$$\mathbb{E} \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t) = \mathbb{E} g(b_t,t)$$

We also have

$$\begin{aligned} \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t)^2 &= \sum_{\substack{\ell \in \mathcal{L}_{t+1} \\ m_t \preceq \ell}} p_{\ell,t} \Big(1 - \frac{1 - (v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\Big)^2 \\ &= \mathbb{P}_{\mathcal{B}_t}(b_t > m_t) \Big(1 - \frac{1 - (v_t - m_t)}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} \mathbb{1}\{b_t > m_t\}\Big)^2 \\ &= \mathbb{P}_{\mathcal{B}_t}(b_t > m_t) + \Big(\frac{\big(1 - (v_t - m_t)\big)^2}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} - 2\big(1 - (v_t - m_t)\big)\Big) \mathbb{1}\{b_t > m_t\} . \end{aligned}$$

Thus we obtain, since $0 \le 1 - (v_t - m_t) \le 2$, that

$$\mathbb{E} \sum_{\ell \in \mathcal{L}_{t+1}} p_{\ell,t} \hat{g}(\ell,t)^2 \le 4$$

Combining the above displays with (A.9) yields

$$\mathbb{E}\log\frac{W_{t+1}}{W_t}\le\eta\mathbb{E}g(b_t,t)+4\eta^2\,.$$

Let $G(b)=\sum_{t=1}^{T}\hat{g}(b,t)$ and $\bar{G}=\sum_{t=1}^{T}g(b_t,t)$. Summing on T, we obtain

$$\mathbb{E}\big[\log\frac{W_T}{W_0}\big]\le\eta\mathbb{E}\bar{G}+4\eta^2T\,.$$

Let $b^\circ\in\operatorname{argmax}_{b\in[0,1]}\sum_{t=1}^{T}\mathbb{E}(v_t-m_t)\mathbb{1}\{b>m_t\}$, and suppose $b^\circ\in\ell^\circ\in\mathcal{L}_T$. We can bound W_T by writing

$$\begin{aligned}\mathbb{E}[\log W_T]=\mathbb{E}\big[\log\sum_{\ell\in\mathcal{L}_T}|\ell|\exp\big(\eta\sum_{t=1}^{T}\hat{g}(\ell,t)\big)\big]&\ge\mathbb{E}\big[\log|\ell^\circ|\exp\big(\eta\sum_{t=1}^{T}\hat{g}(b^\circ,t)\big)\big]\\&=\log\Delta^\circ+\eta G(b^\circ).\end{aligned}$$

Rearranging and noting that $W_0=1$, we obtain

$$\bar{R}_T=\max_{b\in[0,1]}\mathbb{E}G(b)-\mathbb{E}\bar{G}\le4\eta T+\frac{\log(1/\Delta^\circ)}{\eta}\,.$$

Plugging in the given value of η yields the claim. ■

A.2.2. THE GENERIC DOUBLING TRICK

Even though the statement of Theorem 5 appears to require knowledge of several parameters in advance, it is possible to turn EXPTREE into a fully online algorithm at the cost of only slightly worse performance.

We initialize two bounds, $B_T=1$ and $B_\Delta=1$, and run EXPTREE with parameter $\eta=(1/2)\sqrt{B_\Delta/B_T}\wedge1$ until either $t\le B_T$ or $\log1/\Delta\le B_\Delta$ fails to hold. When one of these bounds is breached, we double the bound and restart the algorithm, maintaining the partition $\mathcal{L}_t$ but setting $w_\ell=1$ for all $\ell\in\mathcal{L}_t$. This modified strategy yields the following theorem.

Theorem 9 *The strategy* EXPTREE *run with the above doubling procedure yields an expected regret bound*

$$R_T\le48\sqrt{2T\log(1/\Delta)}\,.$$

Proof Divide the algorithm into stages on which B_T and B_Δ are constant, and denote by $B_T^\star$ and $B_\Delta^\star$ the values of B_T and B_Δ when the algorithm terminates. The proof of Theorem 5 implies that the expected regret incurred during any given stage is at most

$$4\eta T+\frac{\log(1/\Delta)}{\eta}\le4\eta B_T+\frac{B_\Delta}{\eta}\le4\sqrt{T\log(1/\Delta)}\,.$$

It remains to sum these regrets over each stage, since the actual expected regret (which requires a fixed bid across all stages) can only be smaller.

Suppose that the algorithm lasted a total of $\ell + m + 1$ stages, ℓ of which were ended because the bound $t \leq B_T$ was violated and m of which were ended because the bound $\log(1/\Delta) \leq B_\Delta$ was violated. The total regret across all $\ell + m + 1$ stages is bounded by

$$\sum_{i=0}^{\ell} \sum_{j=0}^{m} 4\sqrt{2^i \cdot 2^j} = \frac{4}{(\sqrt{2}-1)^2}(2^{(i+1)/2} - 1)(2^{(j+1)/2} - 1) \leq 48\sqrt{B_T^\star B_\Delta^\star}\,.$$

Moreover, when the algorithm terminates, we have the bounds $B_T^\star \leq T$ and $B_\Delta^\star \leq 2\log(1/\Delta)$. The result follows. ■

A.2.3. PROOF OF THEOREM 6

Denote by $G(b) = \sum_{t=1}^T G(b,t)$ and $\tilde{G}(b) = \sum_{t=1}^T \tilde{g}(b,t)$ the cumulative true and estimated gains for a bet b. Before proving Theorem 6, we establish the following lemma, which shows that $\tilde{G}$ can be viewed as an upper bound on G.

Lemma 10 *With probability at least $1-\delta$ the bound $G(b,T) \leq \tilde{G}(b,T) + \frac{\log T\delta^{-1}}{\beta}$ holds for all $b \in [0,1]$.*

Proof Denote by $\mathbb{E}_t$ expectation with respect to the random choice of b_t, conditioned on the outcomes of rounds $1,\dots,t-1$. Fix $b \in [0,1]$ and define $d(b,t) = g(b,t) - \tilde{g}(b,t)$. Note that

$$\mathbb{E}_t[d(b,t)] = -\frac{\beta \mathbb{1}\{b > m_t\}}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)} - \frac{\beta \mathbb{1}\{b \leq m_t\}}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}$$

so that

$$\begin{aligned}\bar{d}(b,t) := d(b,t) - \mathbb{E}_t[d(b,t)] = v_t \mathbb{1}\{b > m_t\}\big(1 - \frac{\mathbb{1}\{b_t > m_t\}}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}\big) \\ + m_t \mathbb{1}\{b < m_t\}\big(1 - \frac{\mathbb{1}\{b_t < m_t\}}{\mathbb{P}_{\mathcal{B}_t}(b_t < m_t)}\big) \qquad \text{(A.10)}\end{aligned}$$

This immediately immediately yields $\bar{d}(b,t) \leq g(b,t) \leq 1$. Since $\beta \leq 1$, and $e^x \leq 1 + x + x^2$ for $x \leq 1$, we have

$$\mathbb{E}_t\big[e^{\beta d(b,t)}\big] = e^{\beta \mathbb{E}_t[d(b,t)]}\mathbb{E}_t\big[e^{\beta \bar{d}(b,t)}\big] \leq e^{\beta \mathbb{E}_t[d(b,t)]}\big(1 + \beta^2 \mathbb{E}_t[\bar{d}^{\,2}(b,t)]\big).$$

It follows from (A.10) that

$$\begin{aligned}\mathbb{E}_t[\bar{d}^{\,2}(b,t)] &\leq \frac{v_t^2}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}\mathbb{1}\{b > m_t\} + \frac{m_t^2}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}\mathbb{1}\{b \leq m_t\} \\ &\leq \frac{1}{\mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}\mathbb{1}\{b > m_t\} + \frac{1}{1 - \mathbb{P}_{\mathcal{B}_t}(b_t > m_t)}\mathbb{1}\{b \leq m_t\} \\ &= -\frac{1}{\beta}\mathbb{E}_t[d(b,t)].\end{aligned}$$

Combining this with the preceding inequality and the fact that $\beta \mathbb{E}_t[d(b,t)] \leq 1$ yields

$$\mathbb{E}_t\big[e^{\beta d(b,t)}\big] \leq e^{\beta \mathbb{E}_t[d(b,t)]}\big(1 - \beta \mathbb{E}_t[d(b,t)]\big) \leq 1.$$

Let $Z_t = \exp(\beta d(b,t))$. Then

$$\mathbb{E}\big[e^{\beta(G(b)-\tilde{G}(b))}\big] \le \mathbb{E}\big[\exp\big(\beta\sum_{t=1}^{T} d(b,t)\big)\big] = \mathbb{E}\big[\prod_{t=1}^{T} Z_t\big] \le 1,$$

where the last step follows by conditioning on each stage in turn and applying the above bound.

To obtain a uniform bound, we note that the function $b \mapsto G(b) - \tilde{G}(b)$ takes at most T random values $G_1, \dots, G_T$ as b varies across $[0,1]$. Moreover, the proof above establishes that $\max_j \mathbb{E}[\exp(\beta G_j)] \le 1$. Hence

$$\mathbb{E}\big[\exp\big(\beta\big[\max_{b\in[0,1]} G(b) - \tilde{G}(b)\big]\big)\big] = \mathbb{E}\big[\exp\big(\beta \max_{j\in[T]} G_j\big)\big] \le \sum_{j=1}^{T} \mathbb{E}\big[e^{\beta G_j}\big] \le T.$$

Applying the Markov bound yields the claim. ■

We are now in a position to prove Theorem 6.

Proof [Proof of Theorem 6] We proceed as in the proof of Theorem 5. Note that the choice of η guarantees that $\mathcal{B}_t$ is a valid probability distribution.

As above, define $W_t = \sum_{\kappa\in\mathcal{L}_t} |\kappa| w_{\kappa,t}$. We have

$$\log\frac{W_{t+1}}{W_t} = \log\sum_{\ell\in\mathcal{L}_{t+1}} \frac{|\ell| w_{\ell,t}\exp(\eta\tilde{g}(\ell,t))}{W_t} = \log\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}\exp(\eta\tilde{g}(\ell,t)).$$

Since $\eta\tilde{g}(\ell,t) \le \eta\frac{1+\beta}{\gamma} \le 1$, the inequality $e^x \le 1 + x + x^2$ for $x \le 1$ implies

$$\begin{aligned}\log\frac{W_{t+1}}{W_t} &\le \log\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}(1+\eta\tilde{g}(\ell,t)+\eta^2\tilde{g}(\ell,t)^2)\\ &= \log\big(1+\eta\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}\tilde{g}(\ell,t) + \eta^2\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}\tilde{g}(\ell,t)^2\big).\end{aligned}$$

By the same reasoning as in the proof of Theorem 5, we have

$$\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}\tilde{g}(\ell,t) = \frac{1}{1-2\gamma}\sum_{\ell\in\mathcal{L}_{t+1}} \mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)\tilde{g}(\ell,t) \le \frac{1}{1-2\gamma}(g(b_t,t)+2\beta)$$

and similarly

$$\sum_{\ell\in\mathcal{L}_{t+1}} p_{\ell,t}\tilde{g}(\ell,t)^2 = \frac{1}{1-2\gamma}\sum_{\ell\in\mathcal{L}_{t+1}} \mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)\tilde{g}(\ell,t)^2.$$

To compute this last quantity, note that (4.7) implies

$$\begin{aligned}\sum_{\ell\in\mathcal{L}_{t+1}} \mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)\tilde g(\ell,t)^2 &= \sum_{\substack{\ell\in\mathcal{L}_{t+1}\\ m_t\preceq\ell}} \frac{\mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)}{\mathbb{P}_{\mathcal{B}_t}(b_t>m_t)}(v_t\mathbb{1}\{b_t>m_t\}+\beta)\tilde g(\ell,t)\\ &\quad+\sum_{\substack{\ell\in\mathcal{L}_{t+1}\\ m_t\succeq\ell}} \frac{\mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)}{1-\mathbb{P}_{\mathcal{B}_t}(b_t>m_t)}(m_t\mathbb{1}\{b_t\le m_t\}+\beta)\tilde g(\ell,t)\\ &\le g(b_t,t)\tilde g(b_t,t)+\beta\big(\frac{v_t\mathbb{1}\{b_t>m_t\}+\beta}{\mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)}+\frac{m_t\mathbb{1}\{b_t\le m_t\}+\beta}{1-\mathbb{P}_{\mathcal{B}_t}(b_t\in\ell)}\big)\\ &= g(b_t,t)\tilde g(b_t,t)+\beta(\tilde g(1,t)+\tilde g(0,t))\\ &\le (1+\beta)(\tilde g(1,t)+\tilde g(0,t))\,.\end{aligned}$$

where in the last inequality, we used the fact $g(b_t,t)\le 1$ and $\tilde g(b_t,t)\le\tilde g(1,t)+\tilde g(0,t)$. Combining the above bounds yields

$$\log\frac{W_{t+1}}{W_t}\le\frac{\eta}{1-2\gamma}(g(b_t,t)+2\beta)+\frac{\eta^2}{1-2\gamma}(1+\beta)(\tilde g(1,t)+\tilde g(0,t)).$$

Defining $\bar G=\sum_{t=1}^T g(b_t,t)$ and summing on T yields

$$\begin{aligned}\log\frac{W_T}{W_0}&\le\frac{\eta}{1-2\gamma}\bar G+\frac{2T\eta\beta}{1-2\gamma}+\frac{\eta^2}{1-2\gamma}(1+\beta)(\tilde G(1,T)+\tilde G(0,t))\\ &\le\frac{\eta}{1-2\gamma}\bar G+\frac{2T\eta\beta}{1-2\gamma}+\frac{2\eta^2}{1-2\gamma}(1+\beta)\max_{b\in[0,1]}\tilde G(b).\end{aligned}$$

We bound W_T by writing

$$\log W_T\ge\log\Delta^\circ+\eta\max_{b\in[0,1]}\tilde G(b).$$

Rearranging yields

$$(1-2\gamma-2\eta(1+\beta))\max_{b\in[0,1]}\tilde G(b)-\bar G\le 2T\beta+\frac{(1-2\gamma)\log(1/\Delta^\circ)}{\eta}\le 2T\beta+\frac{\log(1/\Delta^\circ)}{\eta}.$$

Applying Lemma 10, with probability $1-\delta$ we have

$$\max_{b\in[0,1]}G(b)\le\max_{b\in[0,1]}\tilde G(b)+\frac{\log(T\delta^{-1})}{\beta}$$

which implies

$$(1-2\gamma-2\eta(1+\beta))\max_{b\in[0,1]}G(b,T)-\bar G\le T\beta+\frac{\log(T\delta^{-1})}{\beta}+\frac{\log(1/\Delta^\circ)}{\eta}$$

since $2\gamma+2\eta(2+\beta)\le 8\eta\le 1$. We obtain

$$\begin{aligned}\max_{b\in[0,1]}G(b)-\bar G&\le 2T\beta+\frac{\log(T\delta^{-1})}{\beta}+\frac{\log(1/\Delta^\circ)}{\eta}+(2\gamma+2\eta(1+\beta))T\\ &\le 2T\beta+8\eta T+\frac{\log(T\delta^{-1})}{\beta}+\frac{\log(1/\Delta^\circ)}{\eta}.\end{aligned}$$

Plugging in the given parameters then yields the claim.

The bound in expectation follows upon integrating the first result. ■

Appendix B. Proofs of lower bounds

B.1. The stochastic case

B.1.1. PROOF OF THEOREM 4

Recall that for $\alpha \in (0,1)$, we let μ_α be the distribution on $[0,1]$ with density

$$g_\alpha(x) = C_\alpha\Big[\big(x-\frac{1}{2}\big)^{\alpha-1}\mathbb{1}\big\{x \in (1/2, 1/2+2\varepsilon]\big\} + \big(x-\frac{1}{2}-2\varepsilon\big)^{\alpha-1}\mathbb{1}\big\{x \in (1/2+2\varepsilon, 1]\big\}\Big]\,,$$

where C_α is a normalizing constant. (In what follows, $C_\alpha > 0$ is a constant that may change from line to line but depends on α only.) When $\alpha \geq 1$, we define μ_α to be the point mass at $1/2+\varepsilon$. With these definitions, μ_α satisfies the margin condition with parameter α around both $1/2$ and $1/2+2\varepsilon$.

We first consider the case where $\alpha < 1$. Recall from (3.1) that the pseudo-regret is given by $\bar{R}_T = \sum_{t=1}^T r_t$ where r_t denotes the *instantaneous regret*, defined by

$$r_t(\nu) = \mathbb{E}_\nu(v-m_t)\mathbb{1}\{v > m_t\} - \mathbb{E}_\nu(v-m_t)\mathbb{1}\{b_t > m_t\}\,.$$

Note first that under ν or ν' we can restrict our attention to strategies that bid $b_t \geq 1/2$. Observe first that since $v = 1/2$ under ν, the definition of the pseudo-regret (3.1) simplifies to

$$\bar{R}_T(\nu) = \sum_{t=1}^T \mathbb{E}_\nu(m_t - v)\mathbb{1}\{b_t > m_t\} = \mathbb{E}_\nu \sum_{t=1}^T \int_{1/2}^{b_t} (x-1/2)g_\alpha(x)dx \tag{B.11}$$

Moreover,

$$\int_{1/2}^{b_t}(x-1/2)g_\alpha(x)dx \geq C_\alpha\big[\bar{b}_t^{\alpha+1}\mathbb{1}\{\bar{b}_t \leq 2\varepsilon\} + \big((2\varepsilon)^{\alpha+1} + (\bar{b}_t - 2\varepsilon)^{\alpha+1}\big)\mathbb{1}\{\bar{b}_t > 2\varepsilon\}\big]$$

where $\bar{b}_t = b_t - 1/2 \geq 0$. Therefore

$$\bar{R}_T(\nu) \geq C_\alpha \mathbb{E}_\nu S_{\alpha+1}\,, \tag{B.12}$$

where

$$S_\alpha = \sum_{t=1}^T \bar{b}_t^\alpha \mathbb{1}\{\bar{b}_t \leq 2\varepsilon\} + \big((2\varepsilon)^\alpha + (\bar{b}_t - 2\varepsilon)^\alpha\big)\mathbb{1}\{\bar{b}_t > 2\varepsilon\}\,.$$

We will use the fact that $\mathbb{E}_\nu S_{\alpha+1} \geq (2\varepsilon)^{\alpha+1}\mathsf{S}(\varepsilon)$ and $\mathbb{E}_\nu S_\alpha \leq (2\varepsilon)^\alpha T + \mathsf{S}(\varepsilon)$, where

$$\mathsf{S}(\varepsilon) = \sum_{t=1}^T \mathbb{P}_\nu\{\bar{b}_t > 2\varepsilon\}\,.$$

Assume first that the bidder uses a deterministic strategy. For any such strategy, define the associated test $\psi_t \in \{\nu, \nu'\}$ by $\psi = \nu$ if $\bar{b}_t \le \varepsilon$ and $\psi = \nu'$ if $\bar{b}_t > \varepsilon$. One the one hand, under ν, the instantaneous regret r_t satisfies

$$r_t(\nu) \ge \mathbb{E}_\nu (m_t - 1/2) \mathbb{1}\{1/2 + \varepsilon > m_t\} \mathbb{1}\{b_t > 1/2 + \varepsilon\} \ge C_\alpha \varepsilon^{\alpha+1} \mathbb{P}_\nu(\psi_t = \nu')\,.$$

On the other hand, under ν', the instantaneous regret r_t satisfies

$$\begin{aligned} r_t(\nu') &\ge \mathbb{E}_{\nu'}\big[\mathbb{1}\{b_t < 1/2 + \varepsilon\}(1/2 + 2\varepsilon - m_t)\big(\mathbb{1}\{1/2 + 2\varepsilon > m_t\} - \mathbb{1}\{b_t > m_t\}\big)\big] \\ &\ge \mathbb{E}_{\nu'}\big[\mathbb{1}\{b_t < 1/2 + \varepsilon\}(1/2 - 2\varepsilon - m_t)\big(\mathbb{1}\{1/2 + 2\varepsilon > m_t\} - \mathbb{1}\{1/2 + \varepsilon > m_t\}\big)\big] \\ &= \mathbb{E}_{\nu'}\big[\mathbb{1}\{b_t < 1/2 + \varepsilon\}(1/2 - 2\varepsilon - m_t)\mathbb{1}\{1/2 + \varepsilon \le m_t < 1/2 + 2\varepsilon\big] \\ &= C_\alpha \mathbb{P}_{\nu'}(\psi_t = \nu) \int_\varepsilon^{2\varepsilon} x^\alpha (2\varepsilon - x) dx \ge C_\alpha \mathbb{P}_{\nu'}(\psi_t = \nu) \varepsilon^{\alpha+1}\,. \end{aligned}$$

The last two displays yield

$$r_t(\nu) + r_t(\nu') \ge C_\alpha \varepsilon^{\alpha+1} \big[\mathbb{P}_\nu(\psi_t = \nu') + \mathbb{P}_{\nu'}(\psi_t = \nu)\big]\,. \tag{B.13}$$

Denote by $\hat{\nu}_t$ and $\hat{\nu}'_t$ the distribution of values *observed by the bidder* during the first $t-1$ rounds under ν and ν', respectively. Since the bidder's strategy is deterministic, the bidder's action at time t and hence the test ψ_t depend only on the observed values for the first $t-1$ rounds. Equation B.13 can therefore be rewritten as

$$r_t(\nu) + r_t(\nu') \ge C_\alpha \varepsilon^{\alpha+1} \big[\mathbb{P}_{\hat{\nu}_t}(\psi_t = \nu') + \mathbb{P}_{\hat{\nu}'_t}(\psi_t = \nu)\big]\,.$$

It follows from Sanov's inequality (see, e.g., (Bubeck et al., 2013, Lemma 4)) that

$$\mathbb{P}_{\hat{\nu}_t}(\psi_t = \nu') + \mathbb{P}_{\hat{\nu}'_t}(\psi_t = \nu) \ge \frac{1}{2} \exp\big[-\mathsf{KL}(\hat{\nu}_t, \hat{\nu}'_t)\big]\,.$$

Moreover, since (i) m_t has the same distribution under both ν and ν' and (ii), v_t is observed only when $b_t \ge m_t$, we get

$$\begin{aligned} \mathsf{KL}(\hat{\nu}_t, \hat{\nu}'_t) &= \mathbb{E}_\nu \sum_{s=1}^{t-1} \mathbb{1}(b_t \ge m_t) \mathsf{KL}\big(\mathsf{Bern}(1/2), \mathsf{Bern}(1/2 + 2\varepsilon)\big) \\ &\le 4\varepsilon^2 \sum_{s=1}^{t} \mathbb{P}_\nu(m_t \le b_t) \\ &\le C_\alpha \varepsilon^2 \mathbb{E}_\nu S_\alpha \\ &\le C_\alpha (2\varepsilon)^{2+\alpha} T + \varepsilon^2 \mathsf{S}(\varepsilon) \end{aligned}$$

where we used the fact that $\varepsilon \le (2\sqrt{2})^{-1}$ in the first inequality. The above display yields the bound

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \ge C_\alpha T \varepsilon^{\alpha+1} \exp\big[-C_\alpha\big((2\varepsilon)^{2+\alpha} T + \varepsilon^2 \mathsf{S}(\varepsilon)\big)\big],$$

and combining with (B.12) yields

$$\begin{aligned} \bar{R}_T(\nu) + \bar{R}_T(\nu') &\ge C_\alpha\big(T\varepsilon^{\alpha+1} \exp\big[-C_\alpha\big((2\varepsilon)^{2+\alpha} T + \varepsilon^2 \mathsf{S}(\varepsilon)\big)\big] + (2\varepsilon)^{\alpha+1} \mathsf{S}(\varepsilon)\big) \\ &\ge C_\alpha\big(T\varepsilon^{\alpha+1} \exp\big[-\varepsilon^2 \mathsf{S}(\varepsilon)\big] + (2\varepsilon)^{\alpha+1} \mathsf{S}(\varepsilon)\big) \end{aligned}$$

for ε such that $(2\varepsilon)^{2+\alpha}T \ = O(1)$. We obtain

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \geq C_\alpha \inf_{S\in[0,T]} \left\{T\varepsilon^{\alpha+1}\exp(-\varepsilon^2 S) + \varepsilon^{\alpha+1}S\right\}$$

The infimum is achieved when

$$S = \frac{\log(T\varepsilon^2)}{\varepsilon^2} \vee 0\,,$$

so in particular

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \geq C_\alpha T\varepsilon^{\alpha+1}$$

for all $\varepsilon = O(T^{-1/2})$. Choosing $\varepsilon = \frac{1}{2}T^{-1/2}$ yields

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \geq C_\alpha T^{\frac{1-\alpha}{2}}\,,$$

as desired.

When $\alpha \geq 1$, we obtain the following analogue to (B.11):

$$\bar{R}_t(\nu) = \sum_{t=1}^{T} \mathbb{E}_\nu \varepsilon \mathbb{1}\{b_t > 1/2 + \varepsilon\} = \varepsilon \mathsf{S}'(\varepsilon)$$

where

$$\mathsf{S}'(\varepsilon) = \sum_{t=1}^{T} \mathbb{P}_\nu\{b_t > 1/2 + \varepsilon\}\,.$$

The rest of the proof is the same apart from some small changes. Since v_t is only observed when $b_t \geq 1/2 + \varepsilon$, we obtain the bound

$$\mathsf{KL}(\hat{\nu}_t, \hat{\nu}_t') \leq C_\alpha \varepsilon^2 \mathsf{S}'(\varepsilon)\,.$$

This yields

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \geq C_\alpha \inf_{S\in[0,T]} \left\{T\varepsilon\exp(-\varepsilon^2 S) + \varepsilon S\right\}.$$

The infimum is again attained at

$$S = \frac{\log(T\varepsilon^2)}{\varepsilon^2} \vee 0\,,$$

which implies

$$\bar{R}_T(\nu) + \bar{R}_T(\nu') \geq C_\alpha \left(1 + \log(T\varepsilon^2)\right)$$

for all $\varepsilon \leq 1/4$. Choosing $\varepsilon = O(1)$ yields the claim.

The same claims hold for randomized strategies upon averaging over the internal randomness of the strategy. ■

B.2. The adversarial case

B.2.1. DISCRETIZATION CANNOT ACHIEVE SUBLINEAR REGRET

A player in the auction game we describe has an uncountable number of actions available at each step. Unlike other settings, naively discretizing the interval and playing on only a finite subset of $[0, 1]$ cannot achieve sublinear regret in this context. To make the following bound comparable to the bounds we achieve in Theorems 5 and 6, we define Δ° to be the smallest positive gap between the opponent's bids over the course of the auctions. We obtain the following.

Proposition 11 *Let $S \subset [0, 1]$ be finite. Let $\mathcal{S}$ be a (possibly randomized) strategy supported only on S. Then there is a sequence of values $\{v_t\}$ and opponent bids $\{m_t\}$ with $\Delta^\circ = \Omega(1/|S|)$ such that the pseudo-regret satisfies*

$$\bar{R}_T = \max_{b \in [0,1]} \sum_{t=1}^{T} \mathbb{E}(v_t - m_t)\mathbb{1}\{b > m_t\} - \sum_{t=1}^{T} \mathbb{E}(v_t - m_t)\mathbb{1}\{b_t > m_t\} \geq \frac{T}{8} \,.$$

Proof We will show that there is a distribution over sequences $\{v_t\}$ and $\{m_t\}$ such that the claimed bound holds in expectation. The existence of specific sequence satisfying the bound immediately follows.

There must exist a pair of points $s_1, s_2 \in [1/4, 3/4]$ such that $s_2 - s_1 \geq \frac{1}{2(|S|+1)}$ and no point in S lies in the interval (s_1, s_2). Choose $\varepsilon = \frac{1}{4}(s_2 - s_1)$. Generate the sequences $\{v_t\}$ and $\{m_t\}$ randomly by setting $(v_t, m_t) = (0, s_2 - \varepsilon)$ with probability $1/2$ and $(v_t, m_t) = (1, s_1 + \varepsilon)$ otherwise, independently for each $t \in \{1, \dots, T\}$.

The optimal bid against this sequence is any point in the interval $(s_1 + \varepsilon, s_2 - \varepsilon)$. On the other hand, any bid outside this interval incurs pseudo-regret of at least $1/8$ at each round. By assumption, no point in S lies in the interval $(s_1 + \varepsilon, s_2 - \varepsilon)$, so the strategy $\mathcal{S}$ incurs pseudo-regret of at least $T/8$. ■

B.2.2. PROOF OF LEMMA 7

We first consider deterministic strategies. Fix an ε. Denote by U the adversary under which $v_t \sim \mathsf{Bern}(m + \varepsilon)$ and by L the adversary under which $v_t \sim \mathsf{Bern}(m - \varepsilon)$.

Given a sequence of bids $b_1, \dots, b_T$, let T_- and T_+ be the number of times t for which $b_t < m$ and $b_t > m$, respectively. Denoting the regret after T rounds by R_T, it is easy to show that

$$\mathbb{E}_U[R_T] \geq \varepsilon \mathbb{E}_U[T_-] \quad \mathbb{E}_L[R_T] \geq \varepsilon \mathbb{E}_L[T_+].$$

Write $\mathbb{P}_U$ and $\mathbb{P}_L$ for the law of T_- under adversary U and L, respectively, and denote by $\mathbb{P}_{\text{av}}$ the distribution of T_- when $v_t \sim \mathsf{Bern}(m)$. Then Pinsker's inequality implies

$$\mathbb{E}_U[T_-] \geq \mathbb{E}_{\text{av}}(T_-) - T\sqrt{\mathsf{KL}(\mathbb{P}_U, \mathbb{P}_{\text{av}})/2}\,, \quad \mathbb{E}_L[T_+] \geq \mathbb{E}_{\text{av}}(T_+) - T\sqrt{\mathsf{KL}(\mathbb{P}_L, \mathbb{P}_{\text{av}})/2}\,.$$

By the data processing inequality,

$$\mathsf{KL}(\mathbb{P}_U, \mathbb{P}_{\text{av}}) \leq T \cdot \mathsf{KL}(\mathsf{Bern}(m + \varepsilon), \mathsf{Bern}(m)) \leq T \frac{\varepsilon^2}{m(1-m)} \leq 8T\varepsilon^2,$$

and likewise for $\mathsf{KL}(\mathbb{P}_L, \mathbb{P}_{\text{av}})$. We therefore obtain

$$\frac{1}{2}\big(\mathbb{E}_U[R_T] + \mathbb{E}_L[R_T]\big) \geq \varepsilon\big(\frac{T}{2} - 2T\varepsilon\sqrt{T}\big).$$

Setting $\varepsilon = \frac{1}{8\sqrt{T}}$ and bounding the average by a maximum yields

$$\max_{A\in\{U,L\}} \max_{b\in[0,1]} \mathbb{E}_A\big[\sum_{t=1}^{T} g(b,t) - \sum_{t=1}^{T} g(b_t,t)\big] \geq \frac{\sqrt{T}}{32}$$

for any deterministic strategy. The claim follows for general strategies by averaging over the bidder's internal randomness and applying Fubini's theorem. ■

B.2.3. PROOF OF THEOREM 8

We can assume without loss of generality that Δ is a power of 2, since this can change the regret by at most a constant.

Set $n = \log_2(1/2\Delta)$. We divide the game into n stages of $\frac{T}{n}$ rounds each and will show that any bidder incurs regret of at least $\frac{1}{32}\sqrt{\frac{T}{n}}$ during each stage by repeatedly applying Lemma 7.

During the first stage, apply Lemma 7 with $m = 1/2$. One of the two adversaries will incur regret in expectation of at least $\frac{1}{32}\sqrt{\frac{T}{n}}$. If that adversary is U, the next stage will use Lemma 7 with $m = 5/8$; if it is L, then the next stage will use $m = 3/8$.

In general, for the ith stage we will apply Lemma 7 with $m = 1/4 + c_i 2^{-i-1}$ for some c_i. If the U adversary has higher regret in expectation at that stage, then $c_{i+1} = 2c_i + 1$; otherwise $c_{i+1} = 2c_i - 1$. Note that during the ith stage, the smallest gap between two of the adversary's bids is 2^{-i-1}.

The structure of the optimum bids for the adversaries U and L guarantees that during each stage, there is an interval within which a fixed bid would be optimal for all previous stages. So after n stages there is a fixed bid that is optimal for all n adversaries. Therefore the regret across the n stages is equal to the sum of the regrets for each stage, and we obtain

$$\max_{b\in[0,1]} \mathbb{E}\sum_{t=1}^{T} g(b,t) - \mathbb{E}\sum_{t=1}^{T} g(b_t,t) \geq n\frac{1}{32}\sqrt{\frac{T}{n}} = \frac{1}{32}\sqrt{Tn} = \frac{1}{32}\sqrt{T\lfloor\log_2(1/2\Delta)\rfloor},$$

as desired. ■