Fast Rate Learning in Stochastic First Price Bidding

Juliette Achddou, Olivier Cappé, Aurélien Garivier

Fast Rate Learning in Stochastic First Price Bidding

Juliette Achddou JULIETTE.ACHDOU@GMAIL.COM
DIENS, INRIA, Université PSL, 1000mercis Group

Olivier Cappé OLIVIER.CAPPE@CNRS.FR
DIENS, CNRS, INRIA, Université PSL,

Aurélien Garivier AURELIEN.GARIVIER@ENS-LYON.FR
UMPA,CNRS, INRIA, ENS Lyon

Editors: Vineeth N Balasubramanian and Ivor Tsang

Abstract

First-price auctions have largely replaced traditional bidding approaches based on Vickrey auctions in programmatic advertising. As far as learning is concerned, first-price auctions are more challenging because the optimal bidding strategy does not only depend on the value of the item but also requires some knowledge of the other bids. They have already given rise to several works in sequential learning, many of which consider models for which the value of the buyer or the opponents' maximal bid is chosen in an adversarial manner. Even in the simplest settings, this gives rise to algorithms whose regret grows as $\sqrt{T}$ with respect to the time horizon T. Focusing on the case where the buyer plays against a stationary stochastic environment, we show how to achieve significantly lower regret: when the opponents' maximal bid distribution is known we provide an algorithm whose regret can be as low as $\log^2(T)$; in the case where the distribution must be learnt sequentially, a generalization of this algorithm can achieve $T^{1/3+\epsilon}$ regret, for any $\epsilon > 0$. To obtain these results, we introduce two novel ideas that can be of interest in their own right. First, by transposing results obtained in the posted price setting, we provide conditions under which the first-price bidding utility is locally quadratic around its optimum. Second, we leverage the observation that, on small sub-intervals, the concentration of the variations of the empirical distribution function may be controlled more accurately than by using the classical Dvoretzky-Kiefer-Wolfowitz inequality. Numerical simulations confirm that our algorithms converge much faster than alternatives proposed in the literature for various bid distributions, including for bids collected on an actual programmatic advertising platform.

Keywords: multi-armed bandits; sequential bidding; auctions

1. Introduction

We consider the problem of setting a bid in repeated first-price auctions. First-price auctions are widely used in practice, partly because they constitute the most natural and simple type of auctions. In particular, they have been largely adopted in the field of programmatic advertising, where they have progressively replaced second-price auctions (Sluis., 2017; Slefo., 2019). This recent transition took place for various reasons. First, whereas second-price auctions have the advantage of being dominant-strategy incentive-compatible and hence allow for simple bidding strategies (Vickrey, 1961), they were made obsolete by the widespread use of *header bidding*, a technology that puts different ad-exchange plat-

forms in competition. With this technology, every participating ad-exchange has to provide the winning bid of the auction organized on its platform; a second-level auction is then organized between all the winners to determine which bidder earns the right of displaying its banner. Second price auctions would hence jeopardize the fairness of the attribution of the placement at sale with header bidding. Second, sellers have benefited from the transition, since many bidders continued to bid as in second-price auctions and despite the automated implementation of so-called *bid shading* by demand-side platforms, meant to adjust their bids to this new situation (Sluis., 2019). The transition to first price auctions raises questions for advertisers who need new bidding strategies. In general, bidders participating in auctions in the context of programmatic advertising do not know the bidding strategies of the other contestants in advance, or anything about the valuations that other bidders attribute to the advertisement slot. Not only do they have to learn other bidders' behavior on the go, but they also need to understand how valuable the placement is for their own use (how many clicks or actions the display of their ad on this placement will lead to), which is usually not the same for all bidders.

In this work, we model the problem faced by a single bidder in repeated stochastic first-price auctions, that is, when the contestants' bids are drawn from a stationary distribution. We consider that the learner's bids will not influence the others' bidding strategies. This approximation is sensible in contexts where the major part of the stakeholders do not have an elaborate bidding strategy. More precisely, many stakeholders never modify their bids or do so at a very low frequency. Moreover, the poll of bidders is very large and each bidder only participates in a fraction of the auctions, which argues in favor of the assumption that the influence of one bidder on the rest of the participants can be neglected.

Model We consider that similar items are sold in T sequential first price auctions. For $t = 1, \dots, T$, the auction mechanism unfolds in the following way. First, the bidder submits her bid B_t for the item that is of unknown value V_t. The other players submit their bids, the maximum of which is called M_t. If $M_t \leq B_t$ (which includes the case of ties), the bidder observes and receives V_t and pays B_t. If $B_t < M_t$, the bidder loses the auction and does not observe V_t.

We make the following additional assumptions: $\{V_t\}_{t\geq 1}$ are independent and identically distributed random variables in the unit interval $[0, 1]$; their expectation is denoted by $v := \mathbb{E}(V_t)$. The $\{M_t\}_{t\geq 1}$ are independent and identically distributed random variables in the unit interval $[0, 1]$ with a cumulative distribution function (CDF) F, independent from the $\{V_t\}_{t\geq 1}$. When applicable, we denote by $f = F'$ the associated probability density function.

Due to the stochastic nature of the setting, we study the first-price utility of the bidder: $U_{v,F}(b) := \mathbb{E}\big[(V_t - b)\mathbb{1}\{M_t \leq b\}\big] = (v - b)F(b)$. The (pseudo-)regret is defined as

$$R_T^{v,F} = T \max_{b\in[0,1]} U_{v,F}(b) - \sum_{t=1}^{T} \mathbb{E}[U_{v,F}(B_t)] \, .$$

We denote by $b^*_{v,F} = \max\left\{ \arg\max_{b\in[0,1]} U_{v,f}(b) \right\}$ the (highest) optimal bid. In the rest of the paper, we will abuse notation and speak about regret although rigorously this quantity should be termed pseudo-regret. Note that the outer max is required as the utility may have multiple maxima (see Section 2 below): in that case, we define the optimal bid as

the one that has the largest winning rate. In the sequel, we exclude the particular case where $F(b^*_{v,F}) = 0$, since in this hopeless situation the contestants always bid above the value of the item and the best strategy is not to bid at all ($B_t \equiv 0$): we thus assume that $F(b^*_{v,F}) > 0$.

In Section 3, we will first assume that F is known to the learner. This setting bears some similarities with the case of second-price auctions considered by (Weed et al., 2016; Achddou et al., 2021): the truthfulness of second-price auctions makes it sufficient for the bidder to learn the value of v and the valuation of the item is the only parameter to estimate in that case. However, an important feature of the second-price auction mechanism is that the utility of the bidder is quadratic in v under very mild assumptions on the bidding distribution F. In the case of first-price auctions, the utility is no longer guaranteed to be unimodal, neither is the optimal bid $b^*_{v,F}$ a regular function of v.

We treat the case, in Section 4, where the CDF F of the opponents' maximal bid is initially unknown to the learner, assuming that the maximal bid M_t is observed for each auction. Note that in this more realistic setting, the bidder could not infer the optimal bid $b^*_{v,F}$ even if she had perfect knowledge of the item value v. The bidder consequently needs to estimate F and v simultaneously, which makes it a clearly harder task. This second setting bears some similarities with the task of fixing a price in the posted price problem (Huang et al., 2018; Kleinberg and Leighton, 2003; Bubeck et al., 2017; Cesa-Bianchi et al., 2019), in which a seller needs to estimate the distribution of the valuations of buyers, in order to set the optimal price in terms of her revenue. However, in contrast to the posted-price setting, there is an additional unknown parameter v that also impacts the utility function.

In both of these settings, the learner is faced with a structured continuously-armed bandit problem with censored feedback. Indeed, the bidder only observes the reward associated with the chosen bid, but she observes the value only when she wins. This introduces a specific exploitation/exploration dilemma, where exploitation is achieved by bidding close to one of the optimal bids but exploration requires that the bids are not set too low. This structure seems to call for algorithms that bid above the optimal bid with high probability, as in (Weed et al., 2016; Achddou et al., 2021) for the second-price case, but we will see in the following that it is not necessarily true.

Related Works A major line of research in the field of online learning in repeated auctions is devoted to fixing a reserve price for second-price auctions or a selling price in posted price auctions, see (Nedelec et al., 2020) for a general survey. In the posted price setting, arbitrarily bad distributions of bids give rise to very hard optimization problems (Roughgarden and Schrijvers, 2016). That is why regularity assumptions are often used, like e.g. the *monotonic hazard rate* (MHR) condition. Most notably, Huang et al. (2018); Cole and Roughgarden (2014); Dhangwatnotai et al. (2015) use this assumption to bound the sample complexity of finding the monopoly price. Regarding online learning in the posted price setting, Kleinberg and Leighton (2003) and Cesa-Bianchi et al. (2019) introduce algorithms for the stochastic case, respectively in the cases where the distribution of the prices are continuous and discrete. Bubeck et al. (2017) study the adversarial counterpart. Blum et al. (2004); Cesa-Bianchi et al. (2014) study online strategies that aim at setting the optimal reserve price in second-price auctions while learning the distribution of the buyer's bids. Cesa-Bianchi et al. (2014) assume that bidders are symmetric, but that the bids distribution

is not necessarily MHR. They introduce an optimistic algorithm based on two ideas. Firstly they observe that exploitation is achieved by submitting a price smaller than the optimal reserve price, and secondly they use the fact that the utility can be bounded in infinite norm, thanks to the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality (Massart, 1990).

The problem of learning in repeated auctions from the point of view of the buyer was originally addressed in the setting of second-price auctions. For the stochastic setting, Weed et al. (2016) propose an algorithm that overbids with high probability, and that is shown to have a regret of the order of $\log^2 T$ under mild assumptions on the distribution of the bids. They also provide algorithms for the adversarial case, that have a regret scaling in $\sqrt{T}$. Achddou et al. (2021) extend their work by proposing tighter optimistic strategies that show better worst case performances. They also analyze non-overbidding strategies, proving that such strategies can perform well on a large class of second-price auctions instances. Flajolet and Jaillet (2017) consider the contextual set-up where the value associated to an item is linear with respect to a context vector associated to the item, and revealed before each action.

Learning in repeated stochastic first price auctions is a difficult problem that has given rise to a number of very different though equally interesting modelizations. Feng et al. (2020) consider auctions in which the values of all the bidders are revealed as a context before each turn, proving that the bids of bidders who use no regret contextual learning strategies in first price auctions converge to Bayes Nash equilibria. Han et al. (2020) also consider the case where the values are assumed to be revealed as an element of context before each auction takes place and the highest bid among others' bids is only shown to the learner when she loses. This setting interestingly introduces a censoring structure that is opposed to the one we consider: in this context, exploitation is achieved by not bidding too high. Han et al. (2020) provide new algorithms for this setting which have a regret of the order of $\sqrt{T}$. A setting somewhat closer to ours is studied by Feng et al. (2018). This work deals with the setting of a bid in an adversarial fashion, when the other bids are revealed at each time step and the value is revealed only upon winning an auction. However the proposed algorithm is based on a discretization of the bidding space which relies on the prior knowledge of the smallest gap between two distinct bids. With this knowledge, the proposed algorithm achieves an adversarial regret of the order of $\sqrt{T}$.

Contributions The highlights of Sections 2–4 are the following. In Section 2 we stress the hardness of the first-price bid optimization task, showing that in general it necessarily leads to high minimax regret rates. We however transplant ideas introduced in the case of posted prices to exhibit natural assumptions ensuring that the first-price utility is smooth, paving the way for faster learning. In Section 3, we consider the case where the learner can assume knowledge of F and propose a new UCB-type algorithm called UCBid1 for learning the optimal bid with low regret. UCBid1 is adaptive to the difficulty of the problem in the sense that its regret is $O(\sqrt{T})$ in difficult cases, but comes down to $O(\log^2 T)$ when the first-price utility is smooth. We also provide lower-bound results suggesting that these rates are nearly optimal. In Section 4, we consider the more general setting where F is initially unknown to the learner. By leveraging the structure of the first-price bidding problem, we are able to propose an algorithm, termed UCBid1+, which is a direct generalization of UCBid1. Interestingly, this algorithm is not optimistic anymore: it does not submit bids

which are with high probability above the (unknown) optimal bid. However, it can still be proved to achieve a regret rate of $O(\sqrt{T})$ in the most general case and, more importantly, a regret rate upper bounded by $O(T^{1/3+\epsilon})$ for every $\epsilon > 0$ when the first-price utility satisfies the regularity assumptions mentioned in Section 2. The latter result relies on an original proof notably based on the use of a local concentration inequality on the empirical CDF. All the proofs corresponding to these three sections are presented in appendix. Section 5 closes the paper with numerical simulations where we compare the proposed algorithms with continuously-armed bandit strategies and tailored strategies from the literature, both using simulated and real-world data.

2. Properties of Stochastic First-Price auctions

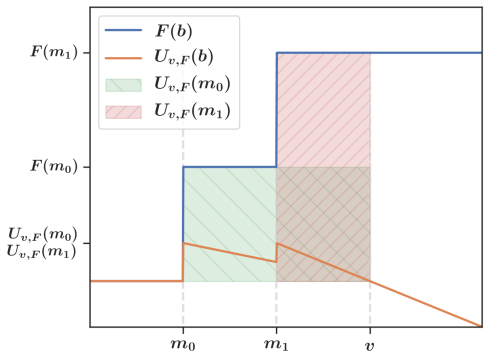


Figure 1: An example with two maximizers

There are two important difficulties with first price auctions. The first one lies in the fact that the utility can have multiple maximizers (or multiple modes with arbitrarily close values) and thus lead to arbitrarily hard optimization problems. To illustrate this, we provide in Figure 1 an example of value v and discrete distribution, supported on two values m_0, m_1, that leads to a utility having two global maximizers. Note that the utility $U_{v,F}(b)$ is the area of the rectangle with vertices $(b, F(b)), (b, 0), (v, F(b)), (v, 0)$. This observation makes it easy to build examples with multiple maxima. Discrete examples like the one in Figure 1 are intuitive because the utility is decreasing between two successive points of the support, but there also exist similar cases with continuous distributions (see for example Appendix A.3). This example also shows that there exist combinations of bids distributions and values for which the utility is not regular around its maximum.

The second difficulty comes from the fact that the mapping from v to the largest maximizer, $\psi_F : v \mapsto b^*_{v,F}$ may also lack regularity. Indeed, keeping the distribution in Figure 1 but setting the value to $v' = v + \Delta$, with a positive Δ (resp. to $v' = -\Delta$) yields that the set of maximizers is $\{m_1\}$ (resp. $\{m_0\}$). Even though ψ_F can not be proved to be regular in all generality, it always holds that ψ_F is increasing. This is intuitive: the optimal bid grows with the private valuation.

Lemma 1 *For any cumulative distribution F, $\psi_F : v \mapsto b^*_{v,F}$ is non decreasing.*

The two aforementioned difficulties contribute to making the problem at hand particularly hard. In the following theorem, we show that any algorithm is bound to have a worst case regret growing at least like $\sqrt{T}$.

Theorem 2 *Let $\mathcal{C}$ denote the class of cumulative distribution functions on $[0,1]$. Any strategy, whether it assumes knowledge of F or not, must satisfy*

$$\liminf_{T\to\infty} \frac{\max_{v\in[0,1],F\in\mathcal{C}} R_T^{v,F}}{\sqrt{T}} \geq \frac{1}{64},$$

Theorem 2 corresponds to Theorem 6 in Han et al. (2020). For completeness, we prove it in Appendix B. The proof relies on specifically hard instances of CDF that are perturbations of the example of Figure 1. It illustrates the complexity of bidding in first-price auctions, when F and v are arbitrary. This complexity stems from specifically hard instances of F and v. We present a natural assumption that avoids these pathological cases.

Assumption 1 *F is continuously differentiable and is strictly log-concave.*

This assumption is reminiscent of the monotonic hazard rate (MHR) condition (see e.g. Cole and Roughgarden (2014)), that appears in the analysis of the posted price problem. While MHR requires $f/(1-F)$ to be increasing, Assumption 1 requires f/F to be decreasing. In particular, this condition is satisfied by truncated exponentials and Beta distributions with f of the form $Cx^{\alpha-1}$ where $\alpha > 1$ or $C(1-x)^{\beta-1}$ where $\beta > 1$, or Beta distributions in which $\alpha + \beta < \alpha\beta$ (see Lemma 15 in Appendix A). Assumption 1 plays roughly the same role for first price auctions than MHR for the posted price setting. It guarantees in particular that there is a unique optimal bid. Note that if F satisfies Assumption 1, F is increasing, and admits an inverse which we denote by F^{-1}.

Lemma 3 *Under Assumption 1, for any $v \in [0,1]$ the mapping $b \mapsto U_{v,F}(b)$ has a unique maximizer.*

As does the MHR assumption for the posted-prices setting, Assumption 1 ensures that the utility is strictly concave *when expressed as a function of the quantile $q = F(b)$ associated with the bid b*. Another important consequence of Assumption 1 is that the mapping from v to the optimal bid $b^*_{v,F}$ is guaranteed to be regular.

Lemma 4 *If Assumption 1 is satisfied and f is continuously differentiable, then $\psi_F : v \mapsto b^*_{v,F}$ is Lipschitz continuous with a Lipschitz constant 1.*

Indeed, if f is continuously differentiable and if f does not vanish on $[0,1[$ (which is implied by Assumption 1), ψ_F is invertible and it inverse ϕ_F writes $\phi_F : b \mapsto b + F(b)/f(b)$. Assumption 1 ensures that ϕ_F admits a derivative that is lower-bounded by $\phi'_F(b) > 1$.

Assumption 1 also implies the important property that the probability of winning the auction at the optimal bid $F(b^*_{v,F})$ cannot be arbitrarily small when compared to $F(v)$.

Lemma 5 *If Assumption 1 is satisfied, then*

$$F(b^*_{v,F}) \geq \frac{F(v)}{e}\ .$$

We conclude this section by additional properties that are essential for obtaining low regret rates: the utility is second-order regular, when expressed as a function of the quantiles. Let $W_{v,F}$ denote the utility expressed as a function of the quantile, $W_{v,F} : q \mapsto U_{v,F}(F^{-1}(q))$, and let $q^*_{v,F} := F(b^*_{v,F})$ be its maximizer. Under Assumption 1, the deviations of $W_{v,F}$ from its maximum are lower-bounded by a quadratic function.

Lemma 6 *Under Assumption 1, for any $q \in [0,1]$,*

$$W_{v,F}(q^*_{v,F}) - W_{v,F}(q) \geq \frac{1}{4}(q^*_{v,F} - q)^2 W_{v,F}(q^*_{v,F}).$$

This property relies, among other arguments, on the observation that

$$W'_{v,F}(q) = v - \phi_F(F^{-1}(q)) = \phi_F(F^{-1}(q^*_{v,F})) - \phi_F(F^{-1}(q))$$

and that ϕ'_F is lower-bounded by 1 under Assumption 1 (see discussion of Lemma 4 above). Similarly, in order to obtain a quadratic lower bound on $W_{v,F}(q)$, one needs to show that ϕ'_F may be upper bounded. This is the purpose of the following regularity assumption.

Assumption 2 *F admits a density f such that $c_f < f(b) < C_f, \forall b \in [b^*_{v,F} - \Delta, b^*_{v,F} + \Delta]$ and $\phi_F : b \mapsto b + F(b)/f(b)$ admits a derivative that is upper-bounded by a constant $\lambda \in \mathbb{R}^+$ on $[b^*_{v,F}, b^*_{v,F} + \Delta]$.*

Assumption 2 holds, in particular, when F is twice differentiable, f is lower-bounded by a positive constant and f' is upper-bounded by a positive constant on a neighborhood of $b^*_{v,F}$. Note that in the field of auction theory, it is common to assume that the utility is approximately quadratic around the maximum, which is a far stronger assumption, as stated in (Nedelec et al., 2020) (see (Kleinberg and Leighton, 2003) for example). Assumption 2 implies the following lower bound for the utility expressed as a function of the quantiles.

Lemma 7 *Under Assumption 2, for any $q \in [q^*_{v,F}, q^*_{v,F} + C_f\Delta]$,*

$$W_{v,F}(q^*_{v,F}) - W_{v,F}(q) \leq \frac{1}{c_f}\lambda(q^*_{v,F} - q)^2.$$

3. Known Bid Distribution

In this section we address the online learning task in the setting where the bid distribution F is known to the learner from the start. In order to set the bid B_t at time t, the available information consists in $N_t := \sum_{s=1}^{t-1} \mathbb{1}\{M_s \leq B_s\}$, the number of observed values before time t, and $\hat{V}_t := \frac{1}{N_t}\sum_{s=1}^{t-1} V_s \mathbb{1}(M_s \leq B_s)$ the average of those values. Let $\epsilon_t := \sqrt{\gamma \log(t-1)/2N_t}$ denote a confidence bonus depending on a parameter $\gamma > 0$ to be specified below.

Algorithm 1 (UCBid1) *Initially set $B_1 = 1$ and, for $t \geq 2$, bid according to*

$$B_t = \max\left\{ \operatorname*{arg\,max}_{b \in [0,1]} (\hat{V}_t + \epsilon_t - b)F(b) \right\}.$$

This algorithm, strongly inspired by UCB-like methods designed for second-price auctions by Weed et al. (2016); Achddou et al. (2021), is a natural approach to first-price auctions. The idea behind this kind of method is that one should rather overestimate the optimal bid, so as to guarantee a sufficient rate of observation. As an UCB-like algorithm, UCBid1 submits an (high probability) upper bound $\psi_F(\hat{V}_t + \epsilon_t)$ of $b^*_{v,F}$, thanks to Lemma 1 and since ψ_F is non decreasing. In practice, the algorithm requires a line search at each step as the utility maximization task is usually non-trivial, as discussed in Section 1.

In the most general case, the regret of UCBid1 admits an upper bound of the order of $\sqrt{T\log(T)}$.

Theorem 8 *When $\gamma > 1$, the regret of UCBid1 is upper-bounded as*

$$R_T^{v,F} \leq \frac{\sqrt{2\gamma}}{F(b^*_{v,F})}\sqrt{T\log T} + O(\log T)\ .$$

Note that $\sqrt{T}$ is the order of the regret of UCB strategies designed for second-price auctions in the absence of regularity assumptions on F (Weed et al., 2016). However, under the regularity assumptions introduced in Section 2, it is possible to achieve faster learning rates.

Theorem 9 *If F satisfies Assumption 1 and 2, then, for any $\gamma > 1$,*

$$R_T^{v,F} \leq \frac{2\gamma\lambda C_f^2}{F(b^*_{v,F})c_f}\log^2(T) + O(\log T).$$

The $\log^2(T)$ rate of the regret comes from the Lipschitz nature of ψ_F, that makes it possible to bound the gap $B_t - b^*_{v,F}$, and from the obervation that the utility is quadratic around its optimum. This explains the similarity with the order of the regret of UCBID in (Weed et al., 2016), when the distribution of the bids admits a bounded density. Indeed, in second-price-auctions, when the distribution of the bids admits a bounded density, the utility is locally quadratic around its maximum and the equivalent of ψ_F is the identity, meaning that the optimal bid is just the value v of the item. The presence of the multiplicative constant $1/F(b^*_{v,F})$ is also expected: it is the average time between two successive observations under the optimal policy. This similarity between the structures of second and first price auctions under Assumptions 1 and 2 also suggest that the constants in the regret may be further improved by using a tighter confidence interval for v based on Kullback-Leibler divergence, proceeding as in (Achddou et al., 2021).

Under Assumption 1, the regret of any optimistic strategy can be shown to satisfy the following lower bound.

Theorem 10 *Consider all environments where V_t follows a Bernoulli distribution with expectation v and F satisfies Assumption 1 and is such that $\phi' \leq \lambda$, and there exists c_f and C_f such that $0 < c_f < f(b) < C_f$, $\forall b \in [0,1]$. If a strategy is such that, for all such environments, $R_T^{v,F} \leq O(T^a)$, for all $a > 0$, and there exists $\gamma > 0$ such that $\mathbb{P}(B_t < b^*) < t^{-\gamma}$, then this strategy must satisfy:*

$$\liminf_{T\to\infty} \frac{R_T^{v,F}}{\log T} \geq c_f^2\lambda^2\left(\frac{v(1-v)(v-b^*_{v,F})}{32}\right).$$

The first assumption, $R_T^{v,F} \leq O(T^a)$, is a common consistency constraint that is used when proving the lower bound of Lai and Robbins (1985) in the well-established theory of multi-armed bandits. The second assumption, $\mathbb{P}(B_t < v) < t^{-\gamma}$, restricts the validity of the lower bound to the class of strategies that overbid with high probability. By construction, this assumption is satisfied for UCBid1.

Note that there is a gap between the rates $\log T$ in the lower bound (Theorem 10) and $\log^2 T$ in the performance bound of UCBid1 (Theorem 9), which we believe is mostly due to the mathematical difficulty of the analysis. The $v(1-v)$ factor may be interpreted as an upper bound on the variance of the value distribution with expectation v. Theorem 10 displays a dependence on v of the order of v^2 when v tends to 0. However this has to be put in perspective with the fact that the value of the optimal utility $U_{v,F}(b^*_{v,F})$ is also quadratic in v, when v tends to zero under the assumptions of Theorem 10 (from Lemma 6).

4. Unknown Bid Distribution

We now turn to the more realistic, but harder, setting where both the parameter v and and the function F need to be estimated simultaneously. For this setting, we propose the following algorithm, which is a natural adaptation of UCBid1, simply plugging in the empirical CDF in place of the unknown F.

It may come as a surprise that we do not add any optimistic bonus to the estimate $\hat{F}_t$: it is not necessary to be optimistic about F since the observation M_t drawn according to F is observed at each time step whatever the bid submitted.

Algorithm 2 (UCBid1+) *Submit a bid equal to* 1 *in the first round, then bid:*

$$B_t = \max\Big\{ \arg\max_{b\in[0,1]} (\hat{V}_t + \epsilon_t - b)\hat{F}_t(b) \Big\},$$

where $\hat{F}_t(b) := \frac{1}{t-1}\sum_{s=1}^{t-1} \mathbb{1}\{M_s < b\}$ *and* $\epsilon_t := \sqrt{\gamma \log(t-1)/2N_t}$.

Although B_t produced by Algorithm 2 could, in principle, be arbitrarily small, it is possible to show that there is no extinction of the observation process. Indeed, after a time that only depends on v and F, $F(B_t)$ is guaranteed to be higher than a strictly positive fraction of $F(b^*_{v,F})$ with high probability (see Lemma 28 in Appendix E). This result implies that the number of successful auctions N_t asymptotically grows at a linear rate (with high probability), making it possible to bound the expected difference between $\hat{V}_t + \epsilon_t$ and v. Combined with the DKW inequality (Massart, 1990), this allows to bound the difference between the utility and $(\hat{V}_t + \epsilon_t - b)\hat{F}_t(b)$ in infinite norm and hence the difference between B_t and $b^*_{v,F}$. Putting all the pieces together (see the complete proof in Appendix E) yields the following upper bound on the regret of UCBid1+.

Theorem 11 *UCBid1+ incurs a regret bounded by*

$$R_T^{v,F} \leq 12\sqrt{\frac{\gamma v}{U_{v,F}(b^*_{v,F})}}\sqrt{T\log T} + O(\log T),$$

provided that $\gamma > 2$.

Note that computing the bid B_t for UCBid1+ is easy, as $(\hat{V}_t+\epsilon_t-b)\hat{F}_t(b)$ necessarily lies among the observed bids because this function is linearly decreasing between observed bids. More precisely, $(\hat{V}_t+\epsilon_t-b)\hat{F}_t(b)) = \hat{F}_t(M^{(i)})(\hat{V}_t+\epsilon_t-b)$, for $b \in [M^{(i)}, M^{(i+1)}[$, where $M^{(i)}$ is the i-th order statistic of the observed bids (obtained by sorting the bids in ascending order). However, as there is no obvious way to update B_t sequentially, this results in a complexity of UCBid1+ that grows quadratically with the time horizon T.

The proof of Theorem 11 relies on the DKW inequality to bound the difference between B_t and b^*. This happens to be very conservative and a little misleading in practice. Indeed, what really matters is the local behavior of the empirical utility, and hence, of $\hat{F}_t$ around b^*. As illustrated by Figure 2, locally, $\hat{F}_t$ is roughly a translation of F plus a negligible perturbation which can be bounded in infinite norm. This intuition is formalized in Lemma 12, a localized version of the DKW inequality. The fact that $\hat{F}_t$ is locally almost parallel to F imposes a constraint on B_t that may be used to bound its distance from b^*, yielding an improved regret rate under Assumptions 1 and 2, as shown by Theorem 13.

Lemma 12 *For any $a, b \in [0,1]$, if F is increasing,*

$$\sup_{a\le x\le b} |\hat{F}_t(x) - F(x) - (\hat{F}_t(a) - F(a))| \le \sqrt{\frac{2(F(b)-F(a))\log\left(\frac{e\sqrt{t}}{\eta\sqrt{2(F(b)-F(a))}}\right)}{t}} + \frac{\log(\frac{t}{2(F(b)-F(a)\eta^2})}{6t},$$

with probability $1-\eta$.

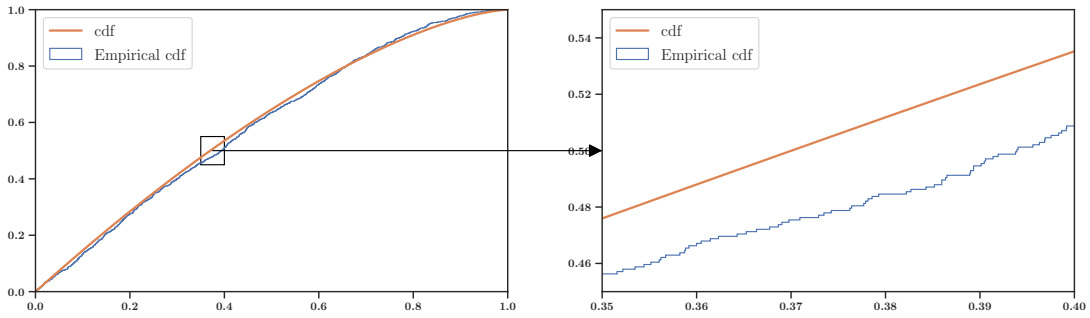


Figure 2: Local behavior of the empirical CDF

Theorem 13 *If F satisfies Assumptions 1 and 2, UCBid1+ incurs a regret bounded by*

$$R_T^{v,F} \le O\big(T^{1/3+\epsilon}\big),$$

for any $\epsilon > 0$, provided that $\gamma > 2$.

UCBid1+ thus retains the adaptivity of UCBid1. In general, its regret is of the order of $\sqrt{T}$ (omitting logarithmic terms), matching the lower bound of Theorem 2. But it is reduced to $T^{1/3+\epsilon}$, for any $\epsilon > 0$, in the smooth case defined by Assumptions 1 and 2. In practice, the improvement over other $\sqrt{T}$-regret algorithms is huge, as shown in the next section.

5. Numerical simulations

5.1. Benchmark Algorithms

Methods pertaining to black box optimization. Sequential black box optimization algorithms, also known as continuously-armed bandits (Kleinberg et al., 2008; Bubeck et al., 2011; Munos, 2011; Valko et al., 2013), are algorithms designed to find the optimum of an unknown function by receiving noisy evaluations of that function at points that are chosen sequentially by the learner. They rely on prior assumptions on the smoothness of the unknown function. For first-price bidding, we may consider that the reward $(v-B_t)\mathbb{1}(M_t \leq B_t)$ is a noisy observation of the utility $U_{v,F}(B_t)$, with a noise bounded by 1. Moreover, when F admits a density f and $f(b) < C_f$, then $-1 < U'_{v,F}(b) = (v-x)f(b) - F(b) < C_f$, which implies that $U_{v,F}$ is Lipschitz with constant $\max(1, C_f)$. As a consequence, all black-box optimization algorithms that consider an objective function with Lispchitz regularity may be used for learning in stochastic first price auctions. HOO (Bubeck et al., 2011) has a parameter ρ related to the level of smoothness of the objective function which we can set to $1/2$, corresponding to the observation that the first-price utility is Lipschitz under the assumptions discussed above. This immediately leads to a first baseline approach with $O(\sqrt{T \log T})$ regret rate. Setting the parameter related to the Lipschitz constant of HOO so that it is larger than C_f is not possible in practice without prior knowledge on F. More generally, knowing the smoothness is considered a challenge most of the time in black-box optimization, so that several methods have been introduced that are adaptive to the smoothness, e.g. stoSOO (Valko et al., 2013).

UCB on a smartly chosen discretization. Combes and Proutiere (2014) prove that when the reward function is unimodal, a discretization based on the smoothness level of this function suffices to achieve a regret of the order of $\sqrt{T}$. If F satisfies Assumption 1, $U_{v,F}$ is unimodal, as shown by the proof of Lemma 3. Hence, using the right discretization while applying UCB, one can achieve a $O(\sqrt{T})$ regret. In particular if the utility is quadratic, the advised discretization is a grid of $O(T^{1/4})$ values.

O-UCBID1. We also implement the following algorithm, that is reminiscent of the method used by (Cesa-Bianchi et al., 2014) to learn reserve prices.

Algorithm 3 (O-UCBid1) *Submit a bid equal to* 1 *in the first round, then bid:*

$$B_t = \max\{b \in [0, \hat{V}_t + \epsilon_t], \hat{U}_t(b) \geq \max_{b \in [0,1]} \hat{U}_t(b) - 2\epsilon_t\},$$

where $\hat{U}_t(b) = (\hat{V}_t - b)\hat{F}_t(b)$.

This algorithm overbids with high probability, by construction. Thanks to the DKW inequality, one can control the difference between the true bid distribution F and its empirical version $\hat{F}_t$ in infinite norm. Because we observe M_t at each round, $\|F - \hat{F}_t\|_\infty$ is at most ϵ_t with high probability. It is easy to show that $\|U_{v,F} - \hat{U}_t\|_\infty$ is bounded by a multiple of ϵ_t showing that B_t is (again with high probability) larger than the unknown optimal bid $b^*_{v,F}$. O-UCBid1 is very close to the method used by (Cesa-Bianchi et al., 2014) to set a reserve price in second-price auctions. While in first-price auctions, a bidder needs to overbid in

order to favor exploration, sellers in second-price auctions are encouraged to offer a lower price than the optimal one, as they can only observe the second highest bid if their reserve price is set lower than the latter. The approach of Cesa-Bianchi et al. (2014) requires successive stages as sellers in second-price auctions can only observe the second-price and need to estimate the distribution of all bids based on this information. In our setting, we have direct access to the opponents' highest bid and successive stages are not required any longer. We prove that the regret incurred by O-UCBid1 is of the order of $\log T\sqrt{T}$ when $\gamma > 1$, which makes it an interesting baseline algorithm, that has guarantees similar to those of black box optimization algorithm, without the need of knowing the smoothness or the horizon. We refer to Theorem 23 in Appendix E for further details.

Methods for discrete distributions We run UCBid1+ on discrete examples. In this case, we compare it to UCB on a discretization of $[0, 1]$ and to WinExp, a generalization of Exp3 for the problem of learning to bid (Feng et al., 2018).

5.2. Experiments On Simulated Data

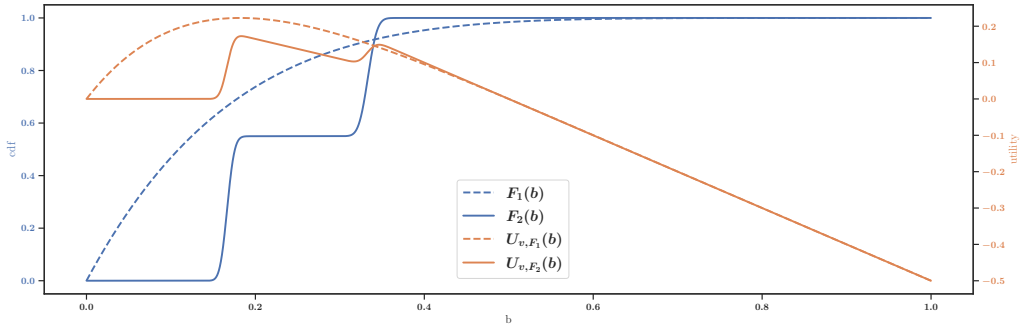
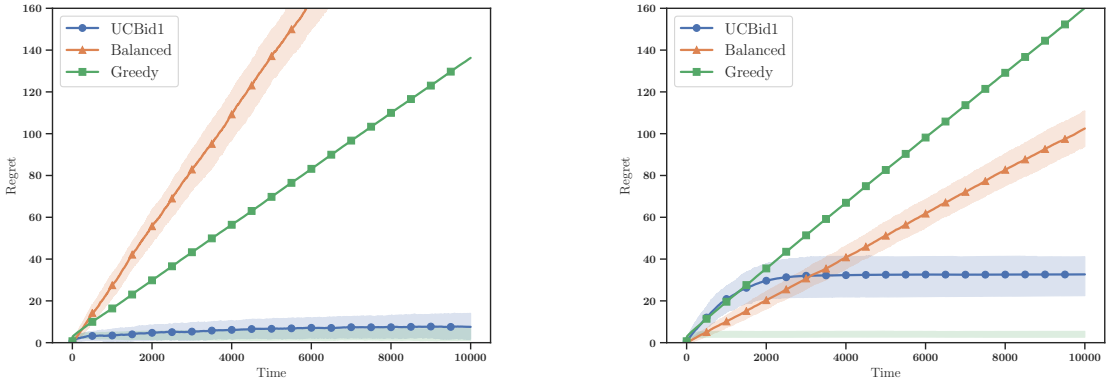


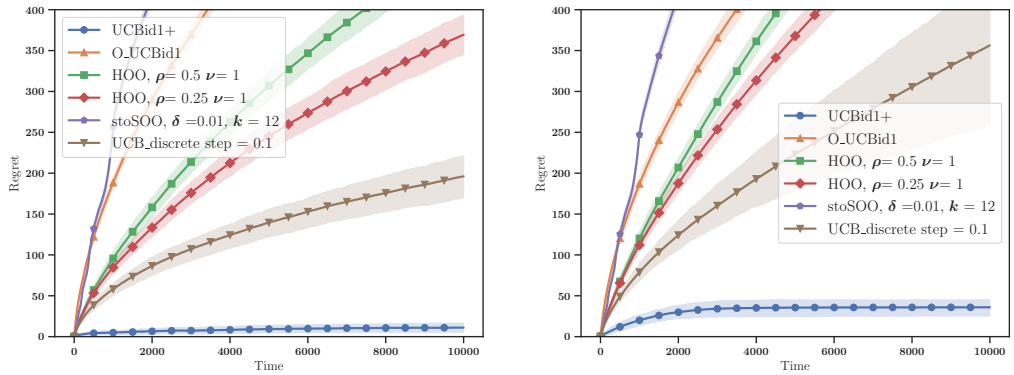
Figure 3: Two choices of F; associated utilities for $v = 1/2$.



(a) Regret plots under the first instance of the problem

(b) Regret plots under the second instance of the problem

Figure 4: Regret plots for known F



(*a*) Regret plots under the first instance of the problem

(*b*) Regret plots under the second instance of the problem

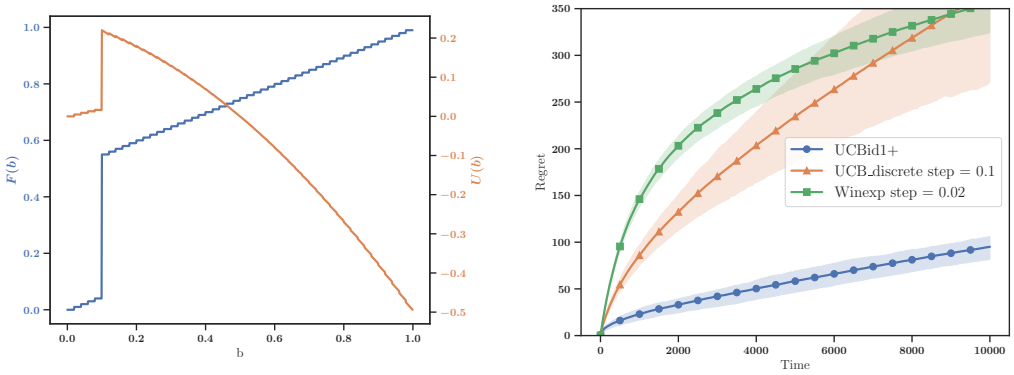
Figure 5: Regret plots for unknown F

In this section we focus on two particular instances of the first price auction learning problem. The first instance is characterized by a value distribution set to a Bernoulli distribution of average 0.5, and a distribution of the highest contestants' bids set to a Beta(1,6). The second instance only differs by the distribution of the highest contestants' bids, which is set to a mixture of two Beta distributions: $0.55 \times Beta(500, 2500) + 0.45 \times Beta(1000, 2000)$. This distribution is very close to that used in the proof of Theorem 2, but is continuous. The cumulative distribution and the matching utility of each instance are plotted on Figure 3. Both distributions are smooth but the first one satisfies Assumption 1, while it is not clear that the second one does.

Figures 4(*a*)subfigure and 4(*b*)subfigure show the regret of various strategies when F is known. The first (respectively second) figure represents the regrets of these strategies under the first (respectively second) instance of the problem described above. The horizon is set to 10000 and the results of 720 Monte Carlo trials are aggregated. The plots represent the average regret over time (shaded areas correspond to the interquartile range). The strategy termed Greedy is a naive strategy that bids $\max \arg\max \hat{U}_t(b)$, whenever it has made more than three observations. It shows a linear regret, which comes from the fact that when it only observes value samples equal to zero during the first three observations, it bids 0 indefinitely, and thus incurs the regret $U_{v,F}(b^*_{v,F}) - U_{v,F}(0)$ at each time step. Observing only 0 three times in a row is not very likely: the third quartile is very small, but the consequences are so terrible that the average is many orders of magnitude higher. The strategy termed Balanced consists in bidding the median of the highest contestants' bids. It guarantees that the learner is able to win half of the rounds. As expected, this strategy, which does not adapt to the instance at hand, shows poor performances in both cases. However, it is a better solution than bidding 0 or 1. Finally, we also plot the regret of UCBid1. Note that in order to implement UCBid1 we would have to compute $\arg\max_{b\in[0,1]}(\hat{V}_t + \epsilon_t - b)F(b)$ at each round; instead we only use an approximation of this quantity by computing the argmax of the function over a grid of 10000 values. UCBid1 outperforms the naive baseline strategies in both cases. Under the more complex second

instance of the problem, it shows a larger regret than under the first one. However, even in this more complex case, the rate of growth of the regret stays very low.

In Figure 5, we analyze the regrets of different algorithms when F is unknown. In this setting, we compare UCB on a discretization of $[0, 1]$ with 10 arms, HOO (Bubeck et al., 2011) with various parameters, O-UCBid1 and UCBid1+ with $\gamma = 1$ and stoSOO (Valko et al., 2013) with the parameters recommended in the latter paper. For efficiency reasons, we also do not allow the tree built by HOO and stoSOO to have a depth larger than $\log_2 T$. The various versions of HOO, UCB, as well as stoSOO show regret plots that could correspond to a $\sqrt{T}$ behavior. UCBid1+ shows a dramatically improved regret plot compared to the black box optimization strategies.



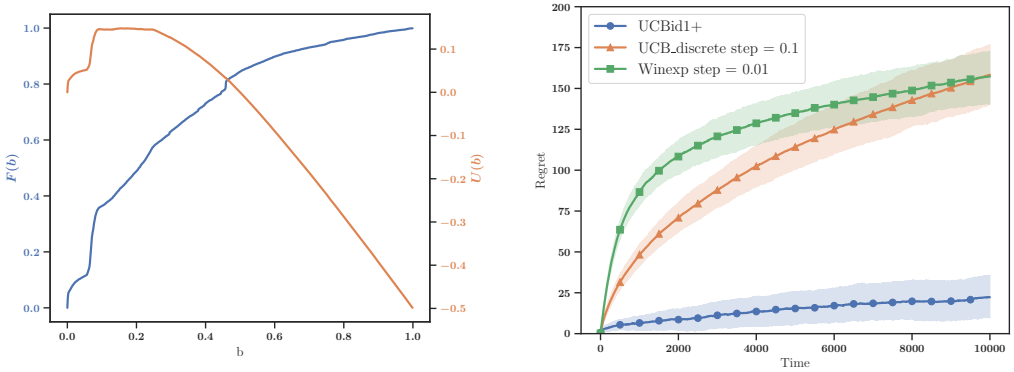
(a) Utility with $v = 1/2$ (b) Regret plots

Figure 6: An example with discrete bids

Figure 6 shows a different example where the distribution of bids is discrete with a probability mass of 0.51 on 0.1 and equal probability masses on $i/50, \forall i \in [1 \dots 4, 6, \dots, 50]$. We compare UCBid1+ with UCB, having operated a discretization into 10 arms and with Winexp with a discretization into 50 arms. UCBid1+ again yields a regret at least 5 times smaller than the other algorithms. In addition, it is important to stress that UCBid1+ and O-UCBid1 are anytime algorithms, while all the alternatives shown on Figures 5 and 6 require, at least, the knowledge of the time horizon.

5.3. Experiments On a Real Bidding Dataset

We also experiment on a real-world bidding dataset representing the highest bids from the contestants of one advertiser on a certain campaign. Thanks to Numberly, a media trading agency, Adverline, an advertising network, and Xandr, a supply and demand-side platform, we collected a set of 56607 bids that were made on a specific placement on Adverline's inventory on auctions that Numberly participated to, for a specific campaign. We keep only the bids smaller than the 90% quantile and we normalize them to get data between 0 and 1 (see Figure 10 in Appendix F for a histogram). The regret plots are represented in Figure 7(b)subfigure. As earlier, with discrete simulated data, we compare UCBid1+ with UCB, having operated a discretization into 10 arms and with Winexp with a discretization into 100 arms. Unsurprisingly, the regret plots are similar to those with simulated data,



(*a*) Utility with $v = 1/2$ (*b*) Regret plots with bidding data

Figure 7: Experiment with real bidding data

since the distributions at hand are similar. UCBid1+ still largely outperforms the baseline algorithms.

Acknowledgments

We would like to thank Adverline for accepting to provide us with the bidding data on their inventories and Xandr for making this data transaction possible. We are very grateful to them for their support on this project. Aurélien Garivier acknowledges the support of the Project IDEXLYON of the University of Lyon, in the framework of the Programme Investissements d'Avenir (ANR-16-IDEX-0005), and Chaire SeqALO (ANR-20-CHIA-0020).

References

Juliette Achddou, Olivier Cappé, and Aurélien Garivier. Efficient algorithms for stochastic repeated second-price auctions. In *Algorithmic Learning Theory*, pages 99–150. PMLR, 2021.

A. Blum, V. Kumar, A. Rudra, and F. Wu. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.

S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011.

S. Bubeck, N. Devanur, Z. Huang, and R. Niazadeh. Multi-scale online learning and its applications to online auctions. *arXiv preprint arXiv:1705.09700*, 2017.

O. Cappé, A. Garivier, O. Maillard, R. Munos, G. Stoltz, et al. Kullback–Leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013.

N. Cesa-Bianchi, C. Gentile, and Y. Mansour. Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, 61(1):549–564, 2014.

N. Cesa-Bianchi, T. Cesari, and V. Perchet. Dynamic pricing with finitely many unknown valuations. In *Algorithmic Learning Theory*, pages 247–273. PMLR, 2019.

R. Cole and T. Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 243–252, 2014.

R. Combes and A. Proutiere. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, pages 521–529. PMLR, 2014.

P. Dhangwatnotai, T. Roughgarden, and Q. Yan. Revenue maximization with a single sample. *Games and Economic Behavior*, 91:318–333, 2015.

Z. Feng, C. Podimata, and V. Syrgkanis. Learning to bid without knowing your value. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 505–522, 2018.

Z. Feng, G. Guruganesh, C. Liaw, A. Mehta, and A. Sethi. Convergence analysis of no-regret bidding algorithms in repeated auctions. *arXiv preprint arXiv:2009.06136*, 2020.

A. Flajolet and P. Jaillet. Real-time bidding with side information. In *Advances in Neural Information Processing Systems*, pages 5168–5178, 2017.

A. Garivier, P. Ménard, and G. Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 44(2):377–399, 2019.

Y. Han, Z. Zhou, and T. Weissman. Optimal no-regret learning in repeated first-price auctions. *arXiv preprint arXiv:2003.09795*, 2020.

Z. Huang, Y. Mansour, and T. Roughgarden. Making the most of your samples. *SIAM Journal on Computing*, 47(3):651–674, 2018.

R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

R. Munos. Optimistic optimization of deterministic functions without the knowledge of its smoothness. In *Advances in neural information processing systems*, 2011.

T. Nedelec, C. Calauzènes, N. El Karoui, and V. Perchet. Learning in repeated auctions. *arXiv preprint arXiv:2011.09365*, 2020.

T. Roughgarden and O. Schrijvers. Ironing in the dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 1–18, 2016.

G. Slefo. Google's ad manager will move to first-price auction., 2019. press.

S. Sluis. Big changes coming to auctions, as exchanges roll the dice on first-price., 2017. press.

S. Sluis. Everything you need to know about bid shading., 2019. press.

M. Valko, A. Carpentier, and R. Munos. Stochastic simultaneous optimistic optimization. In *International Conference on Machine Learning*, pages 19–27. PMLR, 2013.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

J. Weed, V. Perchet, and P. Rigollet. Online learning in repeated auctions. In *Conference on Learning Theory*, pages 1562–1583. PMLR, 2016.

Supplementary Material

Outline. We prove in Appendix A all the results pertaining to Section 2 apart from Theorem 2, which is proved separately in Appendix B. In Appendix C, we introduce preliminary results necessary to analyze the regrets of the algorithms presented in main body of the paper. Appendix D contains all the proofs of the results of Section 3, while the theorems of Section 4 are proved in Appendix E. A figure related to Section 5 is presented in Appendix F.

Notation.

- In the following we write U instead of $U_{v,F}$ (respectively W instead of $W_{v,F}$; b^* instead of $b^*_{v,F}$; q^* instead of $q^*_{v,F}$ and R_T instead of $R_T^{v,F}$) when there is no ambiguity.
- $b(q)$ denotes $F^{-1}(q)$.
- $\hat{V}(n) := 1/n \sum_{s=1}^{n} V(s)$ is the mean of the n first observed values.
- We set $V'_s = V_s$ if $M_s \leq B_s$, and $V'_s = \emptyset$ otherwise.
- We set $\mathcal{F}_t = \sigma((M_s, V'_s)_{s\leq t})$ be the σ-algebra generated by the the bid maxima and the values observed up to time t.
- $S_t := (V_t - b^*)\mathbb{1}(M_t < b^*) - (V_t - B_t)\mathbb{1}(M_t < B_t)$ represents the instantaneous regret.

Appendix A. Properties of first-price auctions

A.1. General properties

Lemma 1 *For any cumulative distribution function F, ψ_F is non decreasing.*

Proof Let $0 < v_1 < v_2 < 1$. We have $U_{v_2,F}(b^*_{v_2,F}) - U_{v_2,F}(b^*_{v_1,F}) \geq 0$ and $U_{v_1,F}(b^*_{v_1,F}) - U_{v_1,F}(b^*_{v_2,F}) \geq 0$, by definition of $b^*_{v_1,F}$ and $b^*_{v_2,F}$.

By summing these two inequalities, $U_{v_2,F}(b^*_{v_2,F}) - U_{v_1,F}(b^*_{v_2,F}) - (U_{v_2,F}(b^*_{v_1,F}) - U_{v_1,F}(b^*_{v_1,F})) \geq 0$. Hence

$$(v_2 - v_1)(F(b^*_{v_2,F}) - F(b^*_{v_1,F})) \geq 0.$$

We then prove the result by contradiction, by assuming that $b^*_{v_1,F} > b^*_{v_2,F}$. Then $F(b^*_{v_1,F}) = F(b^*_{v_2,F})$, since F is non decreasing. In this case,

$$U_{v_1,F}(b^*_{v_1,F}) = (v_1 - b^*_{v_1})F(b^*_{v_1,F}) < (v_1 - b^*_{v_2})F(b^*_{v_2,F}) = U_{v_1,F}(b^*_{v_2,F}).$$

This is impossible, since $b^*_{v_1,F}$ is an optimizer of $U_{v_1,F}$. In conclusion, $b^*_{v_1,F} \leq b^*_{v_2,F}$ ■

A.2. Properties under regularity assumptions

Lemma 3 *If Assumption 1 is satisfied, then for any $v \in [0,1]$, $U_{v,F}$ has a unique maximizer.*

Proof If F satisfies Assumption 1 then $\frac{f}{F}$ is decreasing and $\phi_F : b \mapsto b + \frac{F(b)}{f(b)}$ is increasing and f does not vanish on $]0,1[$.
The derivative of U is $U'(b) = \left(v - b - \frac{F(b)}{f(b)}\right) f(b)$. So $U'(b) = 0$ if and only if $v = b + \frac{F(b)}{f(b)}$. Since ϕ_F is increasing, this can only be satisfied by a single $b \in [0,1]$. Also, since f does not vanish, U is unimodal (increasing then decreasing).
■

Lemma 14 *If Assumption 1 is satisfied, then $W_{v,F}$ is strongly concave.*

If F satisfies Assumption 1 then $\frac{f}{F}$ is decreasing and $\phi_F : b \mapsto b + \frac{F(b)}{f(b)}$ is increasing and f does not vanish on $]0,1[$.
The derivative of U is $U'(b) = \left(v - b - \frac{F(b)}{f(b)}\right) f(b)$. The derivative of W is $W'(q) = \left(v - b - \frac{F(F^{-1}(q))}{f(F^{-1}(q))}\right) = v - \phi'_F(F^{-1}(q))$, since ϕ_F is increasing. Consequently, U' is decreasing, and U' is strongly concave.

Lemma 4 *If Assumption 1 is satisfied and f is differentiable, then $\psi_F : v \mapsto b^*(v,F)$ is Lipschitz continuous with a Lipschitz constant 1.*

Proof If b^* is the optimum of the utility U, then it satisfies $(v - b^*)f(b^*) - F(b^*) = 0$. It satisfies

$$\phi_F(b^*) := b^* + \frac{F(b^*)}{f(b^*)} = v.$$

Since $\phi'_F(b^*) > 1$ thanks to Assumption 1, ϕ_F is invertible and $(\phi_F)^{-1} = \psi_F$ is Lipschitzian with constant 1 . ■

Lemma 5 *If Assumption 1 is satisfied, then*

$$F(b^*) \geq e^{-1} F(v)$$

Proof We know that $b^* < v$ and

$$\log\left(\frac{F(v)}{F(b^*)}\right) = \int_{b^*}^{v} \frac{f(u)}{F(u)} du.$$

Hence

$$\frac{F(v)}{F(b)} = \exp\left(\int_{b^*}^{v} \frac{f(u)}{F(u)} du\right).$$

Since $\frac{f(u)}{F(u)}$ is decreasing, thanks to Assumption 1,

$$\frac{F(v)}{F(b)} \leq \exp\left((v-b^*)\frac{f(b^*)}{F(b^*)}\right).$$

We have $v - b^* = \frac{F(b^*)}{f(b^*)}$, by definition of b^*. Hence $\exp\left(v-b^*)\frac{f(b^*)}{F(b^*)}\right) = \exp(1)$ and

$$F(b^*) \geq \exp(-1)F(v).$$

■

Lemma 6 *If Assumption 1 is satisfied, for any* $0 \leq q' \leq 1$,

$$W(q^*) - W(q') \leq \frac{1}{4}(q^* - q')^2 W(q^*)$$

Proof Note that this proof is an adaptation of the proof of Lemma 3.2 in Huang et al. (2018). In this proof, we denote by $b(q)$ $F^{-1}(q)$.

First of all, let us observe that $U'(b) = (v - \phi_F(b))f(b)$. We have $W'(q) = v - \phi_F(F^{-1}(q))$.

Assumption 1 implies that $\phi'_F(b) > 1, \ \forall b \in [0,1]$.

To prove Lemma 6, we will apply case-based reasoning. There are three cases depending on the relation between q' and q^*: $q' > q^*$, $q' = q^*$, and $q' < q^*$. The second case, i.e., $q' = q^*$, is trivial.

First, consider the case when $q' > q^*$. It holds

$$W(q^*) - W(q') = \int_{q^*}^{q'} -W'(q)dq = \int_{q^*}^{q'} \Big(\phi_F(b(q)) - v\Big)dq.$$

We therefore need to bound $\phi_F(b(q), \forall q \in [q^*, q']$. By definition of q^*, for any q s.t. $q^* \leq q \leq q'$, we have

$$q(v - b(q)) \leq q^*(v - b(q^*)).$$

By rewriting this equation,

$$b(q) \geq \frac{qv - q^*v + q^*b(q^*)}{q} = v\left(\frac{q-q^*}{q}\right) + \frac{q^*}{q}b(q^*) \tag{1}$$

Secondly, by the intermediate value theorem, there exists $b \in [b(q^*), b(q)]$, such that

$$\phi_F(b(q)) - \phi_F(b(q^*)) = \phi'_F(b)\Big(b(q) - b(q^*)\Big) \geq b(q) - b(q^*),$$

for any $q^* \leq q \leq q'$, where the second inequality follows from Assumption 1 that $\frac{d\phi_F(b)}{db} \geq 1$ and F being increasing thanks to Assumption 1. This in turn yields

$$\phi_F(b(q)) \geq v + b(q) - b(q^*),$$

since by definition, $W'(q^*) = \phi_F(b(q^*)) = v$. Combining with Inequality 1, we get that

$$\phi_F(b(q)) - v \geq v(\frac{q-q^*}{q}) + \frac{q^*}{q}b(q^*) - b(q^*) \geq (v - b(q^*))(\frac{q-q^*}{q}) = \frac{W(q^*)}{q^*}(\frac{q-q^*}{q})$$

Therefore, we get that

$$\begin{aligned} W(q^*) - W(q') &= \int_{q^*}^{q'} -W'(q)dq = \int_{q^*}^{q'} \Big(\phi_F(b(q)) - v\Big)dq \geq \frac{W(q^*)}{q^*}\int_{q^*}^{q'} \frac{q-q^*}{q}dq \\ &\geq \frac{W(q^*)}{q^*}\int_{\frac{q'+q^*}{2}}^{q'} \frac{q-q^*}{q}dq, \end{aligned}$$

since $\frac{q-q^*}{q} \geq 0$ for any $q' \leq q \leq q^*$. Moreover, for any $q \geq \frac{q'+q^*}{2}$, we have $\frac{q-q^*}{q} = 1 - \frac{q^*}{q} \geq 1 - \frac{2q^*}{q'+q^*} \geq \frac{q'-q^*}{q'+q^*}$. Hence, we can derive the following inequality

$$W(q^*) - W(q') \geq \int_{\frac{q'+q^*}{2}}^{q'} \frac{q'-q^*}{q'+q^*}\frac{W(q^*)}{q^*}dq = \frac{(q'-q^*)^2}{2(q'+q^*)}\frac{W(q^*)}{q^*} = \frac{(q'-q^*)^2}{2q^*(q'+q^*)}W(q^*) \ .$$

The lemma then follows from the fact that $0 \leq q', q^* \leq 1$.

The second case, $q' > q^*$ has to be treated a little differently than the first, partly because we now need to upper bound $b(q)$ instead of lower-bounding it. We achieve this by using the concavity of W (proved in Lemma 14).

By concavity of the revenue curve, for any $q' \leq q \leq q^*$, we have

$$W(q) \geq \frac{q-q'}{q^*-q'}W(q^*) + \frac{q^*-q}{q^*-q'}W(q') \ ,$$

because W lies above the segment that connects $(q', W(q'))$ and $(q^*, W(q^*))$, between q' and q^*. Hence

$$(v - b(q))q \geq \frac{q-q'}{q^*-q'}(v-b(q^*))q^* + \frac{q^*-q}{q^*-q'}(v-b(q'))q' \geq qv - b(q^*)q^*\frac{q-q'}{q^*-q'} - b(q')q'\frac{q^*-q}{q^*-q'},$$

And

$$-qb(q) \geq \frac{q^*q'}{(q^*-q')}\Big(b(q^*) - b(q')\Big) + q\frac{q'b(q') - q^*b(q^*)}{q^*-q'},$$

which yields

$$qb(q) \leq \frac{q^*q'}{(q^*-q')}\Big(b(q') - b(q^*)\Big) + q\frac{q^*b(q^*) - q'b(q')}{q^*-q'},$$

Dividing both sides by q, we have

$$b(q) \leq \frac{q^*q'}{q(q^*-q')}\Big(b(q') - b(q^*)\Big) + \frac{q^*b(q^*) - q'b(q')}{q^*-q'}, \tag{2}$$

Further, by the intermediate value theorem, there exists $b \in [b(q^*), b(q)]$, such that

$$\phi_F(b(q)) - \phi_F(b(q^*)) = \phi'_F(b)\Big(b(q) - b(q^*)\Big),$$

for any $q^* \leq q \leq q'$. Further, by Assumption 1 that $\frac{d\phi_F(b)}{db} \geq 1$, and because b is increasing thanks to Assumption 1, for any $q' \leq q \leq q^*$,

$$\phi_F(b(q)) - \phi_F(b(q^*)) \leq b(q) - b(q^*)$$

and

$$\phi_F(b(q)) \leq v + b(q) - b(q^*) \;=\; v + b(q) - b(q^*),$$

Combining with Inequality 2, we get that

$$\begin{aligned}\phi_F(b(q)) \;&\leq\; v + \frac{q^* q'}{q(q^* - q')}\Big(b(q') - b(q^*)\Big) + \frac{q^* b(q^*) - q' b(q')}{q^* - q'} - b(q*) \\ &=\; v + \frac{q'(q^* - q)}{q(q^* - q')}\big(b(q') - b(q^*)\big) \;\leq\; v + \frac{q'(q^* - q)}{q^*(q^* - q')}\big(b(q') - b(q^*)\big)\;,\end{aligned}$$

where the last inequality is due to $q \leq q^*$ and $b(q') - b(q^*) < 0$. Hence, we have

$$\begin{aligned}W(q^*) - W(q') &= \int_{q'}^{q^*} W'(q) dq \\ &= \int_{q'}^{q^*} v - \phi_F(b(q)) dq \\ &\geq \int_{q'}^{q^*} \frac{q'(q^* - q)}{q^*(q^* - q')}\big(b(q^*) - b(q')\big) dq \\ &= \frac{q'}{2q^*}(q^* - q')\big(b(q^*) - b(q')\big). \end{aligned} \tag{3}$$

On the other hand, we have

$$W(q^*) - W(q') = (q^* - q')v + q' b(q') - q^* b(q^*). \tag{4}$$

Taking the linear combination $\frac{2q^*}{3q^* - q'} \cdot 3 + \frac{q^* - q'}{3q^* - q'} \cdot 4$, we have

$$\begin{aligned}W(q^*) - W(q') &\geq v\frac{(q^* - q')^2}{3q^* - q')} - \frac{(q^* - q')^2}{3q^* - q'} b(q^*) \\ &= \frac{1}{q^*(3q^* - q')}(q^* - q')^2 W(q^*) \\ &\geq \frac{1}{3}(q^* - q')^2 W(q^*)\;,\end{aligned}$$

where the last inequality holds because $0 \leq q^*, q' \leq 1$. ∎

Lemma 7 *If Assumption 2 is satisfied, for any $F^{-1}(b^*) \leq q' \leq F^{-1}(b^* + \Delta) \leq b^* + C_f \Delta)$,*

$$W(q^*) - W(q') \leq \frac{1}{c_f} \lambda (q^* - q')^2,$$

Proof

$$W(q^*) - W(q') = \int_{q^*}^{q'} -W'(q)dq = \int_{q^*}^{q'} \Big(\phi_F(b(q)) - v\Big)dq.$$

by the intermediate value theorem, there exists $b \in [b(q^*), b(q)]$, such that

$$\phi_F(b(q)) - \phi_F(b(q^*)) = \phi'_F(b)\Big(b(q) - b(q^*)\Big) \geq \lambda(b(q) - b(q^*)),$$

so that $\phi_F(b(q)) - v \leq \lambda(b(q) - b(q^*))$ when $q^* \leq q \leq q'$ and $\phi_F(b(q)) - v \geq \lambda(b(q) - b(q^*))$ when $q' \leq q \leq q^*$. Since f is bounded from below by c_f, and since by the intermediate value theorem $\exists u \in [q, q^*],\ b(q) - b(q^*) = b'(u)(q - q^*) \geq \frac{1}{f(u)}(q - q^*)$, this yields

$$W(q^*) - W(q') \leq \lambda \frac{1}{c_f}(q' - q^*)^2$$

in both cases. ■

Lemma 15 *Beta distributions such that*

$$\alpha + \beta < \alpha\beta$$

satisfy Assumption 1.

Proof The density of a Beta distribution satisfies

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

And

$$f'(x) = \frac{(\alpha-1)x^{\alpha-2}(1-x)^{\beta-1} - (\beta-1)x^{\alpha-1}(1-x)^{\beta-2}}{B(\alpha, \beta)},$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ when Γ denotes the Gamma function. F satisfies assumption 1 if and only if $\left(\frac{f}{F}\right)'(x) = \frac{F(x)f'(x) - f^2(x)}{F^2(x)} < 0,\ \forall x \in]0, 1[$, which is equivalent to:

$$\begin{aligned} f'(x)F(x) - f^2(x) < 0,\ \forall x \in]0,1[&\iff \frac{f'(x)}{f(x)}F(x) < f(x),\ \forall x \in]0,1[) \\ &\iff F(x)B(\alpha,\beta)\left[(\alpha-1)(1-x) - (\beta-1)x\right] < x^\alpha(1-x)^\beta, \\ &\quad \forall x \in]0,1[. \end{aligned}$$

Therefore we study the function $G : x \mapsto F(x)B(\alpha, \beta)\left[(\alpha-1)(1-x) - (\beta-1)x\right] - x^\alpha(1-x)^\beta$. First of all, we observe that $G(0) = 0$. Next, we note that

$$\begin{aligned} G'(x) = &- F(x)(\alpha+\beta-2)B(\alpha,\beta) + ((\alpha-1) - (\alpha+\beta-2)x)x^{\alpha-1}(1-x)^{\beta-1} \\ &- \Big((\alpha(1-x) - \beta x)\, x^{\alpha-1}(1-x)^{\beta-1}\Big) \end{aligned}$$

and $G'(0) = 0$. Now, we compute the second derivative of G:

$$\begin{aligned} G''(x) = &- (\alpha+\beta-2)x^{\alpha-1}(1-x)^{\beta-1} + ((\alpha-1)-(\alpha+\beta-2)x))^2\, x^{\alpha-2}(1-x)^{\beta-2} \\ &- (\alpha+\beta-2)x^{\alpha-1}(1-x)^{\beta-1} - (\alpha-(\alpha+\beta)x)\,((\alpha-1)- \\ &(\alpha+\beta-2)x)x^{\alpha-2}(1-x)^{\beta-2} + (\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$

The sign of $G''(x)$ is the same as that of $P(x) = -((\alpha+\beta)-4)\,(x(1-x)) + (-1+2x)\,((\alpha-1)-(\alpha+\beta-2)x)$.

By simplifying, we get $P(x) = -(\alpha+\beta)x^2 + 2\alpha x - (\alpha-1)$. This polynomial is always negative because its maximum is $P(\frac{\alpha}{\alpha+\beta}) = -\frac{\alpha^2}{\alpha+\beta} + 2\frac{\alpha^2}{\alpha+\beta} - \alpha + 1 = \alpha^2(\frac{2}{\alpha+\beta} - 1) - \alpha + 1 = \frac{\alpha^2}{\alpha+\beta} - \alpha + 1 = \frac{\alpha+\beta-\alpha\beta}{\alpha+\beta}$.

Since $G''(x) < 0, \forall x \in [0,1]$ and $G'(0) = 0$, then $G'(x) < 0,\ \forall x \in [0,1]$. Similarly, $G'(x) < 0, \forall x \in [0,1]$ and $G(0) = 0$, implies $G'(x) < 0,\ \forall x \in [0,1]$, which in turn implies that F satisfies Assumption 1. ■

A.3. Continuous distribution leading to a utility with two global maximizers

Consider a distribution which cumulative distribution function F is piece-wise linear on $[0,v]$ at least. We consider that it changes slope at $a_1 v < v$, and that it is constant on $[a_2 v, v]$, as in Figure 8. We denote by $b_1 = F(a_1 v)$ and $b_2 = F(a_2 v)$. For simplicity we assume that F is constant on $[a_2 v, a_3 v]$ it is linear and does not change slope on $[a_3 v, 1]$ with $a_3 > 1$. We make the following assumptions

$$\begin{cases} a_2 v > v/2, \\ a_2 v \leq \frac{v+a_1 v}{2} - \frac{a_2 v - a_1 v}{b_2 - b_1}\frac{b_1}{2}. \end{cases} \tag{5}$$

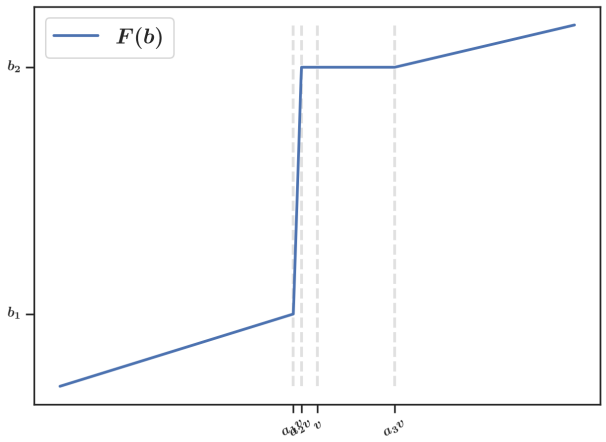


Figure 8: Example of F

Then

- On $[0, a_1 v]$ $U_v(x) = \frac{b_1}{a_1 v} x$, and the optimum on this interval is $v/2$. The optimal value on this interval is $U_v(v/2) = \frac{b_1}{a_1 v} \frac{v^2}{4}$ on this interval.
- On $[a_1 v, a_2 v]$, $U_v(x) = \left(\frac{b_2 - b_1}{a_2 v - a_1 v}(x - a_1 v) + b_1 \right)(v - x)$, and on this interval, $U_v'(x) = \frac{b_2 - b_1}{a_2 v - a_1 v}(v - 2x + a_1 v) - b_1$ and $U_v'(x) = 0 \iff x = \frac{v + a_1 v}{2} - \frac{a_2 v - a_1 v}{b_2 - b_1} \frac{b_1}{2}$. The optimizer on this interval is hence $a_2 v$, if $\frac{v + a_1 v}{2} - \frac{a_2 v - a_1 v}{b_2 - b_1} \frac{b_1}{2} > a_2 v$. Under this condition, the optimal value is $U_v(a_2 v) = b_2(v - a_1 v)$ on this interval. This can also be extended to the whole interval $[a_1 v, v]$, since U is decreasing after $a_2 v$.

Setting

$$\frac{b_1}{a_1} \frac{v}{4} = b_2 \tag{6}$$

leads to the utility having two global maximizers, $v/2$ and $a_2 v$.

To summarize, the utility's argmax is $\{v/2, a_2 v\}$ if the set of Equations 5 holds.

We can for example choose :

$$v = 1/2;\ a_2 = \frac{15}{16};\ a_1 = \frac{29}{32};\ b_2 = \frac{128}{29} b_1; b_1 = 0.5$$

This choice of parameters satisfies Condition 5 and Condition 6. Figure 9 shows the corresponding utility on $[0, v]$.

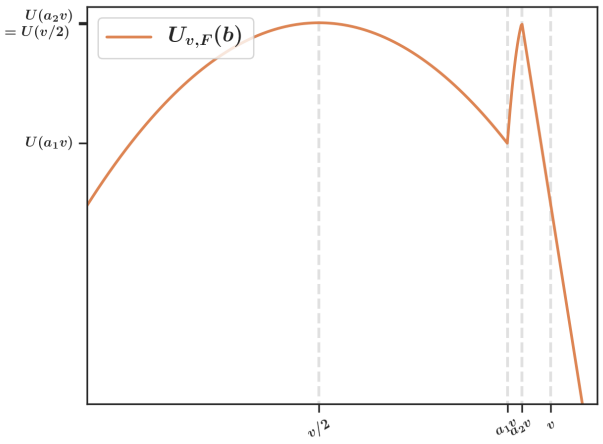


Figure 9: Associated Utility with two maximizers

Appendix B. Lower Bound

Theorem 2 *Let $\mathcal{C}$ denote the class of cumulative distribution functions on $[0, 1]$. Any strategy, whether it assumes knowledge of F or not, must satisfy*

$$\liminf_{T \to \infty} \frac{\max_{v \in [0,1], F \in \mathcal{C}} R_T^{v,F}}{\sqrt{T}} \geq \frac{1}{64},$$

Proof

We exhibit a choice of F, and two alternative Bernoulli value distributions $Ber(v)$ and $Ber(v')$ that are difficult to distinguish but whose difference is large enough so that mistaking one for the other necessarily leads to a regret of the order of $\sqrt{T}$ when the cumulative distribution function is F.

Let $v < 1$ and consider a discrete distribution with support $\left\{\frac{v}{3}, \frac{2v}{3}, 1\right\}$ such that $F(\frac{v}{3}) = A$ and $F(\frac{2v}{3}) = 2A + 3\frac{\Delta_T}{v}$, where Δ_T and A are positive constants, that we will fix later on. A maximizer of the utility can only be a point of the support, since $U_{v,F}$ decreases in the intervals where F is constant. It can not be 1, because $v < 1$. We have $U_{v,F}(\frac{v}{3}) = \frac{2vA}{3}$ and $U_{v,F}(\frac{2v}{3}) = \frac{2vA}{3} + \Delta_T$, while $U_{v,F}(1) \leq 0$. Consequently, when the value is v, the optimum is achieved by bidding $\frac{2v}{3}$ and bidding less than $\frac{2v}{3}$ yields a regret of at least Δ_T. Now let us consider the alternative situation in which the value is $v' = v - \delta_T$, with $\delta_T > 0$. We get $U_{v',F}(\frac{v}{3}) = \frac{2Av}{3} - \delta_T A$ and $U_{v',F}(\frac{2v}{3}) = \frac{2Av}{3} + \Delta_T - \delta_T(2A + \frac{3\Delta_T}{v})$. When $\Delta_T < \delta_T(2A + \frac{3\Delta_T}{v})$, the optimal bid is $\frac{v}{3}$ and the regret incurred by bidding more than $\frac{2v}{3}$ is at least $\delta_T(A + \frac{3\Delta_T}{v}) - \Delta_T$. By setting $\Delta_T = \frac{A\delta_T}{2-3\delta_T/v}$, we ensure that the regret incurred by bidding on the wrong side of $\frac{2v}{3}$ is larger than Δ_T, whether the value is v or v'. Further, by setting $\delta_T = \sqrt{v(1-v)/T}$, we force the error Δ_T to be of the order of $1/\sqrt{T}$.

We also set $A = \frac{1}{4}$, and $v = 1/2$. We can prove that $\forall T > 16$, $2A + 3\frac{\Delta_T}{v} < 1$; Indeed, if $T > 16 > (11/3)^2$, $\frac{4}{3} < 2\sqrt{T} - 6$ hence $\frac{4}{3\sqrt{T}} < 2 - \frac{6}{\sqrt{T}}$ which implies $\frac{2}{3}\frac{\frac{1}{\sqrt{T}}}{2-\frac{6}{\sqrt{T}}} = 6\Delta_T < \frac{1}{2} = 1 - 2A$.

We denote by $\mathbb{P}_{v,F}(\cdot)$ the probability of an event under the first configuration (respectively $\mathbb{E}_{v,F}(\cdot)$ the expectation of a random variable under the first configuration), and by $\mathbb{P}_{v',F}(\cdot)$ the probability of an event under the second configuration (respectively $\mathbb{E}_{v-\delta-T,F}(\cdot)$ the expectation of a random variable under the first configuration). We denote by I_t the information collected up to time $t+1$: $(M_t, V'_t, \dots M_1, V'_1)$. $\mathbb{P}^{I_t}_{v,F}$ (respectively $\mathbb{P}^{I_t}_{v'}$) denotes the law of I_t in the first (respectively second) configuration.

We consider the Kullback Leibler divergence between $\mathbb{P}^{I_t}_{v,F}$ and $\mathbb{P}^{I_t}_{v',F}$. We prove that it is equal to

$$KL(\mathbb{P}^{I_t}_{v}, \mathbb{P}^{I_t}_{v',F}) = kl(v, v')\mathbb{E}[N_t], \tag{7}$$

where $kl(\cdot,\cdot)$ denotes the Kullback Leibler divergence between two Bernoulli distributions. Indeed, thanks to the chain rule for conditional KL,

$$KL(\mathbb{P}^{I_t}_{v,F}, \mathbb{P}^{I_t}_{v',F}) = KL(\mathbb{P}^{I_t}_{v,F}, \mathbb{P}^{I_t}_{v',F}) + KL(\mathbb{P}^{(M_t,V'_t)|I_t}_{v,F}, \mathbb{P}^{(M_t,V'_t)|I_t}_{v',F}),$$

and

$$\begin{aligned} KL(\mathbb{P}^{(M_t,V'_t)|I_t}_{v,F}, \mathbb{P}^{(M_t,V'_t)|I_t}_{v',F}) &= \mathbb{E}[\mathbb{E}[KL(\nu_{I_t} \otimes \mathcal{D}_F, \nu'_{I_t} \otimes \mathcal{D}_F)|I_t]] \\ &= \mathbb{E}[kl(v, v')\mathbb{1}(B_t > M_t)]. \end{aligned}$$

where ν_{I_t}(respectively ν'_{I_t}) denotes the law of V'_t knowing I_t in the first configuration (respectively the second), and $\mathcal{D}_F$ the law of M_t.

By induction, we obtain

$$KL(\mathbb{P}^{I_t}_{v,F}, \mathbb{P}^{I_t}_{v',F}) = kl(v, v')\mathbb{E}_{v,F}[N_t].$$

We stress that in either of the former configurations (under (v, F) or (v', F)), playing on the wrong side of $\frac{2}{3}v$ yields a regret larger than Δ_T. Using this, we get that $\forall T > 16$,

$$\begin{aligned}
\max(R_T^{v,F}, R_T^{v',F}) &\geq \frac{1}{2}(R_T^{v,F} + R_T^{v-\delta,F}) \\
&\geq \frac{1}{2}\sum_{t=1}^{T}\left(\Delta_T \mathbb{P}_{v,F}\left(B_t < \frac{2}{3}v\right) + \Delta_T \mathbb{P}_{v',F}\left(B_t > \frac{2}{3}v\right)\right) \\
&\geq \frac{1}{2}\sum_{t=1}^{T}\left(\Delta_T \mathbb{P}_{v,F}\left(B_t < \frac{2}{3}v\right) + \Delta_T \left(1 - \mathbb{P}_{v',F}(B_t > \frac{2}{3}v)\right)\right) \\
&\geq \frac{1}{2}\sum_{t=1}^{T}\Delta_T \left(1 - TV(\mathbb{P}_{v,F}^{I_t}, \mathbb{P}_{v',F}^{I_t})\right) \\
&\geq \frac{1}{2}\sum_{t=1}^{T}\Delta_T \left(1 - \sqrt{\frac{1}{2}KL(\mathbb{P}_{v,F}^{I_t}, \mathbb{P}_{v',F}^{I_t})}\right) \\
&\geq \frac{1}{2}\sum_{t=1}^{T}\Delta_T \left(1 - \sqrt{\frac{1}{2}\mathbb{E}_{v,F}[N_t]kl(v, v')}\right) \\
&\geq \frac{1}{2}\sum_{t=2}^{T}\Delta_T \left(1 - \sqrt{\frac{1}{2}Tkl(v, v')}\right)
\end{aligned}$$

where we used Pinsker's inequality in the fifth inequality and where $TV(\cdot, \cdot)$ denotes the total variation. Yet, since $kl(v, v') = \frac{(v'-v)^2}{2}\int_0^1 g''(v' + s(v' + s(v - v'))2(1 - s)ds$, where $g(x) = kl(x, v')$ thanks to Taylor's inequality,

$$\begin{aligned}
kl(v, v') &\leq \frac{(v' - v)^2}{2}\int_0^1 2 \max_{u \in [v,v']} g''(u)ds \\
&\leq (v' - v)^2 \frac{1}{\min_{u\in[v,v']} u(1 - u)} \\
&\leq \frac{(v' - v)^2}{v'(1 - v')},
\end{aligned}$$

since $v = \frac{1}{2}$.

Therefore,

$$\begin{aligned}
\max(R_T^{v,F}, R_T^{v',F}) &\geq \frac{1}{2}\sum_{t=1}^{T} \Delta_T \left(1 - \sqrt{\frac{1}{2} T kl(v,v')}\right) \\
&\geq \frac{1}{2}\sum_{t=1}^{T} \Delta_T \left(1 - \sqrt{\frac{1}{8}\frac{1}{(1/2 - \frac{1}{2\sqrt{T}})(1/2 + \frac{1}{2\sqrt{T}})}}\right) \\
&\geq \frac{1}{2} \times \frac{A\delta_T}{2 - 3/2\delta_T} T \left(1 - \sqrt{\frac{1}{8}\frac{1}{(1/2 - \frac{1}{2\sqrt{T}})(1/2 + \frac{1}{2\sqrt{T}})}}\right) \\
&\geq \frac{1}{16 - 12/\sqrt{T}} \sqrt{T} \left(1 - \sqrt{\frac{1}{8}\frac{1}{(1/2 - \frac{1}{2\sqrt{T}})(1/2 + \frac{1}{2\sqrt{T}})}}\right)
\end{aligned}$$

Finally

$$\liminf_{T\to\infty} \frac{\max(R_T^{v,F}, R_T^{v',F})}{\sqrt{T}} \geq \frac{1}{16}\left(1 - \sqrt{\frac{1}{2}}\right) \geq \frac{1}{64}$$

■

Appendix C. Preliminary Results

C.1. Concentration inequalities used for the upper bounds

C.1.1. On the value V_t

Lemma 16 *The following concentration inequality on the values holds*

$$\sum_{t=2}^{T} \mathbb{P}\left((\hat{V}_t - v)^2 \geq \frac{\gamma \log(t-1)}{2N_t}\right) \leq \sum_{t=1}^{T} 2e\sqrt{\gamma}(\log(t))t^{-\gamma}.$$

Proof We have, for all η_{t-1},

$$\begin{aligned}
\sum_{t=2}^{T} \mathbb{P}\left((\hat{V}(N_t) - v)^2 \geq \frac{\eta_{t-1}}{2N_t}\right) &\leq \sum_{t=2}^{T} \mathbb{P}\left(\exists m:\ 1 \leq m \leq t, 2m(\hat{V}(m) - v)^2 \geq \eta_{t-1}\right) \\
&\leq \sum_{t=1}^{T} 2e\sqrt{\eta_{t-1}\log(t-1)}\exp(-\eta_{t-1}) := l_1(T)
\end{aligned}$$

where the second inequality comes from Lemma 11 in (Cappé et al., 2013), and from the fact that V_t is a positive random variable bounded by 1, so $1/2-$ sub-Gaussian.

Therefore, if $\eta_t := \gamma \log t$,

$$l_1(T) = \sum_{t=2}^{T} 2e\sqrt{\gamma}(\log(t-1))(t-1)^{-\gamma}$$

which tends to a finite limit as soon as $\gamma > 1$. ■

C.1.2. On the cumulative distribution function of M_t

Lemma 17 *The following concentration inequality holds on the empirical cumulative distribution $\hat{F}_t$.*

$$\sum_{t=2}^{T} \mathbb{P}\left(\|\hat{F}_t - F\|_\infty \geq \frac{\gamma \log(t-1)}{2(t-1)} \right) \leq 2 \sum_{t=1}^{T} t^{-\gamma}.$$

Proof It holds

$$\begin{aligned}
&\sum_{t=2}^{T} \mathbb{P}\left((\max_{b\in[0,1]} |F_t(b) - F(b)|)^2 \geq \frac{\gamma \log(t-1)}{2(t-1)} \right) \\
&\leq \sum_{t=2}^{T} \mathbb{P}\left(\|\hat{F}_t - F\|_\infty^2 \geq \frac{\gamma \log(t-1)}{2(t-1)} \right) \\
&\leq \sum_{t=1}^{T-1} 2e^{-\frac{2\gamma \log(t}{2t}} \\
&\leq \sum_{t=1}^{T} 2t^{-\gamma},
\end{aligned}$$

according to the Dvoretzky–Kiefer–Wolfowitz inequality (see Massart (1990)).
Note that this also yields

$$\begin{aligned}
\sum_{t=2}^{T} \mathbb{P}\left(\|\hat{F}_t - F\|_\infty \geq \frac{\gamma \log(t-1)}{2N_t} \right) &\leq \sum_{t=2}^{T} \mathbb{P}\left(\|\hat{F}_t - F\|_\infty \geq \frac{\gamma \log(t-1)}{2(t-1)} \right) \\
&\leq 2 \sum_{t=1}^{T} t^{-\gamma}.
\end{aligned}$$

■

C.1.3. Local concentration inequality

This lemma is key for the proof of the upper bound of the regret of UCBid1+. It quantifies the variation of $\hat{F}_t$ on a small interval.

Lemma 12 *For any $a, b \in [0,1]$, if F is continuous and increasing, then*

$$\begin{aligned}
&\sup_{a\leq x\leq b} |\hat{F}_t(x) - F(x) - (\hat{F}_t(a) - F(a))| \\
&\qquad \leq \sqrt{\frac{2(F(b)-F(a)) \log\left(\frac{e\sqrt{t}}{\sqrt{2(F(b)-F(a))}\eta} \right)}{t}} + \frac{\log(\frac{t}{2(F(b)-F(a)\eta^2})}{6t}, \quad (8)
\end{aligned}$$

with probability $1-\eta$

Remark : it follows from the lemma that the the maximal gap between $\hat{F}_t(x) - F(x)$ and $\hat{F}_t(\frac{a+b}{2}) - F(\frac{a+b}{2})$ can easily be bounded by :

$$\sup_{a\leq x\leq b} |\hat{F}_t(x) - F(x) - (\hat{F}_t(\frac{a+b}{2}) - F(\frac{a+b}{2}))| \\ \leq 2\sqrt{\frac{2(F(b)-F(a))\log\left(\frac{e\sqrt{t}}{\sqrt{2\eta(F(b)-F(a))}}\right)}{t}} + 2\frac{\log(\frac{t}{2(F(b)-F(a)\eta^2})}{6t}$$

with probability $1-\eta$.

Proof:

Let $X_1, \dots, X_n \overset{iid}{\sim} dF$. Let $m > 2$ For every $1 \leq i \leq m$, let x_i be such that

$$F(x_i) = F(a) + \frac{i}{m}\big(F(b) - F(a)\big) \ .$$

By Bernstein's inequality, since $t\big(\hat{F}_t(x_i) - \hat{F}_t(a)\big) \sim \mathcal{B}(n, F(x_i) - F(a))$ has a variance bounded by $t\big(F(b) - F(a)\big)$, there is an event A of probability at least $1 - me^{-z}$ on which

$$\max_{0\leq i\leq m} \big|\hat{F}_t(x_i) - \hat{F}_t(a) - (F(x_i) - F(a))\big| \leq \sqrt{\frac{2\big(F(b)-F(a)\big)z}{t}} + \frac{z}{3t} := \delta,$$

by a union bound. Besides, for $i = 0$, $\hat{F}_t(x_i) - \hat{F}_t(a) - (F(x_i) - F(a)) = 0$.

On this event, for every $x_{i-1} \leq x \leq x_i$:

$$\hat{F}_t(x) - \hat{F}_t(a) - (F(x) - F(a)) \leq \hat{F}_t(x_i) - \hat{F}_t(a) - (F(x_i) - F(a)) + F(x_i) - F(x) \leq \delta + \frac{1}{m},$$
$$\hat{F}_t(x) - \hat{F}_t(a) - (F(x) - F(a)) \geq \hat{F}_t(x_{i-1}) - \hat{F}_t(a) - (F(x_{i-1}) - F(a)) + F(x_{i-1}) - F(x) \\ \geq -\delta - \frac{1}{m} \ .$$

and hence

$$\sup_{a\leq t\leq b} \big|\hat{F}_t(x) - \hat{F}_t(a) - (F(x) - F(a))\big| \leq \sqrt{\frac{2\big(F(b)-F(a)\big)z}{t}} + \frac{z}{3t} + \frac{1}{m} \ .$$

Now, take

$$m = \left\lceil \sqrt{\frac{t}{2\big(F(b)-F(a)\big)}} \right\rceil$$

and $z = \log(m/\eta)$: one gets that with probability at least $1-\eta$,

$$\begin{aligned}
&\sup_{a\le t\le b}\left|\hat{F}_t(x) - \hat{F}_t(a) - (F(x)-F(a))\right| \\
&\qquad\le \sqrt{\frac{2\big(F(b)-F(a)\big)\log\left(\frac{\sqrt{\frac{t}{2(F(b)-F(a))}}}{\eta}\right)}{t}} + \frac{\log\left(\frac{\sqrt{\frac{t}{2(F(b)-F(a))}}}{\eta}\right)}{3t} + \sqrt{\frac{2\big(F(b)-F(a)\big)}{t}} \\
&\qquad\le \sqrt{\frac{2\big(F(b)-F(a)\big)\log\left(\frac{e\sqrt{t}}{\sqrt{2(F(b)-F(a))}\eta}\right)}{t}} + \frac{\log\left(\frac{t}{2(F(b)-F(a))\eta^2}\right)}{6t}\,.
\end{aligned}$$

C.2. General bound on the instantaneous regret

In the following, we will repeatedly use the following general bound on the instantaneous regret conditioned on the past and on a current victory.

Lemma 18 *Let A be an $\mathcal{F}_{t-1}$-measurable event. Let S_t denote $(V_t - b^*)\mathbb{1}(M_t < b^*) - (V_t - B_t)\mathbb{1}(M_t < B_t)$. The following inequality holds:*

$$\mathbb{E}\left[S_t\mathbb{1}(B_t > b^*)\mathbb{1}(A)|\mathcal{F}_{t-1}\vee\sigma(\mathbb{1}(B_t > M_t))\right] \le \frac{U(b^*)-U(B_t)}{F(b^*)}\mathbb{1}(M_t\le B_t)\mathbb{1}(A).$$

Proof When $B_t > b^*$, the instantaneous regret can be decomposed as follows

$$S_t\mathbb{1}(B_t > b^*) = (B_t - v)\mathbb{1}(M_t \le b^*)\mathbb{1}(B_t > b^*) + (B_t - b^*)\mathbb{1}\left\{(M_t \le b^* \le B_t)\right\}. \tag{9}$$

Note that in particular, there is no instantaneous regret when $M_t > B_t$. Therefore

$$\begin{aligned}
&\mathbb{E}\left[S_t\mathbb{1}(B_t > b^*)\mathbb{1}(A)|\mathcal{F}_{t-1}\vee\mathbb{1}(B_t > M_t)\right] \\
&\le \frac{(B_t-b^*)F(b^*) + (B_t - v)(F(B_t)-F(b^*))}{F(B_t)}\mathbb{1}(M_t\le B_t)\mathbb{1}(B_t > b^*)\mathbb{1}(A) \\
&\le \frac{U(b^*)-U(B_t)}{F(b^*)}\mathbb{1}(M_t\le B_t)\mathbb{1}(A),
\end{aligned}$$

since $U(b^*) - U(B_t) = (v-b^*)F(b^*) - (v-B_t)F(B_t)$, which also equals $(B_t-b^*)F(b^*) + (B_t - v)(F(B_t)-F(b^*))$. ∎

C.3. Other lemmas

Lemma 19 *The expectations $\mathbb{E}\left[\sum_{t=2}^T \frac{1}{N_t}\mathbb{1}\{M_t\le B_t\}\right]$ and $\mathbb{E}\left[\sum_{t=2}^T \sqrt{\frac{1}{N_t}}\mathbb{1}\{M_t\le B_t\}\right]$ can always be bounded as follows*

$$\begin{cases}
\mathbb{E}\left[\sum_{t=2}^T \frac{1}{N_t}\mathbb{1}\{M_t\le B_t\}\right] \le 1+\log T, \\
\mathbb{E}\left[\sum_{t=2}^T \sqrt{\frac{1}{N_t}}\mathbb{1}\{M_t\le B_t\}\right] \le 1+\sqrt{T}.
\end{cases}$$

Proof Since winning an auction increments the number of observations N_t by 1,

$$\begin{aligned}
\sum_{t=2}^{T} \mathbb{E}\Big[\sqrt{\frac{1}{N_t}}\mathbb{1}(M_t \leq B_t)\Big] &\leq \sum_{t=2}^{T}\sum_{n=1}^{T-1} \sqrt{\frac{1}{n}}\mathbb{1}\{N_t = n,\ N_{t+1} = n+1\} \\
&\leq \sum_{n=1}^{T-1}\sqrt{\frac{1}{n}} \sum_{t=2}^{T} \mathbb{1}\{N_t = n,\ N_t = n+1\} \\
&\leq \sum_{n=1}^{T-1}\sqrt{\frac{1}{n}} \\
&\leq 1 + \sum_{n=2}^{T-1}\int_{n-1}^{n}\sqrt{\frac{1}{u}}du \\
&\leq 1 + \sqrt{T}.
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
\sum_{t=2}^{T} \mathbb{E}\Big[\frac{1}{N_t}\mathbb{1}(M_t \leq B_t)\Big] &\leq \sum_{t=2}^{T}\sum_{n=1}^{T-1} \frac{1}{n}\mathbb{1}\{N_t = n,\ N_{t+1} = n+1\} \\
&\leq \sum_{n=1}^{T-1}\frac{1}{n} \sum_{t=2}^{T} \mathbb{1}\{N_t = n,\ N_t = n+1\} \\
&\leq \sum_{n=1}^{T-1}\frac{1}{n} \\
&\leq 1 + \sum_{n=2}^{T-1}\int_{n-1}^{n}\frac{1}{u}du \\
&\leq 1 + \log T.
\end{aligned}$$

■

Lemma 20 *If g_1 and g_2 are two functions such that $\|g_1 - g_2\|_\infty \leq \delta$, then*

$$g_1(b_1^*) - g_1(b_2^*) \leq 2\delta$$

where $b_1^* = \max(\arg\max_{b\in[0,1]} g_1(b))$ and $b_2^* = \max(\arg\max_{b\in[0,1]} g_2(b))$.

Proof Indeed,

$$\begin{aligned}
0 \leq g_1(b_1^*) - g_1(b_2^*) &\leq g_1(b_1^*) - g_2(b_2^*) + g_2(b_2^*) - g_1(b_2^*) \\
&\leq 2\delta.
\end{aligned}$$

■

Lemma 21 *For any $a > 0$, $t \geq 2a\log(a)$ implies $t \geq a\log t$.*

Proof

$$\begin{aligned} a\log t &\geq a\left(\frac{t}{2a} + \log(2a)\right) \\ &\geq t/2 + a\log(a), \end{aligned}$$

where the first inequality follows from the fact that $\log(x/y) \leq x/y$ for any positive x and y. Hence when $t > 2a\log(a)$, $t \geq t/2 + a\log t \geq a\log t$. ■

Appendix D. Known F

D.1. Upper Bounds of the Regret of UCBid1

We prove the somewhat more precise form of Theorem 8.

Theorem 8 *UCBid1 incurs a regret bounded as follows*

$$R_T \leq \frac{1}{F(b^*)}\sqrt{\gamma\log T}(\sqrt{T}+1) + O(1).$$

Proof We denote by U_t^{UCBid1} the function $b \mapsto (\hat{V}_t + \epsilon_t - b)F(b)$. The regret can be decomposed as follows.

$$R_T \leq 1 + \sum_{t=2}^{T}\mathbb{P}\left(|\hat{V}_t - v| \geq \epsilon_t\right) + \sum_{t=2}^{T}\mathbb{E}\left[S_t\mathbb{1}\left\{|\hat{V}_t - v| \leq \epsilon_t\right\}\right],$$

Lemma 16 yields the following bound on the probability of over-estimating $\hat{V}_t$:

$$\sum_{t=2}^{T}\mathbb{P}(|\hat{V}_t - v| \geq \epsilon_t) \leq \sum_{t=1}^{t} 2e\sqrt{\gamma}(\log t)t^{-\gamma}.$$

Since $F(x) \leq 1, \forall x \in [0,1]$, and $\|U_t^{UCBid1} - U\|_\infty = \|(\hat{V}_t - v + \epsilon_t)F(x)\|_\infty \leq |\hat{V}_t - v + \epsilon_t|$, we can bound the difference between the utility function and its (upper confidence) estimate with high probability:

$$\sum_{t=2}^{T}\mathbb{P}(\|U_t^{UCBid1} - U\|_\infty \geq 2\epsilon_t) \leq \sum_{t=1}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma}.$$

When $\|U_t^{UCBid1} - U\|_\infty \leq 2\epsilon_t$, then

$$|U(b^*) - U(B_t)| \leq 4\epsilon_t,$$

thanks to Lemma 20. Additionally, using Lemma 1, if $\hat{V}_t + \epsilon_t - v \geq 0$, then $B_t \geq b^*$ Therefore,

$$\begin{aligned}
&\sum_{t=2}^{T} \frac{1}{F(b^*)} \mathbb{E}\left[S_t \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{b^* \leq B_t\right\} \mathbb{1}\left\{|\hat{V}_t - v| \leq \epsilon_t\right\}\right] \\
&\leq \sum_{t=2}^{T} \mathbb{E}\left[\frac{U(b^*) - U(B_t)}{F(b^*)} \mathbb{1}\left\{b^* \leq B_t\right\} \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{|\hat{V}_t - v| \leq \epsilon_t\right\}\right] \\
&\leq \sum_{t=2}^{T} \mathbb{E}\left[\frac{U(b^*) - U(B_t)}{F(b^*)} \mathbb{1}\left\{b^* \leq B_t\right\} \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{U(b^*) - U(B_t) \leq 4\epsilon_t\right\}\right] \\
&\leq \sum_{t=2}^{T} \frac{1}{F(b^*)} \mathbb{E}\left[4\epsilon_t \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{(U(b^*) - U(B_t) \leq 4\epsilon_t\right\}\right] \\
&\leq \sum_{t=2}^{T} \frac{1}{F(b^*)} \sqrt{2\frac{\gamma \log T}{N_t}} \\
&\leq \frac{1}{F(b^*)} \sqrt{2\gamma \log T}(1 + \sqrt{T}),
\end{aligned}$$

where the second inequality comes from Lemma 18 (in fact $\left\{|\hat{V}_t - v| \leq \epsilon_t\right\}$ is $\mathcal{F}_{t-1}$-measurable) and the last inequality comes from Lemma 19.

Using Lemma 16 yields

$$\sum_{t=2}^{T} \mathbb{P}(|\hat{V}_t - v| \geq \epsilon_t) \leq \sum_{t=1}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma}.$$

Combining this with the above decomposition of the regret yields

$$R_T \leq 1 + \sum_{t=1}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma} + \frac{1}{F(b^*)} \sqrt{2 \log T}(1 + \sqrt{T}),$$

When $\gamma > 1$, $\sum_{t=1}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma}$ tends to a constant, and

$$R_T \leq \frac{1}{F(b^*)} \sqrt{2\gamma \log T}(1 + \sqrt{T}) + O(1),$$

which concludes the proof.

Theorem 9 *If F satisfies Assumption 1 and 2, then*

$$R_T \leq \frac{2\gamma\lambda C_f^2}{F(b^*)c_f} \log^2(T) + O(\log T),$$

when $\gamma > 1$.

Proof

Thanks to Lemma 1, if $\hat{V}_t + \epsilon_t - v \geq 0$, then $B_t \geq b^*$. Additionally,

$$B_t - b^* \leq (\hat{V}_t + \epsilon_t - v),$$

thanks to Lemma 4. In particular, if $\hat{V}_t + \epsilon_t - v < 2\epsilon_t$,

$$B_t - b^* \leq 2\epsilon_t.$$

The regret can therefore be decomposed as follows :

$$\begin{aligned} R_T \leq 1 + \sum_{t=2}^{T} \mathbb{P}(\hat{V}_t + \epsilon_t - v \leq 0) + \sum_{t=2}^{T} \mathbb{P}(\hat{V}_t - \epsilon_t - v \geq 0) \\ + \mathbb{E}\left[\sum_{t=2}^{T} S_t \mathbb{1}(B_t \in [b^*, b^* + \min(2\epsilon_t, \Delta)]\right] + \sum_{t=2}^{T} \mathbb{E}\left[S_t \mathbb{1}(B_t \in [b^* + \min(2\epsilon_t, \Delta), b^* + \Delta])\right] \end{aligned} \tag{10}$$

Let us bound the third term of this inequality. Thanks to Lemma 18 ,

$$\begin{aligned} &\mathbb{E}\left[S_t \mathbb{1}(B_t \in [b^*, b^* + \epsilon_t])|\mathcal{F}_{t-1} \vee \sigma(\mathbb{1}\left\{M_t \leq B_t\right\})\right] \\ &\qquad \leq \frac{U(b^*) - U(B_t)}{F(b^*)} \times \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{b^* \leq B_t \leq b^* + 2\epsilon_t\right\}, \end{aligned} \tag{11}$$

because $(B_t \in [b^*, b^* + \epsilon_t])$ is $\mathcal{F}_{t-1}$- measurable. This is why

$$\begin{aligned} &\sum_{t=2}^{T} \mathbb{E}\left[\mathbb{E}\left[S_t \mathbb{1}(B_t \in [b^*, b^* + \min(2\epsilon_t, \Delta])|\mathcal{F}_{t-1} \vee \sigma(\{\mathbb{1}\left\{M_t \leq B_t\right\})\}\right]\right] \\ &\leq \sum_{t=2}^{T} \mathbb{E}\left[\frac{U(b^*) - U(B_t)}{F(b^*)} \times \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{b^* \leq B_t \leq b^* + \min(2\epsilon_t, \Delta\right\}\right] \\ &\leq \sum_{t=2}^{T} \mathbb{E}\left[\frac{W(q^*) - W(Q_t)}{F(b^*)} \times \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{q^* \leq Q_t \leq b^* + 2C_f\epsilon_t\right\}\right] \\ &\leq \sum_{t=2}^{T} \mathbb{E}\left[\frac{\lambda(q^* - Q_t)^2}{c_f F(b^*)} \times \mathbb{1}\left\{M_t \leq B_t\right\} \mathbb{1}\left\{q^* \leq Q_t \leq b^* + 2C_f\epsilon_t\right\}\right] \\ &\leq \mathbb{E}\left[\frac{\lambda(2C_f)^2}{c_f F(b^*)} \sum_{t=2}^{T} \left(\frac{\gamma \log T}{2N_t}\right) \mathbb{1}\left\{M_t \leq B_t\right\}\right] \\ &\leq \frac{2\lambda\gamma\bar{C}_f}{c_f F(b^*)} \log T(\log T + 1), \end{aligned}$$

where the third inequality comes from Lemma 7 and the last one follows from Lemma 19.

Thanks to Lemma 16, the sum of the first term and the second term of Equation (10) can be bounded by $\sum_{t=2}^{T} \mathbb{P}(\hat{V}_t - v < \epsilon_t) + \sum_{t=2}^{T} \mathbb{P}(\hat{V}_t - \epsilon_t - v \geq 0) \leq \sum_{t=1}^{T} e\sqrt{\gamma}\frac{\log t}{t^\gamma}$ which is bounded by a constant when $\gamma > 1$.

The last term of Equation (10) can be bounded as follows:

$$\begin{aligned}
\sum_{t=2}^{T} \mathbb{E}\left[S_t \mathbb{1}(B_t \in [b^* + \min(2\epsilon_t, \Delta), b^* + \Delta])\right] &\leq \sum_{t=2}^{T} \mathbb{P}\left[(\Delta > 4\epsilon_t, M_t \leq B_t, B_t > b^*\right] \\
&\leq \sum_{t=2}^{T} \mathbb{P}\left[\Delta^2 > 4\frac{\gamma \log T}{2N_t}, M_t \leq B_t, B_t > b^*\right] \\
&\leq \sum_{t=2}^{T} \sum_{n=1}^{T-1} \mathbb{P}\left[\Delta^2 > 2\frac{\gamma \log T}{2N_t}\right] \mathbb{1}\left[N_t = n,\ N_{t+1} = n+1\right] \\
&\leq \sum_{n=1}^{T-1} \mathbb{1}\left[n < 4\frac{\gamma \log T}{2\Delta^2}\right] \sum_{t=2}^{T} \mathbb{1}\left\{N_t = n,\ N_{t+1} = n+1\right\} \\
&\leq \sum_{n=1}^{T-1} \mathbb{1}\left\{n < 4\frac{\gamma \log T}{2\Delta^2}\right\} \\
&\leq 4\frac{\gamma \log T}{2\Delta^2}
\end{aligned}$$

where the first inequality comes from the fact that when $B_t > b^*$, a positive instantaneous regret can only occur if $M_t \leq B_t$. By summing all components of the regret,

$$R_T \leq 1 + 4\frac{\gamma \log T}{2\Delta^2} + \frac{2\gamma\lambda C_f^2}{F(b^*)c_f}(\log^2(T) + \log T).$$

In conclusion,

$$R_T \leq \frac{2\gamma\lambda C_f^2}{F(b^*)c_f} \log^2(T) + O(\log T)$$

when $\gamma > 1$. ■

D.2. Lower bound of the regret of optimistic strategies

Lemma 10 *Consider all environments where V_t follows a Bernoulli distribution with expectation v and F satisfies Assumption 1 and is such that $\phi' \leq \lambda$, and there exists c_f and C_f such that $0 < c_f < f(b) < C_f$, $\forall b \in [0,1]$. If a strategy is such that, for all such environments, $R_T^{v,F} \leq O(T^a)$, for all $a > 0$, and there exists $\gamma > 0$ such that $\mathbb{P}(B_t < b^*) < t^{-\gamma}$, then this strategy must satisfy:*

$$\liminf_{T\to\infty} \frac{R_T^{v,F}}{\log T} \geq c_f^2 \lambda^2 \left(\frac{v(1-v)(v - b^*_{v,F})}{32}\right).$$

Note that this proof is an adaptation of the proof of the parametric lower bound of (Achddou et al., 2021).

Lemma 22 *If $R_T \leq O(T^a)$, $\forall a > 0$, and F admits a density which is lower bounded by a positive constant and upper bounded. Then,*

$$\lim_{t\to\infty} \mathbb{E}\left[\frac{N_t}{t}\right] = F(b^*).$$

Proof The fraction of won auctions is $\mathbb{E}\left[\frac{N_t}{t}\right] = \mathbb{E}[\frac{1}{t}\sum_{s=1}^t F(B_s]$, by the tower rule. Since F admits a density f, upper bounded by a constant C_f,

$$\mathbb{E}[(F(B_t) - F(b^*))^2]] \leq C_f^2 \mathbb{E}[(B_t - b^*)^2].$$

The consistency assumption implies $\sum_{t=1}^T \mathbb{E}[(B_t - b^*)^2] \leq O(T^a)$, $\forall a > 0$, because of Lemma 6. In particular $\lim_{t\to\infty} \mathbb{E}[(B_t - b^*)^2] = 0$. Combining the two previous arguments yields $\lim_{t\to\infty} \mathbb{E}[(F(B_t) - F(b^*))^2] = 0$. Then, because L_2-convergence implies L_1-convergence, $\lim_{t\to\infty} \mathbb{E}[F(B_t)] = F(b^*)$.
Together with the equality $\mathbb{E}\left[\frac{N_t}{t}\right] = \mathbb{E}[\frac{1}{t}\sum_{s=1}^t F(B_s)]$, and with the Cesaro theorem, this result proves suffices to prove the lemma. ■

We set a time step $t \in [1, T]$. We consider two alternative configurations with identical distributions for M_t but that differ by the distribution of V_t. The value V_t is distributed according to a Bernoulli distribution of expectation v in the first configuration, respectively $v'_t = v + \sqrt{\frac{v(1-v)}{F(b^*)t}}$, in the second configuration.

Notation. We let $\mathbb{P}_v(\cdot)$ denote the probability of an event under the first configuration (respectively $\mathbb{E}_v(\cdot)$ the expectation of a random variable under the first configuration), whereas $\mathbb{P}_{v'_t}(\cdot)$ denotes the probability of an event under the second configuration (respectively $\mathbb{E}_{v'_t}(\cdot)$ the expectation of a random variable under the first configuration). The information collected up to time $t+1$ is denoted $I_t : (M_t, V'_t, \dots M_1, V'_1)$. Finally, $\mathbb{P}_v^{I_t}$ (respectively $\mathbb{P}_{v'_t}^{I_t}$) is the law of I_t in the first (respectively second) configuration.

The Kullback Leibler divergence between $\mathbb{P}_v^{I_t}$ and $\mathbb{P}_{v'_t}^{I_t}$ can be proved to satisfy

$$KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t}) = kl(v, v'_t)\mathbb{E}[N_t],$$

exactly like in Equation 7.

Using Lemma 22, $\forall \epsilon > 0, \exists t_1(\epsilon), \forall t \geq t_1(\epsilon)$,

$$KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t}) \leq kl(v, v'_t)(1+\epsilon)F(b^*).$$

Using the data processing inequality (see for example Garivier et al. (2019)), we get

$$\begin{aligned}
KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t}) &\geq kl\left(\mathbb{P}_v\left(B_t > \frac{b^*_{v,F} + b^*_{v'_t,F}}{2}\right), \mathbb{P}_{v'_t}\left(B_t > \frac{v + b^*_{v'_t,F}}{2}\right)\right) \\
&\geq 2\left(\mathbb{P}_v\left(B_t > \frac{b^*_{v,F} + b^*_{v'_t,F}}{2}\right) - \mathbb{P}_{v'_t}\left(B_t > \frac{b^*_{v,F} + b^*_{v'_t,F}}{2}\right)\right)^2 \\
&\geq 2\left(\mathbb{P}_v\left(B_t > \frac{b^*_{v,F} + b^*_{v'_t,F}}{2}\right) + \mathbb{P}_{v'_t}\left(B_t \leq \frac{b^*_{v,F} + b^*_{v'_t,F}}{2}\right) - 1\right)^2,
\end{aligned}$$

where the second inequality comes from Pinsker inequality. Consequently, we get

$$\mathbb{P}_v\left(B_t > \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) + \mathbb{P}_{v'_t}\left(B_t \leq \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) \geq 1 - \sqrt{\frac{1}{2}KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}.$$

Specifically, $\forall t > t_0(\epsilon)$,

$$\mathbb{P}_v\left(B_t > \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) + \mathbb{P}_{v'_t}\left(B_t \leq \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) \geq 1 - \sqrt{\frac{1}{2}kl(v,v'_t)(1+\epsilon)F(b^*_{v,F})t}.$$

Using the fact that $\mathbb{E}_v[(B_t - b^*_{v,F})^2] \geq \left(b^*_{v,F} - \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right)^2 \mathbb{P}_v\left(B_t > \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right)$ yields

$$\begin{aligned}
\mathbb{E}_v[(B_t - b^*_{v,F})^2] &\geq \left(\frac{b^*_{v,F}-b^*_{v'_t,F}}{2}\right)^2 \mathbb{P}_v\left(B_t > \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) \\
&\geq \left(\lambda\frac{v-v'_t}{2}\right)^2 \mathbb{P}_v\left(B_t > \frac{b^*_{v,F}+b^*_{v'_t,F}}{2}\right) \\
&\geq \lambda^2 \frac{v(1-v)}{4F(b^*_{v,F})t}\left(1 - \sqrt{\frac{1}{2}(1+\epsilon)kl(v,v'_t)F(b^*_{v,F})t} - 1/t^\gamma\right),
\end{aligned}$$

where the second inequality comes from the fact that $v = \phi_F(b^*_{v,F})$ (resp. $v'_t = \phi_F(b^*_{v'_t,F})$) and that $\phi'_F \leq \lambda$ and the the second inequality stems from the assumption that the algorithm outputs a bid that does not underestimate $b^*_{v'_t,F}$ with high probability: $\mathbb{P}_{v'_t}(B_t < b^*_{v'_t,F}) < \frac{1}{t^\gamma}$.

We use the fact that $\forall \epsilon > 0,\ \exists t_2(v,\epsilon),\ \forall t \geq t_2(v,\epsilon),\ kl\left(v, v + \sqrt{\frac{v(1-v)}{F(b^*_{v,F})t}}\right) \leq \frac{1+\epsilon}{2F(b^*_{v,F})t}$ which is proved by observing that $kl(v,v') = \frac{(v'-v)^2}{2}\int_0^1 g''(v'+s(v'+s(v-v'))2(1-s)ds$, where $g(x) = kl(x,v')$; and that thanks to Taylor's inequality,

$$\begin{aligned}
kl(v,v') &\leq \frac{(v'-v)^2}{2}\int_0^1 2\max_{u\in[v,v']} g''(u)ds \\
&\leq (v'-v)^2 \frac{1}{\min_{u\in[v,v']} u(1-u)}
\end{aligned}$$

and that $\forall \epsilon > 0,\ \exists t_2(v,\epsilon)$, such that $\min_{u\in[v,v']} u(1-u) < \frac{1+\epsilon}{v(1-v)}$. Putting all the pieces together yields
$\forall t \geq \max(t_1(\epsilon), t_2(v,\epsilon))$,

$$\mathbb{E}_v[(B_t - b^*_{v,F})^2] \geq \frac{v(1-v)}{4F(b^*_{v,F})t}\left(1 - \sqrt{\frac{1}{4}(1+\epsilon)^2} - 1/t^\gamma\right).$$

Let $t_0(v,\epsilon) = \max(t_1(\epsilon), t_2(v,\epsilon))$. We obtain

$$\sum_{t=1}^T \mathbb{E}_v[(B_t - b^*_{v,F})^2] \geq \sum_{t=t_0(v,\epsilon)}^T \lambda^2 \frac{v(1-v)}{4F(b^*_{v,F})t}\left(1 - \frac{1}{2}(1+\epsilon) - 1/t^\gamma\right).$$

Recall that, according to Lemma 6,

$$R_T(v) = \sum_{t=1}^{T} \mathbb{E}\left[U(b^*_{v,F}) - U(B_t)\right] \geq \frac{U(b^*_{v,F})}{4} \sum_{t=1}^{T} \mathbb{E}_v[(Q_t - q^*)^2] \geq \frac{c_f^2 U(b^*_{v,F})}{4} \sum_{t=1}^{T} \mathbb{E}_v[(B_t - b^*_{v,F})^2].$$

Hence, $\forall \epsilon > 0$,

$$R_T(v) \geq \lambda^2 \frac{c_f^2 U(b^*_{v,F})}{4} \left(\frac{v(1-v)}{4} \left(1 - \frac{1}{2}(1+\epsilon) \right) \right) \log \frac{T}{t_0(v,\epsilon)} - O(1).$$

And $\forall \epsilon > 0$,

$$\liminf_{T\to\infty} \frac{R_T(v)}{\log T} \geq \frac{c_f^2 \lambda^2 U(b^*_{v,F})}{4} \left(\frac{v(1-v)}{4F(b^*_{v,F})} \left(1 - \frac{1}{2}(1+\epsilon) \right) \right).$$

Since this holds for all ϵ,

$$\liminf_{T\to\infty} \frac{R_T(v)}{\log T} \geq \lambda^2 c_f^2 \left(\frac{v(1-v)(v - b^*_{v,F})}{32} \right).$$

■

Appendix E. Unknown F

E.1. Upper Bound of the Regret of O-UCBid1

Theorem 23 *O-UCBid1 incurs a regret bounded by*

$$R_T \leq \frac{4\sqrt{2}}{F(b^*)} \sqrt{\gamma \log T} (\sqrt{T} + 1) + O(1).$$

We first observe that the algorithm overbids ($B_t > b^*$) when F and v belong to their confidence regions $\mathbb{F}_t = \{\tilde{F}, \|F - \hat{F}_t\| \leq \epsilon_t\}$ and $\mathbb{V}_t = [v - \epsilon_t, v + \epsilon_t]$.

Lemma 24 *The bid submitted by O-UCBid1 is an upper bound of b^* when $\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t$.*

$$\left\{ \ \|\hat{U}_t - U\|_\infty \leq 2\epsilon_t \ \right\} \text{ implies } b^* \leq B_t.$$

Proof Let us pick $\underline{b} \in \arg\max \hat{U}_t$.

$$\hat{U}_t(\underline{b}) - \hat{U}_t(b^*) = \hat{U}_t(\underline{b}) - U(b^*) + U(b^*) - \hat{U}_t(b^*) \leq 4\epsilon_t.$$

We deduce that $\hat{U}_t(b^*) \geq \hat{U}_t(\underline{b}) - 4\epsilon_t \geq \max \hat{U}_t - 4\epsilon_t$.
Hence, $b^* \in \left\{ b \in [0,1], \hat{U}_t(b) \geq \max \hat{U}_t - 2\epsilon_t \right\}$. By definition of B_t, this yields $B_t \geq b^*$. ■

Next we observe that if F and v lie in their confidence regions $\mathbb{F}_t$ and $\mathbb{V}_t$, then $\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t$. (Recall that $\hat{U}_t(b) = (\hat{V}_t - b)\hat{F}_t(b)$.) Indeed, we have

$$\begin{aligned}\hat{U}_t(b) - U(b) &= (\hat{V}_t - b)\hat{F}_t(b) - (v-b)F(b) \\ &= (\hat{V}_t - v)F(b) + \hat{V}_t(\hat{F}_t(b) - F(b)) + b(F(b) - \hat{F}_b) \\ &= (\hat{V}_t - v)F(b) + (\hat{V}_t - b)(\hat{F}_t(b) - F(b))\end{aligned}$$

which yields

$$|\hat{U}_t(b) - U(b)| \leq |\hat{V}_t - v| + \|F(b) - \hat{F}_t(b)\|_\infty. \tag{12}$$

We then decompose the regret into

$$\begin{aligned}E(R_T) &= \sum_{t=1}^{T} \mathbb{E}(U(b^*) - U(B_t)) \\ &\leq 1 + \sum_{t=}^{T} \mathbb{P}(F \notin \mathbb{F}_t \text{ or } v \notin \mathbb{V}_t) + \sum_{t=2}^{T} \mathbb{E}\left(S_t \mathbb{1}(B_t > b^*)\mathbb{1}(\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t\right).\end{aligned} \tag{13}$$

The second term of the second hand side of Equation 13 is easily bounded thanks to the concentration inequalities in Lemmas 16 and 17. In fact, combining these latter lemmas yields the following bound.

Lemma 25

$$\sum_{t=2}^{T} \mathbb{P}(F \notin \mathbb{F}_t \text{ or } v \notin \mathbb{V}_t) \leq 2 \sum_{t=1}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma}$$

We apply Lemma 18 to bound the third term of the second hand side of Equation 13 as follows:

$$\begin{aligned}&\mathbb{E}\Big[]S_t \mathbb{1}(B_t > b^*)\mathbb{1}(\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t)\Big] \\ &\leq \frac{1}{F(b^*)}\mathbb{E}\Big[U(b^*) - U(B_t)) \times \mathbb{1}(M_t \leq B_t)\mathbb{1}(\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t)\mathbb{1}(B_t > b^*)\Big],\end{aligned} \tag{14}$$

because $\mathbb{1}(B_t > b^*)\mathbb{1}(\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t)$ is $\mathcal{F}_{t-1}$-measurable. We then bound the deviation $(U(b^*) - U(B_t))\mathbb{1}(M_t \leq B_t)$ by $8\epsilon_t$ by using Lemma 20.

Lemma 26 *When applying the O-UCBid1 strategy, if $\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t$, then*

$$|U(B_t) - U(b^*)| \leq 8\epsilon_t.$$

Proof Assume $\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t$. Note that $\hat{U}_t(B_t) - \hat{U}_t(b^*) = \hat{U}_t(B_t) - \hat{U}_t(\hat{b}) + \hat{U}_t(\hat{b}) - \hat{U}_t(b^*)$, where $\hat{b} = \max \arg\max_{b \in [0,1]} (\hat{V}_t - b)\hat{F}_t(b)$.

By design , we have $\hat{U}_t(B_t) - \hat{U}_t(\hat{b}) = -2\epsilon_t$. Thanks to Lemma 20, and because $\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t$ we know that $0 \leq \hat{U}_t(\hat{b}) - \hat{U}_t(b^*) \leq 4\epsilon_t$. This yields $|\hat{U}_t(B_t) - \hat{U}_t(b^*)| \leq 4\epsilon_t$.

Finally

$$|U(B_t) - U(b^*)| \leq 8\epsilon_t.$$

■

Then, by summing, we get

$$\begin{aligned}
&\sum_{t=2}^{T} \mathbb{E}\Big[]S_t \mathbb{1}(B_t > b^*)\mathbb{1}(\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t)\Big] \\
&\leq \sum_{t=2}^{T} \frac{1}{F(b^*)} \mathbb{E}\Big(U(b^*) - U(B_t)) \times \mathbb{1}(M_t \leq B_t)\mathbb{1}(B_t > b^*)\mathbb{1}(\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t,\ F \in \mathbb{F}_t,\ v \in \mathbb{V}_t)\Big) \\
&\leq \sum_{t=2}^{T} \frac{1}{F(b^*)} \mathbb{E}\Big[8\epsilon_t \times \mathbb{1}(M_t \leq B_t)\mathbb{1}(\|U - \hat{U}_t\|_\infty \leq 2\epsilon_t)\mathbb{1}(B_t > b^*)\Big] \\
&\leq \sum_{t=2}^{T} \frac{1}{F(b^*)} \mathbb{E}\Big[8\sqrt{\frac{\log T}{2N_t}}\mathbb{1}(M_t \leq B_t)\Big] \\
&\leq \frac{1}{F(b^*)} 4\sqrt{2\log T}(\sqrt{T}+1),
\end{aligned}$$

where the last inequality comes from Lemma 19. Using Equation 13 and Lemma 25 yields

$$R_T \leq \frac{1}{F(b^*)} 4\sqrt{2\log T}(\sqrt{T}+1) + \sum_{t=2}^{T} 2e\sqrt{\gamma}(\log t)t^{-\gamma}.$$

Consequently, when $\gamma > 1$,

$$R_T \leq \frac{1}{F(b^*)} 4\sqrt{2\log T}(\sqrt{T}+1) + O(1).$$

E.2. General Upper Bound of the Regret of UCBid1+

We prove a slightly different version of Theorem 2 than that of the main paper.

Theorem 2 *UCBid1+ incurs a regret bounded by*

$$\begin{aligned}
R_T &\leq 12\sqrt{\frac{\gamma\alpha}{F(b^*)}}\sqrt{\log T}\sqrt{T} + O(\log T) \\
&\leq 12\frac{1}{U(b^*)}\sqrt{v\gamma}\sqrt{\log T}\sqrt{T} + O(\log T),
\end{aligned}$$

where $\alpha := \frac{v}{v-b^*}$, *provided that* $\gamma > 2$.

Proof We denote by $\mathcal{E}$ the event $\{\forall t_0 < t < T,\ |\hat{V}_t - v| \leq \epsilon_t, \|F - \hat{F}_t\|_\infty \leq \sqrt{\frac{\gamma\log(t-1)}{2(t-1)}}\}$, where $t_0 := \min(3, 1 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right))$.

Using Lemmas 16 and 17, this event happens with high probability, when $\gamma > 2$.

Lemma 27 *The probability of the complementary of $\mathcal{E}$ is bounded as follows*

$$\mathbb{P}\left(\mathcal{E}^C\right) \leq 4e(\gamma-1)(\log T)(T)^{1-\gamma}.$$

provided that $\gamma > 2$.

Proof
We have

$$\begin{aligned}
\mathbb{P}\left(\exists t \in [t_0, T],\ (\hat{V}(N_t) - v)^2 \geq \frac{\gamma \log(t-1)}{2N_t}\right) &\leq \mathbb{P}\left(\exists t \in [2, T],\ (\hat{V}(N_t) - v)^2 \geq \frac{\gamma \log(t-1)}{2N_t}\right) \\
&\leq \sum_{t=2}^{T} \mathbb{P}\left((\hat{V}(N_t) - v)^2 \geq \frac{\gamma \log(t-1)}{2N_t},\right) \\
&\leq \sum_{t=1}^{T} 2e \log(t) t^{-\gamma} \\
&\leq \int_{u=1}^{T} 2e \log(t) u^{-\gamma} du \\
&\leq 2e(\gamma-1)\log(T)(T)^{1-\gamma},
\end{aligned}$$

thanks to Lemma 16. Similarly,

$$\begin{aligned}
\mathbb{P}\left(\exists t \in [t_0, T],\ \|F - \hat{F}\|_\infty \geq \sqrt{\frac{\gamma \log(t-1)}{2N_t}}\right) &\leq \sum_{t=t_0}^{T} \mathbb{P}\left(\ \|F - \hat{F}\|_\infty \geq \sqrt{\frac{\gamma \log(t-1)}{2N_t}}\right) \\
&\leq 2 \sum_{t=t_0}^{T} t^{-\gamma} \\
&\leq \int_{u=2}^{T} 2u^{-\gamma} du \\
&\leq 2(\gamma-1)(T)^{1-\gamma}
\end{aligned}$$

thanks to Lemma 17. ■

When $\mathcal{E}$ occurs, it is possible to prove that $F(B_t)$ is lower-bounded by a positive constant as soon as t is large enough.

Lemma 28 *On $\mathcal{E}$, provided that $t > t_0 := \min\left(3, 1 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2} \log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right)\right)$, $F(B_t)$ is lower bounded by*

$$F(B_t) > \frac{F(b^*)}{2\alpha},$$

where $\alpha = \frac{v}{v-b^*}$.

Proof $b^* = \frac{\alpha-1}{\alpha} v$. Since we are on $\mathcal{E}$,

$$b^* \leq \frac{\alpha-1}{\alpha}(\hat{V}_t + \epsilon_t).$$

Hence

$$\hat{V}_t + \epsilon_t \leq \alpha(\hat{V}_t + \epsilon_t - b^*).$$

Since $B_t > 0$,

$$\hat{V}_t + \epsilon_t - B_t \leq \alpha(\hat{V}_t + \epsilon_t - b^*).$$

And

$$\frac{\hat{V}_t + \epsilon_t - B_t}{\hat{V}_t + \epsilon_t - b^*} \leq \alpha.$$

By definition of B_t,

$$(\hat{V}_t + \epsilon_t - B_t)\hat{F}_t(B_t) \geq (\hat{V}_t + \epsilon_t - b^*)\hat{F}_t(b^*)$$

which implies

$$\hat{F}_t(B_t) \geq \frac{\hat{V}_t + \epsilon_t - b^*}{\hat{V}_t + \epsilon_t - B_t}\hat{F}_t(b^*) \geq \frac{1}{\alpha}\hat{F}_t(b^*)$$

Now,

$$\begin{aligned} F(B_t) &\geq \hat{F}_t(B_t) - \sqrt{\frac{\gamma \log(t-1)}{2(t-1)}} \\ &\geq \frac{1}{\alpha}\hat{F}_t(b^*) - \sqrt{\frac{\gamma \log(t-1)}{2(t-1)}} \\ &\geq \frac{1}{\alpha}F(b^*) - \left(\frac{1}{\alpha} + 1\right)\sqrt{\frac{\gamma \log(t-1)}{2(t-1)}}, \end{aligned}$$

because we assume that we are on $\mathcal{E}$. Note that if $t > t_0$, then

$$\frac{4\gamma(\alpha+1)^2}{F(b^*)^2} < \frac{(t-1)}{\log(t-1)},$$

thanks to Lemma 21, and

$$\left(\frac{1}{\alpha} + 1\right)\sqrt{\frac{\gamma \log(t-1)}{2(t-1)}} < \frac{1}{2\alpha}F(b^*),$$

so that

$$F(B_t) \geq \frac{F(b^*)}{2\alpha},$$

which concludes the proof. ∎

Lemma 29 $\forall t > t_0$,

$$\mathbb{P}\left(N_t < \frac{1}{4\alpha}F(b^*)(t-t_0), \mathcal{E}\right) \leq \exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right).$$

Proof Indeed if $t \geq t_0$, then N_t is larger than the sum N'_t of $t - t_0$ samples from a Bernoulli distribution with average $\frac{1}{2\alpha}F(b^*)$, hence the probability that $N_t < \frac{1}{4\alpha}F(b^*)(t-t_0)$ intersected with $\mathcal{E}$ can be bounded as follows.

$$\begin{aligned}
&\mathbb{P}\left(N_t < \frac{1}{4\alpha}F(b^*)(t-t_0), \mathcal{E}\right) \\
&\leq \mathbb{P}\left(N'_t < +\frac{1}{4\alpha}F(b^*)(t-t_0)\right) \\
&\leq \mathbb{P}\left(\frac{1}{2\alpha}F(b^*)(t-t_0) - (N'_t - t_0) > \frac{1}{4\alpha}F(b^*)(t-t_0)\right) \\
&\leq \exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right) \\
&\leq \exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right),
\end{aligned}$$

where we used Hoeffding's inequality for the third inequality. ■

Finally, we can prove that the expected instantaneous regret conditioned on B_t is bounded by a multiple of ϵ_t.

Lemma 30

$$U(B_t) - U(b^*) \leq 6\epsilon_t$$

Proof

Thanks to Equation 12, we have $\|\hat{U}_t - U\|_\infty \leq 2\epsilon_t$. Very similarly we have

$$\|U_t^{UCBid1+} - \hat{U}\|_\infty = \max_{b\in[0,1]} |\epsilon_t \hat{F}_t(b)| \leq \epsilon_t,$$

where $U^{UCBid1+} : b \mapsto (\hat{V}_t + \epsilon_t - b)\hat{F}_t(b)$. Hence,

$$\|U_t^{UCBid1+} - U\|_\infty \leq 3\epsilon_t.$$

By Lemma 20, this yields

$$U(B_t) - U(b^*) \leq 6\epsilon_t$$

■

Proof of the Theorem We use the following decomposition

$$\begin{aligned}
R_T &\leq T \times \mathbb{P}(\mathcal{E}^c) + \sum_{t=1}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E}\}] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + t_0 + \sum_{t=t_0}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E}\}]
\end{aligned}$$

Thanks to Lemma 28, and when $t > t_0$, $F(B_t) \geq \frac{1}{2\alpha}F(b^*)$. Using this, we get $N_t > \frac{1}{4\alpha}F(b^*)(t-t_0), \forall t > t_0$ with high probability.

Thanks to Lemma 29,

$$\begin{aligned}\mathbb{E}[S_t\mathbb{1}\{\mathcal{E}\}] &\leq \exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right) + \mathbb{E}\left[S_t\mathbb{1}\{N_t \geq \frac{1}{4\alpha}F(b^*)(t-t_0)\}\right] \\ &\leq \exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right) + \mathbb{E}\left[6\sqrt{\frac{4\alpha\gamma\log T}{F(b^*)(t-t_0)}}\mathbb{1}\{N_t \geq \frac{1}{4\alpha}F(b^*)(t-t_0)\}\right];\end{aligned}$$

By summing,

$$\begin{aligned}\sum_{t=t_0}^{T}\mathbb{E}[S_t\mathbb{1}\{\mathcal{E}\}] &\leq \sum_{t=t_0}^{T}\exp\left(-\frac{2((\frac{1}{2\alpha}F(b^*))^2}{4}(t-t_0)\right) + \sum_{t=t_0}^{T}6\sqrt{\frac{4\alpha\gamma\log T}{F(b^*)(t-t_0)}} \\ &\leq \frac{1}{1-\exp(-\frac{2(\frac{1}{2\alpha}F(b^*))^2}{4})} + 6\sqrt{\frac{4\alpha\gamma}{F(b^*)}}\sqrt{\log T}\sqrt{T} \\ &\leq \frac{4}{\frac{1}{2\alpha}F(b^*)} + 6\sqrt{\frac{4\alpha\gamma}{F(b^*)}}\sqrt{\log T}(\sqrt{T}),\end{aligned}$$

where the last inequality comes from $1-\exp(-u) \geq 2/u$, for any positive u. Using the decomposition of the regret yields

$$\begin{aligned}R_T &\leq t_0 + T\mathbb{P}(\mathcal{E}^C) + \frac{4}{\frac{1}{2\alpha}F(b^*)} + 6\sqrt{\frac{4\alpha}{F(b^*)}}\sqrt{\log T}\sqrt{T} \\ &\leq 4 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right) + 4e(\gamma-1)\log T(T)^{2-\gamma} + \frac{8\alpha}{F(b^*)} + 12\sqrt{\frac{\alpha\gamma}{F(b^*)}}\sqrt{\log T}\sqrt{T} \\ &\leq 4 + \frac{8\alpha}{F(b^*)} + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right) + 4e(\gamma-1)\log T + 12\sqrt{\frac{\alpha\gamma}{F(b^*)}}\sqrt{\log T}(\sqrt{T}),\end{aligned}$$

which concludes the proof. ■

E.3. Proof of an Intermediary Regret Rate under Assumptions 1 and 2

In this section, we prove an easier version of Theorem 13. We will use lemmas of the previous subsection for this version as well as for the more complex version. In particular we have already proven that $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$, occurs with high probability. Under Assumptions 1 and 2 and on this event, we prove the following result.

Lemma 31 *Under Assumptions 1 and 2 and if* $t > \max(t_0, t_1)$,

- $\|F - \hat{F}_t\|_\infty \leq \epsilon_t^+$ *and* $|v - \hat{V}_t| \leq \epsilon_t^+$,
- $|U(b^*) - U(B_t)| \leq 6\epsilon_t^+$

- $|b^* - B_t| \leq \Delta$,
- $|b^* - B_t| \leq 1/\sqrt{c_U}\sqrt{6\epsilon_t^+}$.
- $|U(b^*) - U(B_t)| \leq C_U(b^* - B_t)^2$

on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$, *where* $\begin{cases} t_0 = \min\left(3, 1 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right)\right) \\ t_1 = 2\sqrt{C_u}\Delta^{1/4}\frac{\gamma\alpha}{F(b^*)}\log T, \\ \epsilon_t^+ = \sqrt{\frac{2\alpha\gamma\log t}{F(b^*)t}}, \\ c_U = c_f\frac{1}{4}U(b^*), \\ C_U = \frac{C_f}{c_f}\lambda. \end{cases}$

Proof On the event $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$, $\|F - \hat{F}_t\|_\infty \leq \epsilon_t^+$ and $|v - \hat{V}_t| \leq \epsilon_t^+$ where $\epsilon_t^+ = \sqrt{\frac{2\alpha\gamma\log t}{F(b^*)t}}$ from Lemmas, 16,17 29 and $|U(b^*) - U(B_t)| \leq 6\epsilon_t \leq 6\epsilon_t^+$ from Lemmas 30 and 29.

Under Assumptions 1 and 2, we prove that after t_1, we have $|B_t - b^*| \leq \Delta$ on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$, so that we will be able to use the boundedness of the density after this time step.

When F satisfies assumption 1, U is unimodal, as shown in the proof of Lemma 3, and so if

$$U(b^*) - U(b) \leq \min(U(b^*) - U(b^* - \Delta), U(b^*) - U(b^* + \Delta)),$$

then

$$b \in [b^* - \Delta, b^* + \Delta].$$

It follows that if

$$6\epsilon_t^+ \leq \min(U(b^*) - U(b^* - \Delta), U(b^*) - U(b^* + \Delta))$$

and therefore $6\epsilon_t^+ \leq C_u\Delta$ where $C_u := \lambda C_f/c_f$ (see Lemma 7), then

$$|b^* - B_t| \leq \Delta$$

on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$. Then, for all $t > 2\sqrt{c_u}\Delta^{1/4}\frac{\gamma\alpha}{F(b^*)}\log T := t_1$, we have $|B_t - b^*| \leq \Delta$ on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$.

Under Assumption 1, for any $q \in [0,1]$, $W_{v,F}(q^*_{v,F}) - W_{v,F}(q) \geq \frac{1}{4}(q^*_{v,F} - q)^2 W_{v,F}(q^*_{v,F})$. We have $U = W \circ F$, so that if $t > t_1$, then $B_t \in [b^* - \Delta, b^* + \Delta]$ and $U(b^*) - U(B_t) \geq c_f\frac{1}{4}(b^* - B_t)^2 U(b*) := c_U(b^* - B_t)^2$. In this case, we can also prove that $|b^* - B_t| \leq 1/\sqrt{c_U}\sqrt{6\epsilon_t^+}$, under $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$. ■

Proposition 32 *Under Assumptions 1 and 2 and if* $t > \max(t_0, t_1)$, $\delta_t < \Delta$, $|B_t - b^*| \leq \delta_t$, *and* $\epsilon_t^+ \leq M\delta_t$,
Then

$$|B_t - b^*|^2 \leq \frac{6}{c_U}\sqrt{\frac{C_f\delta_t\log\left(\frac{Me^2t\sqrt{2t}}{2c_f\eta^2}\right)}{t}} + \frac{2\log(\frac{Mt\sqrt{2t}}{2c_f\eta^2})}{c_U t} + \frac{2}{c_U}(2C_f + 1)\delta_t\sqrt{\frac{2\alpha\gamma\log T}{F(b^*)t}},$$

with probability $1-\eta$ *on* $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$.

Proof It is clear from Lemma 12 that

$$\sup_{b^*-\delta_t \leq b \leq b^*+\delta_t} |\hat{F}_t(b) - F(b) - (\hat{F}_t(b^*) - F(b^*))| \leq 2\sqrt{\frac{2C_f\delta_t \log\left(\frac{e\sqrt{t}}{\sqrt{2c_f\delta_t}\eta}\right)}{t}} + 2\frac{\log(\frac{t}{2c_f\delta_t\eta^2})}{6t} := \beta_t,$$

with probability $1-\eta$. We can also decompose $U(b) - U_t^{UCBid1+}(b) - (U_t^{UCBid1+}(b^*) - U(b^*))$ into

$$\begin{aligned}
&U(b) - U_t^{UCBid1+}(b) - (U_t^{UCBid1+}(b^*) - U(b^*)) \\
&= (v-b)F(b) - (\hat{V}_t + \epsilon_t - b)\hat{F}_t(b) - \left((v-b^*)F(b^*) - (\hat{V}_t + \epsilon_t - b^*)\hat{F^*}_t(b)\right) \\
&= (v-b)F(b) - (v-b)\hat{F}_t(b) - \left((v-b^*)F(b^*) - (v-b^*)\hat{F^*}_t(b)\right) - (\hat{V}_t + \epsilon_t - v)\left(\hat{F}_t(b) - \hat{F}_t(b^*)\right) \\
&= (v-b^*)\left(F(b) - \hat{F}_t(b) - \left(F(b^*) - \hat{F^*}_t(b)\right)\right) - (\hat{V}_t + \epsilon_t - v)\left(\hat{F}_t(b) - \hat{F}_t(b^*)\right) \\
&\quad + (b^*-b)(\hat{F}(b) - \hat{F}_t(b))
\end{aligned}$$

which in turn proves that

$$\begin{aligned}
|U(b) - U_t^{UCBid1+}(b) - (U_t^{UCBid1+}(b^*) - U(b^*))| &\leq \beta_t + 2\epsilon_t|\hat{F}_t(b) - \hat{F}_t(b^*)| + \delta_t|\hat{F}_t(b) - \hat{F}_t(b)| \\
&\leq \beta_t + 2\epsilon_t^+(C_f\delta_t + \beta_t) + \delta_t\epsilon_t^+ \\
&\leq \beta_t + 2\epsilon_t^+\beta_t + (2C_f+1)\delta_t\epsilon_t^+ \\
&\leq 3\beta_t + (2C_f+1)\delta_t\epsilon_t^+ := \gamma_t,
\end{aligned}$$

for all b in $[b^*-\delta_t, b^*+\delta_t]$.

Now, we know that $U(b^*) - U(b)$ is lower bounded by $c_U(b^*-b)^2$, on this interval and $\|U_t^{UCBid1+}(b) - U(b) + U_t^{UCBid1+}(b^*) - U(b^*)\|_\infty \leq \gamma_t$ on $[b^*-\delta_t, b^*+\delta_t]$. We call G the shifted version of U defined by $G(b) = U(b) + U_t^{UCBid1+}(b^*) - U(b^*)$. Its argmax is b^* and $G(b^*) - G(b)$ is lower bounded by $c_U(b^*-b)^2$
then $c_U(B_t - b^*)^2 \leq G(b^*) - G(B_t) \leq 2\gamma_t$ (see Lemma 20).

Then , by definition of γ_t and β_t:

$$\begin{aligned}
(B_t - b^*)^2 &\leq \frac{6}{c_U}\sqrt{\frac{C_f\delta_t \log\left(\frac{e^2t}{2c_f\delta_t\eta^2}\right)}{t}} + \frac{2\log(\frac{t}{2c_f\delta_t\eta^2})}{c_U t} + \frac{2}{c_U}(2C_f+1)\delta_t\epsilon_t^+ \\
&\leq \frac{6}{c_U}\sqrt{\frac{C_f\delta_t \log\left(M\frac{e^2t}{2c_f\epsilon_t^+\eta^2}\right)}{t}} + \frac{2\log(\frac{Mt}{2c_f\epsilon_t^+\eta^2})}{c_U t} + \frac{2}{c_U}(2C_f+1)\delta_t\epsilon_t^+ \\
&\leq \frac{6}{c_U}\sqrt{\frac{C_f\delta_t \log\left(\frac{Me^2t\sqrt{t}}{2c_f\eta^2}\right)}{t}} + \frac{2\log(\frac{Mt\sqrt{t}}{2c_f\eta^2})}{c_U t} + \frac{2}{c_U}(2C_f+1)\delta_t\sqrt{\frac{2\alpha\gamma\log T}{F(b^*)t}}.
\end{aligned}$$

where the last inequality stems from that fact that $1/\epsilon_t^+ = \sqrt{\frac{F(b^*)t}{2\alpha\gamma\log t}} \leq \sqrt{t}$ since $\alpha, \gamma \geq 1$. ■

Theorem 33 *Under Assumptions 1 and 2,*

$$R_T \leq O(T^{3/8} \log T).$$

Proof

From Lemma 31, we have that $|b^* - B_t| \leq 1/\sqrt{c_U}\sqrt{6\epsilon_t^+}$, on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$. Therefore, we can apply Proposition 32 with $\delta_t = \frac{1}{\sqrt{c_U}}\sqrt{6\epsilon_t^+}$ with $M = \frac{\sqrt{c_U}}{\sqrt{6}}$, and $\eta = \frac{1}{t}$.

We use the general fact that $\log(At^\alpha) \leq 2\alpha \log t$ as soon as $t^\alpha > A$, for all $A, a > 0$, to derive the following two inequalities :

$\forall t \geq \left(\frac{Me^2}{2c_f}\right)^{\frac{1}{4}}$,

$$\frac{6}{c_U}\sqrt{\frac{C_f \delta_t \log\left(\frac{Me^2 t\sqrt{t}}{2c_f\eta^2}\right)}{t}} \leq \frac{6\sqrt{8}\sqrt{C_f}}{c_U^{\frac{5}{4}}}\sqrt{\frac{\delta_t \log t}{t}} = \frac{24(72\alpha\gamma)^{\frac{1}{8}}\sqrt{C_f}}{c_U^{\frac{5}{4}}F(b^*)^{\frac{1}{8}}}\sqrt{\frac{\log^2 t}{t^{\frac{5}{4}}}}.$$

$\forall t \geq (\frac{M}{2c_f})^{\frac{1}{4}}$,

$$\frac{2\log(\frac{Mt^2 t\sqrt{t}}{2c_f})}{c_U t} \leq \frac{16}{c_U}\frac{\log t}{t}.$$

We also have, for all t,

$$\begin{aligned}\frac{2}{c_U}(2C_f+1)\delta_t\sqrt{\frac{2\gamma\alpha\log T}{F(b^*)t}} &\leq \frac{2}{c_U}(2C_f+1)\sqrt{\frac{2\gamma\alpha}{F(b^*)}}\delta_t\sqrt{\frac{\log t}{t}} \\ &= \frac{2(72\alpha\gamma)^{\frac{1}{4}}}{c_U^{\frac{3}{2}}F(b^*)^{\frac{1}{4}}}(2C_f+1)\sqrt{\frac{2\gamma\alpha}{F(b^*)}}\frac{(\log t)^{\frac{1}{4}}}{t^{\frac{1}{4}}}\sqrt{\frac{\log t}{t}}\end{aligned}$$

Therefore $|B_t - b^*|^2 \leq \left(\frac{24(72\alpha\gamma)^{\frac{1}{8}}\sqrt{C_f}}{c_U^{\frac{5}{4}}F(b^*)^{\frac{1}{8}}} + \frac{16}{c_U} + \frac{2(72\alpha\gamma)^{\frac{1}{4}}}{c_U^{\frac{3}{2}}F(b^*)^{\frac{1}{4}}}(2C_f+1)\sqrt{\frac{2\gamma\alpha}{F(b^*)^{\frac{1}{8}}}}\right)\frac{\log t}{t^{\frac{5}{8}}}$ with probability $1 - \frac{1}{t}$, for $t \geq \max(\left(\frac{Me^2}{2c_f}\right)^{\frac{1}{4}}, (\frac{M}{2c_f})^{\frac{1}{4}}) := t_2$ on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha}F(b^*)t\}$. On this event, $U(b^*) - U(B_t) \leq C_U(b^* - B_t)^2$

We use the following decomposition

$$
\begin{aligned}
R_T &\leq T \times \mathbb{P}(\mathcal{E}^c) + \sum_{t=1}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E}\}] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_2) + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E}\}] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_2) + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha} F(b^*) t\}\}] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_2) + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_2)}^{T} C_U \mathbb{E}[(b^* - B_t)^2] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_2) + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_2)}^{T} C_0 \frac{\log t}{t^{\frac{5}{8}}} + \sum_{t=\max(t_0,t_1,t_2)}^{T} \frac{1}{t} \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_2) + \sum_{t=\max(t_0,t_1,t_2)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + C_0 \frac{8}{3} T^{\frac{3}{8}} \log T + \log T \\
&\leq \log T + 4e(\gamma - 1) \log T (T)^{2-\gamma} + \max(t_0, t_1, t_2) + \frac{8\alpha}{F(b^*)} + \frac{8}{3} C_0 T^{\frac{3}{8}} \log T.
\end{aligned}
$$

where
$$
\begin{cases}
t_0 = \min\left(3, 1 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2} \log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right)\right) \\
t_1 = 2\sqrt{C_u}\Delta^{1/4}\frac{\gamma\alpha}{F(b^*)} \log T, \\
t_2 = \max(\left(\frac{\sqrt{c_U}e^2}{2c_f\sqrt{6}}\right)^{\frac{1}{4}}, (\frac{\sqrt{c_U}}{2c_f\sqrt{6}})^{\frac{1}{4}}) = (\frac{\sqrt{c_U}e^2}{2c_f\sqrt{6}})^{\frac{1}{4}}, \\
C_0 = \left(\frac{24(72\alpha\gamma)^{\frac{1}{8}}\sqrt{C_f}}{c_U^{\frac{5}{4}} F(b^*)^{\frac{1}{8}}} + \frac{16}{c_U} + \frac{2(72\alpha\gamma)^{\frac{1}{4}}}{c_U^{\frac{3}{2}} F(b^*)^{\frac{1}{4}}}(2C_f + 1)\sqrt{\frac{2\gamma\alpha}{F(b^*)}}\right) C_U.
\end{cases}
$$

Therefore

$$
R_T \leq \frac{8}{3} C_0 T^{\frac{3}{8}} \log T + o(T^{\frac{3}{8}} \log T).
$$

■

E.4. Proof of Theorem 13

Theorem 33 is proved by applying Proposition 32 once. By iterating the argument, we can actually achieve a regret of the order of T^a, for any $a > \frac{1}{3}$. The proof involves an induction argument. The following lemma is the main element of the proof of the induction.

Lemma 34 *Assume that t and F satisfy the assumptions of Proposition 32. Assume that $|B_t - b^*|$ is bounded by $\delta_t^{(k)}$ such that $\delta_t^{(k)} = \min(1, C^{(k)} \log(t) t^{-u_k})$ with probability $1 - \eta^{(k)}$, and $u_k < 2/3$, $C^{(k)} \geq 1$. Then $|B_t - b^*|$ is bounded by $\delta_t^{(k+1)}$ such that $\delta_t^{(k+1)} = \min(1, C^{(k+1)} \log(t) t^{-\frac{1}{4}(1+u_k)})$ with probability $1 - \eta^{(k)} - \frac{1}{Kt}$,*
where $C^{(k+1)} = C \left(C^{(k)}\right)^{\frac{1}{4}}$ and where $C = \max\left(1, \frac{12\sqrt{2C_f}}{c_U} + \frac{16}{c_U} + \frac{2}{c_U}(2C_f+1)\sqrt{\frac{2\gamma\alpha}{F(b^*}}\right)$.

Proof

We use Proposition 32, and the fact that $\epsilon_t^+ \leq \sqrt{2\alpha\gamma/F(b^*)}\frac{\log t}{t^{-u_k}} \leq \sqrt{2\alpha\gamma/F(b^*)}\delta_t^{(k)}$ to prove that

$$|B_t - b^*|^2 \leq \frac{6}{c_U}\sqrt{\frac{C_f \delta_t^{(k)} \log\left(\frac{Me^2 t\sqrt{2t}K^2t^2}{2c_f}\right)}{t}} + \frac{2\log(\frac{MK^2t^2t\sqrt{2t}}{2c_f})}{c_U t} + \frac{2}{c_U}(2C_f+1)\delta_t^{(k)}\sqrt{\frac{2\alpha\gamma\log t}{F(b^*)t}},$$

with probability $(1-\eta^{(k)})(1-\frac{1}{Kt})$ and with $M = \sqrt{2\alpha\gamma/F(b^*)}$

We use the general fact that $\log(At^\alpha) \leq 2\alpha \log t$ as soon as $t^\alpha > A$, for all $A, a > 0$, to derive the following two inequalities :

$\forall t \geq \left(\frac{Me^2K^2}{2c_f}\right)^{\frac{1}{4}}$,

$$\frac{6}{c_U}\sqrt{\frac{C_f \delta_t^{(k)} \log\left(\frac{Me^2 t\sqrt{2t}K^2t^2}{2c_f}\right)}{t}} \leq \frac{6\sqrt{8C_f}}{c_U}\sqrt{\frac{\delta_t^{(k)}\log t}{t}} := C_1\sqrt{\frac{\delta_t^{(k)}\log t}{t}} := C_1\beta_{1,t}.$$

$\forall t \geq (\frac{MK^2}{2c_f})^{\frac{1}{4}}$,

$$\frac{2\log(\frac{MK^2t^2t\sqrt{2t}}{2c_f})}{c_U t} \leq \frac{16}{c_U}\frac{\log t}{t} := C_2\frac{\log t}{t} := C_2\beta_{2,t}.$$

We also have, for all t,

$$\frac{2}{c_U}(2C_f+1)\delta_t^{(k)}\sqrt{\frac{2\alpha\gamma\log T}{F(b^*)t}} \leq \frac{2}{c_U}(2C_f+1)\sqrt{\frac{2\gamma\alpha}{F(b^*}}\delta_t^{(k)}\sqrt{\frac{\log t}{t}} := C_3\delta_t^{(k)}\sqrt{\frac{\log t}{t}} := C_3\beta_{3,t}$$

We can derive the following bounds

- $\beta_{3,t} \leq \beta_{1,t}$ since $\delta_t^{(k)} \leq 1$.
- $\beta_{2,t} \leq \beta_{1,t}$ since $\delta_t^{(k)} = \min(1, C^{(k)}\log(t)t^{-u_k}) \geq \frac{\log t}{t}$.

Hence

$$|B_t - b^*|^2 \leq (C_1 + C_2 + C_3)\beta_{1,t} = (C_1 + C_2 + C_3)\sqrt{\frac{\delta_t^{(k)} \log t}{t}},$$

with probability $1 - \eta^{(k)} \frac{1}{Kt}$. This yields

$$\begin{aligned}
|B_t - b^*| &\leq \sqrt{(C_1 + C_2 + C_3)} \left(\frac{\delta_t^{(k)} \log t}{t} \right)^{\frac{1}{4}} \\
&\leq \sqrt{(C_1 + C_2 + C_3)} \left(\frac{\min(1, C^{(k)} \log^2(t) t^{-u_k})}{t} \right)^{\frac{1}{4}} \\
&\leq \sqrt{(C_1 + C_2 + C_3)} (C^{(k)})^{1/4} t^{-\frac{1}{4}(1+u_k)} \log t \\
&\leq C \left(C^{(k)} \right)^{1/4} t^{-\frac{1}{4}(1+u_k)} \log t,
\end{aligned}$$

■

Proposition 35 *Assume that t and F satisfy the assumptions of Proposition 32. If $t > t_3 = max\left((\frac{\sqrt{2\alpha\gamma/F(b^*)}K^2}{2c_f})^{\frac{1}{4}}, (\frac{\sqrt{2\alpha\gamma/F(b^*)}e^2K^2}{2c_f})^{\frac{1}{4}} \right)$, then on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha} F(b^*)t\}$,*

$$|B_t - b^*| \leq C^{(0)} C^{\frac{1}{3}} \log(t) t^{-\frac{1}{3} + \frac{1}{3 \times 4^K} + \frac{1}{4^{K+1}}},$$

with probability $1 - \frac{1}{t}$ where $C = \max\left(1, \frac{12\sqrt{2C_f}}{c_U} + \frac{16}{c_U} + \frac{2}{c_U}(2C_f + 1)\sqrt{\frac{2\gamma\alpha}{F(b^*)}}\right)$, and $C^{(0)} = \max\left(1, \sqrt{\frac{1}{c_U}} \left(\frac{72\gamma\alpha}{F(b^*)}\right)^{\frac{1}{4}}\right)$

Proof The proposition follows from using an induction argument based on Lemma 34. We can initiate an induction argument with $\delta_t^{(0)}$ such that

$$\delta_t^{(0)} = \min(1, C^{(0)} \log(t) t^{-u_k}),$$

writing $u_0 = \frac{1}{4}$ and $C^{(0)} = \max(1, \sqrt{\frac{1}{c_U}} \left(\frac{72\alpha\gamma}{F(b^*)}\right)^{1/4})$, thanks to Lemma 31. The fact that u_k and $C^{(k)}$ as defined as in Lemma 34 satisfy $u_{k+1} = \frac{1}{4}(1 + u_k)$ which yields

$$u_K = \left(\frac{1}{4}\right)^K u_0 + \sum_{i=1}^{K} \frac{1}{4^i} = \left(\frac{1}{4}\right)^K u_0 + 4\frac{1/4 - (1/4)^{K+1}}{3}$$

and $C^{(k+1)} = C \times (C^{(k)})^{\frac{1}{4}}$ which yields

$$C^{(K)} = \left(C^{(0)}\right)^{\frac{1}{4^K}} C^{\sum_{i=1}^{K} \frac{1}{4^i}} \leq C^{\frac{1}{3}},$$

suffices to complete the induction.

■

We recall Theorem 13.

Theorem 13 *Under Assumptions 1 and 2,*

$$R_T \leq O(T^{1/3+\epsilon}),$$

for any $\epsilon > 0$ *as long as* $\gamma > 2$.

We choose K such that $\frac{1}{3} + \frac{2}{3\times 4^K} + \frac{2}{4^{K+1}} < \frac{1}{3} + \epsilon$. (We can choose $K = \left\lceil \log_4 \left(\frac{3}{14} \frac{1}{\epsilon} \right) \right\rceil) + 1$ for example). Then, thanks to proposition 35, for all $t > t_3$, on $\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha} F(b^*)t\}$,

$$|B_t - b^*| \leq C^{(0)} C^{\frac{1}{3}} \log(t) t^{-\frac{1}{3} + \frac{1}{3\times 4^K} + \frac{1}{4^{K+1}}},$$

with probability $1 - \frac{1}{t}$. We can therefore do the same decomposition as in the proof of Theorem 33.

$$
\begin{aligned}
R_T &\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_3) + \sum_{t=\max(t_0,t_1,t_3)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_3)}^{T} \mathbb{E}[S_t \mathbb{1}\{\mathcal{E} \cap \{N_t \geq \frac{1}{4\alpha} F(b^*) t\}\}] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_3) + \sum_{t=\max(t_0,t_1,t_3)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_3)}^{T} C_U \mathbb{E}[(b^* - B_t)^2] \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_3) + \sum_{t=\max(t_0,t_1,t_3)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + \sum_{t=\max(t_0,t_1,t_3)}^{T} C^{(0)} C_U C^{\frac{1}{3}} (\log t) t^{-\frac{2}{3} + \frac{2}{3 \times 4^K} + \frac{2}{4^{K+1}}} \\
&\quad + \sum_{t=\max(t_0,t_1,t_3)}^{T} \frac{1}{t} \\
&\leq T \times \mathbb{P}(\mathcal{E}^c) + \max(t_0, t_1, t_3) + \sum_{t=\max(t_0,t_1,t_3)}^{T} \mathbb{P}(\mathcal{E} \cap \{N_t < \frac{1}{4\alpha} F(b^*) t\}) \\
&\quad + C^{(0)} C_U C^{\frac{1}{3}} \frac{1}{\frac{1}{3} + \frac{2}{3 \times 4^K} + \frac{2}{4^{K+1}}} T^{\frac{1}{3} + \frac{2}{3 \times 4^K} + \frac{2}{4^{K+1}}} \log T + \log T \\
&\leq \log T + 4e(\gamma - 1) \log T (T)^{2-\gamma} + \max(t_0, t_1, t_3) \\
&\quad + \frac{8\alpha}{F(b^*)} + 3C^{(0)} C_U C^{\frac{1}{3}} T^{\frac{1}{3} + \epsilon}.
\end{aligned}
$$

where
$$
\begin{cases}
t_0 = \min\left(3, 1 + 8\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2} \log\left(4\frac{\gamma(\alpha+1)^2}{\alpha(F(b^*))^2}\right)\right) \\
t_1 = 2\sqrt{c_u} \Delta^{1/4} \frac{\gamma\alpha}{F(b^*)} \log T, \\
t_3 = (\frac{\sqrt{2\alpha\gamma / F(b^*) e^2 K^2}}{2c_f})^{\frac{1}{4}}, \\
C^{(0)} = \max\left(1, \sqrt{\frac{1}{c_U}} \left(\frac{72\gamma\alpha}{F(b^*)}\right)^{\frac{1}{4}}\right) \\
C = \max\left(1, \frac{12\sqrt{2C_f}}{c_U} + \frac{16}{c_U} + \frac{2}{c_U}(2C_f + 1)\sqrt{\frac{2\gamma\alpha}{F(b^*)}}\right)
\end{cases}.
$$

Hence

$$
R_T \leq O(T^{1/3+\epsilon}).
$$

Appendix F. Further figures

We present in Figure 10 the histogram of the normalized data used to simulate the real-world experiment.

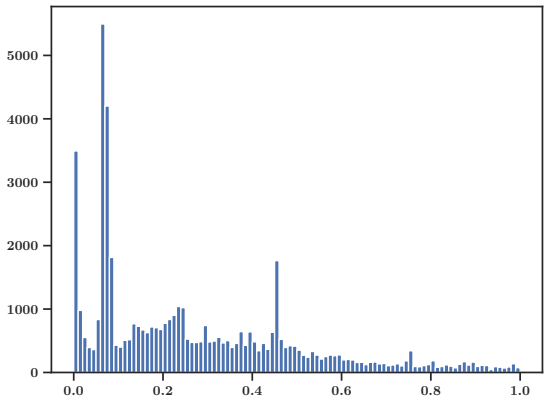


Figure 10: Bidding Data histogram