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

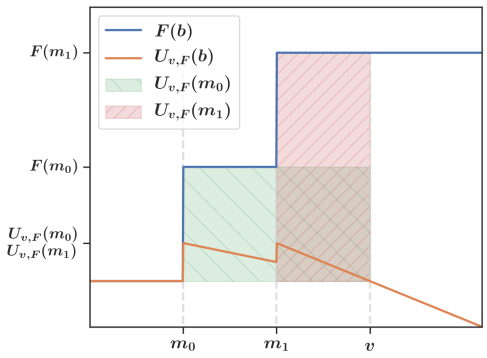


Figure 1: An example with two maximizers

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*

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

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 $1-\eta$.

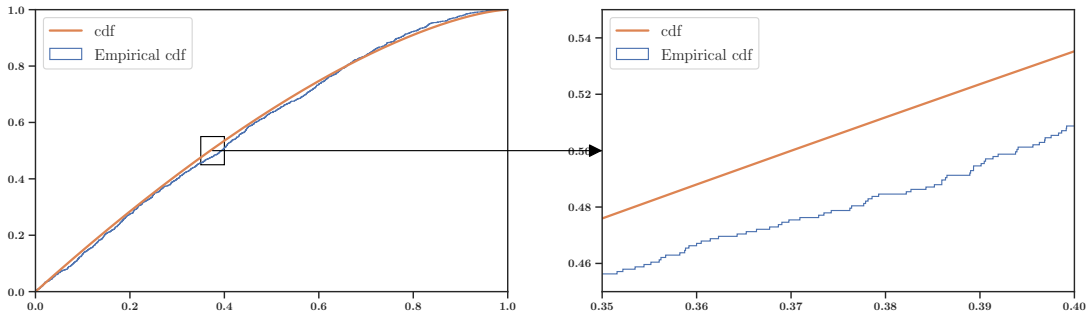


Figure 2: Local behavior of the empirical CDF

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

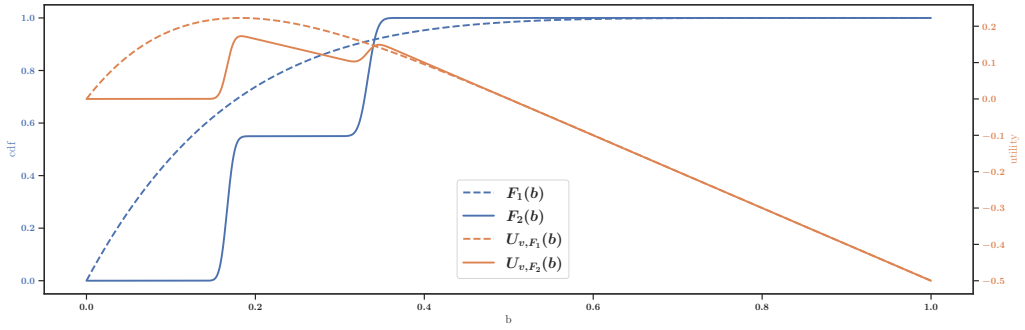
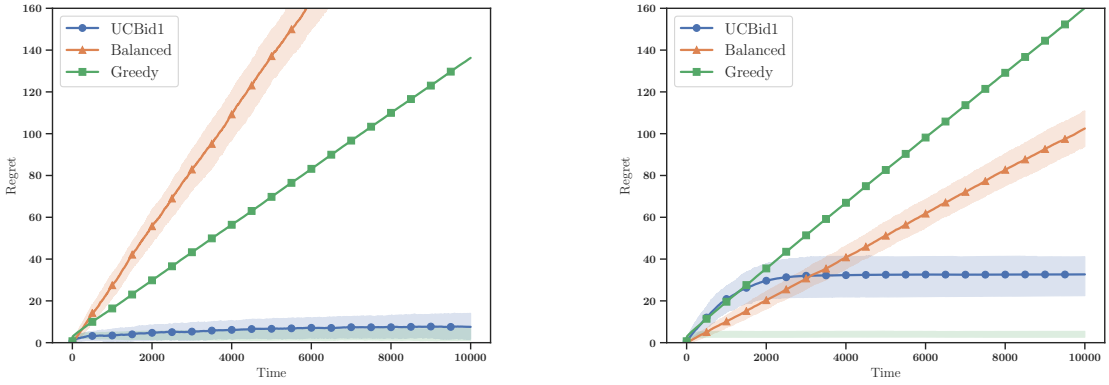


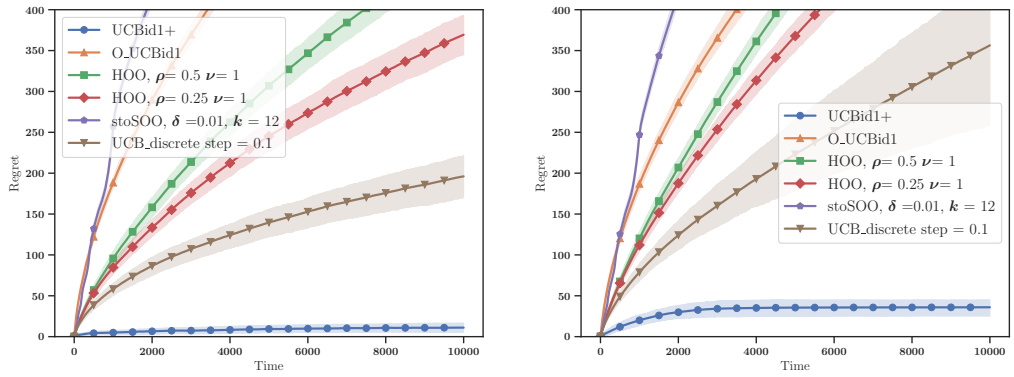
Figure 3: Two choices of F; associated utilities for $v = 1/2$.



(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

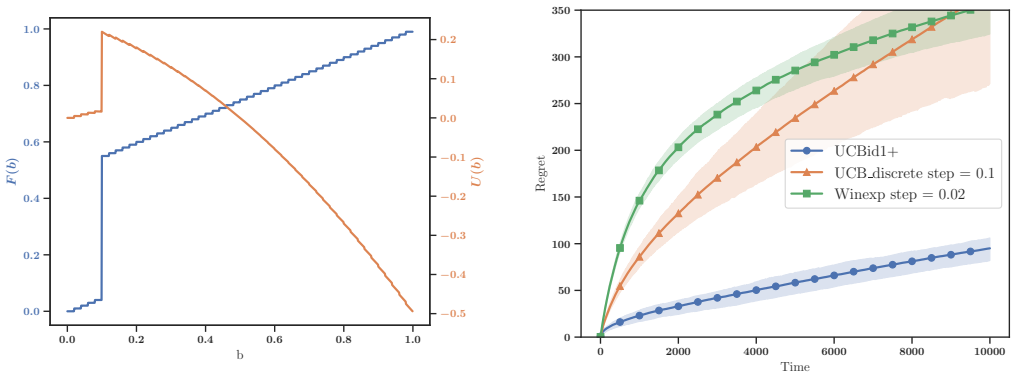
Figure 5: Regret plots for unknown F

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*) and 4(*b*) 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.



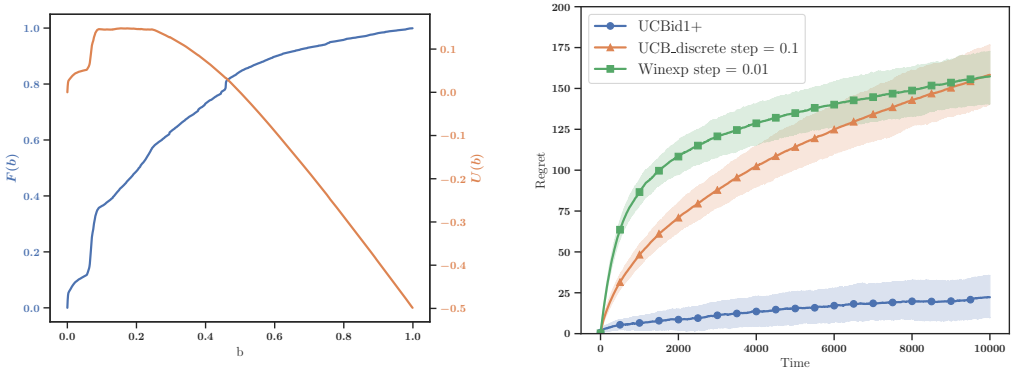
(a) Utility with $v = 1/2$ (b) Regret plots

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



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

Figure 7: Experiment with real bidding data

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.