

Regret Minimization for Reserve Prices in Second-Price Auctions

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour

Abstract—We show a regret minimization algorithm for setting the reserve price in a sequence of second-price auctions, under the assumption that all bids are independently drawn from the same unknown and arbitrary distribution. Our algorithm is computationally efficient, and achieves a regret of $\tilde{O}(\sqrt{T})$ in a sequence of T auctions. This holds even when the number of bidders is stochastic with a known distribution.

Index Terms—Prediction theory, sequential analysis, statistical learning, semi-supervised learning.

I. INTRODUCTION

CONSIDER a merchant selling items through e-Bay auctions. The sell price in each auction is the second-highest bid, and the merchant knows the price at which the item was sold, but not the individual bids from the bidders that participated in the auction. How can the merchant set a reserve price in order to optimize revenues? Similarly, consider a publisher selling advertisement space through Ad Exchange (such as AdX) or Supply Side Platform (such as AdSense), where advertisers bid for the advertisement slot and the price is the second-highest bid. With no access to the number of bidders that participate in the auction, and knowing only the actual price that was charged, how can the publisher set an optimal reserve price?

We abstract this scenario by considering the following problem: A seller is faced with repeated auctions, where each auction has a (different) set of bidders, and each bidder draws bids from some fixed unknown distribution which is the same for all bidders. It is important to remark that we need not assume that the bidders indeed bid their private value. Our assumption on the bidders' behavior, a priori, implies that if they bid using the same strategy, their bid distribution is identical.¹ The sell price is the second-highest bid, and the

seller's goal is to maximize the revenue by only relying on information regarding revenues on past auctions.

The issue of revenue maximization in second-price auctions has received a significant attention in the economics literature. The Revenue Equivalence theorem shows that truthful mechanisms² that allocate identically have identical revenue (see [15]). Myerson [14], for the case of monotone hazard rate distributions, characterized the optimal revenue maximization truthful mechanism as a second-price auction with a seller's *reserve price*, i.e., with a minimum price disqualifying any bid below it.

In addition to their theoretical relevance, reserve prices are to a large extent the main mechanism through which a seller can directly influence the auction revenue in today's electronic markets. The examples of e-Bay, AdX and AdSense are just a few in a large collection of such settings. The practical significance of optimizing reserve prices in sponsored search was reported in [16], where optimization produced a significant impact on Yahoo!'s revenue.

We stress that unlike much of the mechanism design literature (see [15]), we are not searching for the optimal revenue maximization truthful mechanism. Rather, our goal is to maximize the seller's revenue in a given, yet very popular, mechanism of second-price auction with a reserve price. In our model, the seller has only information about the auction price (and possibly about the number of bidders that participated in the auction). We assume all buyers have the same unknown bid distribution, but we make no assumptions about this distribution, only that the bids are from a bounded domain. In particular, we do not assume that the distribution has a monotone hazard rate, a traditional assumption in the economics literature. The main modeling assumption we rely upon is that buyers draw their value independently from the same distribution (i.e., bids are independent and identically distributed). This is a reasonable assumption when the auction is open to a wide audience of potential buyers. In this case, it is plausible that the seller's strategy of choosing reserve prices has no influence on the distribution of bids.

A. Our Results

The focus of our work is on setting the reserve price in a second-price auction, in order to maximize the seller's revenue. Our main result is an online algorithm that optimizes the seller's reserve price based only on the observation of the seller's actual revenue at each step. We show that after T steps

²A mechanism is truthful if it is a dominant action for the bidders to bid their private value.

Manuscript received July 11, 2013; revised August 8, 2014; accepted October 16, 2014. Date of publication October 29, 2014; date of current version December 22, 2014. N. Cesa-Bianchi was supported in part by the MIUR through the ARS TechnoMedia Project within PRIN 2010-2011 under Grant 2010N5K7EB 003. Y. Mansour was supported in part by the Israel Science Foundation, in part by the U.S.-Israel Binational Science Foundation, in part by the Ministry of Science and Technology, Israel, and in part by the Israeli Centers of Research Excellence Program under Grant 4/11.

N. Cesa-Bianchi is with the Dipartimento di Informatica, Università degli Studi di Milano, Milan 20122, Italy (e-mail: nicolo.cesa-bianchi@unimi.it).

C. Gentile is with the Dipartimento di Scienze Teoriche ed Applicate, Università dell'Insubria, Como 22100, Italy (e-mail: claudio.gentile@uninsubria.it).

Y. Mansour is with the School of Computer Science, Tel Aviv University, Tel Aviv 6997801, Israel (e-mail: mansour@tau.ac.il).

Communicated by V. Borkar, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2014.2365772

¹For example, if we had considered a first-price auction, then assuming that bidders use the same strategy to map their private value to a bid would result in the same bid distribution.

(T repetitions of the auction) our algorithm has a regret of only $\tilde{\mathcal{O}}(\sqrt{T})$. Namely, using our online algorithm the seller has an average revenue per auction that differs from that of the optimal reserve price by at most $\tilde{\mathcal{O}}(1/\sqrt{T})$, assuming the value of any bid lies in a bounded range interval.

Our algorithm is rather easy to explain and motivate at a high level. Let us start with a simple $\mathcal{O}(T^{2/3})$ regret minimization algorithm, similar to [12]. The algorithm discretizes the range of reserve prices to $\Theta(T^{1/3})$ price bins, and uses some efficient multi-armed bandit algorithm (see [5]) over the bins. It is easy to see that lowering the optimal reserve price by ϵ will result in an average loss of at most ϵ .³ This already shows that vanishing average regret is achievable, specifically, a regret of $\mathcal{O}(T^{2/3})$. Our main objective is to improve over this basic algorithm and achieve a regret of $\tilde{\mathcal{O}}(\sqrt{T})$.

An important observation to understand our algorithm is that by setting the reserve price low (say, zero) we observe the second-highest bid, since this will be the price in the auction. Hence, with enough observations, we can reconstruct the distribution of the second-highest bid. Given the assumption that the bidders' bid distributions are identical, we can recover the bid distribution of an individual bidder, and the distribution of the highest bid. Clearly, a good approximation to this distribution results in a good approximation to the optimal reserve price. Unfortunately, this simple method does not improve the regret, since a good approximation of the second-highest bid distribution incurs a significant loss in the exploration, and results in a regret of $\mathcal{O}(T^{2/3})$, similar to the regret of the discretization approach.

Our main solution is to perform only a rough estimate of the second-highest bid distribution. Using this rough estimate, we can set a better reserve price. In order to facilitate future exploration, it is important to set the new reserve price to the lowest potentially optimal reserve price. The main benefit is that our new reserve price has a lower regret with respect to the optimal reserve price, and we can bound this improved regret. We continue in this process, getting improved approximations to the optimal reserve price, and accumulating lower regret (per time step) in each successive iteration, resulting in a total regret of $\tilde{\mathcal{O}}(\sqrt{T})$ for T time steps.

Our ability to reconstruct the bid distribution depends on our knowledge about the number of participating bidders in the auction. Our simpler case involves a known number of bidders (Section II). We later extend the algorithm and analysis to the case where there is stochasticity in the number of bidders through a known distribution (Section III). In both cases we prove a regret bound of $\tilde{\mathcal{O}}(\sqrt{T})$. This bound is optimal up to logarithmic factors. In fact, simple choices of the bid distribution exist that force any algorithm to have order \sqrt{T} regret, even when there are only two bidders whose bids are revealed to the algorithm at the end of each auction.

Finally, in Section IV we present two extensions. One is for the case when the regret analysis refers to the stronger notion of realized regret (Section IV-A), the other extension

is a standard twist that removes any prior knowledge on the time horizon T (Section IV-B).

B. Related Work

There is a vast literature in Algorithmic Game Theory on second price auctions, with sponsored search as a motivating application. An important thread of research concerns the design of truthful mechanisms to maximize the revenue in the worst case, and the derivation of competitive ratio bounds, see [10]. A recent related work [8] discusses revenue maximization in a Bayesian setting. Their main result is a mechanism that achieves a constant approximation ratio with respect to *any* prior distribution using a single sample. They also show that with additional samples, the approximation ratio improves, and in some settings they even achieve a $1 - \epsilon$ approximation. In contrast, we assume a fixed but unknown prior distribution, and consider the rate at which we can approximate the optimal reserve price. In our setting, as we mentioned before, achieving a $1 - \epsilon$ approximation, even for $\epsilon = T^{-1/3}$, is straightforward, and the main focus of this paper is to show that a rate of $\epsilon = T^{-1/2}$ is attainable.

Item pricing, which is related to regret minimization under partial observation [5], has also received significant attention. A specific related work is [12], where the effect of knowing the demand curve is studied. (The demand curve is equivalent to the bid distribution.) The mechanism discussed in [12] is a *posted price* mechanism, and the regret is computed in both stochastic and adversarial settings. In the stochastic setting they assume that the expected revenue function is strictly concave, and use the UCB algorithm of [3] over discretized bid values to derive their strategy. Again, we do not make such assumptions in our work.

The question of the identification of the buyers' utilities given the auction outcome has been studied in the economics literature. The main goal is to recover in the limit the buyers' private value distribution (i.e., the buyers' utility function), given access to the resulting auction price (i.e., the auction outcome) and assuming that bidders utilities are independent and identically distributed [1], [9]. It is well known in the economics literature that given a bid distribution that has a monotone hazard rate, there is a unique reserve price maximizing the expected revenue in a second-price auction, and this optimal price is independent of the number of bidders [14]. As we do not make the monotone hazard rate assumption, in our case the optimal price for each auction might depend on the actual (varying) number of bidders. Because the seller does not observe the number of bidders before setting the reserve price (Section III), we prove our results using the regret to the best reserve price, with respect to a known prior over the number of bidders. As we just argued, depending on the bid distribution, this best reserve price need not be the same as the optimal reserve price one could set when knowing the actual number of bidders in advance.

There have been some works [7], [11], [20] on optimizing the reserve price, concentrating on more involved issues that arise in practice, such as discrete bids, nonstationary behavior, hidden bids, and more. While we are definitely not the first

³Note that the setting is not symmetric, and increasing by ϵ might lower the revenue significantly, by disqualifying many attractive bids.

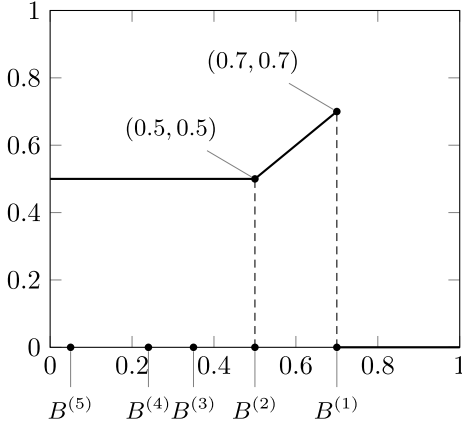


Fig. 1. The revenue function R for $m = 5$ bids of value $B^{(1)} = 0.7$, $B^{(2)} = 0.5$, $B^{(3)} = 0.35$, $B^{(4)} = 0.24$, $B^{(5)} = 0.05$. For $p \in [0, B^{(2)}]$ the revenue is constant, $R(p) = B^{(2)}$. For $p \in [B^{(2)}, B^{(1)}]$ the revenue grows linearly, $R(p) = p$. For $p \in [B^{(1)}, 1]$ the revenue is null, $R(p) = 0$.

ones to consider approximating optimal reserve prices in a second-price auction, to the best of our knowledge this is the first work that derives formal and concrete convergence rates.

Finally, note that any algorithm for one-dimensional stochastic bandit optimization could potentially be applied to solve our revenue maximization problem. Indeed, whenever a certain reserve price is chosen, the algorithm observes a realization of the associated stochastic revenue. While many algorithms exist that guarantee low regret in this setting, they all rely on specific assumptions on the function to optimize (in our case, the expected revenue function). See [6] obtains a regret of order \sqrt{T} under smoothness and strong concavity. The authors of [2] achieve a regret worse only by logarithmic factors without concavity, but assuming other conditions on the derivatives. The work [21] shows a bound of the same order just assuming unimodality. The work [4] also obtains the same asymptotics $\tilde{O}(\sqrt{T})$ on the regret using a local Lipschitz condition. The approach developed in this paper avoids making any assumption on the expected revenue function, such as Lipschitzness or bounded number of maxima. Instead, it exploits the specific feedback model provided by the second-price auction in order gain information about the optimum.

II. KNOWN NUMBER OF BIDDERS

We first show our results for the case where the number of bidders m is known and fixed. In Section III we will remove this assumption, and extend the results to the case when the number of bidders is a random variable with a known distribution. Fortunately, most of the ideas of the algorithm can be explained and nicely analyzed in the simpler case.

A. Preliminaries

The auctioneer organizes an auction about an item to be sold. He collects $m \geq 2$ bids B_1, B_2, \dots, B_m which are i.i.d. bounded random variables (for definiteness, we let $B_i \in [0, 1]$ for $i = 1, \dots, m$) whose common cumulative distribution function F is *arbitrary* and unknown. We let

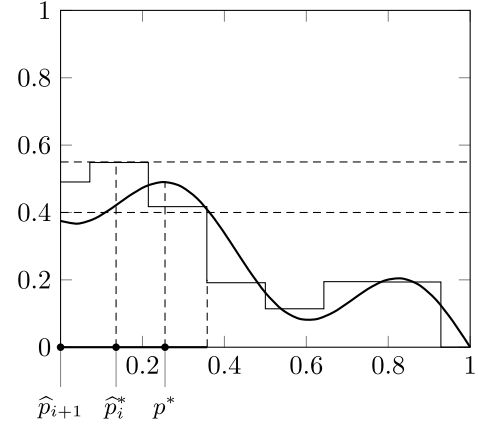


Fig. 2. At the beginning of Stage $i + 1$, Algorithm 1 has at its disposal an estimate $\hat{\mu}_i$ (here represented by the piecewise constant solid line) of the actual expected revenue function μ (the thick solid line). The upper horizontal dashed line indicates the estimate $\hat{\mu}_i(\hat{p}_i^*)$ of the actual maximum $\mu(p^*)$ (recall that \hat{p}_i^* is a maximizer of $\hat{\mu}_i(\cdot)$). The lower horizontal dashed line indicates the lower end of the confidence interval for $\hat{\mu}_i(\hat{p}_i^*)$. This defines the next set P_{i+1} of candidate optimal reserve prices, here marked by the thick solid line on the price axis, and the next reserve price \hat{p}_{i+1} , which is the lowest price in P_{i+1} . In this figure, $\hat{p}_{i+1} = 0$. Also, for simplicity, we have disregarded the further constraint $\hat{F}_{2,i}(p) \leq 1 - \alpha$.

$B^{(1)}, B^{(2)}, \dots, B^{(m)}$ denote the corresponding order statistics $B^{(1)} \geq B^{(2)} \geq \dots \geq B^{(m)}$.

In this simplified setting, we consider a protocol in which a learning algorithm (or a “mechanism”) is setting a *reserve price* (i.e., a *minimal price*) $p \in [0, 1]$ for the auction. The algorithm then observes a revenue $R(p) = R(p; B_1, \dots, B_m)$ defined as follows:

$$R(p) = \begin{cases} B^{(2)} & \text{if } p \leq B^{(2)} \\ p & \text{if } B^{(2)} < p \leq B^{(1)} \\ 0 & \text{if } p > B^{(1)}. \end{cases}$$

In words, if the reserve price p is higher than the highest bid $B^{(1)}$, the item is not sold, and the auctioneer’s revenue is zero; if p is lower than $B^{(1)}$ but higher than the second-highest bid $B^{(2)}$ then we sell at the reserve price p (i.e., the revenue is p); finally, if p is lower than $B^{(2)}$ we sell the item to the bidder who issued the highest bid $B^{(1)}$ at the price of the second-highest bid $B^{(2)}$ (hence the revenue is $B^{(2)}$). Figure 1 gives a pictorial illustration of the revenue function $R(p)$.

The expected revenue $\mu(p) = \mathbb{E}[R(p)]$ is the expected value of the revenue gathered by the auctioneer when the algorithm plays price p , the expectation being over the bids B_1, B_2, \dots, B_m . Let

$$p^* = \operatorname{argmax}_{p \in [0, 1]} \mu(p)$$

be the optimal price for the bid distribution F . We also write F_2 to denote the cumulative distribution function of $B^{(2)}$. We can write the expected revenue as

$$\begin{aligned} \mu(p) &= \mathbb{E}[B^{(2)}] + \mathbb{E}[p - B^{(2)} | B^{(2)} < p \leq B^{(1)}] \\ &\quad \times \mathbb{P}[B^{(2)} < p \leq B^{(1)}] \\ &\quad - \mathbb{E}[B^{(2)} | p > B^{(1)}] \mathbb{P}[p > B^{(1)}] \end{aligned}$$

where the first term is the baseline, the revenue of a second-price auction with no reserve price. The second term is the

gain due to the reserve price (increasing the revenue beyond the second-highest bid). The third term is the loss due to the possibility that we will not sell (when the reserve price is higher than the highest bid). The following fact streamlines the computation of $\mu(p)$. All proofs are given in the appendices.

Fact 1: With the notation introduced so far, we have

$$\mu(p) = \mathbb{E}[B^{(2)}] + \int_0^p F_2(t) dt - p(F(p))^m$$

where the expectation $\mathbb{E}[\cdot]$ is over the m bids B_1, B_2, \dots, B_m . The algorithm interacts with its environment (the bidders) in a sequential fashion. At each time step $t = 1, 2, \dots$ the algorithm sets a price p_t and receives revenue $R_t(p_t) = R(p_t; B_{t,1}, \dots, B_{t,m})$ which is a function of the random bids $B_{t,1}, \dots, B_{t,m}$ at time t . The price p_t depends on past revenues $R_s(p_s)$ for $s < t$, and therefore on past bids. Given a sequence of reserve prices p_1, \dots, p_T , we define the (cumulative) expected regret as

$$\sum_{t=1}^T (\mathbb{E}[R_t(p^*)] - \mathbb{E}_t[R_t(p_t)]) = \sum_{t=1}^T (\mu(p^*) - \mu(p_t)) \quad (1)$$

where the expectation $\mathbb{E}_t = \mathbb{E}_t[\cdot \mid p_1, \dots, p_{t-1}]$ is over the random bids at time t , conditioned on all past prices p_1, \dots, p_{t-1} (i.e., conditioned on the past history of the bidding process). This implies that the expected regret (1) is indeed a random variable, as each p_t depends on the past random revenues. Our goal is to devise an algorithm whose regret after T steps is $\tilde{O}(\sqrt{T})$ with high probability, and with as few assumptions as possible on F . We see in the sequel that, when T is large, this goal can actually be achieved with *no assumptions whatsoever* on the underlying distribution F . Moreover, in Section IV-A we use a uniform convergence argument to show that the same regret bound $\tilde{O}(\sqrt{T})$ holds with high probability for the realized regret

$$\max_{p \in [0,1]} \sum_{t=1}^T (R_t(p) - R_t(p_t)).$$

Note that here the realized revenue of the seller is compared against the best reserve price *on each sequence of bid realizations*. Therefore, the realized regret is a much stronger notion of regret than the expected regret (1).

It is well known that from the distribution of any order statistics one can reconstruct the underlying distribution. Unfortunately, we do not have access to the true distribution of order statistics, but only to an approximation thereof. We first need to show that a small deviation in our approximation will have a small effect on our final result. The following preliminary lemma will be of great importance in our approximations. It shows that if we have a small error in the approximation of $F_2(p)$ we can recover $\mu(p)$ with a small error. The function $\beta(\cdot)$ therein maps $(F(\cdot))^m$ to $F_2(\cdot)$. In fact, since the bids are independent with the same distribution, we have $F_2(p) = m(F(p))^{m-1}(1-F(p)) + (F(p))^m = \beta((F(p))^m)$. The main technical difficulty arises from the fact that the function $\beta^{-1}(\cdot)$ we use in reconstructing $(F(\cdot))^m$ from $F_2(\cdot)$ —see pseudocode in Algorithm 1, is not a Lipschitz function.

Lemma 1: Fix an integer $m \geq 2$ and consider the function

$$\beta(x) = m x^{\frac{m-1}{m}} - (m-1)x, \quad x \in [0, 1].$$

Then $\beta^{-1}(\cdot)$ exists in $[0, 1]$. Moreover, if $a \in (0, 1)$ and $x \in [0, 1]$ are such that $a - \epsilon \leq \beta(x) \leq a + \epsilon$ for some $\epsilon \geq 0$, then

$$\beta^{-1}(a) - \frac{2\epsilon}{\sqrt{1-a}} \leq x \leq \beta^{-1}(a) + \frac{2\epsilon}{\sqrt{1-a}}. \quad (2)$$

In a nutshell, this lemma shows how approximations in the value of $\beta(\cdot)$ turn into approximations in the value of $\beta^{-1}(\cdot)$. Because the derivative of β^{-1} is infinite at 1, we cannot hope to get a good approximation unless a is bounded away from 1. For this very reason, we need to make sure that our function approximations are only applied to cases where the arguments are not too close to 1. The approximation parameter α in the pseudocode of Algorithm 1 serves this purpose.

B. The Algorithm

Our algorithm works in *stages*, where the same price is consistently played during each stage.⁴ Stage 1 lasts T_1 steps, during which the algorithm plays $p_t = \hat{p}_1$ for all $t = 1, \dots, T_1$. Stage 2 lasts T_2 steps, during which the algorithm plays $p_t = \hat{p}_2$ for all $t = T_1 + 1, \dots, T_1 + T_2$, and so on, up to S stages. Overall, the regret suffered by this algorithm can be written as

$$\sum_{i=1}^S (\mu(p^*) - \mu(\hat{p}_i)) T_i$$

where the sum is over the S stages. The length T_i of each stage will be set later on, as a function of the total number of steps T . The reserve prices $\hat{p}_1, \hat{p}_2, \dots$ are set such that $0 = \hat{p}_1 \leq \hat{p}_2 \leq \dots \leq 1$. At the end of each stage i , the algorithm computes a new estimate $\hat{\mu}_i$ of the expected revenue function μ in the interval $[\hat{p}_i, 1]$, where p^* is likely to lie. This estimate depends on the empirical cumulative distribution function $\hat{F}_{2,i}$ of F_2 computed during stage i in the interval $[\hat{p}_i, 1]$. The algorithm's pseudocode is given in Algorithm 1. The quantity $C_{\delta,i}(p)$ therein is defined as

$$C_{\delta,i}(p) = p \sqrt{\frac{2}{(1 - \hat{F}_{2,i}(p))T_i} \ln \frac{6S}{\delta}}.$$

$C_{\delta,i}(p)$ is a confidence interval (at confidence level $1 - \delta/(3S)$) for the point estimate $\hat{\mu}_i(p)$ in stage i , where $S = S(T)$ is either the total number of stages or an upper bound thereof.

Stage 1 is a seed stage, where the algorithm computes a first approximation $\hat{\mu}_1$ of μ . Since the algorithm plays $\hat{p}_1 = 0$, and $R(0) = B^{(2)}$, during this stage T_1 independent realizations of the second-bid variable $B^{(2)}$ are observed. Hence the empirical distribution $\hat{F}_{2,1}$ in Algorithm 1 is a standard cumulative empirical distribution function based on i.i.d. realizations of $B^{(2)}$. The approximation $\hat{\mu}_1$ is based on the corresponding expected revenue μ contained in Fact 1, where $\beta(\cdot)$ is the function defined in Lemma 1, mapping $(F(p))^m$ to $F_2(p)$.

⁴For simplicity, we have disregarded rounding effects in the computation of the integer stage lengths T_i .

Algorithm 1 Regret Minimizer

Input: Confidence level $\delta \in (0, 1]$, approximation parameter $\alpha \in (0, 1]$, time horizon T ;

Let stage lengths $T_i = T^{1-2^{-i}}$ for $i = 1, 2, \dots$

Stage 1:

- For all $t = 1, \dots, T_1$, play $p_t = \hat{p}_1 = 0$ and observe revenues $R_1(0), \dots, R_{T_1}(0)$.
- Compute, for $x \in [0, 1]$, empirical distribution

$$\hat{F}_{2,1}(x) = \frac{1}{T_1} \left| \{t = 1, \dots, T_1 : R_t(0) \leq x\} \right|.$$

- Compute, for $p \in [0, 1]$, approximation

$$\hat{\mu}_1(p) = \mathbb{E}[B^{(2)}] + \int_0^p \hat{F}_{2,1}(t) dt - p \beta^{-1}(\hat{F}_{2,1}(p)).$$

For Stage $i = 2, 3, \dots$

- For all $t = 1 + \sum_{j=1}^{i-1} T_j, \dots, \sum_{j=1}^i T_j$, play $p_t = \hat{p}_i$, and observe revenues $R_1(\hat{p}_i), \dots, R_{T_i}(\hat{p}_i)$, where \hat{p}_i is computed as follows:

- Compute maximizer

$$\hat{p}_{i-1}^* = \operatorname{argmax}_{p \in [\hat{p}_{i-1}, 1] : \hat{F}_{2,i-1}(p) \leq 1-\alpha} \hat{\mu}_{i-1}(p).$$

- Let $P_i = \{p \in [\hat{p}_{i-1}, 1] : \hat{\mu}_{i-1}(p) \geq \hat{\mu}_{i-1}(\hat{p}_{i-1}^*) - 2C_{\delta,i-1}(\hat{p}_{i-1}^*) - 2C_{\delta,i-1}(p)\}$.

- Set $\hat{p}_i = \min P_i \cap \{p : \hat{F}_{2,i-1}(p) \leq 1 - \alpha\}$.

- Compute, for $x \in [\hat{p}_i, 1]$, empirical distribution

$$\hat{F}_{2,i}(x) = \frac{1}{T_i} \left| \{t = 1, \dots, T_i : R_t(\hat{p}_i) \leq x\} \right|.$$

- Compute, for $p \in [\hat{p}_i, 1]$, approximation

$$\hat{\mu}_i(p) = \mathbb{E}[B^{(2)}] + \int_0^{\hat{p}_i} F_2(t) dt + \int_{\hat{p}_i}^p \hat{F}_{2,i}(t) dt - p \beta^{-1}(\hat{F}_{2,i}(p)).$$

Note that if β^{-1} is available, maximizing the above function (done in Stage 2) can easily be computed from the data. The presence of the unknown constant $\mathbb{E}[B^{(2)}]$ is not a problem for this computation.⁵ In Stage 2 (encompassing trials $t = T_1 + 1, \dots, T_1 + T_2$) the algorithm calculates the empirical maximizer

$$\hat{p}_1^* = \operatorname{argmax}_{p \in [0, 1] : \hat{F}_{2,1}(p) \leq 1-\alpha} \hat{\mu}_1(p)$$

then computes the set of candidate optimal reserve prices

$$P_2 = \{p \in [0, 1] : \hat{\mu}_1(p) \geq \hat{\mu}_1(\hat{p}_1^*) - 2C_{\delta,1}(\hat{p}_1^*) - 2C_{\delta,1}(p)\}$$

and sets the reserve price \hat{p}_2 to be the lowest one in P_2 , subject to the additional constraint that⁶ $\hat{F}_{2,1}(p) \leq 1 - \alpha$. Price \hat{p}_2 is played during all trials within Stage 2. The corresponding revenues $R_t(\hat{p}_2)$, for $t = 1, \dots, T_2$, are gathered and used to construct an empirical cumulative distribution $\hat{F}_{2,2}$ and an approximate expected revenue function $\hat{\mu}_2$ to be used only in

⁵Note that in the algorithm (subsequent Stage 2) we either take the difference of two values $\hat{\mu}_1(p_1) - \hat{\mu}_1(p_2)$, in which case the constant cancels, or maximize over $\hat{\mu}_1(p)$, in which case the constant does not change the outcome.

⁶Note that the intersection is not empty, since \hat{p}_1^* is in the intersection.

the subinterval⁷ $[\hat{p}_2, 1]$.

In order to see why $\hat{F}_{2,2}$ and $\hat{\mu}_2$ are useful only on $[\hat{p}_2, 1]$, observe that

$$R(\hat{p}_2) = \begin{cases} \hat{p}_2 \text{ or } 0 & \text{if } B^{(2)} < \hat{p}_2 \\ B^{(2)} & \text{if } B^{(2)} \geq \hat{p}_2. \end{cases}$$

Thus, for any $x \geq \hat{p}_2$ we have that

$$\mathbb{P}(R(\hat{p}_2) \leq x) = \mathbb{P}(B^{(2)} \leq x).$$

Hence, if we denote by $R_1(\hat{p}_2), \dots, R_{T_2}(\hat{p}_2)$ the revenues observed by the algorithm during Stage 2, the empirical distribution function

$$\hat{F}_{2,2}(x) = \frac{1}{T_2} \left| \{t = 1, \dots, T_2 : R_t(\hat{p}_2) \leq x\} \right|$$

approximates $F_2(x)$ only for $x \in [\hat{p}_2, 1]$.

All other stages $i > 2$ proceed similarly, each stage i relying on the existence of empirical estimates $\hat{F}_{2,i-1}$, $\hat{\mu}_{i-1}$, and \hat{p}_{i-1} delivered by the previous stage $i - 1$. Figure 2 gives a pictorial explanation of the way the algorithm works.

⁷Once again, computing the argmax of $\hat{\mu}_2$ over $[\hat{p}_2, 1]$ as well as the set of candidates P_3 (done in the subsequent Stage 3) is not prevented by the presence of the unknown constants $\mathbb{E}[B^{(2)}]$ and $\int_0^{\hat{p}_2} F_2(t) dt$ therein.

C. Regret Analysis

We start by showing that for all stages i the term $1 - \widehat{F}_{2,i}(p)$ in the denominator of $C_{\delta,i}(p)$ can be controlled for all p such that $\mu(p)$ is bounded away from zero. Recall that $S = S(T)$ denotes (an upper bound on) the total number of stages.

Lemma 2: *With the notation introduced so far, for any fixed stage i ,*

$$1 - \widehat{F}_{2,i}(p) \geq \frac{\mu(p)^2}{6} - \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}$$

holds with probability at least $1 - \delta/(3S)$, uniformly over $p \in [\widehat{p}_i, 1]$, conditioned on all past stages.

In the sequel, we use Lemma 2 with $p = p^*$ and assume that $1 - \widehat{F}_{2,i}(p^*) \geq \alpha$ holds for each stage i with probability at least $1 - \delta/(3S)$, where the approximation parameter α is defined as

$$\alpha = \frac{\mu(p^*)^2}{6} - \sqrt{\frac{1}{2 \min_i T_i} \ln \frac{6S}{\delta}}$$

provided $p^* \in [\widehat{p}_i, 1]$. In order to ensure that $\alpha > 0$, it suffices to have $\mu(p^*) > 0$ and T large enough—see Theorem 1 below. Recall that it is important to guarantee that $\widehat{F}_{2,i}(p)$ be bounded away from 1 for all arguments p which we happen to evaluate $\widehat{F}_{2,i}$ at. This is because the function β^{-1} has an infinite derivative at 1.

The following lemma is crucial to control the regret of Algorithm 1. It states that the approximation in stage i is accurate. In addition, it bounds the empirical regret in stage i , provided our current reserve price is lower than the optimal reserve price. The proof is a probabilistic induction over stages.

Lemma 3: The event

$$|\mu(p) - \widehat{\mu}_i(p)| \leq 2C_{\delta,i}(p) \quad \text{for all } p \in [\widehat{p}_i, 1] \quad (3)$$

holds with probability at least $1 - \delta/3$ simultaneously in all stages $i = 1, \dots, S$. Moreover, the events

$$\begin{aligned} p^* &\geq \widehat{p}_i \\ 0 &\leq \widehat{\mu}_i(\widehat{p}_i^*) - \widehat{\mu}_i(p^*) \leq 2C_{\delta,i}(\widehat{p}_i^*) + 2C_{\delta,i}(p^*) \end{aligned} \quad (4)$$

both hold with probability at least $1 - \delta$ simultaneously in all stages $i = 1, \dots, S$.

The next theorem proves our regret bound under the assumption that $\mu(p^*)$ is nonzero. Note that $\mu(p^*) = 0$ corresponds to the degenerate case $\mu(p) = 0$ for all $p \in [0, 1]$. Under the above assumption, the theorem states that when the horizon T is sufficiently large, then with high probability the regret of Algorithm 1 is $\mathcal{O}(\sqrt{T \log \log \log T} (\log \log T)) = \widetilde{\mathcal{O}}(\sqrt{T})$. It is important to remark that in this bound there is no explicit dependence on the number m of bidders.

Theorem 1: For any distribution F of the bids and any $m \geq 2$ such that $\mu(p^) > 0$, we have that Algorithm 1 operating on any time horizon T such that*

$$T > \frac{1}{\mu(p^*)^8} \left(72 \ln \frac{6(1 + \log_2 \log_2 T)}{\delta} \right)^2$$

using approximation parameter $\alpha \geq \mu(p^)^2/12$ has regret*

$$\begin{aligned} &\mathcal{O} \left(\frac{\sqrt{T (\log \log \log T + \log 1/\delta)} (\log \log T)}{\mu(p^*)} \right) \\ &= \widetilde{\mathcal{O}} \left(\frac{\sqrt{T \log 1/\delta}}{\mu(p^*)} \right) \end{aligned}$$

with probability at least $1 - \delta$.

The proof of this theorem follows by applying at each stage i the uniform approximation delivered by Lemma 3 on the accuracy of empirical to true regret. This would bound the regret in stage i by $8 \sqrt{\frac{1}{\alpha T_{i-1}} \ln \frac{6S}{\delta}}$ —see the proof in Appendix V. We then set the length T_i of stage i as $T_i = T^{1-2^{-i}}$, i.e., $T_1 = \sqrt{T}$, $T_2 = T^{3/4}$, $T_3 = T^{7/8}$, \dots , which implies that the total number of stages S is $\mathcal{O}(\log \log T)$. Finally, we sum the regret over the stages to derive the theorem.

Two remarks are in order at this point. First, the reader should observe that the bound in Theorem 1 does not explicitly contain a dependence on m ; this is mostly due to the fact that the approximation result in Lemma 1 is in turn independent of m . The number of bidders m shows up implicitly only through $\mu(p^*)$ —see also the discussion at the end of Section V. Second, the way we presented it makes Algorithm 1 depending on the time horizon T , though this prior knowledge is not strictly required: In Section IV-B we show a standard “doubling trick” for making Algorithm 1 independent of the time horizon T .

D. Lower Bounds

The next result shows that the \sqrt{T} dependence of the regret on the time horizon T is not a consequence of our partial information setting. Indeed, this dependence cannot be removed even if the mechanism is allowed to observe the actual bids after setting the reserve price in each repetition of the auction.

Theorem 2: There exists a distribution of bids such that any deterministic algorithm operating with $m = 2$ bidders is forced to have expected regret

$$\sum_{t=1}^T (\mu(p^*) - \mu(p_t)) = \Omega(\sqrt{T}).$$

Although the result is proven for deterministic algorithms, it can easily be extended to randomized algorithms through a standard argument.

III. RANDOM NUMBER OF BIDDERS

We now consider the case when the number of bidders m in each trial is a random variable M distributed according to a *known* discrete distribution Q over $\{2, 3, 4, \dots\}$. The assumption that Q is known is realistic: one can think of estimating it from historical data that might be provided by the auctioneer. On each trial, the value $M = m$ is randomly generated according to Q , and the auctioneer collects m bids B_1, B_2, \dots, B_m . For given m , these bids are i.i.d. bounded random variables $B \in [0, 1]$ with unknown cumulative distribution F , which is the setting considered in Section II.

For simplicity, we assume that M is independent of the random variables B_i . For fixed $M = m$, we denote by $B_m^{(1)} \geq B_m^{(2)} \geq \dots \geq B_m^{(m)}$ the corresponding order statistics.

Our learning algorithm is the same as before: In each time step, the algorithm is requested to set reserve price $p \in [0, 1]$ and, for the given realization of $M = m$, only observes the value of the revenue function $R^m(p) = R(p; B_1, B_2, \dots, B_m)$ defined as

$$R^m(p) = \begin{cases} B_m^{(2)} & \text{if } p \leq B_m^{(2)} \\ p & \text{if } B_m^{(2)} < p \leq B_m^{(1)} \\ 0 & \text{if } p > B_m^{(1)} \end{cases}$$

without knowing the specific value of m that generated this revenue. Namely, after playing price p the algorithm is observing an independent realization of the random variable $R^M(p)$. The expected revenue $\mu(p)$ is now

$$\mu(p) = \mathbb{E}_M \mathbb{E}[R^M(p)] = \sum_{m=2}^{\infty} Q(m) \mathbb{E}[R^m(p) \mid M = m]$$

where the inner expectation $\mathbb{E}[\cdot \mid M = m]$ is over the random bids B_1, B_2, \dots, B_m .

Again, we want to minimize the expected regret with respect to the optimal reserve price

$$p^* = \operatorname{argmax}_{p \in [0, 1]} \mu(p)$$

for the bid distribution F , averaged over the distribution Q over the number of bidders M , where the expected regret over T time steps is

$$\sum_{t=1}^T (\mu(p^*) - \mu(p_t))$$

and p_t is the price set by the algorithm at time t . In Section IV-A we show that the same regret bound holds for the realized regret

$$\max_{p \in [0, 1]} \sum_{t=1}^T (R_t^{M_t}(p) - R_t^{M_t}(p_t))$$

where M_t is the number of bidders at time t .

Let $F_{2,m}$ denote the cumulative distribution function of $B_m^{(2)}$. We use $\mathbb{E}_M[F_{2,M}](x)$ to denote the mixture distribution $\sum_{m=2}^{\infty} Q(m) F_{2,m}(x)$. Likewise,

$$\mathbb{E}_M[F^M](x) = \sum_{m=2}^{\infty} Q(m) (F(x))^m.$$

Relying on Fact 1, one can easily see that

$$\begin{aligned} \mu(p) &= \mathbb{E}_M \mathbb{E}[B_M^{(2)}] + \int_0^p \mathbb{E}_M[F_{2,M}](t) dt \\ &\quad - p \mathbb{E}_M[F^M](p). \end{aligned} \quad (5)$$

As in Section II, our goal is to devise an online algorithm whose expected regret is of the order \sqrt{T} , with as few assumptions as possible on F and Q .

We first extend Lemma 1 to handle this more general setting.⁸

Lemma 4: Let T be the probability generating function of M ,

$$T(x) = \sum_{m=2}^{\infty} Q(m) x^m$$

and define the auxiliary function

$$A(x) = T(x) + (1-x)T'(x)$$

where, for both functions, we let the argument x range in $[0, 1]$. Then T and A are bijective mappings from $[0, 1]$ onto $[0, 1]$ and both T^{-1} and A^{-1} exist in $[0, 1]$. Moreover, letting $a \in (0, 1)$, and $0 \leq \epsilon < 1 - a$, if x is such that

$$a - \epsilon \leq A(T^{-1}(x)) \leq a + \epsilon$$

then

$$T(A^{-1}(a)) - \frac{\epsilon \mathbb{E}[M]}{1 - (a + \epsilon)} \leq x \leq T(A^{-1}(a)) + \frac{\epsilon \mathbb{E}[M]}{1 - (a + \epsilon)}. \quad (6)$$

In addition, if⁹

$$(T''(x))^2(1-x) + T'(x)T''(x) \geq T'(x)T'''(x)(1-x) \quad (7)$$

holds for all $x \in [0, 1]$ then, for any $a \in (0, 1)$ and $\epsilon \geq 0$,

$$T(A^{-1}(a)) - \frac{\epsilon}{1-a} \leq x \leq T(A^{-1}(a)) + \frac{\epsilon}{1-a}. \quad (8)$$

Observe that $T(\cdot)$ and $A(\cdot)$ in this lemma have been defined in such a way that¹⁰

$$\mathbb{E}_M[F_{2,M}](p) = A(F(p))$$

and

$$\mathbb{E}_M[F^M](p) = T(F(p)).$$

Hence, $\mathbb{E}_M[F^M](p)$ in (5) satisfies

$$\mathbb{E}_M[F^M](p) = T\left(A^{-1}\left(\mathbb{E}_M[F_{2,M}](p)\right)\right).$$

In particular, when $\mathbb{P}(M = m) = 1$ as in Section II, we obtain $T(x) = x^m$ and $A(x) = m x^{m-1} - (m-1)x^m$. Thus, in this case $A(T^{-1}(\cdot))$ is the function $\beta(\cdot)$ defined in Lemma 1, and the reconstruction function $\beta^{-1}(\cdot)$ we used throughout Section II is $T(A^{-1}(\cdot))$. Because this is a more general setting than the one in Section II, we do still have the technical issue of insuring that the argument of this reconstruction function is not too close to 1.

As in the fixed m case, the algorithm proceeds in stages. In each stage i the algorithm samples the function $\mathbb{E}_M[F_{2,M}]$

⁸More precisely, in dealing with a more general setting we only obtain a slightly looser result than Lemma 1.

⁹Condition (7) is a bit hard to interpret: It is equivalent to the convexity of the function $T(A^{-1}(x))$ for $x \in [0, 1]$ (see the proof of Lemma 4 in Appendix V), and it can be shown to be satisfied by many standard parametric families of discrete distributions Q , e.g., Uniform, Binomial, Poisson, Geometric. There are, however, examples where this condition does not hold. For instance, the distribution Q , where $Q(2) = 0.4$, $Q(8) = 0.6$, and $Q(m) = 0$ for any $m \neq 2, 8$ does not satisfy (7) for $x = 0.6$, i.e., it yields a function $T(A^{-1}(x))$ which is not convex on $x = 0.6$.

¹⁰Recall from Section II-A that, for any fixed $M = m$, we have $F_{2,m}(p) = m(F(p))^{m-1}(1-F(p)) + (F(p))^m$.

by sampling $R^M(p)$ at appropriate values of p . This allows it to build an empirical distribution $\widehat{F}_{2,i}$ and to reconstruct the two unknown functions $\mathbb{E}_M[F_{2,M}]$ and $\mathbb{E}_M[F^M]$ occurring in (5) over an interval of reserve prices that is likely to contain p^* . Whereas $\mathbb{E}_M[F_{2,M}]$ is handled directly, the reconstruction of $\mathbb{E}_M[F^M]$ requires us to step through the functions T and A according to the following scheme:

$$\begin{aligned}\widehat{F}_{2,i}(p) &\approx \mathbb{E}_M[F_{2,M}](p) = A(F(p)) \\ &\iff A^{-1}(\widehat{F}_{2,i}(p)) \approx F(p) \\ &\iff T(A^{-1}(\widehat{F}_{2,i}(p))) \approx T(F(p)) = \mathbb{E}_M[F^M](p).\end{aligned}$$

Namely, in stage i we sample $\mathbb{E}_M[F_{2,M}]$ to obtain the empirical distribution $\widehat{F}_{2,i}$, and then estimate $\mathbb{E}_M[F^M]$ in (5) through $T(A^{-1}(\widehat{F}_{2,i}(\cdot)))$.

In order to emphasize that the role played by the composite function $A(T^{-1}(\cdot))$ here is the very same as the function $\beta(\cdot)$ in Section II, we overload the notation and define in this section $\beta(x) = A(T^{-1}(x))$, where T and A are given in Lemma 4. Moreover, we define for brevity $\bar{F}_2(x) = \mathbb{E}_M[F_{2,M}](x)$.

With this notation in hand, the detailed description of the algorithm becomes very similar to the one in Section II-B. Hence, in what follows we only emphasize the differences, which are essentially due to the modified confidence interval delivered by Lemma 4, as compared to Lemma 1.

In particular, if we rely on (6), the new confidence interval size for Stage i depends on the empirical distribution $\widehat{F}_{2,i}$ through the quantity (we again overload the notation)

$$C_{\delta,i}(p) = \frac{p \mathbb{E}[M]}{1 - \widehat{F}_{2,i}(p) - \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}} \sqrt{\frac{2}{T_i} \ln \frac{6S}{\delta}}$$

with

$$T_i > \frac{1}{2(1 - \widehat{F}_{2,i}(p))^2} \ln \frac{6S}{\delta}.$$

Similarly, if we rely on (8), we have instead

$$C_{\delta,i}(p) = \frac{p}{1 - \widehat{F}_{2,i}(p)} \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}.$$

The resulting pseudocode is the same as in Algorithm 1, where the observations $R_t(\widehat{p}_i)$ therein have to be interpreted as distributed i.i.d. as $R^M(\widehat{p}_i)$, and $\mathbb{E}[B^{(2)}]$ and F_2 in $\widehat{\mu}_i$ are replaced by their M -average counterparts $\mathbb{E}_M \mathbb{E}[B_M^{(2)}]$ and \bar{F}_2 . We call the resulting algorithm the *Generalized Algorithm 1*.

As for the analysis, Lemma 2 is replaced by the following (because of notation overloading, the statement is the same as that of Lemma 2, but the involved quantities are different, and so is the proof in the appendix).

Lemma 5: *With the notation introduced at the beginning of this section, if $S = S(T)$ is (an upper bound on) the total number of stages, we have that, for any fixed stage i ,*

$$1 - \widehat{F}_{2,i}(p) \geq \frac{\mu(p)^2}{6} - \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}$$

holds with probability at least $1 - \delta/(3S)$, uniformly over $p \in [\widehat{p}_i, 1]$, conditioned on all past stages.

Then an easy adaptation of Lemma 3 leads to the following expected regret bound. The proof is very similar to the proof of Theorem 1, and is therefore omitted.

Theorem 3: *With the notation introduced at the beginning of this section, for any pair of distributions F and Q such that $\mu(p^*) > 0$ we have that the Generalized Algorithm 1, operating on any time horizon T satisfying*

$$T > \frac{1}{\mu(p^*)^8} \left(288 \ln \frac{6(1 + \log_2 \log_2 T)}{\delta} \right)^2$$

and with approximation parameter $\alpha \geq \mu(p^)^2/12$, has regret*

$$\begin{aligned}\frac{A}{\mu(p^*)^2} \times \mathcal{O} \left(\sqrt{T(\log \log \log T + \log 1/\delta)} (\log \log T) \right) \\ = \frac{A}{\mu(p^*)^2} \times \tilde{\mathcal{O}} \left(\sqrt{T \log 1/\delta} \right)\end{aligned}$$

with probability at least $1 - \delta$, where $A = \mathbb{E}[M]$ if (6) holds and $A = 1$ if (8) holds.

IV. EXTENSIONS

This section further extends the results contained in the previous two sections. First, we show how to bound with high probability the realized regret (Section IV-A). Second, we show how to turn Algorithm 1 into an algorithm that does not rely on prior knowledge of the time horizon T (Section IV-B).

A. Bounding the Realized Regret

In this section, we show how to bound in probability the realized regret

$$\max_{p \in [0,1]} \sum_{t=1}^T R_t^{M_t}(p) - \sum_{t=1}^T R_t^{M_t}(p_t)$$

suffered by the Generalized Algorithm 1. As a special case, this clearly applies to Algorithm 1, too.

We need the following definitions and results from empirical process theory—see [19]. Let \mathcal{F} be a set of $[0, 1]$ -valued functions defined on a common domain \mathbb{X} . We say that \mathcal{F} shatters a sequence $x_1, \dots, x_n \in \mathbb{X}$ if there exists $r_1, \dots, r_n \in \mathbb{R}$ such that for each $(a_1, \dots, a_n) \in \{0, 1\}^n$ there exists $f \in \mathcal{F}$ for which $f(x_i) > r_i$ iff $a_i = 1$ for all $i = 1, \dots, n$. The pseudo-dimension [17] of \mathcal{F} , which is defined as the length of the longest sequence shattered by \mathcal{F} , controls the rate of uniform convergence of means to expectations in \mathcal{F} . This is established by the following known lemma, which combines Dudley's entropy bound with a bound on the metric entropy of \mathcal{F} in terms of the pseudo-dimension—see [18], [19].

Lemma 6: *Let X_1, X_2, \dots be i.i.d. random variables defined on a common probability space and taking values in \mathbb{X} . There exists a universal constant $C > 0$ such that, for any fixed T and δ ,*

$$\sup_{f \in \mathcal{F}} \left| \sum_{t=1}^T f(X_t) - T \mathbb{E}[f] \right| \leq C \sqrt{dT \ln \frac{1}{\delta}}$$

Algorithm 2 Anytime Regret Minimizer**Input:** Confidence level $\delta \in (0, 1]$, approximation parameter $\alpha \in (0, 1]$;**For Block** $r = 0, 1, 2, \dots$

Run Algorithm 1 with input parameters

$$\text{block } r \text{ confidence } \delta_r = \frac{\delta}{(r+1)(r+2)}, \quad \text{approximation parameter } \alpha, \quad \text{block } r \text{ horizon } T_r = 2^r.$$

with probability at least $1 - \delta$, where d is the pseudo-dimension of \mathcal{F} .

Recall that $\mathbb{E}_M \mathbb{E}[R^M(p)] = \mu(p)$ for all $p \in [0, 1]$. Let $\mathcal{R} = \{R^M(p) : p \in [0, 1]\}$ be the class of revenue functions indexed by reserve prices $p \in [0, 1]$. Hence, for each p , $R^M(p)$ is a $[0, 1]$ -valued function of the number M of bidders and the bids B_1, \dots, B_M . In the appendix we prove the following bound.

Lemma 7: The pseudo-dimension of the class \mathcal{R} is 2.

As announced, the following is the main result of this section, whose proof combines Lemma 6, Lemma 7, together with a standard martingale argument.

Theorem 4: Under the assumptions of Theorem 3 (Section III), the actual regret of Generalized Algorithm 1 satisfies

$$\max_{p \in [0, 1]} \sum_{t=1}^T R_t^{M_t}(p) - \sum_{t=1}^T R_t^{M_t}(p_t) = \frac{A}{\mu(p^*)^2} \times \tilde{O}\left(\sqrt{T \log \frac{1}{\delta}}\right)$$

with probability at least $1 - \delta$, where $A = \mathbb{E}[M]$ if (6) holds and $A = 1$ if (8) holds.

B. The Case of Unknown Time Horizon T

We use a standard “doubling trick” argument—see [5, Sec. 2.3] applied to Algorithm 1 (the same argument applies to the Generalized Algorithm 1). The idea is to partition the sequence $1, 2, \dots$ of time steps into blocks of geometrically growing length, where each block $r = 0, 1, \dots$ starts at time 2^r and ends at time $2^{r+1} - 1$. At the beginning of each new block r , we restart Algorithm 1 from scratch, using $T_r = 2^{r+1} - 1 - 2^r + 1 = 2^r$ as new horizon parameter, and setting the current confidence level as $\delta_r = \delta / (r+1)(r+2)$, where δ is the desired confidence level. The algorithm’s pseudocode is given in Algorithm 2.

Using the standard analysis for the doubling trick, Algorithm 2 is easily seen to achieve the following bound on the expected regret.

Theorem 5: For any distribution F of the bids and any $m \geq 2$ such that $\mu(p^*) > 0$, we have that Algorithm 2 using approximation parameter $\alpha \geq \mu(p^*)^2/12$ has regret

$$O\left(\frac{1}{\mu(p^*)^8} \log^2\left(\frac{T \log \log T}{\delta}\right) + \frac{\sqrt{T} \log \log T}{\mu(p^*)} \sqrt{\log \frac{\log T}{\delta}}\right)$$

with probability at least $1 - \delta$ simultaneously over all T .

Clearly enough, combining with Theorem 4 a similar statement can be given for the realized regret as well.

V. CONCLUSIONS AND DISCUSSION

Optimizing the reserve price in a second-price auction is an important theoretical and practical concern. We introduced a regret minimization algorithm to optimize the reserve price incurring a regret of only $\tilde{O}(\sqrt{T})$. We showed the result both for the case where the number of bidders is known, and for the case where the number of bidders is drawn from a known distribution. The former assumption, of known fixed number of bidders, is applicable when the number of bidders is given as the outcome of the auction. The assumption that the distribution over the number of bidders is known is rather realistic, even in the case where the number of participating bidders is not given explicitly. For example, one can hope to estimate such data from historical data that might be made available from the auctioneer.

Our optimization of the reserve prices depends only on observable outcomes of the auction. Specifically, we need only observe the seller’s actual revenue at each step. This is important in many applications, such as e-Bay, AdX or AdSense, where the auctioneer is a different entity from the seller, and provides the seller with only a limited amount of information regarding the actual auction. It is also important that we make no assumptions about the distribution of the bidder’s bid (or its relationship to the bidder’s valuation) since many such assumptions are violated in reality. The only assumption that we do make is that the distributions of the bidders are identical. This assumption is a fairly good approximation of reality in many cases where the seller conducts a large number of auctions and bidders rarely participate in a large number of them.

The resulting algorithm is very simple at a high level, and potentially attractive to implement in practice. Conceptually, we would like to estimate the optimal reserve price. The main issue is that if we simply exploit the current best estimate, we might miss essential exploration. This is why, instead of playing the current best estimate, the algorithm plays a minimal ϵ -optimal reserve price, where ϵ shrinks over time. The importance of playing the minimal near-optimal reserve price is that it allows for efficient exploration of the prices, due to the specific feedback model provided by the second-price auction setting.

An interesting direction for extending our results is the generalized second price auction model, when multiple items of different quality are sold at each step. Here the problem of estimating the expected revenue function becomes more involved due to the presence of terms that depend on the correlation of order statistics.

A different open issue, of more technical nature, is whether the inverse dependence on $\mu(p^*)$ in Theorem 1 (and on $\mu(p^*)^2$ in Theorem 3) can somehow be removed. Indeed, these factors do not seem to be inherent to the problem itself, but only to the kind of algorithms we use.

In a similar vein, because the number of bidders (Section II) or the distribution on the number of bidders (Section III under assumption (8)) does not explicitly show up in the regret bounds, one may wonder whether our algorithm really needs to know this information. Unfortunately, the answer seems to be affirmative, as our algorithm hinges on reconstructing the underlying bid distribution $F(\cdot)$ from the distribution of the second-highest bid $F_2(\cdot)$, and we are currently unaware of how this could be done without knowing m (or its distribution). A simple attempt to remove the dependence on m from Algorithm 1 is to let $m \rightarrow \infty$ in the reconstruction function $\beta(\cdot)$. The resulting $\beta(x)$ would still be well defined, since $\lim_{m \rightarrow \infty} \beta(x) = x - x \log x$, uniformly over $x \in [0, 1]$, and Lemma 1 would still hold since its statement is independent of m . However, we would no longer be optimizing the “right” function $\hat{\mu}_i(p)$ but an approximation thereof, the error in this approximation propagating across time steps in an *additive* manner, so that at the end of T steps we would obtain a *linear* regret bound of the form $\tilde{O}(d_m T + \sqrt{T})$, where d_m is constant with T , and goes to zero as the true underlying m goes to infinity. The question whether it is possible to refine this simple argument so as to achieve a nontrivial (i.e., sublinear in T) cumulative regret bound without knowing anything about m remains open.

APPENDIX MAIN PROOFS

Proof of Fact 1: By definition of $R(p)$ we can write

$$\mu(p) = \int_p^1 x dF_2(x) + p \mathbb{P}(B^{(2)} < p \leq B^{(1)}). \quad (9)$$

By applying the identity $\mathbb{E}[X] = \int \mathbb{P}(X > x) dx$ to the nonnegative random variable $B^{(2)} \mathbb{1}_{\{B^{(2)} > p\}}$ we obtain

$$\begin{aligned} \int_p^1 x dF_2(x) &= p(1 - F_2(p)) + \int_p^1 (1 - F_2(x)) dx \\ &= p - pF_2(p) + \mathbb{E}[B^{(2)}] \\ &\quad - \int_0^p (1 - F_2(x)) dx \\ &= \mathbb{E}[B^{(2)}] - pF_2(p) + \int_0^p F_2(t) dt. \end{aligned}$$

Moreover,

$$F_2(p) = m(F(p))^{m-1}(1 - F(p)) + (F(p))^m$$

and

$$\mathbb{P}(B^{(2)} < p \leq B^{(1)}) = m(1 - F(p))(F(p))^{m-1}.$$

Substituting the above into (9) and simplifying concludes the proof. \square

Proof of Lemma 1: A simple derivative argument shows that the function $\beta(\cdot)$ is a strictly increasing and concave mapping

from $[0, 1]$ onto $[0, 1]$. Hence its inverse $\beta^{-1}(\cdot)$ exists and is strictly increasing and *convex* on $[0, 1]$. From our assumptions we immediately have: (i) $x \leq \beta^{-1}(a + \epsilon)$ for any $\epsilon \in [0, 1 - a]$, and (ii) $\beta^{-1}(a - \epsilon) \leq x$ for any $\epsilon \in [0, a]$.

In turn, because of the convexity of $\beta^{-1}(\cdot)$, we have

$$\beta^{-1}(a + \epsilon) \leq \beta^{-1}(a) + \frac{1 - \beta^{-1}(a)}{1 - a} \epsilon, \quad \forall \epsilon \in [0, 1 - a]. \quad (10)$$

Similarly, by the convexity and the monotonicity of $\beta^{-1}(\cdot)$ we can write

$$\begin{aligned} \beta^{-1}(a - \epsilon) &\geq \beta^{-1}(a) - \frac{d\beta^{-1}(x)}{dx} \Big|_{x=a} \epsilon \\ &\geq \beta^{-1}(a) - \frac{1 - \beta^{-1}(a)}{1 - a} \epsilon, \quad \forall \epsilon \in [0, a]. \end{aligned} \quad (11)$$

At this point, we need the following technical claim.

Claim 1:

$$1 - \beta^{-1}(a) \leq 2\sqrt{1 - a}, \quad \forall a \in [0, 1].$$

Proof of Claim: Note that the case $a \in [0, 3/4]$ holds trivially since $2\sqrt{1 - a} \geq 1$, and therefore we need to consider only $a \in [3/4, 1]$. Introduce the auxiliary function $f(a) = 1 - 2\sqrt{1 - a}$. The claim is proven by showing that $\beta(f(a)) \leq a$ for all $a \in [3/4, 1]$. We prove the claim by showing that $\beta(f(a))$ is a concave function of $a \in [3/4, 1]$, and that $\frac{d\beta(f(a))}{da} \Big|_{a=1} \geq 1$, while $\beta(f(1)) = 1$. We have

$$\frac{d\beta(f(a))}{da} = -2(m-1) \frac{1 - (f(a))^{-1/m}}{1 - f(a)}.$$

Hence, using L'Hopital's rule,

$$\frac{d\beta(f(a))}{da} \Big|_{a=1} = \frac{2(m-1)}{m} \geq 1$$

since $m \geq 2$. Moreover,

$$\begin{aligned} \frac{d^2\beta(f(a))}{da^2} &= -\left(\frac{m-1}{m}\right) (f(a))^{-\frac{m+1}{m}} \times \frac{1}{1-a} \\ &\quad + \frac{m-1}{2} \times \frac{(f(a))^{-1/m} - 1}{(1-a)^{3/2}} \end{aligned}$$

which is nonpositive if and only if $m((f(a))^{-1/m} - 1) \leq (1 - f(a))(f(a))^{-\frac{m+1}{m}}$ holds for any $a \in [3/4, 1]$. Since $f(a)$ ranges in $[0, 1]$ when $a \in [3/4, 1]$, after some simplifications, one can see that the above inequality is equivalent to

$$(m+1)x \leq m x^{\frac{m+1}{m}} + 1, \quad \forall x \in [0, 1].$$

In turn, this inequality can be seen to hold by showing via a simple derivative argument that the function $g(x) = m x^{\frac{m+1}{m}} + 1$ is convex and increasing for $x \in [0, 1]$, while $g(0) = 1 > 0$ and $g'(1) = m + 1$. \square

The claim together with (10) and (11) allows us to conclude the proof of Lemma 1. Specifically, the second inequality in (2) is obtained by (10) and extended to any $\epsilon \geq 0$ just by observing that, by the claim, for $\epsilon > 1 - a$ the right-most side of (2) is larger than 1. Moreover, the first inequality in (2) is obtained by (11) and extended to any $\epsilon \geq 0$ by observing that for $\epsilon > a$ the left-most side of (2) is smaller than $\beta^{-1}(a) - \frac{2a}{\sqrt{1-a}} \leq a - \frac{2a}{\sqrt{1-a}} \leq 0$ for any $a \in [0, 1]$, where we have used the fact that $\beta^{-1}(a) \leq a$. \square

Proof of Lemma 2: Let $B_k^{(1)}$ and $B_k^{(2)}$ denote the maximum and the second-maximum of k i.i.d. bids B_1, \dots, B_k . Set for brevity $A = \mathbb{P}(B_m^{(1)} > p)$. Then we have

$$A \leq 2 \mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p)$$

and

$$A \leq 2 \mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p) + \mathbb{P}(B_1 > p) \leq 3 \mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p).$$

Hence

$$\begin{aligned} 1 - F_2(p) &= \mathbb{P}(B_m^{(2)} > p) \\ &\geq \mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p) \times \mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p) \\ &\geq \frac{A}{3} \times \frac{A}{2}. \end{aligned}$$

In turn, $A \geq \mu(p)$, since each time all the bids are less than p the revenue is zero. Therefore we have obtained that

$$1 - F_2(p) \geq \frac{\mu^2(p)}{6}$$

holds for all $p \in [0, 1]$. Finally, since $\widehat{F}_{2,i}$ is the empirical version of F_2 based on the observed revenues during stage i (see Section II-C), the classical Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [13] implies that with probability at least $1 - \delta/3S$, conditioned on all past stages,

$$\max_{p \in [\widehat{p}_i, 1]} |\widehat{F}_{2,i}(p) - F_2(p)| \leq \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}.$$

□

Proof of Lemma 3: We start by proving (3). Fix any stage i and write

$$\begin{aligned} |\mu(p) - \widehat{\mu}_i(p)| &\leq \left| \int_0^p F_2(t) dt - \int_0^p \widehat{F}_{2,i}(t) dt \right| \\ &\quad + p \left| (F(p))^m - \beta^{-1}(\widehat{F}_{2,i}(p)) \right| \\ &\leq \int_0^p |F_2(t) - \widehat{F}_{2,i}(t)| dt \\ &\quad + p \left| (F(p))^m - \beta^{-1}(\widehat{F}_{2,i}(p)) \right| \\ &\leq p \max_{t \in [0, p]} |F_2(t) - \widehat{F}_{2,i}(t)| \\ &\quad + p \left| (F(p))^m - \beta^{-1}(\widehat{F}_{2,i}(p)) \right|. \end{aligned} \quad (12)$$

The DKW inequality implies that

$$p \max_{t \in [0, p]} |F_2(t) - \widehat{F}_{2,i}(t)| \leq p \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}} \leq C_{\delta,i}(p) \quad (13)$$

holds with probability at least $1 - \delta/(3S)$. As for the second term in (12) we apply again the DKW inequality in combination with Lemma 1 with $x = (F(p))^m = \beta^{-1}(F_2(p))$, $a = \widehat{F}_{2,i}(p)$, and $\epsilon = \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}$. This yields

$$p \left| \beta^{-1}(F_2(p)) - \beta^{-1}(\widehat{F}_{2,i}(p)) \right| \leq C_{\delta,i}(p)$$

with the same probability of at least $1 - \delta/(3S)$. Putting together and using the union bound over the S stages gives (3).

We prove (4) by induction on $i = 1, \dots, S$. We first show that the base case $i = 1$ holds with probability at least $1 - \delta/S$.

Then we show that if (4) holds for $i - 1$, then it holds for i with probability at least $1 - \delta/S$ over all random events in stage i . Therefore, using a union bound over $i = 1, \dots, S$ we get that (4) holds simultaneously for all i with probability at least $1 - \delta$.

For the base case $i = 1$ note that $\widehat{\mu}_1(p^*) \leq \widehat{\mu}_1(\widehat{p}_1^*)$ holds with probability at least $1 - \delta/(3S)$ because we are assuming (Lemma 2) that $\widehat{F}_2(p^*) \leq 1 - \alpha$ holds with the same probability, and so \widehat{p}_1^* maximizes $\widehat{\mu}_1$ over a range that with probability at least $1 - \delta/(3S)$ contains p^* . Moreover, using (3) we obtain

$$\mu(p^*) - \widehat{\mu}_1(p^*) \leq 2C_{\delta,1}(p^*)$$

and

$$\widehat{\mu}_1(\widehat{p}_1^*) - \mu(\widehat{p}_1^*) \leq 2C_{\delta,1}(\widehat{p}_1^*).$$

Since $\mu(\widehat{p}_1^*) - \mu(p^*) \leq 0$ by definition of p^* , we obtain

$$0 \leq \widehat{\mu}_1(\widehat{p}_1^*) - \widehat{\mu}_1(p^*) \leq 2C_{\delta,1}(\widehat{p}_1^*) + 2C_{\delta,1}(p^*)$$

as required. Finally, $p^* \geq \widehat{p}_1$ trivially holds because $\widehat{p}_1 = 0$.

We now prove (4) for $i > 1$ using the inductive assumption $p^* \geq \widehat{p}_{i-1}$ and

$$0 \leq \widehat{\mu}_{i-1}(\widehat{p}_{i-1}^*) - \widehat{\mu}_{i-1}(p^*) \leq 2C_{\delta,i-1}(\widehat{p}_{i-1}^*) + 2C_{\delta,i-1}(p^*).$$

The inductive assumption and $\widehat{F}_{2,i}(p^*) \leq 1 - \alpha$ directly imply $p^* \in P_i \cap \{p : \widehat{F}_{2,i-1}(p) \leq 1 - \alpha\}$ (recall the definition of the set of candidate prices P_i given in Algorithm 1). Thus we have $p^* \geq \widehat{p}_i$ and $\widehat{\mu}_i(\widehat{p}_i^*) \geq \widehat{\mu}_i(p^*)$, because \widehat{p}_i^* maximizes $\widehat{\mu}_i$ over a range that contains p^* . The rest of the proof closely follows that of (4) for the base case $i = 1$. □

Proof of Theorem 1: If $S = S(T)$ is the total number of stages, then the regret of our algorithm is

$$\begin{aligned} (\mu(p^*) - \mu(\widehat{p}_0)) T_1 + \sum_{i=2}^S (\mu(p^*) - \mu(\widehat{p}_i)) T_i \\ \leq T_1 + \sum_{i=2}^S (\mu(p^*) - \mu(\widehat{p}_i)) T_i. \end{aligned} \quad (14)$$

For all stages $i > 1$ the following chain on inequalities jointly hold with probability at least $1 - \delta$ uniformly over $i = 2, \dots, S$,

$$\begin{aligned} \mu(p^*) - \mu(\widehat{p}_i) \\ &\leq \mu(p^*) - \widehat{\mu}_{i-1}(\widehat{p}_i) + 2C_{\delta,i-1}(\widehat{p}_i) \\ &\quad \text{(by (3) —note that } \widehat{p}_i \geq \widehat{p}_{i-1}\text{)} \\ &\leq \mu(p^*) - \widehat{\mu}_{i-1}(\widehat{p}_{i-1}^*) + 2C_{\delta,i-1}(\widehat{p}_{i-1}^*) + 4C_{\delta,i-1}(\widehat{p}_i) \\ &\quad \text{(since } \widehat{p}_i \in P_i\text{)} \\ &\leq \mu(p^*) - \widehat{\mu}_{i-1}(p^*) + 2C_{\delta,i-1}(\widehat{p}_{i-1}^*) + 4C_{\delta,i-1}(\widehat{p}_i) \\ &\quad \text{(since } \widehat{\mu}_{i-1}(\widehat{p}_{i-1}^*) \geq \widehat{\mu}_{i-1}(p^*) \text{ —see (4))} \\ &\leq 2C_{\delta,i-1}(p^*) + 2C_{\delta,i-1}(\widehat{p}_{i-1}^*) + 4C_{\delta,i-1}(\widehat{p}_i) \\ &\quad \text{(by } p^* \geq \widehat{p}_i \text{ combined with (3))} \\ &\leq 8 \sqrt{\frac{1}{\alpha T_{i-1}} \ln \frac{6S}{\delta}} \end{aligned}$$

where in the last step we used the fact that $\widehat{F}_{2,i-1}(p^*) \leq 1 - \alpha$ holds by Lemma 2, and that $\widehat{F}_{2,i-1}(p) \leq 1 - \alpha$ for $p = \widehat{p}_i$ and

$p = \widehat{p}_{i-1}^*$ by the very definitions of \widehat{p}_i and \widehat{p}_{i-1}^* , respectively. Substituting back into (14) we see that with probability at least $1 - \delta$ the regret of our algorithm is at most

$$T_1 + 8 \sum_{i=2}^S T_i \sqrt{\frac{1}{\alpha T_{i-1}} \ln \frac{6S}{\delta}}.$$

Our setting $T_i = T^{1-2^{-i}}$ for $i = 1, 2, \dots$ implies that S is upper bounded by the minimum integer n such that

$$\sum_{i=1}^n T^{1-2^{-i}} \geq T.$$

Since $i \geq \log_2 \log_2 T$ makes $T_i \geq \frac{T}{2}$, then $S \leq \lceil 2 \log_2 \log_2 T \rceil = \mathcal{O}(\log \log T)$. Moreover, observe that $T_i = T^{1-2^{-i}}$ is equivalent to $T_1 = \sqrt{T}$ and $\frac{T_i}{\sqrt{T_{i-1}}} = \sqrt{T}$, for $i > 1$. We therefore have the upper bound

$$(14) \leq \sqrt{T} + 8\sqrt{T}S \sqrt{\frac{1}{\alpha} \ln \frac{6S}{\delta}}. \quad (15)$$

If $\mu(p^*) > 0$ and

$$\min_i T_i = T_1 = \sqrt{T} \geq \frac{72 \ln(6S/\delta)}{\mu(p^*)^4}$$

then $\alpha \geq \mu(p^*)^2/12$, and the above is of order

$$\frac{\sqrt{T(\log \log \log T + \log 1/\delta)} (\log \log T)}{\mu(p^*)}$$

as claimed. \square

Proof of Lemma 4: We start by observing that $T(0) = A(0) = 0$, $T(1) = A(1) = 1$, $T'(x) \geq 0$ for $x \in [0, 1]$, and $A'(x) = (1-x)T''(x) \geq 0$ when $x \in [0, 1]$. Hence both $T(x)$ and $A(x)$ are strictly increasing mappings from $[0, 1]$ onto $[0, 1]$, and so are $T^{-1}(x)$, $A^{-1}(x)$ and $A(T^{-1}(x))$. Hence our assumptions on x can be rewritten as

$$T(A^{-1}(a - \epsilon)) \leq x \leq T(A^{-1}(a + \epsilon)).$$

Moreover, since $T(\cdot)$ and $A(\cdot)$ are both $C^\infty(0, 1)$, so is $T(A^{-1}(\cdot))$. Let $\epsilon < 1 - a$. We can write

$$T(A^{-1}(a + \epsilon)) = T(A^{-1}(a)) + \epsilon \left. \frac{dT(A^{-1}(x))}{dx} \right|_{x=\zeta}$$

for some $\zeta \in (a, a + \epsilon)$, where

$$\frac{dT(A^{-1}(x))}{dx} = \frac{T'(y)}{A'(y)} = \frac{T'(y)}{(1-y)T''(y)}$$

and we set for brevity $y = A^{-1}(x) \in [0, 1]$. Now, for any $y \in [0, 1]$,

$$\frac{T'(y)}{T''(y)} = \frac{y \sum_{m \geq 2} m Q(m) y^{m-2}}{\sum_{m \geq 2} m(m-1) Q(m) y^{m-2}} \leq y \leq 1.$$

As a consequence, since A^{-1} is a nondecreasing function, we can write

$$\left. \frac{dT(A^{-1}(x))}{dx} \right|_{x=\zeta} \leq \frac{1}{1 - A^{-1}(a + \epsilon)} \leq \frac{1}{1 - T^{-1}(a + \epsilon)} \quad (16)$$

the last inequality deriving from¹¹ $A(x) \geq T(x)$ for all $x \in [0, 1]$. Finally, from the convexity of T we have $T(x) \geq T(1) + (x-1)T'(1) = 1 + (x-1)\mathbb{E}[M]$. Thus $T^{-1}(x) \leq 1 - \frac{1-x}{\mathbb{E}[M]}$, $x \in [0, 1]$, which we plug back into (16) to see that

$$\left. \frac{dT(A^{-1}(x))}{dx} \right|_{x=\zeta} \leq \frac{\mathbb{E}[M]}{1 - (a + \epsilon)}.$$

Replacing backwards, this yields the second inequality of (6).

To prove the first inequality of (6), we start off showing it to hold for $\epsilon < \min\{a, 1-a\}$, and then extend it to $\epsilon < 1-a$. Set $\epsilon < a$. Then proceeding as above we can see that, for some $\zeta \in (a - \epsilon, a)$,

$$\begin{aligned} T(A^{-1}(a)) &= T(A^{-1}(a - \epsilon)) + \epsilon \left. \frac{dT(A^{-1}(x))}{dx} \right|_{x=\zeta} \\ &\leq T(A^{-1}(a - \epsilon)) + \frac{\epsilon}{1 - T^{-1}(a)} \\ &\leq T(A^{-1}(a - \epsilon)) + \frac{\epsilon \mathbb{E}[M]}{1 - a} \\ &\leq T(A^{-1}(a - \epsilon)) + \frac{\epsilon \mathbb{E}[M]}{1 - (a + \epsilon)} \end{aligned}$$

the last inequality requiring also $\epsilon < 1 - a$. If now ϵ satisfies $a \leq \epsilon < 1 - a$ (assuming $a < 1/2$) then the first inequality of (6) is trivially fulfilled. In fact,

$$\begin{aligned} T(A^{-1}(a)) - \frac{\epsilon \mathbb{E}[M]}{1 - (a + \epsilon)} &\leq A(A^{-1}(a)) - \frac{a \mathbb{E}[M]}{1 - 2a} \\ &= a \left(1 - \frac{\mathbb{E}[M]}{1 - 2a} \right) \\ &< 0 \end{aligned}$$

since $\mathbb{E}[M] \geq 2$. This concludes the proof of (6).

In order to prove (8), we set for brevity $y = A^{-1}(x)$, and using the rules of differentiating inverse functions, we see that

$$\begin{aligned} \frac{d^2 T(A^{-1}(x))}{dx^2} &= \frac{1}{(1-y)^2 T''(y)} + \frac{T'(y)}{(1-y)^3 (T''(y))^2} \\ &\quad - \frac{T'(y)T'''(y)}{(1-y)^2 (T''(y))^3}. \end{aligned}$$

Thus $\frac{d^2}{dx^2} T(A^{-1}(x)) \geq 0$ for $x \in [0, 1]$ is equivalent to

$$\begin{aligned} (T''(y))^2 (1-y) + T'(y)T''(y) \\ \geq T'(y)T'''(y)(1-y), \quad \forall x \in [0, 1]. \end{aligned}$$

Since y ranges over $[0, 1]$ when x does, (7) is actually equivalent to the convexity of $T(A^{-1}(x))$ on $x \in [0, 1]$. Under the above convexity assumption, we can write, for $\epsilon \leq 1 - a$,

$$\begin{aligned} T(A^{-1}(a + \epsilon)) &\leq T(A^{-1}(a)) + \frac{1 - T(A^{-1}(a))}{1 - a} \epsilon \\ &\leq T(A^{-1}(a)) + \frac{\epsilon}{1 - a}. \end{aligned}$$

On the other hand, if $\epsilon > 1 - a$ the above inequality vacuously holds, since the right-hand side is larger than one, while $T(A^{-1}(x)) \leq 1$ for any $x \in [0, 1]$. This proves the

¹¹Whereas the function $A(\cdot)$ is, in general, neither convex nor concave, $T(\cdot)$ is a convex lower bound on $A(\cdot)$.

second inequality in (8). Similarly, by the convexity and the monotonicity of $T(A^{-1}(\cdot))$ we can write, for all $\epsilon \in [0, a]$,

$$\begin{aligned} T(A^{-1}(a - \epsilon)) &\geq T(A^{-1}(a)) - \left. \frac{dT(A^{-1}(x))}{dx} \right|_{x=a} \epsilon \\ &\geq T(A^{-1}(a)) - \frac{1 - T(A^{-1}(a))}{1 - a} \epsilon \\ &\geq T(A^{-1}(a)) - \frac{\epsilon}{1 - a} \end{aligned}$$

which gives the first inequality in (8). We extend the above to any $\epsilon \geq 0$ by simply observing that $\epsilon > a$ implies that $T(A^{-1}(a)) - \frac{\epsilon}{1-a} < a - \frac{a}{1-a} < 0$, where $T(A^{-1}(a)) \leq a$ follows from the convexity of $T(A^{-1}(\cdot))$. This makes (8) trivially fulfilled. \square

Proof of Lemma 5: Let $B_m^{(1)}$ and $B_m^{(2)}$ denote the highest and the second-highest of m i.i.d. bids B_1, \dots, B_m . Recall from the proof of Lemma 2 that, for any $m \geq 2$

$$\mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p) \geq \frac{1}{2} \mathbb{P}(B_m^{(1)} > p)$$

and

$$\mathbb{P}(B_{\lfloor m/2 \rfloor}^{(1)} > p) \geq \frac{1}{3} \mathbb{P}(B_m^{(1)} > p).$$

Moreover,

$$\begin{aligned} 1 - \bar{F}_2(p) &= \mathbb{E}_M \left[\mathbb{P}(B_M^{(2)} > p \mid M) \right] \\ &\geq \mathbb{E}_M \left[\mathbb{P}(B_{\lfloor M/2 \rfloor}^{(1)} > p \mid M) \times \mathbb{P}(B_{\lfloor M/2 \rfloor}^{(1)} > p \mid M) \right] \\ &\geq \frac{1}{6} \mathbb{E}_M \left[\left(\mathbb{P}(B_M^{(1)} > p \mid M) \right)^2 \right] \\ &\geq \frac{1}{6} \left(\mathbb{E}_M \left[\mathbb{P}(B_M^{(1)} > p \mid M) \right] \right)^2 \geq \frac{1}{6} \mu^2(p) \end{aligned}$$

the second-last inequality being Jensen's, and the last one deriving from $\mathbb{I}_{\{B_m^{(1)} > p\}} \geq R^m(p)$ for all $m \geq 2$ and $p \in [0, 1]$.

We then conclude as in the proof of Lemma 2 by applying DKW on the uniform convergence of $\hat{F}_{2,i}$ to \bar{F}_2 . \square

Proof of Theorem 2: Consider a setting with two bidders ($m = 2$) where both bids B_1, B_2 are revealed at the end of each auction, irrespective of the chosen reserve price. Note that a lower bound in this setting implies a lower bound in the harder setting of Theorem 1, in which only the revenue is revealed.

Consider bid distributions of the form

$$\mathbb{P}(B = \frac{1}{2}) = \frac{1}{2} + \epsilon \quad \text{and} \quad \mathbb{P}(B = \frac{3}{4}) = \frac{1}{2} - \epsilon.$$

Since the bid distribution is supported on $\{\frac{1}{2}, \frac{3}{4}\}$, the expected revenue of a reserve price $0 < p < \frac{1}{2}$ is never greater than that of $p = \frac{1}{2}$. Similarly, the expected revenue of a reserve price $\frac{1}{2} < p < \frac{3}{4}$ is never greater than that of $p = \frac{3}{4}$. Therefore, without loss of generality we may restrict our attention to strategies that select their prices from $\{\frac{1}{2}, \frac{3}{4}\}$.

We now compute the expected revenue of $p = \frac{1}{2}$ and $p = \frac{3}{4}$,

$$\begin{aligned} \mu(\frac{1}{2}) &= \frac{1}{2} \left(\frac{1}{2} + \epsilon \right)^2 + 2 \times \frac{1}{2} \left(\frac{1}{4} - \epsilon^2 \right) + \frac{3}{4} \left(\frac{1}{2} - \epsilon \right)^2 \\ &= \left(\frac{3}{4} \right)^2 - \frac{1}{4} \epsilon + \frac{1}{4} \epsilon^2 \\ \mu(\frac{3}{4}) &= 2 \times \frac{3}{4} \left(\frac{1}{4} - \epsilon^2 \right) + \frac{3}{4} \left(\frac{1}{2} - \epsilon \right)^2 \\ &= \left(\frac{3}{4} \right)^2 - \frac{3}{4} \epsilon - \frac{3}{4} \epsilon^2. \end{aligned}$$

Therefore

$$\mu(\frac{1}{2}) - \mu(\frac{3}{4}) = \frac{\epsilon}{2} + \epsilon^2. \quad (17)$$

For any fixed $\epsilon \in (0, \frac{1}{4})$, consider now the random variable $Z \in \{-1, +1\}$ such that $\mathbb{P}(Z = +1) = \mathbb{P}(Z = -1) = \frac{1}{2}$ and let μ_P be the expected revenue function of the bid distribution

$$\mathbb{P}(B = a) = \begin{cases} \frac{1}{2} + Z\epsilon & \text{if } a = \frac{1}{2} \\ \frac{1}{2} - Z\epsilon & \text{if } a = \frac{3}{4}. \end{cases}$$

We now prove that no deterministic mechanism can have small regret on both conditional bid distributions $\mathbb{P}(\cdot \mid Z = +1)$ and $\mathbb{P}(\cdot \mid Z = -1)$

According to (17),

$$\mu_P(\frac{1}{2} \mid Z = +1) = \mu_P(\frac{3}{4} \mid Z = +1) + \frac{\epsilon}{2} + \epsilon^2.$$

Moreover, switching ϵ to $-\epsilon$ gives

$$\mu_P(\frac{3}{4} \mid Z = -1) = \mu_P(\frac{1}{2} \mid Z = -1) + \frac{\epsilon}{2} - \epsilon^2.$$

Since ϵ is chosen of the order of $T^{-1/2}$, in the rest of the proof we may ignore the term ϵ^2 appearing in the expected revenue function μ_P . This adds a constant to the regret, which is taken into account by the asymptotic notation. Now let $p^* = p^*(Z)$ be the optimal reserve price for the conditional bid distribution. That is, $p^* = \frac{1}{2}$ if $Z = +1$ and $p^* = \frac{3}{4}$ if $Z = -1$.

Fix any deterministic algorithm choosing reserve prices p_1, p_2, \dots . Let $T_{1/2}$ and $T_{3/4}$ be the number of times $p_t = \frac{1}{2}$ and $p_t = \frac{3}{4}$, respectively. Finally, let T^* be the number of times $p_t = p^*$. Because the regret increases by $\frac{\epsilon}{2}$ every time $p_t \neq p^*$ (recall that we are ignoring the ϵ^2 term in μ_P), the expected regret of the algorithm with respect to the worst-case choice of Z is

$$\begin{aligned} &\max_{z \in \{-1, +1\}} \sum_{t=1}^T \mathbb{E}[\mu(p^*) - \mu(p_t) \mid Z = z] \\ &\geq \sum_{t=1}^T \mathbb{E}[\mu(p^*) - \mu(p_t)] \\ &= \frac{\epsilon}{2} (T - \mathbb{E}[T^*]) \\ &= \frac{\epsilon}{2} \left(T - \frac{1}{2} (\mathbb{E}[T_{1/2} \mid Z = +1] + \mathbb{E}[T_{3/4} \mid Z = -1]) \right). \end{aligned}$$

In this simplified setting, where the mechanism can observe the individual bids, each p_t is determined by the independent bid pairs $(B_1^{(1)}, B_1^{(2)}), \dots, (B_T^{(1)}, B_T^{(2)})$. Let P_T^+ and P_T^- be the joint distributions of the bid pairs when $Z = +1$ and

$Z = -1$, respectively. Finally, let $P_T = \frac{1}{2}(P_T^+ + P_T^-)$. Then, Pinsker's inequality implies

$$\begin{aligned} P_T^+(p_t = \frac{1}{2}) &\leq P_T(p_t = \frac{1}{2}) + \sqrt{\frac{1}{2}\text{KL}(P_T \| P_T^+)} \\ &\leq P_T(p_t = \frac{1}{2}) + \sqrt{\frac{1}{4}\text{KL}(P_T^- \| P_T^+)} \end{aligned}$$

where we used the convexity of the relative entropy $\text{KL}(\cdot \| P_T^+)$ in the last step.

We now recognize that P_T^+ and P_T^- are product distributions of T pairs of (shifted and scaled) independent Bernoulli random variables $B_{\frac{1}{2}+\epsilon}$ and $B_{\frac{1}{2}-\epsilon}$, with parameters $\frac{1}{2} + \epsilon$ and $\frac{1}{2} - \epsilon$, respectively. Therefore, taking the scaling factor into account and using the chain rule for relative entropy gives

$$\begin{aligned} \text{KL}(P_T^- \| P_T^+) &\leq \text{KL}(P_{T-1}^- \| P_{T-1}^+) + 2\text{KL}\left(B_{\frac{1}{2}-\epsilon} \| B_{\frac{1}{2}+\epsilon}\right) \\ &\leq 2T\text{KL}\left(B_{\frac{1}{2}-\epsilon} \| B_{\frac{1}{2}+\epsilon}\right) \\ &\leq 32\epsilon^2 T \end{aligned}$$

the last inequality holding if $\epsilon \in [0, 0.47]$. Hence

$$\begin{aligned} \mathbb{E}[T_{1/2} | Z = +1] &\leq \mathbb{E}[T_{1/2}] + T\sqrt{\frac{1}{4} \cdot 32\epsilon^2 T} \\ &= \mathbb{E}[T_{1/2}] + \epsilon T^{3/2}\sqrt{8}. \end{aligned}$$

Similarly,

$$\mathbb{E}[T_{3/4} | Z = -1] \leq \mathbb{E}[T_{3/4}] + \epsilon T^{3/2}\sqrt{8}.$$

Therefore

$$\begin{aligned} &\frac{1}{2}\left(\mathbb{E}[T_{1/2} | Z = +1] + \mathbb{E}[T_{3/4} | Z = -1]\right) \\ &\leq \frac{1}{2}\left(\mathbb{E}[T_{1/2}] + \mathbb{E}[T_{3/4}]\right) + \epsilon T^{3/2}\sqrt{8} \\ &= \frac{T}{2} + \epsilon T^{3/2}\sqrt{8}. \end{aligned}$$

This implies

$$\sum_{t=1}^T \mathbb{E}[\mu(p^*) - \mu(p_t)] \geq \frac{\epsilon}{2}\left(T - \frac{T}{2} - \epsilon T^{3/2}\sqrt{8}\right).$$

Choosing $\epsilon = \Theta(T^{-1/2})$ concludes the proof of the theorem. \square

Proof of Lemma 7: Since the revenue $R^M(p)$ is determined by $B_M^{(1)}$ and $B_M^{(2)}$ only, we use the notation $R_p(b_1, b_2)$ to denote the revenue $R^M(p)$ when $B_M^{(1)} = b_1$ and $B_M^{(2)} = b_2$. Since $b_1 \geq b_2$, in order to compute the pseudo-dimension of \mathcal{F} we have to determine the largest number of points shattered in the region $S = \{(b_1, b_2) : 0 \leq b_2 \leq b_1 \leq 1\} \subset \mathbb{R}^2$ where the functions R_p are defined as

$$R_p(b_1, b_2) = \begin{cases} b_2 & \text{if } b_2 \geq p \\ p & \text{if } b_2 < p \leq b_1 \\ 0 & \text{if } b_1 < p. \end{cases}$$

Note that each function R_p defines an axis-parallel rectangle with corners (p, p) , $(p, 0)$, $(1, p)$ and $(1, 0)$. Inside the rectangle $R_p = p$, to the left of the rectangle $R_p = 0$, and points $(b_1, b_2) \in S$ above it satisfy $R_p(b_1, b_2) = b_2$.

We now show that \mathcal{F} shatters any two points $x_1 = (b_1, b_2)$ and $x_2 = (b_1 + \epsilon, b_2 + \epsilon)$ in the region S such that $\epsilon > 0$ and $b_2 + \epsilon < b_1$. This is shown by the following case analysis where we use the values $r_1 = b_2$ and $r_2 = b_2 + \epsilon$.

- $p > b_1 + \epsilon$, this realizes the pattern $(0, 0)$ because $R_p(x_1) = 0 \leq b_2$ and $R_p(x_2) = 0 \leq b_2 + \epsilon$;
- $b_1 < p < b_1 + \epsilon$, this realizes the pattern $(0, 1)$ because $R_p(x_1) = 0 \leq b_2$ and $R_p(x_2) = p > b_2 + \epsilon$;
- $b_2 + \epsilon < p < b_1$, this realizes the pattern $(1, 1)$ because $R_p(x_1) = p > b_2$ and $R_p(x_2) = p > b_2 + \epsilon$;
- $b_2 < p < b_2 + \epsilon$, this realizes the pattern $(1, 0)$ because $R_p(x_1) = p > b_2$ and $R_p(x_2) = b_2 + \epsilon$.

In order to prove that \mathcal{F} can not shatter any three points in S , arrange on the real line the six coordinate values of these three points. These six numbers define seven intervals. When p ranges within any such interval, the value of R_p must remain constant on all the three points. This is because the value of $R_p(b_1, b_2)$ changes only when p crosses b_1 or b_2 . But then, \mathcal{F} can only realize at most seven of the eight patterns needed to shatter the three points. \square

Proof of Theorem 4: For the sake of brevity, let $R_t(p)$ denote $R_t^{M_t}(p)$. Also, let $\mathbb{E}_t[\cdot]$ be the conditional expectation $\mathbb{E}_t[\cdot | p_1, \dots, p_{t-1}]$, i.e., the expectation of the random variable at argument conditioned on all past bids and all past number of bidders. Let p_T^* be the random variable defined as

$$p_T^* = \operatorname{argmax}_{p \in [0, 1]} \sum_{t=1}^T R_t(p).$$

Then

$$\begin{aligned} &\sum_{t=1}^T R_t(p_T^*) - \sum_{t=1}^T R_t(p_t) \\ &= \sum_{t=1}^T R_t(p_T^*) - T\mu(p_T^*) + \sum_{t=1}^T \mathbb{E}_t[R_t(p_t)] - \sum_{t=1}^T R_t(p_t) \\ &\quad + T\mu(p_T^*) - \sum_{t=1}^T \mathbb{E}_t[R_t(p_t)] \\ &\leq \max_{p \in [0, 1]} \left(\sum_{t=1}^T R_t(p) - T\mu(p) \right) \end{aligned} \quad (18)$$

$$+ \sum_{t=1}^T \left(\mathbb{E}_t[R_t(p_t)] - R_t(p_t) \right) \quad (19)$$

$$+ T\mu(p_T^*) - \sum_{t=1}^T \mathbb{E}_t[R_t(p_t)]. \quad (20)$$

In order to bound (18) we combine Lemma 6 with Lemma 7. This gives

$$\max_{p \in [0, 1]} \left(\sum_{t=1}^T R_t(p) - T\mu(p) \right) \leq C\sqrt{2T \ln \frac{1}{\delta}}$$

with probability at least $1 - \delta$, where C is the constant mentioned in Lemma 6.

In order to bound (19), note that $Z_t = \mathbb{E}_t[R_t(p_t)] - R_t(p_t)$ for $t = 1, 2, \dots$ is a martingale difference sequence with

bounded increments, $\mathbb{E}_t[Z_t] = 0$ with $Z_t \in [-1, 1]$ for each t . Therefore, the Hoeffding-Azuma inequality for martingales establishes that

$$\sum_{t=1}^T \left(\mathbb{E}_t[R_t(p_t)] - R_t(p_t) \right) \leq \sqrt{2T \ln \frac{1}{\delta}}$$

with probability at least $1 - \delta$.

Finally, term (20) is bounded via Theorem 3 after observing that $\mu(p_T^*) \leq \mu(p^*)$, where $p^* = \operatorname{argmax}_{p \in [0,1]} \mu(p)$ is the maximizer of the expected revenue. This concludes the proof. \square

Proof of Theorem 5: Following (15), the regret accumulated in block r is upper bounded by

$$\sqrt{T_r} \left(1 + 8 S_r \sqrt{\frac{1}{\alpha} \ln \frac{6S_r}{\delta_r}} \right)$$

with probability at least $1 - \delta_r$, where $S_r \leq \lceil 2 \log_2 \log_2 T_r \rceil = \mathcal{O}(\log r)$, provided $\mu(p^*) > 0$ and

$$\sqrt{T_r} \geq \frac{72 \ln(6S_r/\delta_r)}{\mu(p^*)^4}. \quad (21)$$

On the other hand, if (21) is false, then we can simply upper bound the cumulative regret in block r by its length T_r .

Because the algorithm is restarted at the beginning of each block, these cumulative regrets have to be summed over blocks. Let r_T be the index of the final block. Clearly, if the total number of steps is T then $T_{r_T} \leq 2T$, and $r_T \leq 1 + \log_2 T$. Moreover, denote by \bar{r} the larger r such that (21) is false. Taking a union bound over blocks (and noting that $\sum_r \delta_r \leq \delta$), we have that with probability at least $1 - \delta$ the cumulative regret of Algorithm 1 is upper bounded by

$$\sum_{r=0}^{\bar{r}} T_r + \sum_{r=\bar{r}+1}^{r_T} \sqrt{T_r} \left(1 + 8 S_r \sqrt{\frac{1}{\alpha} \ln \frac{6S_r}{\delta_r}} \right).$$

We upper bound the two sums in the above expression separately. We have

$$\sum_{r=0}^{\bar{r}} T_r = \mathcal{O} \left(2^{\bar{r}} \right).$$

But by its very definition \bar{r} also satisfies

$$2^{\bar{r}/2} \leq \frac{72 \ln(6S_{r_T}/\delta_{r_T})}{\mu(p^*)^4}$$

so that the first sum can be overapproximated as

$$\sum_{r=0}^{\bar{r}} T_r = \mathcal{O} \left(\frac{1}{\mu(p^*)^8} \log^2 \left(\frac{(\log \log T)(T+1)(T+2)}{\delta} \right) \right).$$

As for the second sum, we can write

$$\begin{aligned} & \sum_{r=\bar{r}+1}^{r_T} \sqrt{T_r} \left(1 + 8 S_r \sqrt{\frac{1}{\alpha} \ln \frac{6S_r}{\delta_r}} \right) \\ & \leq \sum_{r=0}^{r_T} \sqrt{T_r} \left(1 + 8 S_r \sqrt{\frac{1}{\alpha} \ln \frac{6S_r}{\delta_r}} \right) \end{aligned}$$

$$\begin{aligned} & \leq \sum_{r=0}^{r_T} \sqrt{T_r} \left(1 + 8 S_{r_T} \sqrt{\frac{1}{\alpha} \ln \frac{6S_{r_T}}{\delta_{r_T}}} \right) \\ & = \sum_{r=0}^{r_T} \sqrt{T_r} \times \mathcal{O} \left(\frac{\log \log T}{\mu(p^*)} \sqrt{\log \log T + \log 1/\delta} \right) \\ & = \mathcal{O} \left(2^{r_T/2} \right) \times \mathcal{O} \left(\frac{\log \log T}{\mu(p^*)} \sqrt{\log \log T + \log 1/\delta} \right) \\ & = \mathcal{O} \left(\frac{\sqrt{T} \log \log T}{\mu(p^*)} \sqrt{\log \log T + \log 1/\delta} \right). \end{aligned}$$

Putting together proves the claimed bound. \square

REFERENCES

- [1] S. Athey and P. A. Haile, "Identification of standard auction models," *Econometrica*, vol. 70, no. 6, pp. 2107–2140, 2002.
- [2] P. Auer, R. Ortner, and C. Szepesvári, "Improved rates for the stochastic continuum-armed bandit problem," in *Proc. 20th Annu. Conf. Learn. Theory*, 2007, pp. 454–468.
- [3] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, nos. 2–3, pp. 235–256, 2002.
- [4] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári, "X-armed bandits," *J. Mach. Learn. Res.*, vol. 12, pp. 1587–1627, Feb. 2011.
- [5] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
- [6] E. W. Cope, "Regret and convergence bounds for a class of continuum-armed bandit problems," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1243–1253, Jun. 2009.
- [7] E. David, A. Rogers, N. R. Jennings, J. Schiff, S. Kraus, and M. H. Rothkopf, "Optimal design of English auctions with discrete bid levels," *ACM Trans. Internet Technol.*, vol. 7, no. 2, 2007, Art. ID 12.
- [8] P. Dhangwatnotai, T. Roughgarden, and Q. Yan, "Revenue maximization with a single sample," in *Proc. 11th ACM Conf. Electron. Commerce*, 2010, pp. 129–138.
- [9] P. A. Haile and E. Tamer, "Inference with an incomplete model of English auctions," *J. Political Economy*, vol. 111, no. 1, pp. 1–51, 2003.
- [10] J. D. Hartline and A. R. Karlin, "Profit maximization in mechanism design," in *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [11] A. X. Jiang and K. Leyton-Brown, "Bidding agents for online auctions with hidden bids," *Mach. Learn.*, vol. 67, nos. 1–2, pp. 117–143, 2007.
- [12] R. Kleinberg and T. Leighton, "The value of knowing a demand curve: Bounds on regret for online posted-price auctions," in *Proc. 44th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2003, pp. 594–605.
- [13] P. Massart, "The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality," *Ann. Probab.*, vol. 18, no. 3, pp. 1269–1283, 1990.
- [14] R. B. Myerson, "Optimal auction design," *Math. Oper. Res.*, vol. 6, no. 1, pp. 58–73, 1981.
- [15] N. Nisan, "Introduction to mechanism design (for computer scientists)," in *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [16] M. Ostrovsky and M. Schwarz, "Reserve prices in internet advertising auctions: A field experiment," in *Proc. 12th ACM Conf. Electron. Commerce*, 2011, pp. 59–60.
- [17] D. Pollard, *Empirical Processes: Theory and Applications* (Probability and Statistics), vol. 2. Hayward, CA, USA: Institute of Mathematical Statistics, 1990.
- [18] A. W. van der Vaart and J. A. Wellner, *Stochastic Convergence and Empirical Processes*, 2nd ed. New York, NY, USA: Springer-Verlag, 2011.
- [19] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*. New York, NY, USA: Springer-Verlag, 1996.
- [20] W. E. Walsh, D. C. Parkes, T. Sandholm, and C. Boutilier, "Computing reserve prices and identifying the value distribution in real-world auctions with market disruptions," in *Proc. 23rd AAAI Conf. Artif. Intell.*, 2008, pp. 1499–1502.
- [21] J. Y. Yu and S. Mannor, "Unimodal bandits," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 41–48.

Nicolò Cesa-Bianchi is professor of Computer Science at the University of Milano, Italy, where he is currently director of the Computer Science programs. He was President of the Association for Computational Learning and member of the steering committee of the EC-funded Network of Excellence PASCAL2. He served as action editor for the Machine Learning Journal, for IEEE TRANSACTIONS ON INFORMATION THEORY, and for the *Journal of Machine Learning Research*. He is currently associate editor for the *Journal of Information and Inference* and member of the Board of Directors of the Association for Computational Learning. He was program chair of the 13th Annual Conference on Computational Learning Theory and of the 13th International Conference on Algorithmic Learning Theory. He held visiting positions with UC Santa Cruz, Graz Technical University, Ecole Normale Supérieure in Paris, Google, and Microsoft Research. His main research interests include statistical learning theory, online learning, sequential prediction. He is coauthor of the monographs *Prediction, Learning, and Games* and *Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems*. He is recipient of a Google Research Award and of a Xerox Foundation UAC Award.

Yishay Mansour got his PhD from MIT in 1990, following it he was a postdoctoral fellow in Harvard and a Research Staff Member in IBM T.J. Watson Research Center. Since 1992 he is at Tel-Aviv University, where he is currently a Professor of Computer Science and has served as the first head of the Blavatnik School of Computer Science during 2000-2002. He was the director of the Israeli Center of Research Excellence in Algorithms. Prof. Mansour joined Microsoft Research in Israel in 2014. Before that he held visiting positions with Microsoft, Bell Labs, AT&T research Labs, IBM Research, and Google Research. He has mentored start-ups as Riverhead, which was acquired by Cisco, Ghooonet and Verix. Prof. Mansour has published numerous journal and proceeding papers in various areas of computer science with special emphasis on communication networks, machine learning, and algorithmic game theory, and has supervised over a dozen graduate students in those areas. Prof. Mansour is currently an associate editor in a multiple distinguished journals and has been on numerous conference program committees. He was the program chair of COLT (1998) and serves on the COLT steering committee.

Claudio Gentile is currently associate professor in Computer Science at the University of Insubria, Italy, where he is also coordinating the PhD program in Computer Science and Computational Mathematics. His current research interests are in Online Learning, Big Data analysis, Machine Learning on networked data, and applications thereof. He is member of the editorial board of the *Machine Learning Journal* from 2006, member of the editorial board of the *Journal of Machine Learning Research* from 2009, steering committee member of EU-funded Networks of Excellence projects PASCAL and PASCAL2, PC chair, Area Chair or Senior PC member of several international conferences in the field of Machine Learning/Artificial Intelligence/Data Mining (e.g., COLT, NIPS, IJCAI, ECML/PKDD). He was visiting researcher at UC Santa Cruz, Microsoft Research, Technion, University College London, Tel-Aviv University, INRIA Lille, Telefonica, Amazon.