Regret Minimization for Reserve Prices in Second-Price Auctions*

Nicolò Cesa-Bianchi† Claudio Gentile‡ Yishay Mansour§

Abstract

We show a regret minimization algorithm for setting the reserve price in second-price auctions. We make the assumption that all bidders draw their bids from the same unknown and arbitrary distribution. Our algorithm is computationally efficient, and achieves a regret of $\widetilde{\mathcal{O}}(\sqrt{T})$, even when the number of bidders is stochastic with a known distribution.

1 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 AdX or 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. Notice that we need not assume that the bidders indeed bid their value. Our assumption on the bidders symmetry, 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 mechanisms 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. Yet, 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 strong impact on Yahoo!'s revenue.

We stress that unlike much of the mechanism design literature, 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.

1.1 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

*Supported in part by the PASCAL2 Network of Excellence under EC grant no. 216886. This publication only reflects the authors' views.

†Dept. of Computer Science, Università degli Studi di Milano, Italy.

‡Dipartimento di Scienze Teoriche ed Applicate, Università dell'Insubria, Italy.

§School of Computer Science, Tel Aviv University, Israel.

[1] For 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.

revenue at each step. We show that after T steps (T repetitions of the auction) our algorithm has a regret of only $\widetilde{\mathcal{O}}(\sqrt{T})$. 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 $\widetilde{\mathcal{O}}(1/\sqrt{T})$, assuming the bids are in the range $[0, 1]$.

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 algorithm, similar to [12]. The algorithm discretizes the range of reserve prices to $\Theta(T^{1/3})$ price bins, and uses some efficient bandit algorithm over the bins. It is easy to see that lowering the optimal reserve price by ϵ will result in an average loss[2] of at most ϵ. 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 $\widetilde{\mathcal{O}}(\sqrt{T})$.

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. 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 to the revenue of 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 $\widetilde{\mathcal{O}}(\sqrt{T})$ 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 2). We later extend the algorithm and analysis to the case where there is stochasticity in the number of bidders through a known distribution (Section 3). In both cases we prove a regret bound of $\widetilde{\mathcal{O}}(\sqrt{T})$. This bound is optimal up to logarithmic factors. In fact, simple choices of the bid distribution exist that force any algorithm to have order of $\sqrt{T}$ regret, even when there are only two bidders whose bids are revealed to the algorithm at the end of each auction.

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

1.2 Related work. There is a vast literature in Algorithmic Game Theory on the topic of second price auction with sponsored search as a motivation. 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 w.r.t. *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 [9, 1]. 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. 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 3), we prove our results using the regret to the best reserve price, w.r.t. 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, 20, 11] on optimizing the reserve price, concentrating on more involved issues that arise in practice, such as discrete bids, non-stationary behavior, hidden bids, and more. While we are definitely not the first 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). For example, [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 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 $\widetilde{\mathcal{O}}(\sqrt{T})$ 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.

2 Known number of bidders.

We first show our results for the case where the number of bidders m is known and fixed. In Section 3 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.

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

In this simplified setting, we consider a protocol in which a learning algorithm (or a "mechanism") is setting a *reserve* 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 $B^{(2)}$).

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

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

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 2.1. *With the notation introduced so far, we have*

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

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 $t = 1, 2, \ldots$ the algorithm sets a price p_t and receives revenue $R_t(p_t) = R(p_t\,;\, B_{t,1}, \ldots, B_{t,m})$ which is a function of the random bids $B_{t,1}, \ldots, B_{t,m}$ 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 $p_1, \ldots, p_T$, we define the (cumulative) regret as

$$(2.1)\quad \sum_{t=1}^{T}\big(\mathbb{E}\left[R_t(p^*)\right]-\mathbb{E}_t[R_t(p_t)]\big) = \sum_{t=1}^{T}\big(\mu(p^*)-\mu(p_t)\big)$$

where the expectation $\mathbb{E}_t = \mathbb{E}_t[\,\cdot \mid p_1, \ldots, p_{t-1}]$ is over the random bids at time t, conditioned on all past prices $p_1, \ldots, p_{t-1}$ (i.e., conditioned on the past history of the bidding process). This implies that the regret $\sum_{t=1}^{T}\big(\mu(p^*) - \mu(p_t)\big)$ 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 $\widetilde{\mathcal{O}}(\sqrt{T})$ 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 Appendix B we use a uniform convergence argument to show that the same regret bound $\widetilde{\mathcal{O}}(\sqrt{T})$ holds with high probability for the actual regret

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

Note that here the actual revenue of the seller is compared against the best reserve price *on each sequence of bid realizations*. Therefore, this is a much harder notion of regret than (2.1).

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 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. Note that $\beta(\cdot)$ is defined there in such a way that $F_2(p) = \beta((F(p))^m)$. 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 2.1. *Fix an integer $m \ge 2$ and consider the function*

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

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

$$(2.2)\quad \beta^{-1}(a) - \frac{2\,\epsilon}{\sqrt{1-a}} \le x \le \beta^{-1}(a) + \frac{2\,\epsilon}{\sqrt{1-a}}\ .$$

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 will serve this purpose.

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

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

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 $\widehat{p}_1, \widehat{p}_2, \ldots$ are set such that $0 = \widehat{p}_1 \le \widehat{p}_2 \le \cdots \le 1$. At the end of each stage i, the algorithm computes a new estimate $\widehat{\mu}_i$ of the expected revenue function μ in the interval $[\widehat{p}_i, 1]$, where p^* is likely to lie. This estimate depends on the empirical cumulative distribution function $\widehat{F}_{2,i}$ of F_2 computed during stage i in the interval $[\widehat{p}_i, 1]$. The algorithm pseudocode is given in Algorithm 1. The quantity $C_{\delta,i}(p)$ therein is defined as

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

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

Stage 1 is a seed stage where the algorithm computes a first approximation $\widehat{\mu}_1$ of μ. Since the algorithm plays $\widehat{p}_1 = 0$, 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 $\widehat{F}_{2,1}$ in Algorithm 1 is a standard cumulative empirical distribution function based on i.i.d. realizations of $B^{(2)}$.

Algorithm 1: Regret Minimizer

Input: confidence level $\delta \in (0,1]$, approximation parameter $\alpha \in (0,1]$;

Stage 1:

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

$$\widehat{F}_{2,1}(x) = \frac{1}{T_1}\left|\left\{t = 1, \ldots, T_1 : R_t(0) \le x\right\}\right|;$$

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

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

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

- For all $t = 1 + \sum_{j=1}^{i-1} T_j, \ldots, \sum_{j=1}^{i} T_j$, play $p_t = \widehat{p}_i$, and observe revenues $R_1(\widehat{p}_i), \ldots, R_{T_i}(\widehat{p}_i)$, where $\widehat{p}_i$ is computed as follows:
  - Compute maximizer
  $$\widehat{p}^*_{i-1} = \operatorname*{argmax}_{p \in [\widehat{p}_{i-1},1]\,:\,\widehat{F}_{2,i-1}(p) \le 1-\alpha} \widehat{\mu}_{i-1}(p)\ ,$$
  - Let $P_i = \Big\{p \in [\widehat{p}_{i-1}, 1]\ :\ \widehat{\mu}_{i-1}(p) \ge \widehat{\mu}_{i-1}(\widehat{p}^*_{i-1}) - 2C_{\delta,i-1}(\widehat{p}^*_{i-1}) - 2C_{\delta,i-1}(p)\Big\}$,
  - Set $\widehat{p}_i = \min P_i \bigcap \Big\{p\ :\ \widehat{F}_{2,i-1}(p) \le 1 - \alpha\Big\}$;
- Compute, for $x \in [\widehat{p}_i, 1]$, empirical distribution

$$\widehat{F}_{2,i}(x) = \frac{1}{T_i}\left|\Big\{t = 1, \ldots, T_i : R_t(\widehat{p}_i) \le x\Big\}\right|;$$

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

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

The approximation $\widehat{\mu}_1$ is based on the corresponding expected revenue μ contained in Fact 2.1, where $\beta(\cdot)$ is the function defined in Lemma 2.1. 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 $\mathbb{E}\big[B^{(2)}\big]$ is not a problem for this computation.[3] In Stage 2 (encompassing trials $t = T_1 + 1, \ldots, T_1 + T_2$) the algorithm calculates the empirical maximizer

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

then computes the set of candidate optimal reserve prices

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

and sets the reserve price $\widehat{p}_2$ to be the lowest one in P_2, subject to the additional constraint that[4] $\widehat{F}_{2,1}(p) \le$

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

[4] Note that the intersection is not empty, since $\widehat{p}^*_1$ is in the intersection.

$1-\alpha$. Price $\widehat{p}_2$ is played during all trials within Stage 2. The corresponding revenues $R_t(\widehat{p}_2)$, for $t = 1, \dots, T_2$, are gathered and used to construct an empirical cumulative distribution $\widehat{F}_{2,2}$ and an approximate expected revenue function $\widehat{\mu}_2$ to be used only in the subinterval[5] $[\widehat{p}_2, 1]$.

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

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

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

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

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

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

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

All other stages $i > 2$ proceed similarly, each stage i relying on the existence of empirical estimates $\widehat{F}_{2,i-1}$, $\widehat{\mu}_{i-1}$, and $\widehat{p}_{i-1}$ delivered by the previous stage $i-1$.

2.3 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 the total number of stages.

LEMMA 2.2. *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 $1 - \delta/(3S)$, conditioned on all past stages and uniformly over $p \in [\widehat{p}_i, 1]$.

In the sequel, we use Lemma 2.2 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 such that

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

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 2.1 below. Recall that it is important to guarantee that $\widehat{F}_{2,i}(p)$ is 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 in 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 i.

LEMMA 2.3. *The event*

$$\left| \mu(p) - \widehat{\mu}_i(p) \right| \leq 2C_{\delta,i}(p) \qquad \text{for all } p \in [\widehat{p}_i, 1]. \tag{2.3}$$

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

$$\begin{aligned} & p^* \geq \widehat{p}_i \\ & 0 \leq \widehat{\mu}_i(\widehat{p}_i^*) - \widehat{\mu}_i(p^*) \leq 2C_{\delta,i}(\widehat{p}_i^*) + 2C_{\delta,i}(p^*) \end{aligned} \tag{2.4}$$

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 $\mathcal{O}\big(\sqrt{T \log\log\log T}(\log\log T)\big) = \widetilde{\mathcal{O}}\big(\sqrt{T}\big)$. It is important to remark that in this bound there is no explicit dependence on the number m of bidders.

THEOREM 2.1. *With the notation introduced so far, 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 stage lengths $T_i = T^{1-2^{-i}}$ for $i = 1, 2, \dots$ (for simplicity, we have disregarded rounding effects in the computation of the integer stage lengths T_i), and approximation parameter $\alpha \geq \mu(p^*)^2/12$ has regret

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

with probability at least $1 - \delta$.

[5] Once again, computing the argmax of $\widehat{\mu}_2$ over $[\widehat{p}_2, 1]$ 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 $\mathbb{E}\left[B^{(2)}\right]$ and $\int_0^{\widehat{p}_2} F_2(t)\,dt$ therein.

The proof of this theorem follows by applying at each stage i the uniform approximation delivered by Lemma 2.3 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 A. We then set the length T_i of stage i as $T_i = T^{1-2^{-i}}$, 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.

2.4 Lower bounds. The next theorem shows that the $\sqrt{T}$ dependence on the time horizon T cannot be removed even in very special cases.

THEOREM 2.2. *With the notation introduced so far, for any T large enough, there exists a distribution of the bids such that any algorithm operating with $m = 2$ bidders is forced to have regret*

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

with constant probability.

The proof, which is omitted from this version of the paper, constructs a simple family of bid distributions that the are statistically indistinguishable unless $T = \Omega\big(1/\epsilon^2\big)$ realizations of the bid distribution are observed. Since by the maxima of the revenue functions for the two distributions are designed to be ϵ-apart, the regret accumulated during these T observations is of the order of $\epsilon\, T = \Omega\big(\sqrt{T}\big)$, and the result follows.

Notice that this result is not a consequence of our partial information setting. The above lower bound holds even when the algorithm is afforded to observe the actual bids at each round, rather than only the revenue.

3 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 $\{2, 3, 4, \ldots\}$. 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 $B_1, B_2, \ldots, B_m$. 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 2. For simplicity, we assume that M is independent of the random variables B_i. For fixed $M = m$, we denote by $B_m^{(1)} \geq B_m^{(2)} \geq \ldots \geq B_m^{(m)}$ 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 $R^m(p) = R(p; B_1, B_2, \ldots, B_m)$ 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

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

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

Again, we want to minimize the 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 regret over T time steps is

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

and p_t is the price set by the algorithm at time t. Again, in Appendix B we show that the same regret bound holds for the actual regret

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

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 $\mathbb{E}_M\big[F_{2,M}\big](x)$ to denote the mixture distribution $\sum_{m=2}^{\infty} Q(m)F_{2,m}(x)$. Likewise,

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

Relying on Fact 2.1, one can easily see that

$$\mu(p) = \mathbb{E}_M\mathbb{E}\big[B_M^{(2)}\big] + \int_0^p \mathbb{E}_M\big[F_{2,M}\big](t)\,dt - p\,\mathbb{E}_M\big[F^M\big](p) \ . \tag{3.5}$$

As in Section 2, our goal is to devise an online algorithm whose regret is of the form $\sqrt{T}$ with as few assumptions as possible on F and Q.

We first extend[6] Lemma 2.1 to handle this more general setting.

LEMMA 3.1. *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 $0 \le \epsilon < 1-a$, if x is such that

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

then

$$T(A^{-1}(a)) - \frac{\epsilon\, \mathbb{E}[M]}{1-(a+\epsilon)} \le x \le T(A^{-1}(a)) + \frac{\epsilon\, \mathbb{E}[M]}{1-(a+\epsilon)}. \tag{3.6}$$

In addition, if[7]

$$(T''(x))^2\,(1-x) + T'(x)T''(x) \ge T'(x)T'''(x)(1-x) \tag{3.7}$$

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

$$T(A^{-1}(a)) - \frac{\epsilon}{1-a} \le x \le T(A^{-1}(a)) + \frac{\epsilon}{1-a}\ . \tag{3.8}$$

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

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

and

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

Hence, $\mathbb{E}_M[F^M](p)$ in (3.5) satisfies $\mathbb{E}_M[F^M](p) = T\Big(A^{-1}\big(\mathbb{E}_M[F_{2,M}](p)\big)\Big)$. In particular, when $\mathbb{P}(M = m) = 1$ as in Section 2, 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 2.1, and the reconstruction function $\beta^{-1}(\cdot)$ we used throughout Section 2 is $T(A^{-1}(\cdot))$. Because this is a more general setting then the one in Section 2, we do still have the technical issue of insuring that the argument of this recostruction 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 $\mathbb{E}_M[F_{2,M}]$ 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 (3.5) 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:

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

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 (3.5) 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 2, we overload the notation and define in this section $\beta(x) = A(T^{-1}(x))$, where T and A are given in Lemma 3.1. 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 2.2. Hence, in what follows we only emphasize the differences, which are essentially due to the modified confidence interval delivered by Lemma 3.1, as compared to Lemma 2.1.

In particular, if we rely on (3.6), 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)

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

with

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

Similarly, if we rely on (3.8), we have instead

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

[6] In fact, dealing with a more general setting only allowed us to obtained a looser result than Lemma 2.1.

[7] Condition (3.7) 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 3.1 in Appendix A), 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 (3.7) for $x = 0.6$, i.e., it yields a function $T(A^{-1}(x))$ which is not convex on $x = 0.6$.

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 $\mathbb{E}_M\mathbb{E}\big[B_M^{(2)}\big]$ and $\bar{F}_2$. We call the resulting algorithm the *Generalized Algorithm* 1.

As for the analysis, Lemma 2.2 is replaced by the following (because of notation overloading, the statement is the same as that of Lemma 2.2, but the involved quantities are different, and so is the proof in the appendix).

LEMMA 3.2. *With the notation introduced at the beginning of this section, if $S = S(T)$ is 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 $1-\delta/(3S)$, conditioned on all past stages, uniformly over $p \in [\widehat{p}_i, 1]$.

Then an easy adaptation of Lemma 2.3 leads to the following regret bound. The proof is very similar to the proof of Theorem 2.1, and is therefore omitted.

THEOREM 3.1. *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*

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

with stage lengths $T_i = T^{1-2^{-i}}$ for $i = 1, 2, \ldots,$ and approximation parameter $\alpha \geq \mu(p^*)^2/12$ has regret

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

with probability at least $1-\delta$, where $A = \mathbb{E}[M]$ if (3.6) holds and $A = 1$ if (3.8) holds.

4 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 $\widetilde{\mathcal{O}}(\sqrt{T})$. 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 to the seller only 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 of those 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.

We are currently working on extending our results to 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. We are also trying to see whether the inverse dependence on $\mu(p^*)$ in Theorem 2.1 (and on $\mu(p^*)^2$ in Theorem 3.1) 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.

References

[1] Susan Athey and Philip A. Haile. Identification of standard auction models. *Econometrica*, 70(6):2107–2140, 2002.

[2] P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. *Proceedings of the 20th Annual Conference on Learning Theory*, pages 454–468, 2007.
[3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.
[4] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12:1587–1627, 2011.
[5] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, 2006.
[6] E.W. Cope. Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Transactions on Automatic Control*, 54(6):1243–1253, 2009.
[7] Esther David, Alex Rogers, Nicholas R. Jennings, Jeremy Schiff, Sarit Kraus, and Michael H. Rothkopf. Optimal design of English auctions with discrete bid levels. *ACM Trans. Internet Techn.*, 7(2), 2007.
[8] P. Dhangwotnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. In *ACM Conference on Electronic Commerce*, 2010.
[9] Philip A. Haile and Elie Tamer. Inference with an incomplete model of English auctions. *J. of Political Economy*, 111(1):1–51, 2003.
[10] J. Hartline and A. Karlin. Profit maximization in machanism design. In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, editors, *Algorithmic Game Theory.* Cambridge University Press, 2007.
[11] Albert Xin Jiang and Kevin Leyton-Brown. Bidding agents for online auctions with hidden bids. *Machine Learning*, 67:117–143, 2007.
[12] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: bounds on regret for online posted-price auctions. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 594–605, 2003.
[13] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990.
[14] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.
[15] N. Nisan. Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, editors, *Algorithmic Game Theory.* Cambridge University Press, 2007.
[16] Michael Ostrovsky and Michael Schwarz. Reserve prices in Internet advertising auctions: a field experiment. In *ACM Conference on Electronic Commerce*, pages 59–60, 2011.
[17] D. Pollard. Empirical processes: theory and applications. In *NSF-CBMS regional conference series in probability and statistics*, 1990.
[18] A.W. van der Vaart. Empirical processes and statistical learning. Technical report, Vrije Universiteit Amsterdam, 2011.
[19] A.W. Van Der Vaart and J.A. Wellner. *Weak convergence and empirical processes.* 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 *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1499–1502, 2008.
[21] J.Y. Yu and S. Mannor. Unimodal bandits. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

A Proofs.

Proof of Fact 2.1: By definition of $R(p)$ we can write

$$\mu(p) = \int_p^1 x\, dF_2(x) + p\,\mathbb{P}\big(B^{(2)} < p \le B^{(1)}\big) \ . \tag{A.1}$$

By applying the identity $\mathbb{E}[X] = \int \mathbb{P}(X > x)\,dx$ to the nonnegative r.v. $B^{(2)}\,\mathbb{I}_{\{B^{(2)}>p\}}$ we obtain

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

Moreover,

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

and

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

Substituting the above into (A.1) and simplifying concludes the proof. □

Proof of Lemma 2.1: 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) $x \le \beta^{-1}(a+\epsilon)$ for any $\epsilon \in [0, 1-a]$, and (ii) $\beta^{-1}(a-\epsilon) \le x$ for any $\epsilon \in [0,a]$. In turn, because of the convexity of $\beta^{-1}(\cdot)$, we have

$$\beta^{-1}(a+\epsilon) \le \beta^{-1}(a) + \frac{1-\beta^{-1}(a)}{1-a}\,\epsilon,\ \forall \epsilon \in [0, 1-a] \ . \tag{A.2}$$

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

$$\begin{aligned}\beta^{-1}(a-\epsilon) &\ge \beta^{-1}(a) - \frac{d\beta^{-1}(x)}{dx}\Big|_{x=a}\,\epsilon \\ &\ge \beta^{-1}(a) - \frac{1-\beta^{-1}(a)}{1-a}\,\epsilon,\ \forall \epsilon \in [0,a] \ .\end{aligned} \tag{A.3}$$

We need the following technical claim.

Claim 1.

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

Proof of Claim: Introduce the auxiliary function $f(a) = 1-2\sqrt{1-a}$. The claim is proven by showing that $\beta(f(a)) \le a$ for all $a \in [3/4,1]$. Note that the case $a \in [0,3/4)$ is trivially verified. We prove the claim by showing that $\beta(f(a))$ is a concave function of $a \in [3/4,1]$, and that $\frac{d\,\beta(f(a))}{d\,a}\big|_{a=1} \ge 1$, while $\beta(f(1)) = 1$. We have

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

Hence, using L'Hopital's rule,

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

since $m \ge 2$. Moreover,

$$\frac{d^2\beta(f(a))}{d\,a^2} = -\left(\frac{m-1}{m}\right)(f(a))^{-\frac{m+1}{m}} \times \frac{1}{1-a} + \frac{m-1}{2} \times \frac{(f(a))^{-1/m}-1}{(1-a)^{3/2}}$$

which is nonpositive if and only if $m((f(a))^{-1/m}-1) \le (1-f(a))\,(f(a))^{-\frac{m+1}{m}}$ 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

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

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 $g'(1) = m+1$. □

The claim together with (A.2) and (A.3) allows us to conclude the proof of Lemma 2.1. Specifically, the second inequality in (2.2) is obtained by (A.2) and extended to any $\epsilon \ge 0$ just by observing that, by the claim, for $\epsilon > 1-a$ the right-most side of (2.2) is larger than 1. Moreover, the first inequality in (2.2) is obtained by (A.3) and extended to any $\epsilon \ge 0$ by observing that for $\epsilon > a$ the left-most side of (2.2) is smaller than $\beta^{-1}(a) - \frac{2a}{\sqrt{1-a}} \le a - \frac{2a}{\sqrt{1-a}} \le 0$ for any $a \in [0,1]$, where we have used the fact that $\beta^{-1}(a) \le a$. □

Proof of Lemma 2.2: Let $B_k^{(1)}$ and $B_k^{(2)}$ denote the maximum and the second-maximum of k i.i.d. bids $B_1,\ldots,B_k$. Set for brevity $A = \mathbb{P}\big(B_m^{(1)} > p\big)$. Then we have

$$A \le 2\,\mathbb{P}\big(B_{\lceil m/2\rceil}^{(1)} > p\big)$$

and

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

Hence

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

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

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

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 2.3), the classical Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [13] implies that with probability at least $1-\delta/3S$, conditioned on all past stages,

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

□

Proof of Lemma 2.3: We start by proving (2.3). Fix any stage i and write

$$\begin{aligned} \big|\mu(p) - \widehat{\mu}_i(p)\big| &\le \left|\int_0^p F_2(t)\,dt - \int_0^p \widehat{F}_{2,i}(t)\,dt\right| \\ &\quad + p\left|(F(p))^m - \beta^{-1}\big(\widehat{F}_{2,i}(p)\big)\right| \\ &\le \int_0^p \big|F_2(t) - \widehat{F}_{2,i}(t)\big|\,dt \\ &\quad + p\left|(F(p))^m - \beta^{-1}\big(\widehat{F}_{2,i}(p)\big)\right| \\ &\le p\max_{t\in[0,p]} \big|F_2(t) - \widehat{F}_{2,i}(t)\big| \\ &\quad + p\left|(F(p))^m - \beta^{-1}\big(\widehat{F}_{2,i}(p)\big)\right|\ . \end{aligned} \tag{A.4}$$

The DKW inequality implies that

$$p\max_{t\in[0,p]} \big|F_2(t) - \widehat{F}_{2,i}(t)\big| \le p\sqrt{\frac{1}{2T_i}\,\ln\frac{6S}{\delta}} \le C_{\delta,i}(p) \tag{A.5}$$

holds with probability at least $1-\delta/(3S)$. As for the second term in (A.4) we apply again the DKW inequality in combination with Lemma 2.1 with $x = (F(p))^m = \beta^{-1}\big(F_2(p)\big)$, $a = \widehat{F}_{2,i}(p)$, and $\epsilon = \sqrt{\frac{1}{2T_i} \ln \frac{6S}{\delta}}$. This yields

$$p\Big|\beta^{-1}\big(F_2(p)\big) - \beta^{-1}\big(\widehat{F}_{2,i}(p)\big)\Big| \le C_{\delta,i}(p)$$

with the same probability of at least $1-\delta/(3S)$. Putting together and using the union bound over the S stages gives (2.3).

We prove (2.4) by induction on $i = 1, \ldots, S$. We first show that the base case $i = 1$ holds with probability at least $1 - \delta/S$. Then we show that if (2.4) 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 $i = 1, \ldots, S$ we get that (2.4) holds simultaneously for all i with probability at least $1 - \delta$.

For the base case $i = 1$ note that $\widehat{\mu}_1(p^*) \le \widehat{\mu}_1(\widehat{p}^*_1)$ holds with probability at least $1-\delta/(3S)$ because we are assuming (Lemma 2.2) that $\widehat{F}_2(p^*) \le 1 - \alpha$ 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 (2.3) we obtain

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

and

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

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

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

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

We now prove (2.4) for $i > 1$ using the inductive assumption $p^* \ge \widehat{p}_{i-1}$ and

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

The inductive assumption and $\widehat{F}_{2,i}(p^*) \le 1 - \alpha$ directly imply $p^* \in P_i \bigcap \Big\{p : \widehat{F}_{2,i-1}(p) \le 1 - \alpha\Big\}$ (recall the definition of the set of candidate prices P_i given in Algorithm 1). Thus we have $p^* \ge \widehat{p}_i$ and $\widehat{\mu}_i(\widehat{p}^*_i) \ge \widehat{\mu}_i(p^*)$, because $\widehat{p}^*_i$ maximizes $\widehat{\mu}_i$ over a range that contains p^*. The rest of the proof closely follows that of (2.4) for the base case $i = 1$. □

Proof of Theorem 2.1: If $S = S(T)$ is the total number of stages, then the regret of our algorithm is

$$\big(\mu(p^*) - \mu(\widehat{p}_0)\big)\, T_1 + \sum_{i=2}^{S} \big(\mu(p^*) - \mu(\widehat{p}_i)\big)\, T_i$$

$$\le T_1 + \sum_{i=2}^{S} \big(\mu(p^*) - \mu(\widehat{p}_i)\big)\, T_i\ . \tag{A.6}$$

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

$$\begin{aligned}
&\mu(p^*) - \mu(\widehat{p}_i)\\
&\le \mu(p^*) - \widehat{\mu}_{i-1}(\widehat{p}_i) + 2C_{\delta,i-1}(\widehat{p}_i)\\
&\qquad \text{(by (2.3) —note that } \widehat{p}_i \ge \widehat{p}_{i-1})\\
&\le \mu(p^*) - \widehat{\mu}_{i-1}\big(\widehat{p}^*_{i-1}\big) + 2C_{\delta,i-1}(\widehat{p}^*_{i-1}) + 4C_{\delta,i-1}(\widehat{p}_i)\\
&\qquad \text{(since } \widehat{p}_i \in P_i)\\
&\le \mu(p^*) - \widehat{\mu}_{i-1}(p^*) + 2C_{\delta,i-1}(\widehat{p}^*_{i-1}) + 4C_{\delta,i-1}(\widehat{p}_i)\\
&\qquad \text{(since } \widehat{\mu}_{i-1}(\widehat{p}^*_{i-1}) \ge \widehat{\mu}_{i-1}(p^*) \text{ —see (2.4))}\\
&\le 2C_{\delta,i-1}(p^*) + 2C_{\delta,i-1}(\widehat{p}^*_{i-1}) + 4C_{\delta,i-1}(\widehat{p}_i)\\
&\qquad \text{(by } p^* \ge \widehat{p}_i \text{ combined with (2.3))}\\
&\le 8\sqrt{\frac{1}{\alpha T_{i-1}} \ln \frac{6S}{\delta}}
\end{aligned}$$

where in the last step we used the fact that $\widehat{F}_{2,i-1}(p^*) \le 1 - \alpha$ holds by Lemma 2.2, and that $\widehat{F}_{2,i-1}(p) \le 1 - \alpha$ 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 (A.6) 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 $i = 1, 2, \ldots$ implies that S is upper bounded by the minimum integer n such that

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

Since $i \ge \log_2 \log_2 T$ makes $T_i \ge \frac{T}{2}$, then $S \le \lceil 2 \log_2 \log_2 T\rceil = \mathcal{O}(\log\log T)$. 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

$$(A.6) \le \sqrt{T} + 8\sqrt{T}\, S \sqrt{\frac{1}{\alpha} \ln \frac{6S}{\delta}}\ .$$

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

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

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

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

as claimed. □

Proof of Lemma 3.1: 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

$$T(A^{-1}(a+\epsilon)) = T(A^{-1}(a)) + \epsilon \; \frac{d\,T(A^{-1}(x))}{dx}\bigg|_{x=\xi}$$

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

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

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

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

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

(A.7)
$$\frac{d\,T(A^{-1}(x))}{dx}\bigg|_{x=\xi} \leq \frac{1}{1-A^{-1}(a+\epsilon)} \leq \frac{1}{1-T^{-1}(a+\epsilon)}$$

the last inequality deriving from[8] $A(x) \geq T(x)$ 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 (A.7) to see that

$$\frac{d\,T(A^{-1}(x))}{dx}\bigg|_{x=\xi} \leq \frac{\mathbb{E}[M]}{1-(a+\epsilon)} \; .$$

Replacing backwards, this yields the second inequality of (3.6).

To prove the first inequality of (3.6), 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 $\xi \in (a-\epsilon, a)$,

$$\begin{aligned} T(A^{-1}(a)) &= T(A^{-1}(a-\epsilon)) + \epsilon \; \frac{d\,T(A^{-1}(x))}{dx}\bigg|_{x=\xi} \\ &\leq T(A^{-1}(a-\epsilon)) + \frac{\epsilon}{1-T^{-1}(a)} \\ &\leq T(A^{-1}(a-\epsilon)) + \frac{\epsilon\,\mathbb{E}[M]}{1-a} \\ &\leq T(A^{-1}(a-\epsilon)) + \frac{\epsilon\,\mathbb{E}[M]}{1-(a+\epsilon)} \end{aligned}$$

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 (3.6) is trivially fulfilled. In fact,

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

since $\mathbb{E}[M] \geq 2$. This concludes the proof of (3.6).

In order to prove (3.8), 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

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

Since y ranges over [0,1] when x does, (3.7) 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 second inequality in (3.8). Similarly, by the convexity and the monotonicity of $T(A^{-1}(\cdot))$ we can

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

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 (3.8). 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 (3.8) trivially fulfilled. □

Proof of Lemma 3.2: Let $B_m^{(1)}$ and $B_m^{(2)}$ denote the highest and the second-highest of m i.i.d. bids $B_1, \ldots, B_m$. Recall from the proof of Lemma 2.2 that, for any $m \geq 2$

$$\mathbb{P}\big(B_{\lceil m/2\rceil}^{(1)} > p\big) \geq \frac{1}{2}\,\mathbb{P}\big(B_m^{(1)} > p\big)$$

and

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

Moreover,

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

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 2.2 by applying DKW on the uniform convergence of $\widehat{F}_{2,i}$ to $\bar{F}_2$. □

B Bounding the actual regret.

We show how to bound in probability the actual 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. We need the following definitions and results from empirical process theory —see, e.g., [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 $x_1, \ldots, x_n \in \mathbb{X}$ if there exists $r_1, \ldots, r_n \in \mathbb{R}$ such that for each $(a_1, \ldots, a_n) \in \{0, 1\}^n$ there exists $f \in \mathcal{F}$ for which $f(x_i) \geq r_i$ iff $a_i = 1$ for all $i = 1, \ldots, n$. 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, e.g., [19, 18].

Lemma B.1. *Let $X_1, X_2, \ldots$ 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,*

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

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

Recall that $\mathbb{E}_M\mathbb{E}\big[R^M(p)\big] = \mu(p)$ for all $p \in [0, 1]$. Let $\mathcal{R} = \big\{R^M(p) : p \in [0, 1]\big\}$ 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 $B_1, \ldots, B_M$.

Lemma B.2. *The pseudo-dimension of the class $\mathcal{R}$ is 2 .*

Proof. [Proof sketch] 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

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

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) above it satisfy $R_p(b_1, b_2) = b_2$. It is easy to verify that $\mathcal{F}$ shatters any two points (b_1, b_2) and $(b_1 - \epsilon, b_2 + \epsilon)$ in the region S: the pattern $(0, 0)$ is achieved for a p such that both points are inside the rectangle, the pattern $(1, 1)$ is achieved for a p such that both points are above the rectangle, and the patterns $(0,1)$ and $(1,0)$ are achieved for values of p that have either point inside the rectangle. It is now easy to realize that no three points can be shattered using rectangles of this form.

THEOREM B.1. *Under the assumptions of Theorem 3.1, the actual regret of Generalized Algorithm 1 satisfies*

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

with probability at least $1-\delta$*, where* $A = \mathbb{E}[M]$ *if (3.6) holds and* $A = 1$ *if (3.8) holds.*

Proof. For the sake of brevity, let $R_t(p)$ denote $R_t^{M_t}(p)$. Also, let the conditional expectation $\mathbb{E}_t[\cdot]$ denote $\mathbb{E}_t[\cdot \,|\, p_1, \ldots, p_{t-1}]$, i.e., the expectation of the random variable at argument w.r.t. M and the bids $B_1, \ldots, B_M$, conditioned on all past bids and 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

$$\begin{aligned}
&\sum_{t=1}^{T} R_t(p_T^*) - \sum_{t=1}^{T} R_t(p_t) \\
&\quad = \sum_{t=1}^{T} R_t(p_T^*) - T\mu(p_T^*) \\
&\quad\quad + T\mu(p_T^*) - \sum_{t=1}^{T} R_t(p_t) \\
&\quad\quad + \sum_{t=1}^{T} \mathbb{E}_t\big[R_t(p_t)\big] - \sum_{t=1}^{T} \mathbb{E}_t\big[R_t(p_t)\big] \\
&\quad \le \max_{p\in[0,1]} \left(\sum_{t=1}^{T} R_t(p) - T\mu(p)\right) && \text{(B.8)} \\
&\quad\quad + \sum_{t=1}^{T} \Big(\mathbb{E}_t\big[R_t(p_t)\big] - R_t(p_t)\Big) && \text{(B.9)} \\
&\quad\quad + T\mu(p_T^*) - \sum_{t=1}^{T} \mathbb{E}_t\big[R_t(p_t)\big] \ . && \text{(B.10)}
\end{aligned}$$

In order to bound (B.8) we combine Lemma B.1 with Lemma B.2. This gives

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

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

In order to bound (B.9), note that $Z_t = \mathbb{E}_t\big[R_t(p_t)\big] - R_t(p_t)$ for $t = 1, 2, \ldots$ 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

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

with probability at least $1-\delta$.

Finally, term (B.10) is bounded via Theorem 3.1 after observing that $\mu(p_T^*) \le \mu(p^*)$, where $p^* = \operatorname{argmax}_{p\in[0,1]} \mu(p)$ is the maximizer of the expected revenue. This concludes the proof.