Learning to Bid in Repeated Auctions with Hidden Reserve Prices

Shri Prathaa M - EE22B144

Contents of this Presentation

01

Introduction

Understanding the Setup of First and Second Price Auction with Single Bidder and Single Seller

First Price Auction

- Online Sellers often use **hidden reserve prices**
- Buyer cannot observe reserves → must learn to bid optimally
- Incorrect bidding
  - ❑ Overbidding → overspending
  - ❑ Underbidding → lost opportunities

Objective: Minimize regret over entire time horizon

Second Price Auction

- Seller has a Hidden Reserve Price
  - Fixed - bidding once above reserve price reveals it
  - Stochastic - each winning bid reveals the current reserve price drawn from unknown distribution, F
- Bid should be below Valuation (v) to get non-negative utility
- Bid shouldn’t go below Reserve price (hidden) to get positive utility
- **Assumption**: Buyer doesn’t want to reveal his valuation
- **Objective**: Learn Empirical Distribution Function and use it to bid optimally

Single Buyer Fixed Hidden Reserve Price FPA

Truthful Seller

- Fixes the true Reserve Price (p) before auction
- Buyer wins if he bids above p and pays his bid amount

Strategic Seller

- Seller strategically chooses to increase reserve price from original value
- Seller could be assumed adversarial (He knows the algorithm the buyer would use to bid and optimise his reserve price)

Stochastic Reserve Price Auction

SPA

Every winning round reveals the underlying reserve price that can be used to build the empirical distribution of reserve price

FPA

The winning bids can be used to build the empirical distribution of winning bids

02

Online Learning in Auctions

Multi-armed bandits

- **Stochastic bandits**

  Regret Upper bound $O(M \log T)$

  Lower bound $\Omega\left(\frac{\log T}{\mathrm{KL}(p_j \,\|\, p^*)}\right)$

- **Distribution-independent (Stochastic)**

  Upper bound $O(\sqrt{MT \log T})$

  Lower bound $\Omega(\sqrt{MT})$

- **Adversarial bandits**

  Upper bound $O(\sqrt{MT \log M})$

  Lower bound $\Omega(\sqrt{MT})$

Bandit learning involves selecting one of M actions and observing only the chosen reward.
Let p_j be the mean reward of arm j and $p^* = \max_j p_j$.

Online Repeated Auctions

- Online advertising relies on repeated first-price (FPA) and second-price (SPA) auctions.
- Buyers shade bids in FPA, while truthful bidding is optimal in SPA.
- Real ad markets show complex bidder behavior due to budgets and strategic (myopic or non-myopic) bidding.
- Myerson–Satterthwaite bilateral trade model: sellers post above their valuation; buyers bid below their valuation.

Online Learning in Auctions

- Prior work mostly targets seller-side learning:
  Posted-price auctions, dynamic pricing, and discounted valuation models.
- Known regret rates for sellers : $\widetilde{O}(\sqrt{T})$ for truthful buyers
  $\Theta(T^{2/3})$ for strategic buyers.
- Buyer-side learning is less studied, especially with hidden reserve prices.
- Feedback is heavily censored in FPA, resembling binary bandit feedback.

03

Fixed Hidden Reserve Price FPA

Monotone Bid Price algorithm and Fast Search algorithm

Truthful Seller

Monotone Bid Price Algorithm

- Begins at low bid and increases monotonically until it wins once at bid, b^*
- Bids b^* for the rest of the rounds
- Attains a regret bounds of $O(\sqrt{T})$ and $\Omega(\sqrt{T})$

Upper bound :

Assume that the bid prices are monotone and it follows

$$b_k - b_{k-1} \leq b_{k+1} - b_k.$$

Then

$$\mathrm{RegretB}(\mathcal{M}, v, p) \leq 2\sqrt{T\left(\tilde{p} - b_{k_{\min}}\right)(v-p) - \left(\tilde{p} - b\right)}.$$

Here, $k_{\min} = k^*(\epsilon_1)$.

Lower Bound : Let k^* be the first time bid price is accepted and accepted bid price be b^*

- $\mathrm{RegretB}(\mathcal{M}, v, p) \geq \frac{1}{2}\sqrt{T}$ *for* $k^* \geq \sqrt{T}$
- $\mathrm{RegretB}(\mathcal{M}, v, p) \geq c\sqrt{T - \sqrt{T}}$ *for* $k^* < \sqrt{T}$ *and* $c > 1$.

Strategic Seller

Monotone Bid Price Algorithm

- Begins at low bid and increases geometrically until it wins once at bid, b^*
- Bids b^* for the rest of the rounds
- Attains a regret bounds of $O(\sqrt{T})$ and $\Omega(\sqrt{T})$

Lower Bound:

Similar to proof of upper bound in truthful seller; bound b^*- p using $(b_{k_{\min}+1} - b_{k_{\min}})$

And optimise k^* using AM-GM inequalities

Upper Bound:

Let k^* be 1^{st} time the strategic seller accepts bid. Bidder bids $b_t = b_1 \cdot (1+\beta)^{t-1}$.
The seller maximises $\sum_{k=k^*}^{T} b_k = (T - k^* + 1) b_{k^*}$ and gets $k^* = T - \frac{1}{\ln(1+\beta)}$.
Using $b_{k^*} - p \le b_{k^*} - b_{k^*-1} = b_{k^*-1} \cdot \beta \le \beta \cdot v$,

$$\text{Regret} \le \left(T - \frac{1}{\beta}\right)(v - p) + \left(1 + \frac{1}{\beta}\right) \cdot \beta v$$

At $\beta = \frac{1}{T - \sqrt{T}}$, Regret $\le \sqrt{T} \cdot (v - p) + v\left(1 + \frac{1}{T - \sqrt{T}}\right)$

Truthful Seller

Fast Search Algorithm

- Work in phases, shrinking a feasible interval $[a, b]$ of reserve prices from $\frac{1}{2}$ to $\frac{1}{T}$.
- In each phase, the bids are a, $a + \epsilon$, $a + 2\epsilon$, $\ldots$, $a + k\epsilon$ until a bid is accepted.
- Set new lower bound as $a + (k-1)\epsilon$ and upper bound as $a + k\epsilon$ for next phase.
- Update the step size for the next phase by setting the new ϵ as ϵ^2.

Proof

- The number of phases is $\log\log T$. Each phase has at most one accepted bid

  So, acceptance regret is $O(\log\log T)$.
- In a phase, interval length is $b - a = \sqrt{\epsilon}$ and so each rejected bid has regret at most $\sqrt{\epsilon}$. Bids are spaced by ϵ, giving $\sqrt{\epsilon}/\epsilon = 1/\sqrt{\epsilon}$ rounds.

  So, total rejection regret per phase is 1.
- Total regret over $\log\log T$ phases is $O(\log\log T)$.

Randomized Fast Search Algorithm

First Phase:
Begin with initial interval $[a_1,b_1]=[0,v]$ and split it into n sub intervals. Randomly select one sub-interval and offer log logT bid prices within it. Select k such sub-intervals

Acceptance case in first phase:
Identify the smallest sub-interval $[a_i,b_i]$ containing an accepted price. This ensures true reserve price p lies in $[0,b_i]$

Rejection case in first phase:
If no bid is accepted in a chosen sub-interval, repeat the first-phase selection at most k_1 times

Refinement phase:
Use the accepted interval from the first phase, divide it again into n sub intervals, and randomly choose k sub intervals. Offer log logT bids and find the minimum interval from which a price is accepted. Repeat this bidding for k_2 phases.

Final FS stage:
Apply Fast Search normally for another log logT phases to produce a final estimate of the reserve price.

FS Algorithm produces O(log log T) regret. The Randomised FS includes $k.k_1.k_2$ $(\log \log T)^2$ rounds before FS, giving $O(\log^3 T (\log \log T)^2)$ regret

04

Reserve Prices drawn from Distribution

Explore-Then-Commit and Explore-Exploit Multi Stage algorithms for SPA

Explore then Commit Algorithm

Stochastic SPA

- **Exploration Phase:**
  Set the bid price b to 1 for auctions upto time (T_1), and record the accepted reserve prices $\{p_1, p_2, .., p_{T1}\}$.

- **Exploitation Phase:**
  Compute an estimate μ from the observed prices and use μ as the bid price for the remaining ($T - T_1$) auctions.

Stochastic FPA

- **Exploration Phase:**
  Set the bid price to $\{b_i\}_{i=1}^{M}$ each for T_i times and $\widehat{T} = \sum_{i=1}^{M} T_i$

- **Exploitation Phase:**
  Compute an estimate μ of the winning bids and use μ as the bid price for the remaining ($T - \widehat{T}$) auctions.

Using DKW Inequality, $\Pr\left(\sup_{x\in[0,1]} |\widetilde{F}_n(x) - F(x)| > \epsilon\right) \leq \delta$ for $\epsilon = \sqrt{\frac{\ln(2/\delta)}{2T_1}}$ and Regret $\leq^1 T_1 + 2(T - T_1)\sqrt{\frac{\ln(2/\delta)}{2T_1}}$

(If Exploration phase length is T_1)

For SPA, $T_1=T^{2/3}$ gives regret of $O(T^{2/3} \ln T)$ by taking $\delta=1/T^2$

For FPA, exploration phase is $MT^{2/3}$ and so regret is $O(M \ln T\, T^{2/3})=O((\ln T)^2.T^{2/3})$ for $M= \ln T$

Explore Exploit Multi Stage Algorithm

Stochastic SPA

- **Stage Initialization:** Set stage length $T_i = T^{1-2^{-i}}$ and begin Stage 1 with bid price $b_1 = 1$ for T_1 auction rounds.
- **Exploration:** Observe reserve prices $\{p_1, p_2, \ldots, p_{T_1}\}$ from accepted auctions. Construct empirical CDF $\widetilde{F}_1$ and empirical mean $\widehat{\mu}_1$.
- **Exploitation:** Compute

$$\widehat{b}_1^* = \max_{\zeta \in [\widehat{\mu}_1,\, \widehat{\mu}_1 + 2C_{\delta,i}(\widehat{\mu}_1)]} \zeta \quad \text{s.t. } \widetilde{F}_1(\zeta) \geq \alpha.$$

- **Refinement:** Define feasible set

$$S_2 = \{b \in [0, \widehat{b}_1^*] : \widehat{L}_1(b) \geq \widehat{L}_1(\widehat{b}_1^*) - 2C_{\delta,i}(\widehat{b}_1^*) - 2C_{\delta,i}(b)\},$$

then choose $\widehat{b}_2 = \max_{b \in S_2}\{\, b : \widetilde{F}_1(b) \geq \alpha \,\}$ and repeat for remaining stages.

Stochastic FPA

- **Initialisation:** Initialize stage lengths $T_i = T^{1-2^{-i}}$. Bids are drawn from $S_1 \in \{b_1, b_2, \ldots, b_M\}$.
- **Exploration:** Set the bid price $b_1 \in S_1$ for T_1 rounds. Observe and store pairs of bids and indicators of winning to construct empirical distribution of winning bids.
- **Exploitation:** Let the empirical distribution of winning bids after stage 1 be $\widetilde{F}_1$ and the empirical mean $\widehat{\mu}_1$. Compute

$$\widehat{b}_1^* = \max_{\zeta} \text{ s.t. } \zeta \in [\widehat{\mu}_1, \widehat{\mu}_1 + C_{\delta,i}(\widehat{\mu}_1)] \text{ and } \widetilde{F}_1(\zeta) \geq \alpha.$$

- **Refinement:** Define

$$S_2 = \{\, b \in [0, \widehat{b}_1^*] : \ \widehat{L}_1(b) \geq \widehat{L}_1(\widehat{b}_1^*) - C_{\delta,i}(\widehat{b}_1^*) - C_{\delta,i}(b) \text{ and } \widetilde{F}_1(b) \geq \alpha \,\},$$

and repeat for remaining stages (same as Stage 1).

Comparative Study

Model Type	Seller Type	Algorithm	Regret Upper Bound
Repeated FPA (Fixed RP, p)	Truthful	Monotone Bid Price	$O(\sqrt{T})$
Repeated FPA (Fixed RP, p)	Truthful	Fast Search	$O(\log\log T)$
Repeated FPA (RP distribution, F)	Truthful	Explore-Then-Commit	$\tilde{O}(M^{1/3}T^{2/3})$
Repeated FPA (RP distribution, F)	Truthful	Explore-Exploit Multi-Stage	$\tilde{O}(\sqrt{MT})$
Repeated FPA (Fixed RP, p)	Strategic	Monotone Bid Price	$O(\sqrt{T})$
Repeated FPA (Fixed RP, p)	Strategic	Randomized Fast Search	$\tilde{O}(\log^3 T)$
Repeated SPA (RP distribution, F)	Truthful	Explore-Then-Commit	$\tilde{O}(T^{2/3})$
Repeated SPA (RP distribution, F)	Truthful	Explore-Exploit Multi-Stage	$\tilde{O}(\sqrt{T})$

Here we assume that the bid, $b \in [0, 1]$ and M is the number of bid prices we explore in FPA.

Exploratory Questions

- **Strategic seller in Stochastic Reserve Price**

  In repeated auction, seller draws $p_t \sim F$ and inflates it to p_t' adversarially

- **Budget Constrained Buyers**

  Extend stochastic reserve-price models to budgeted buyers using

  - Bandits-with-Knapsacks ideas
  - Shadow-price methods: Index rules like UCB(reward) − λ□·LCB(cost)

Appendix

Monotone Bid Price Algorithm

Proof Sketch :

Upper bound :

We assume that $b_k - b_{k-1} \leq b_{k+1} - b_k$ and $b_{k^*-1} < p < b_{k^*}$ and so $(b_{k^*} - p) \leq b_{k_{\max}+1} - b_{k_{\max}}$

Regret $= T(v-p) - (T-k^*)(v - b_{k^*}) = k^*(v-p) + (T-k^*)(b_{k^*} - p)$

We bound $(b_{k_{\max}+1} - b_{k_{\max}}) \leq \frac{\Delta}{k^*}$ using $b_{k_{\max}+1} = \sum_{k=k_{\min}}^{k_{\max}+1} (b_k - b_{k-1}) + b_{k_{\min}} \leq k_{\max}(b_{k_{\max}+1} - b_{k_{\max}}) + b_{k_{\min}}$

Now we take $\infty > \widetilde{p} > 0$ such that $k_{\max}(b_{k_{\max}+1} - b_{k_{\max}}) + b_{k_{\min}} \leq \widetilde{p}$

$b_{k^*} - p \leq (b_{k_{\max}+1} - b_{k_{\max}}) \leq \frac{\widetilde{p} - b_{k_{\min}}}{k_{\max}} \leq \frac{\widetilde{p} - b_{k_{\min}}}{k^*}$ and so Regret $\leq k^*(v-p) + (T-k^*)\frac{(\widetilde{p} - b_{k_{\min}})}{k^*}$

Regret $\leq 2\sqrt{T(\widetilde{p} - b_{k_{\min}})(v-p)} - (\widetilde{p} - b_{k_{\min}})$.

Monotone Bid Price Algorithm

Lower bound :

For k* $\geq \sqrt{T}$, Regret $= T(v-p) - (T-k^*)(v-b^*)$ and for $b^* - p \approx \epsilon$, and $1 >> \epsilon > 0$

$$\geq \sqrt{T}\left[(v-p)\sqrt{T} - (v-b^*)\sqrt{T} + (v-b^*)\right] \geq \sqrt{T}\left[\epsilon\sqrt{T} + (v-b^*)\right] \geq \sqrt{T}(v-b^*)$$

For $k^* < \sqrt{T}.$, E[Regret] = $(v - \frac{\epsilon_1+\epsilon_2}{2})\mathbb{E}[k^*] + \mathbb{E}\left[(T-k^*)(b^*-p)\right] \geq (v - \frac{\epsilon_1+\epsilon_2}{2})\mathbb{E}[k^*] + (T-\sqrt{T})\frac{\delta}{\mathbb{E}[k^*]}$

Regret $\geq c\sqrt{T-\sqrt{T}}$ where $c = 2\sqrt{\delta\left(v - \frac{\epsilon_1+\epsilon_2}{2}\right)}$. and $\mathbb{E}\left[(b^*-p)\right] \geq \frac{\delta}{\mathbb{E}[k^*]}$ using Cauchy Schwarz

$$\sum_{i=2}^{k^*}(b_i - b_{i-1}) \leq \sqrt{k^*\left(\sum_{i=2}^{k^*}(b_i-b_{i-1})^2\right)}$$

By substituting $\mathbb{E}\left[(b^*-p)\right] = \sum_{k=2}^{k_{max}} \mathbb{E}\left[1_{\{p\in[b_k, b_{k-1}]\}}(b_k - p)\right] = \sum_{k=2}^{k_{\max}} \frac{(b_k-b_{k-1})^2}{2}.$

(We have assumed that p is uniformly distributed in $[\epsilon_1, \epsilon_2]$ where $\epsilon_1 < 1/2$ and $\epsilon_2 > 1/2$)

Monotone Bid Price Algorithm

Proof of Lower Bound:

Using $\sum_{t=k^*}^{T} b_{k^*} \geq \sum_{t=k^*+1}^{T} b_{k^*+1}$, to get $(T-k^*)b_{k^*} \geq (T-(k^*+1))b_{k^*+1}$

Regret $= k^*(v-p) + (T-k^*)(b_{k^*} - p) \geq k^*(v-p) + (T-k^*)(b_{k^*+1} - p) \; - b_{k^*+1}$

$\geq k^*(v-p) + (T-k^*)(b_{k_{\min+1}} - b_{k_{\min}}) \; - b_{k^*+1}$

By taking $\Delta \leq k_{\min}(b_{k_{\min}+1} - b_{k_{\min}}) + b_{k_{\min}}$ and thus $\frac{\Delta - b_{k_{\min}}}{k^*} \leq \frac{\Delta - b_{k_{\min}}}{k_{\min}} \leq (b_{k_{\min}+1} - b_{k_{\min}})$.

Regret $\geq k^*(v-p) + (T-k^*)\frac{(\Delta - b_{k_{\min}})}{k^*} \; - b_{k^*+1}$

By optimising k*, Regret $\geq \sqrt{2T(\Delta - b_{k_{\min}})} + b_{k_{\min}} - 2\Delta$.

Monotone Bid Price Algorithm

Upper bound

Let k^* be 1^{st} time the strategic seller accepts bid. Bidder bids $b_t = b_1 \cdot (1+\beta)^{t-1}$.
The seller maximises $\sum_{k=k^*}^{T} b_k = (T - k^* + 1) b_{k^*}$ and gets $k^* = T - \frac{1}{\ln(1+\beta)}$.
Using $b_{k^*} - p \leq b_{k^*} - b_{k^*-1} = b_{k^*-1} \cdot \beta \leq \beta \cdot v$,

$$\text{Regret} = k^*(v-p) + (T-k^*) \cdot (b_{k^*} - p) \leq \left(T - \frac{1}{\ln(1+\beta)}\right)(v-p) + \left(\frac{1}{\ln(1+\beta)}\right) \cdot (b_{k^*} - p)$$

$$\text{Regret} \leq \left(T - \frac{1}{\beta}\right)(v-p) + \left(1 + \frac{1}{\beta}\right) \cdot \beta v$$

At $\beta = \frac{1}{T - \sqrt{T}}$, Regret $\leq \sqrt{T} \cdot (v-p) + v\left(1 + \frac{1}{T-\sqrt{T}}\right)$

Explore Exploit Multi Stage Algorithm for Stochastic SPA

1. Lower Bound on the Estimated Win Probability

With probability $1 - \frac{\delta}{3S}$, uniformly for all $b \in [0, b_i]$,

$$\tilde{F}_i(b) \ \geq \ \alpha(b) - \sqrt{\frac{1}{2T_i} \ln\left(\frac{6S}{\delta}\right)}.$$

2. Concentration Event $\mathcal{E}_i$

Define

$$\mathcal{E}_i = \left\{ \max_{b \in [0, b_i]} \left| \tilde{F}_i(b) - F(b) \right| \leq \sqrt{\frac{1}{2T_i} \ln\left(\frac{6S}{\delta}\right)} \right\}.$$

Then

$$\Pr(\mathcal{E}_i) \geq 1 - \frac{\delta}{3S}.$$

3. Uniform Concentration Over All Stages

Let

$$\mathcal{E} := \bigcap_{i=1}^{S} \mathcal{E}_i.$$

By the union bound,

$$\Pr(\mathcal{E}) \geq 1 - \sum_{i=1}^{S} \Pr(\mathcal{E}_i^c) = 1 - \frac{\delta}{3}.$$

4. Revenue Functions and Confidence Radius

Define the true and empirical revenue curves:

$$L(b) = \int_0^b s \, dF(s), \qquad \tilde{L}_i(b) = \int_0^b s \, d\tilde{F}_i(s).$$

The confidence radius is

$$C_{\delta,i}(b) = b \sqrt{\frac{1}{2\tilde{F}_i(b) T_i} \ln\left(\frac{6S}{\delta}\right)}.$$

On event $\mathcal{E}_i$,

$$b \max_{c \in [0, b]} \left| \tilde{F}_i(c) - F(c) \right| \ \leq \ C_{\delta,i}(b).$$

5. Deviation of Revenue Integrals

For any $b \in [0, b_i]$,

$$|L(b) - \tilde{L}_i(b)| \leq 2b \max_{c \in [0, b]} \left| \tilde{F}_i(c) - F(c) \right|.$$

For FPA the Loss term is bounded by $C_{\delta,i}$ while its $2.C_{\delta,i}$ for SPA

Explore Exploit Multi Stage Algorithm for Stochastic SPA

6. Bounding the Loss

Using the confidence radius at stage $i-1$,

$$L(b^*) - L(\hat{b}_i) \le 8C\,\delta_{i-1}(b_{\max}),$$

and therefore

$$R_S \le T_1 + 8\sum_{i=2}^{S} T_i b_{\max} \sqrt{\frac{\ln(\frac{6S}{\delta})}{2\tilde{F}_{i-1}(b_{\max})T_{i-1}}}.$$

7. Choice of Epoch Lengths

Let

$$T_i = T^{1-2^{-i}}, \qquad i = 1, \ldots, n.$$

Then

$$T_1 = \sqrt{T}, \qquad T_2 = T^{3/4}, \qquad T_n = T^{1-2^{-n}}.$$

Choose

$$n = \lceil \log_2 \log T \rceil,$$

so that $2^n \approx \log T$.

8. Proof of Theorem: Regret Bound

We aim to show

$$R_S = \sum_{i=1}^{S} |L(b^*) - L(\hat{b}_i)|\, T_i = \tilde{O}(\sqrt{T}).$$

From the previous bounds,

$$R_S \le T_1 + 8\sum_{i=2}^{S} T_i b_{\max} \sqrt{\frac{\ln(\frac{6S}{\delta})}{2\tilde{F}_{i-1}(b_{\max})T_{i-1}}}.$$

Using $T_1 = \sqrt{T}$ and the growth of T_i,

$$R_S = O\left(\sqrt{T} + \sqrt{T}\log T\sqrt{\ln\log\log T}\right) = \tilde{O}(\sqrt{T})$$

where $\tilde{O}(\cdot)$ hides polylogarithmic factors.

In regret expression, $T_i = T^{1-(½)^i} / |S_i|$ for round, i for each element in S_i ,
8 becomes 4 and a factor of √M gets added in 2nd term of regret

Thank You