Learning to Bid in Repeated First-price Auctions



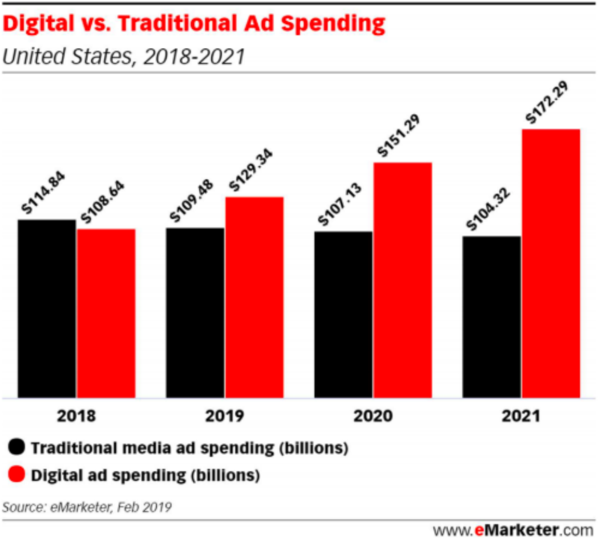


Yanjun Han
(Stanford EE)

Tsachy Weissman (Stanford), Zhengyuan Zhou (NYU),
Aaron Flores & Erik Ordentlich (Yahoo! Research)

Stanford Information Theory Forum

Success of digital ads



Online auctions

Advertiser (bidder)   Ad exchange   Publisher (seller)

Some popular auction designs:

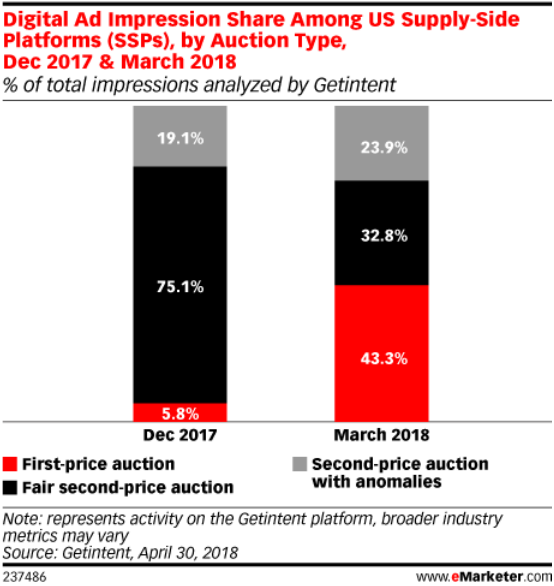
- second-price auction: the bidder with the highest bid wins the auction, and pays the price equal to the second highest bid
- first-price auction: the bidder with the highest bid wins the auction, and pays the price equal to the highest bid

From second-price to first-price

There is a recent industrial shift to first-price auctions:

- greater transparency to bidders
- enhanced monetization for sellers
- preferable model for header-bidding

From second-price to first-price



Bidder's challenge

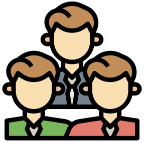
How to bid in first-price auctions where it is no longer optimal to bid truthfully?

- unknown characteristics of others' bids
- possibly censored feedback

Bidder's sequential decision model

private source

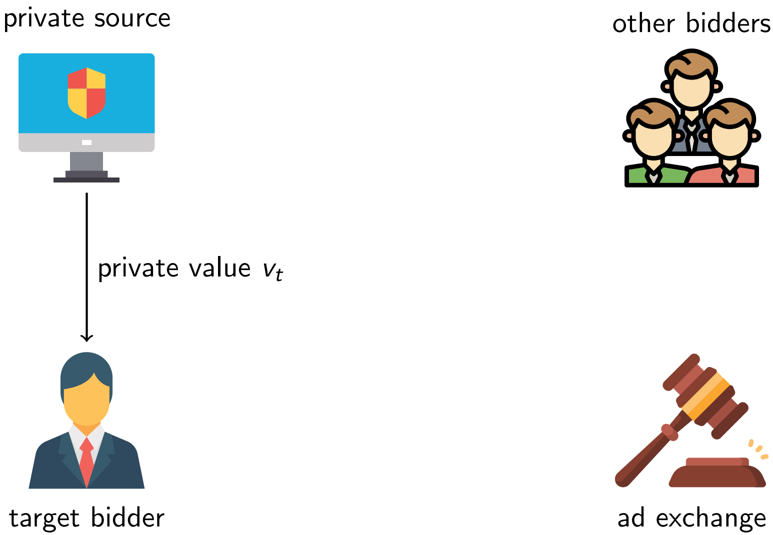
other bidders



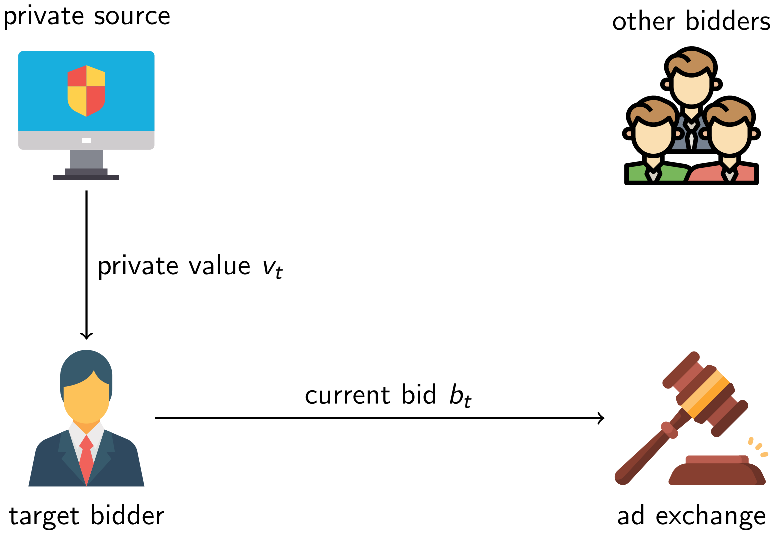
target bidder

ad exchange

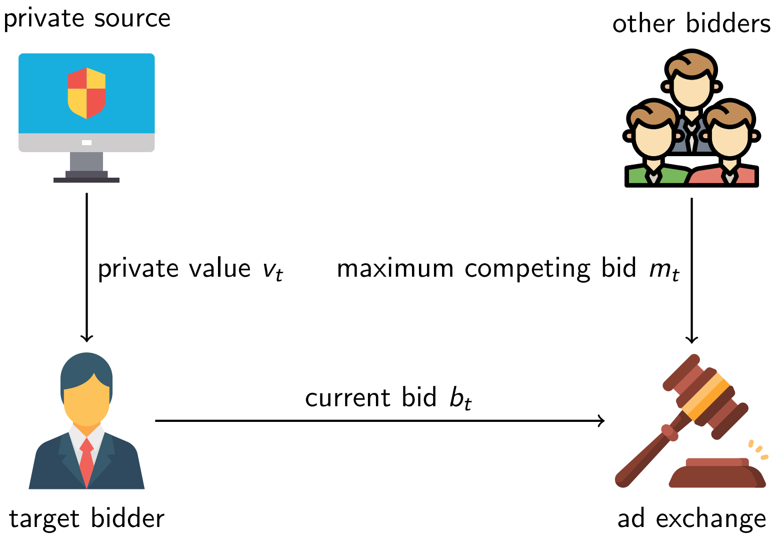
Bidder's sequential decision model



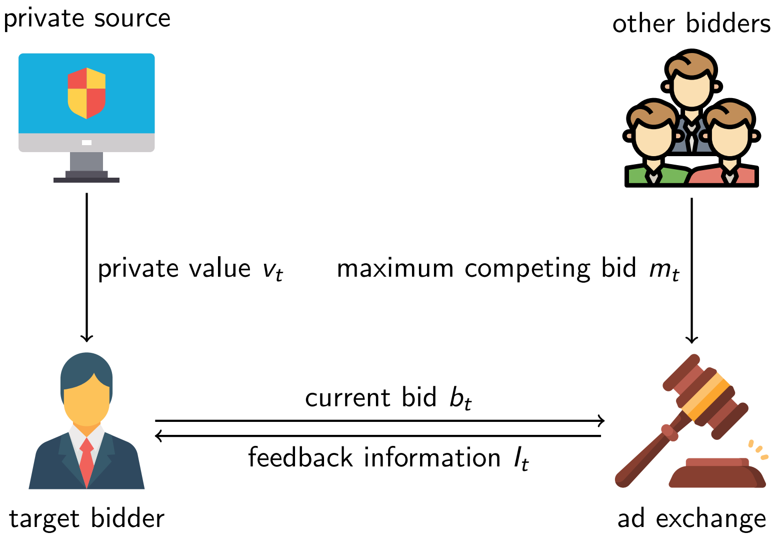
Bidder's sequential decision model



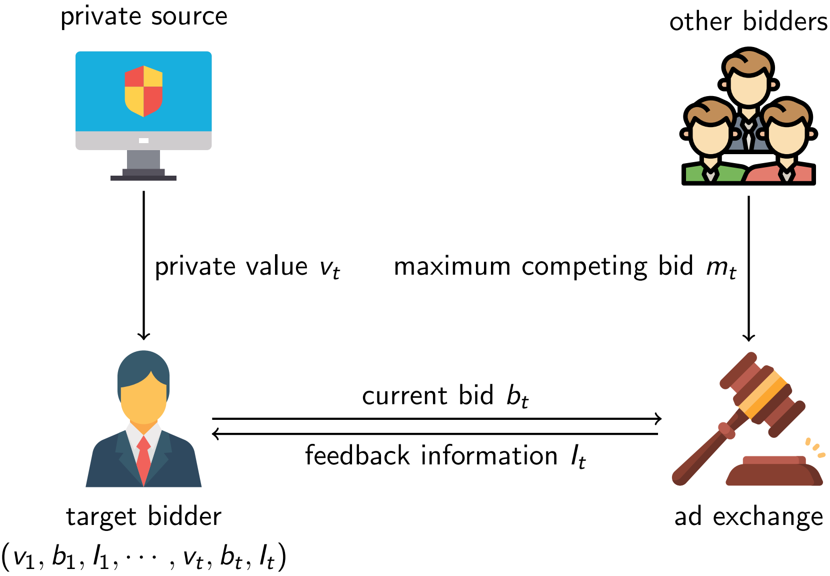
Bidder's sequential decision model



Bidder's sequential decision model



Bidder's sequential decision model



Reward and regret

Important notations:

- time horizon: T
- private valuation: $v_t \in [0, 1]$
- bidder's bid: $b_t \in [0, 1]$
- maximum competing bid: $m_t \in [0, 1]$
- instantaneous reward: $r(b_t; v_t, m_t) = (v_t - b_t) \cdot \mathbb{1}(b_t \geq m_t)$

Bidder's goal

Devise a bidding policy $\pi = (b_t)_{t=1}^T$ to minimize the regret:

$$R_T(\pi) \triangleq \max_{f \in \mathcal{F}} \sum_{t=1}^{T} r(f(v_t); v_t, m_t) - \mathbb{E}\left[\sum_{t=1}^{T} r(b_t; v_t, m_t)\right],$$

with $\mathcal{F}$ being a reasonable and rich family of bidding strategies.

Feedback structures

- Unobservable bids: the bidder only knows whether he/she wins or not, i.e. $I_t = 1(b_t \geq m_t)$ (studied in [Balseiro et al. 2019])
- Winner-only observable bids: the bidder only knows the winner's bid, i.e. $I_t = \max\{b_t, m_t\}$ (Setting I)
- Observable bids: the bidder knows the minimum bid to win, i.e. $I_t = m_t$ (Setting II)

Setting I: stochastic auctions

Assumptions:

- modeling of private value: $v_t \overset{\text{i.i.d.}}{\sim} F$ or adversarial
- modeling of others' bids: $m_t \overset{\text{i.i.d.}}{\sim} G$ with unknown CDF $G(\cdot)$
- feedback structure: only the winning bid $\max\{b_t, m_t\}$ is revealed

Regret in stochastic auctions

$$R_T(\pi) \triangleq \sum_{t=1}^{T} \left(\max_b (v_t - b)G(b) - \mathbb{E}[(v_t - b_t)G(b_t)] \right).$$

Key features:

- whenever the bidder wins the auction, he/she loses the information
- requires learning of G based on censored feedback

Setting II: adversarial auctions

Assumptions:

- modeling of private value: v_t adversarial
- modeling of others' bids: m_t adversarial
- feedback structure: m_t is always revealed

Regret in adversarial auctions

$$R_T(\pi) \triangleq \max_{f \in \mathcal{F}_{\text{Lip}}} \sum_{t=1}^{T} r(f(v_t); v_t, m_t) - \mathbb{E}\left[\sum_{t=1}^{T} r(b_t; v_t, m_t)\right],$$

where $\mathcal{F}_{\text{Lip}}$ is the set of all 1-Lipschitz functions $f : [0,1] \to [0,1]$.

Key features:

- no distributional assumption on others' bids
- robust to others' strategic or even adversarial moves

This talk

Main theorem

In both settings, there exist efficiently computable bidding strategies π such that

$$R_T(\pi) \lesssim \sqrt{T} \cdot \mathrm{polylog}(T).$$

Theoretical highlights:

- discontinuous reward function
- strong time-variant oracle
- stochastic setting: learning with censored feedback
- adversarial setting: efficient tracking of large set of experts

Part I: Stochastic Auctions

Yanjun Han
Stanford EE

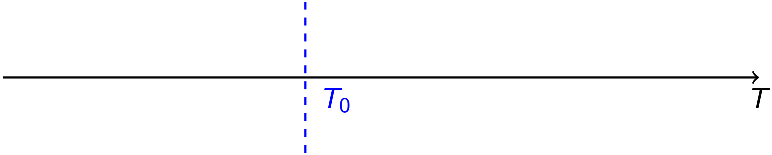
Zhengyuan Zhou
NYU Stern

Tsachy Weissman
Stanford EE

"Optimal No-regret Learning in Repeated First-price Auctions"
arXiv: 2003.09795

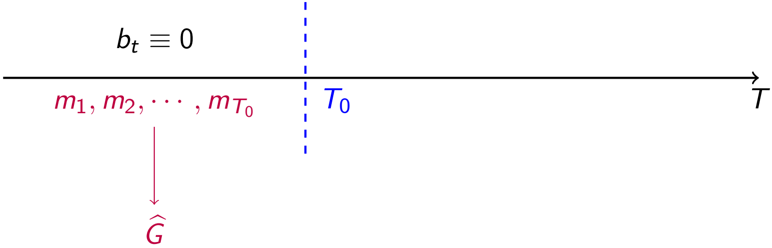
A simple $O(T^{2/3})$-regret strategy

Explore-then-commit strategy:



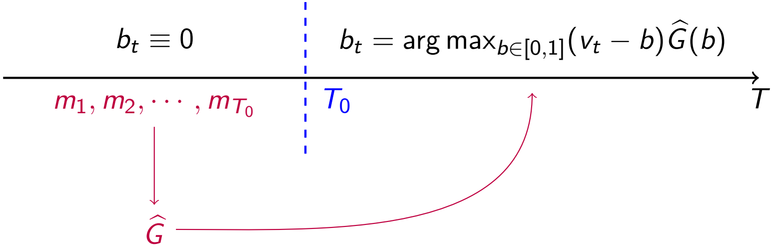
A simple $O(T^{2/3})$-regret strategy

Explore-then-commit strategy:



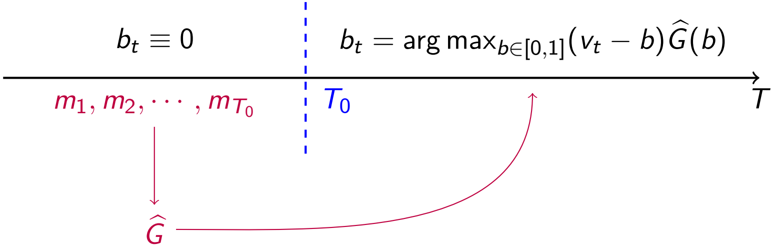
A simple $O(T^{2/3})$-regret strategy

Explore-then-commit strategy:



A simple $O(T^{2/3})$-regret strategy

Explore-then-commit strategy:

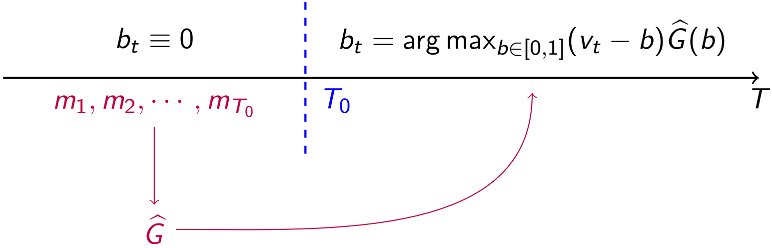


Regret analysis:

$$R_T(\pi^{\mathrm{ETC}}) = O\left(T_0 + \frac{T}{\sqrt{T_0}}\right) \overset{T_0 \sim T^{2/3}}{=} O(T^{2/3})$$

A simple $O(T^{2/3})$-regret strategy

Explore-then-commit strategy:



Regret analysis:

$$R_T(\pi^{\mathrm{ETC}}) = O\left(T_0 + \frac{T}{\sqrt{T_0}}\right) \overset{T_0 \sim T^{2/3}}{=} O(T^{2/3})$$

Question

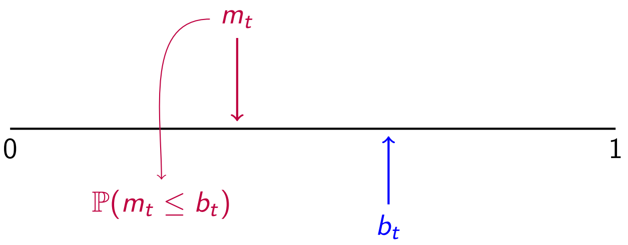
Can the regret bound be improved to $\widetilde{O}(\sqrt{T})$?

Challenges

- shown in [Balseiro et al. 2019] that $\widetilde{\Theta}(T^{2/3})$ regret is optimal when feedback is binary
- is better performance attainable with our richer feedback?
- note our selection bias...

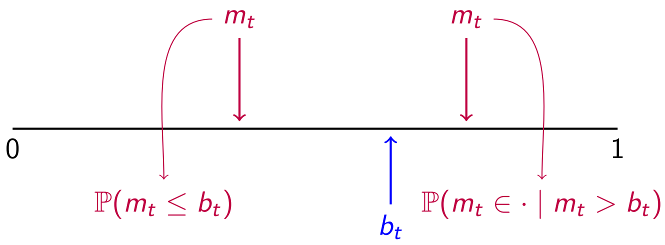
Challenges

- shown in [Balseiro et al. 2019] that $\widetilde{\Theta}(T^{2/3})$ regret is optimal when feedback is binary
- is better performance attainable with our richer feedback?
- note our selection bias...



Challenges

- shown in [Balseiro et al. 2019] that $\widetilde{\Theta}(T^{2/3})$ regret is optimal when feedback is binary
- is better performance attainable with our richer feedback?
- note our selection bias...



Monotone Group Contextual Bandit

Multi-armed bandit

- sequential decision making with horizon T and K actions
- aim to maximize the cumulative reward
- bandit feedback: only the reward of each chosen action is revealed

Action \ Time	1	2	3	4	5	6	7	...	T
1						✓			
2		✓							
3	✓				✓				
4									✓
5				✓					
...								✓	
K			✓				✓		

Optimal regret relative to the best fixed action is $\Theta(\sqrt{KT})$.

Contextual multi-armed bandit

- multi-armed bandit with C contexts
- each context corresponds to a different environment on the rewards
- bandit feedback: only the reward of each chosen action under the given environment is revealed

Action \ Time	1	2	3	4	5	6	7	...	T
1									
2									
3	✓								
4									
5				✓					
...								✓	
K			✓						

Environment under context c_1

Action \ Time	1	2	3	4	5	6	7	...	T
1						✓			
2		✓							
3					✓				
4									✓
5									
...									
K							✓		

Environment under context c_2

Optimal regret relative to the best context-specific action is $\Theta(\sqrt{CKT})$.

Relationships to optimal bidding

Correspondence between contextual bandits and bidding:

- bidder's bid $\longleftrightarrow$ action
- private value $\longleftrightarrow$ context

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1									
b_2									
b_3	✓								
b_4									
b_5				✓					
...								✓	
b_K			✓						

Environment under private value $v = v_1$

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1						✓			
b_2		✓							
b_3					✓				
b_4									✓
b_5									
...									
b_K							✓		

Environment under private value $v = v_2$

Regret analysis:

$$\text{Regret} = O\left(\sqrt{CKT} + \frac{T}{C} + \frac{T}{K}\right) \overset{C=K=T^{1/4}}{=} O(T^{3/4}).$$

Monotone feedback

Question

Does bandit feedback really hold, i.e. each action (bid) only provides information about the reward of that only action?

Answer

No! Each bid provides a monotone feedback, i.e. information about the rewards of all larger bids given all contexts.

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1						$\checkmark$			
b_2		$\checkmark$				$\checkmark$			
b_3	$\checkmark$	$\checkmark$			$\checkmark$	$\checkmark$			
b_4	$\checkmark$	$\checkmark$			$\checkmark$	$\checkmark$			$\checkmark$
b_5	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$			$\checkmark$
...	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$
b_K	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

Environment under private value $v = v_1$

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1						$\checkmark$			
b_2		$\checkmark$				$\checkmark$			
b_3	$\checkmark$	$\checkmark$			$\checkmark$	$\checkmark$			
b_4	$\checkmark$	$\checkmark$			$\checkmark$	$\checkmark$			$\checkmark$
b_5	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$			$\checkmark$
...	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$
b_K	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

Environment under private value $v = v_2$

Monotone optimal action

Lemma

Let $b^\star(v) = \arg\max_{b\in[0,1]}(v-b)G(b)$ be the optimal bid given private value v. Then the map $v \mapsto b^\star(v)$ is non-decreasing.

Implication: the optimal action under each context, albeit unknown to the learner, is known to be monotone in the context.

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1						✓			
b_2		✓				✓			
b_3	✓	✓			✓	✓			
b_4	✓	✓			✓	✓			✓
b_5	✓	✓		✓	✓	✓			✓
...	✓	✓		✓	✓	✓		✓	✓
b_K	✓	✓	✓	✓	✓	✓	✓	✓	✓

Environment under private value $v = v_1$

Bid \ Time	1	2	3	4	5	6	7	...	T
b_1						✓			
b_2		✓				✓			
b_3	✓	✓			✓	✓			
b_4	✓	✓			✓	✓			✓
b_5	✓	✓		✓	✓	✓			✓
...	✓	✓		✓	✓	✓		✓	✓
b_K	✓	✓	✓	✓	✓	✓	✓	✓	✓

Environment under private value $v = v_2$

Monotone group contextual bandit

Definition

A monotone group contextual bandit is a contextual bandit with C contexts, K actions, and time horizon T satisfying both monotone feedback and monotone optimal action properties.

Theorem (Upper Bound with Stochastic Context)

If the contexts are i.i.d. across time, then there is a policy π with

$$\mathbb{E}[R_T(\pi)] \lesssim \sqrt{T}\log(T)\log(CKT).$$

- in stochastic first-price auctions, there is a bidding policy achieving an $O(\sqrt{T}\log^2 T)$ expected regret when the private values are i.i.d.

Policy: monotone successive elimination

High-level description:

- successively eliminate probably bad actions under each context
- by eliminating more actions if necessary, ensure that the smallest active action under each context is non-decreasing over contexts
- choose the smallest active action given the current context

Limitations

Number of available observations at time t is about $\sum_{s<t} \mathbb{1}(v_s \leq v_t)$:

- $\Theta(t)$ in the best scenario $v_1 \leq v_2 \leq \cdots \leq v_T$
- $\Theta(1)$ in the worst scenario $v_1 > v_2 > \cdots > v_T$
- $\Theta(t)$ in expectation for any i.i.d. distribution

Theorem (Lower Bound)

There exists an instance of a monotone group contextual bandit and an adversarially chosen sequence of contexts such that, any policy incurs a worst-case regret at least $\Omega(T^{2/3})$.

- $\widetilde{O}(\sqrt{T})$ regret on average, but $\Omega(T^{2/3})$ again for worst-case contexts
- this framework does not extend to adversarial private values!

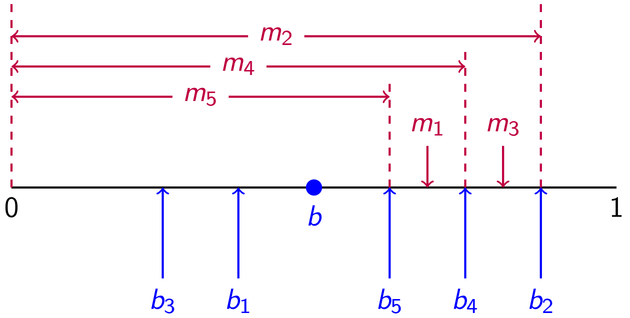
An Interval-Splitting Scheme

Correlated reward

- Specializing to first-price auctions, reward estimation is equivalent to the estimation of $\mathbb{P}(m_t > b)$ for each bid b
- For two bids $b < b'$:
  - monotone feedback: bidding price b gives a fresh observation for the estimation of $\mathbb{P}(m_t > b')$
  - partial feedback: bidding price b' also gives partial information for the estimation of $\mathbb{P}(m_t > b)$
  - partial feedback possible due to correlated reward

$$\mathbb{P}(m_t > b) = \underbrace{\mathbb{P}(m_t > b')}_{\text{one more observation}} + \underbrace{\mathbb{P}(b < m_t \le b')}_{\text{smaller target quantity}}$$

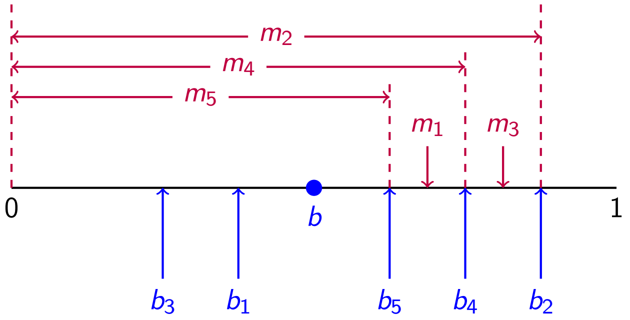
Interval-splitting estimation



$$
\begin{aligned}
\widehat{\mathbb{P}}(m_t > b) &= \widehat{\mathbb{P}}(b < m_t \leq b_5) + \widehat{\mathbb{P}}(b_5 < m_t \leq b_4) + \widehat{\mathbb{P}}(b_4 < m_t \leq b_2) + \widehat{\mathbb{P}}(m_t > b_2) \\
&= \frac{0}{2} + \frac{1}{3} + \frac{1}{4} + \frac{0}{5}
\end{aligned}
$$

Different sample sizes in different intervals.

Confidence bound



$$\mathrm{sd}(b) \approx \sqrt{\frac{\mathbb{P}(b < m_t \leq b_5)}{2} + \frac{\mathbb{P}(b_5 < m_t \leq b_4)}{3} + \frac{\mathbb{P}(b_4 < m_t \leq b_2)}{4} + \frac{\mathbb{P}(m_t > b_2)}{5}}$$

$$\widehat{\mathrm{sd}(b)} \approx \sqrt{\frac{\widehat{\mathbb{P}}(b < m_t \leq b_5)}{2} + \frac{\widehat{\mathbb{P}}(b_5 < m_t \leq b_4)}{3} + \frac{\widehat{\mathbb{P}}(b_4 < m_t \leq b_2)}{4} + \frac{\widehat{\mathbb{P}}(m_t > b_2)}{5}}$$

UCB policy

Bidding strategy: at each round, the bidder selects the bid $b \in [0, 1]$ which maximizes the upper confidence bound of the reward

$$b_t = \arg\max_{b \in [0,1]} \quad (v_t - b) \cdot \left(\widehat{\mathbb{P}}_t(m_t > b) + \widehat{\mathrm{sd}_t(b)} \right) .$$

Some other catches:

- dependence across different intervals
- dependence across time
- estimation error of $\widehat{\mathrm{sd}_t(b)}$

Solution

A multi-stage algorithm; see full paper for details.

Performance of UCB

Theorem (Upper Bound with Adversarial Private Value)

Even for adversarially chosen private values, the (multi-stage version of) UCB algorithm achieves

$$R_T(\pi^{\mathrm{UCB}}) \lesssim \sqrt{T} \log^3 T.$$

Summary of Part I

- censored feedback in first-price auctions modeled as a monotone group contextual bandit
- $\widetilde{O}(\sqrt{T})$ regret on average, but $\Omega(T^{2/3})$ in worst case
- an additional nature of correlated rewards in first-price auctions leads to $\widetilde{O}(\sqrt{T})$ regret in worst case

Part II: Adversarial Auctions

Yanjun Han
Stanford EE

Zhengyuan Zhou
NYU Stern

Aaron Flores
Yahoo! Research

Erik Ordentlich
Yahoo! Research

Tsachy Weissman
Stanford EE

"Learning to Bid Optimally and Efficiently in Adversarial First-price Auctions"
arXiv: 2007.04568

Setting

Assumptions:

- modeling of private value: v_t adversarial
- modeling of others' bids: m_t adversarial
- feedback structure: m_t is always revealed

Regret in adversarial auctions

$$R_T(\pi) \triangleq \max_{f \in \mathcal{F}_{\mathrm{Lip}}} \sum_{t=1}^{T} r(f(v_t); v_t, m_t) - \mathbb{E}\left[\sum_{t=1}^{T} r(b_t; v_t, m_t)\right],$$

where $\mathcal{F}_{\mathrm{Lip}}$ is the set of all 1-Lipschitz functions $f : [0,1] \to [0,1]$.

Main results

Theorem (Adversarial First-price Auction)

There exists a bidding strategy π such that

$$R_T(\pi) \lesssim \sqrt{T} \log T.$$

Furthermore, this regret can be attained via an efficient algorithm requiring $O(T)$ space and $O(T^{1.5})$ time.

A Statistically Optimal Policy

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	...	T
1									
2									
3									
4									
5									
...									
K									

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	...	T
1									
2									
3	✓								
4									
5									
...									
K									

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	...	T
1	✓								
2	✓								
3	✓								
4	✓								
5	✓								
...	✓								
K	✓								

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	$\dots$	T
1	✓								
2	✓	✓							
3	✓								
4	✓								
5	✓								
$\dots$	✓								
K	✓								

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	...	T
1	✓	✓							
2	✓	✓							
3	✓	✓							
4	✓	✓							
5	✓	✓							
...	✓	✓							
K	✓	✓							

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	$\dots$	T
1	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\dots$	✓	✓	✓	✓	✓	✓	✓	✓	✓
K	✓	✓	✓	✓	✓	✓	✓	✓	✓

Prediction with expert advice

- sequential decision making with horizon T and K experts (actions)
- aim to maximize the cumulative reward
- full-information feedback: rewards of all experts are revealed

Expert \ Time	1	2	3	4	5	6	7	...	T
1	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓	✓
...	✓	✓	✓	✓	✓	✓	✓	✓	✓
K	✓	✓	✓	✓	✓	✓	✓	✓	✓

Optimal regret relative to the best fixed expert is $\Theta(\sqrt{T \log K})$.

A continuous set of experts

What is the advice of each expert?

- of course, not to bid a constant price...
- instead, each expert suggests a bidding strategy $f \in \mathcal{F}_{\text{Lip}}$
- however, $|\mathcal{F}_{\text{Lip}}| = +\infty$

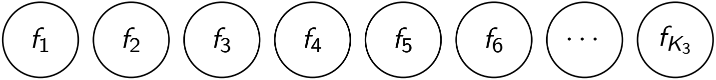
Covering Lemma (Kolmogorov–Tikhomirov'59)

For any $\varepsilon > 0$, one can choose $\exp(O(1/\varepsilon))$ candidates in $\mathcal{F}_{\text{Lip}}$ such that any element of $\mathcal{F}_{\text{Lip}}$ is ε-close to one of the candidates.

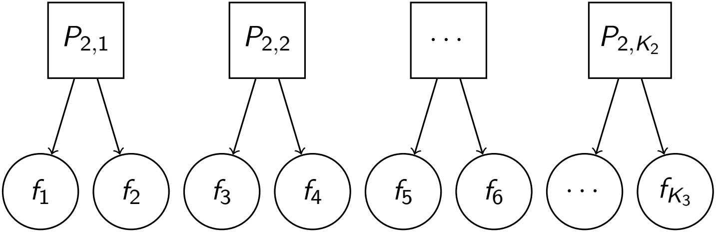
Restricting to the above candidates:

- approximation error: $O(T\varepsilon)$
- regret against the best candidate: $O(\sqrt{T/\varepsilon})$
- best achievable regret: $O(T^{2/3})$

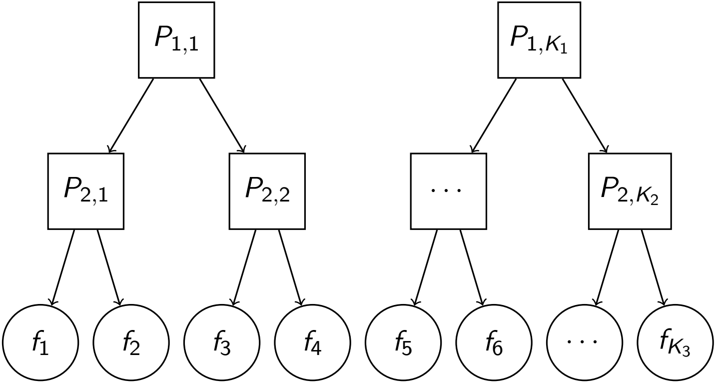
A hierarchical chaining of experts



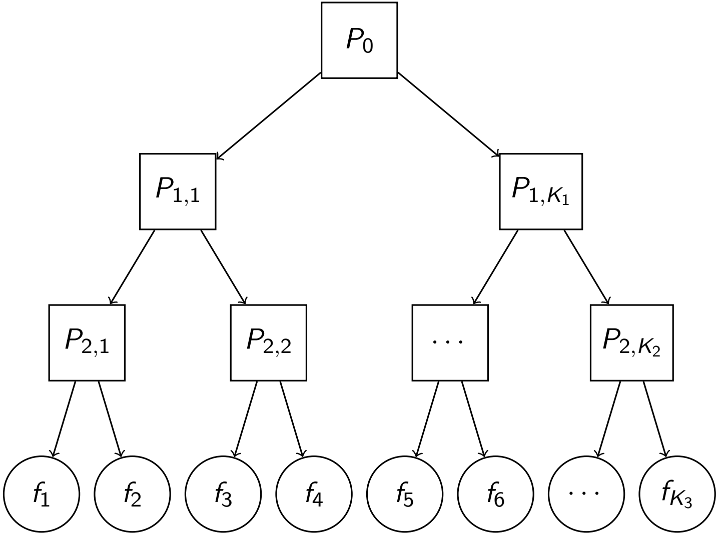
A hierarchical chaining of experts



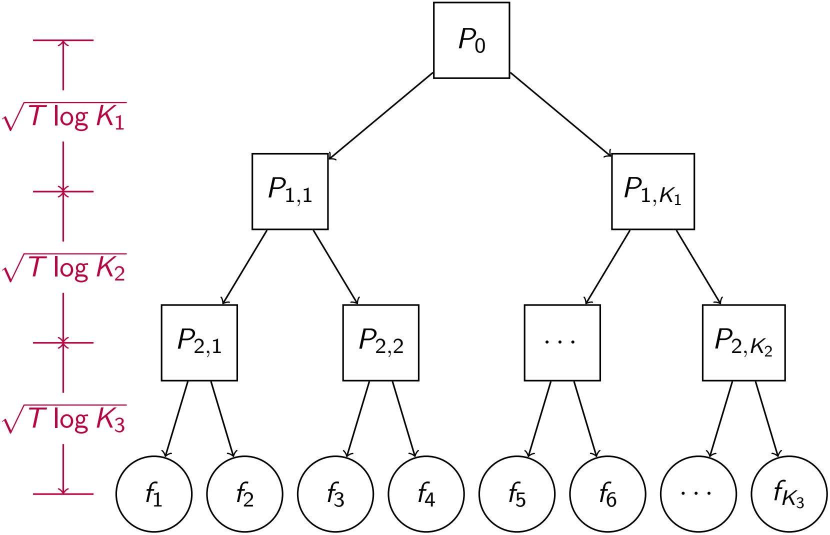
A hierarchical chaining of experts



A hierarchical chaining of experts



A hierarchical chaining of experts



Help from a good expert

Definition (Good Expert)

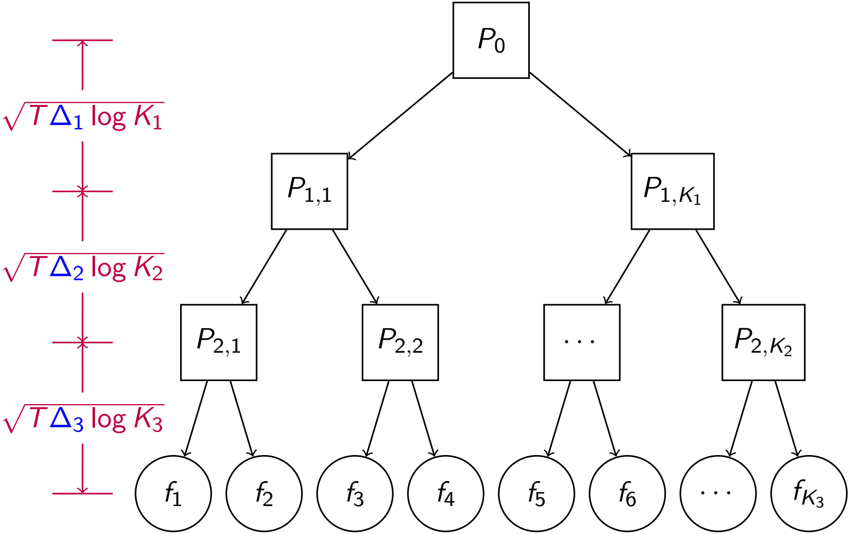
In prediction with expert advice, an expert is Δ-good if at each time, the reward of that expert is Δ-close to the reward of the best expert.

- naïvely, a regret bound $O(T\Delta)$ is achievable with a good expert
- however, a bad regret bound for $\Delta = O(1)$

Theorem (Optimal Regret with Good Expert)

For $\Delta \in [T^{-1} \log K, 1]$, the optimal regret in prediction with expert advice and a Δ-good expert is $\Theta(\sqrt{T\Delta \log K})$.

Policy: ChEW (Chained Exponential Weighting)



Analysis of ChEW

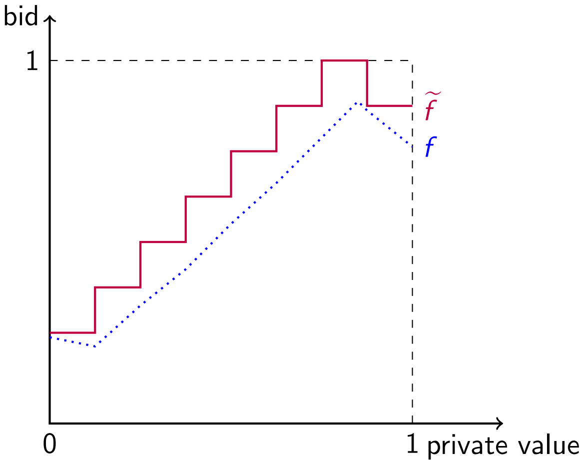
Theorem (A Statistically Optimal Policy)

The ChEW policy satisfies

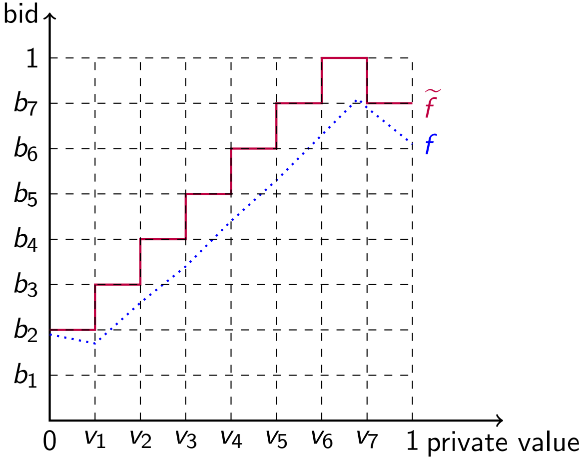
$$R_T(\pi^{\mathrm{ChEW}}) \lesssim \sqrt{T} \log T.$$

A Computationally Efficient Policy

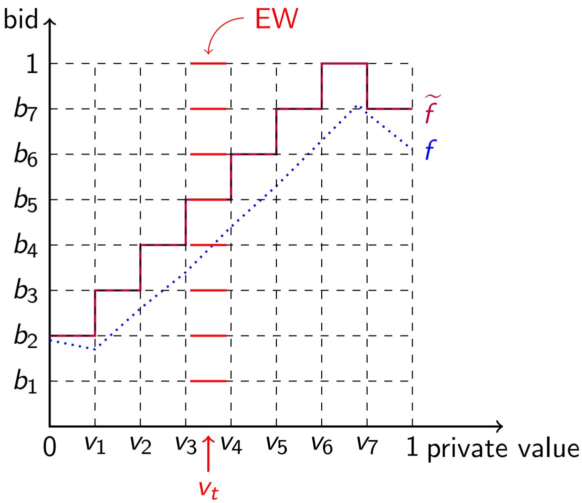
Help from product structure



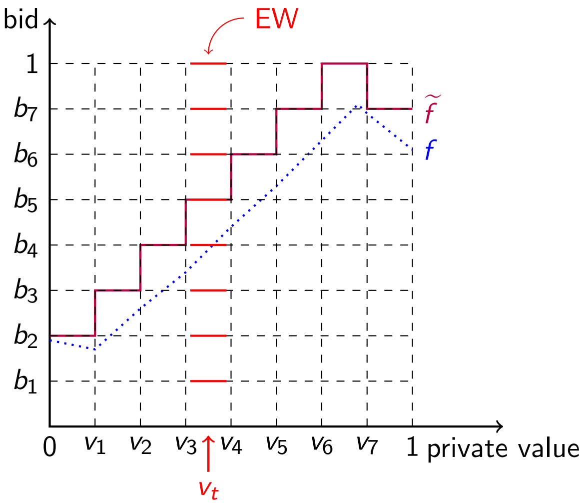
Help from product structure



Help from product structure

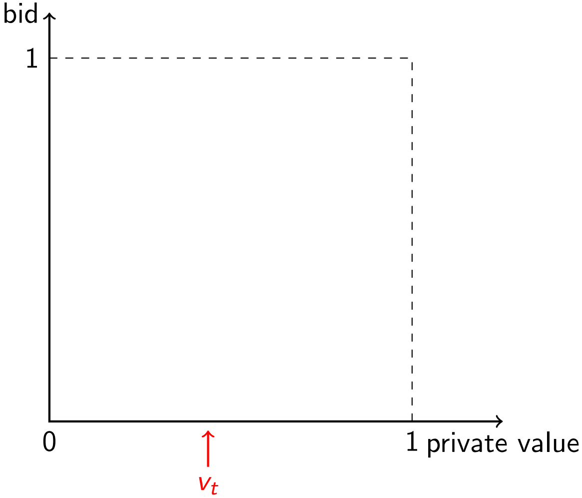


Help from product structure

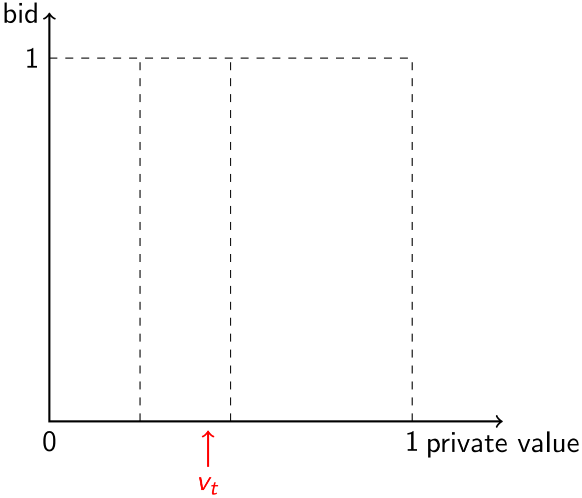


Efficient computation possible with a product structure.

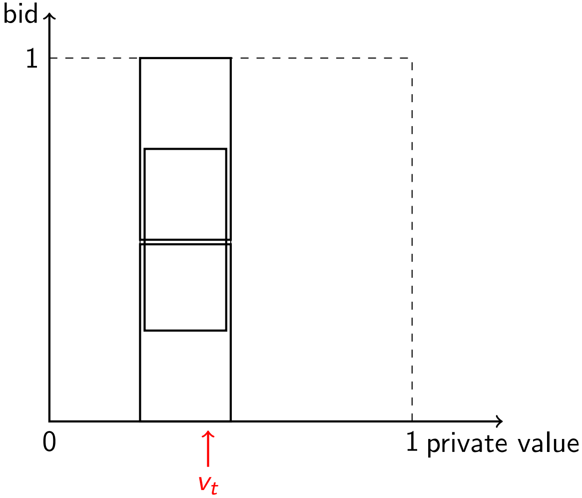
Policy: SEW (Successive Exponential Weighting)



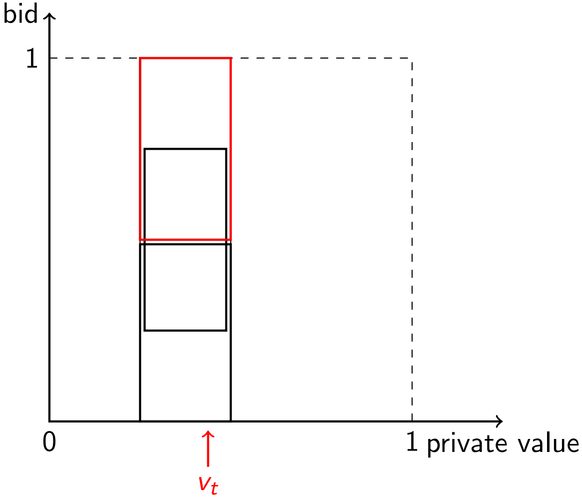
Policy: SEW (Successive Exponential Weighting)



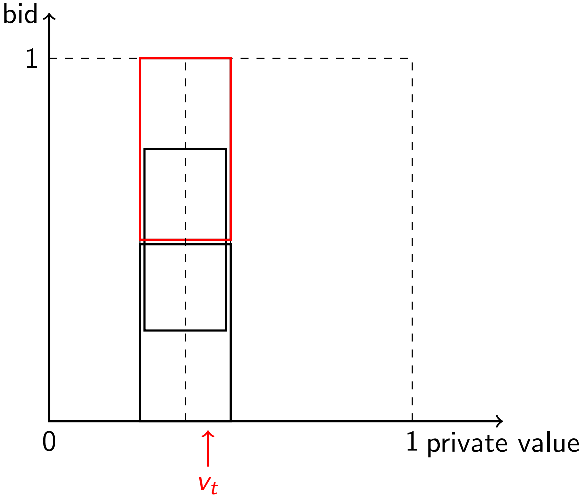
Policy: SEW (Successive Exponential Weighting)



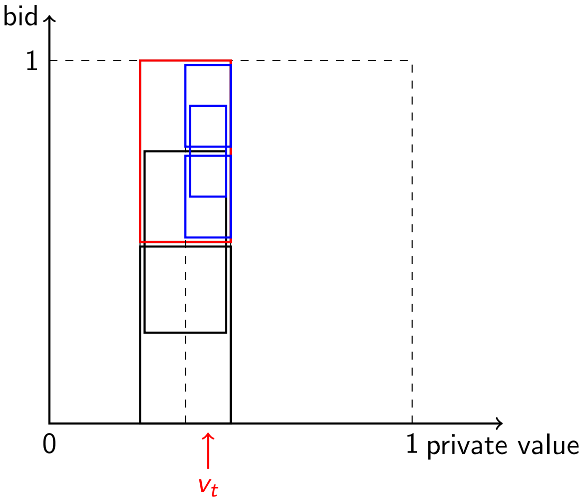
Policy: SEW (Successive Exponential Weighting)



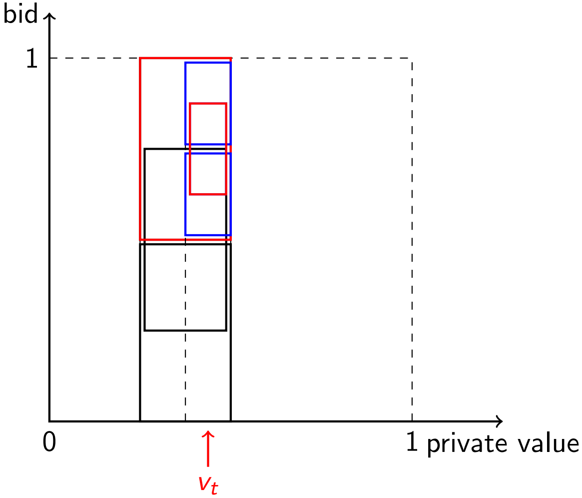
Policy: SEW (Successive Exponential Weighting)



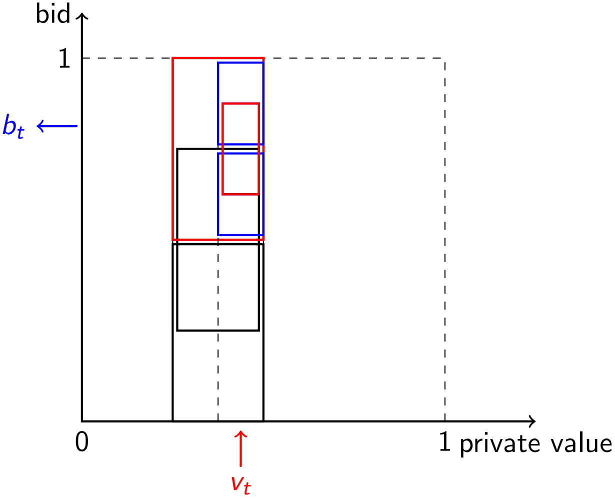
Policy: SEW (Successive Exponential Weighting)



Policy: SEW (Successive Exponential Weighting)

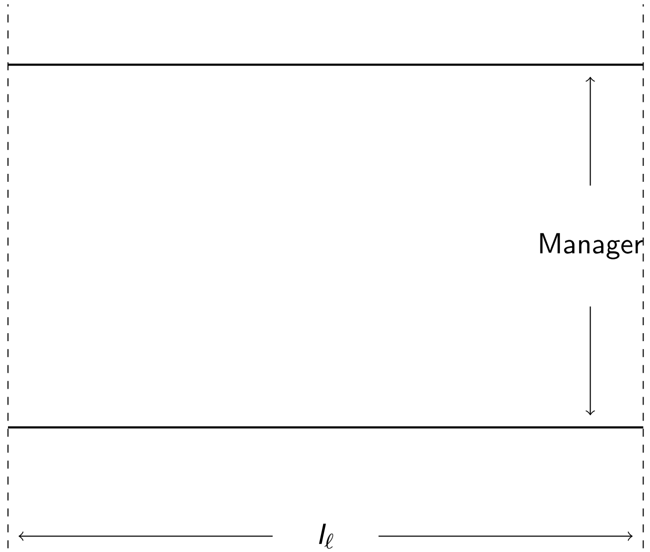


Policy: SEW (Successive Exponential Weighting)

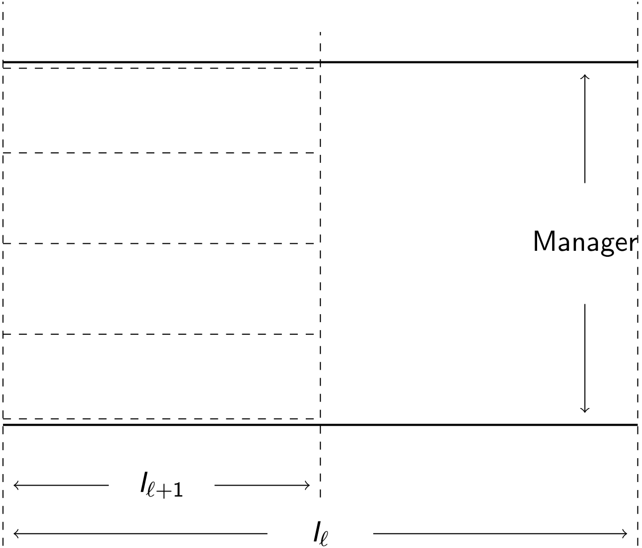


Different layers of experts correspond to different resolutions.

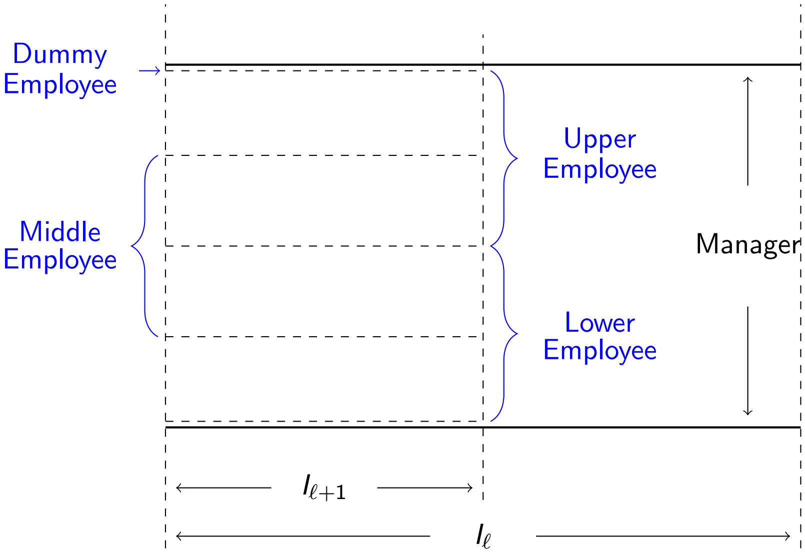
Product structure at each level



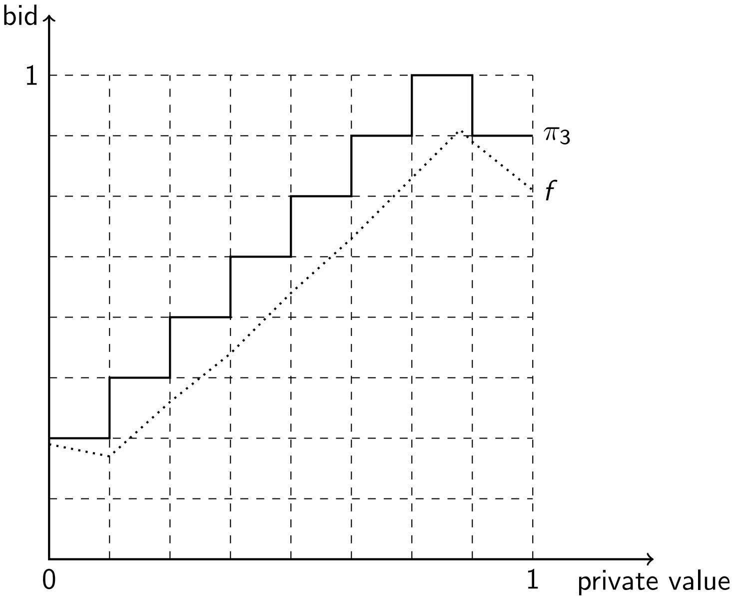
Product structure at each level



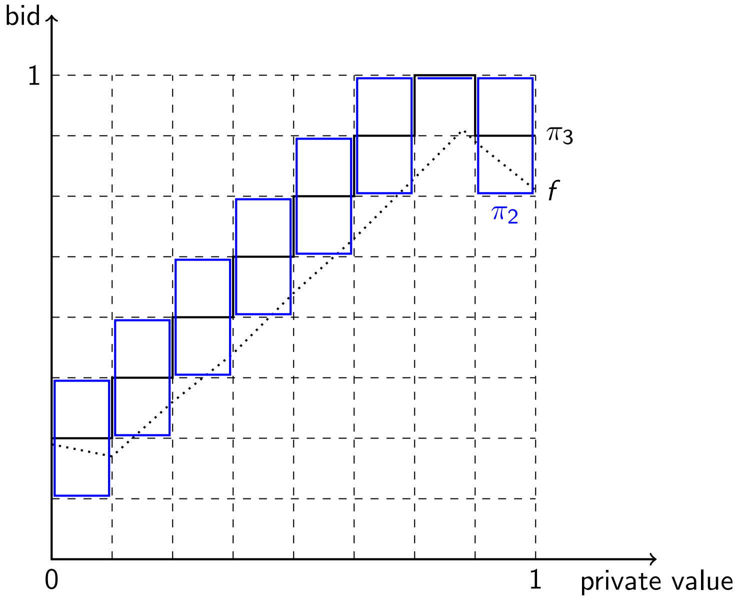
Product structure at each level



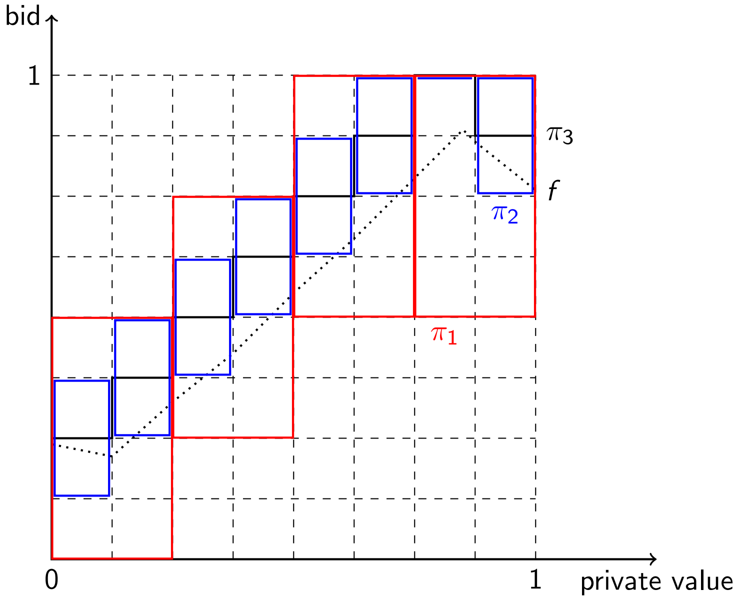
Analysis of SEW



Analysis of SEW



Analysis of SEW



A computationally efficient policy

Theorem (A Computationally Efficient Policy)

The SEW policy takes $O(T)$ space and $O(T^{1.5})$ time, and satisfies

$$R_T(\pi^{\mathsf{SEW}}) \lesssim \sqrt{T} \log T.$$

Real-data Experiments

Datasets and competing policies

Datasets:

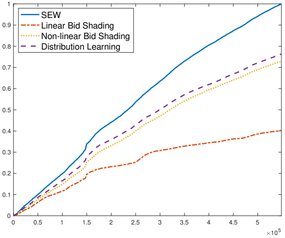
- three real datasets from Verizon Media
- each consists of two sequences $\{v_t\}$ and $\{m_t\}$
- duration: from March 24, 2020 to April 22, 2020
- sample size: 0.54M, 1.00M, and 1.57M

Competing policies:

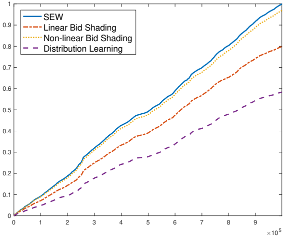
- linear bid-shading: $b_t = \theta \cdot v_t$
- non-linear bid-shading: $b_t = f(v_t; \theta)$ with non-linear f
- distribution learning:

$$b_t = \arg\max_b \ \mathbb{E}_{m_t \sim \widehat{P}_t}[r(b; v_t, m_t)].$$

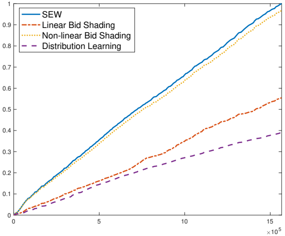
Experimental results



Dataset A



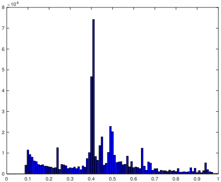
Dataset B



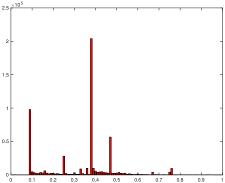
Dataset C

Adaptation to different data nature

Visualization of Dataset A:

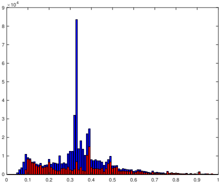


Private values

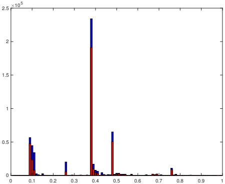


Competing bids

Bidder's bids:



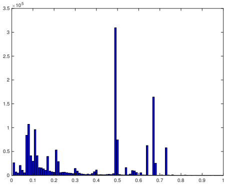
Non-linear bid shading



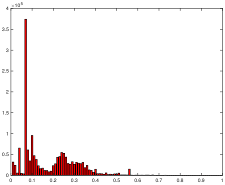
SEW

Adaptation to different data nature (cont.)

Visualization of Dataset C:

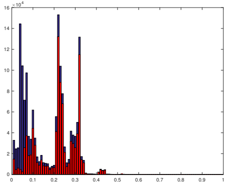


Private values

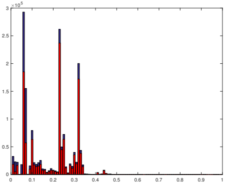


Competing bids

Bidder's bids:



Non-linear bid shading



SEW

Summary of Part II

- statistical optimality by hierarchical chaining
- efficient implementation by product structure
- superior empirical performances on all datasets

Concluding remarks

Optimal regret efficiently achievable for a single bidder in various scenarios with different assumptions on:

- characteristics of the other bidders' bids
- characteristics of the bidder's private valuation
- feedback structure of the auction
- reference policies with which our bidder competes

Future directions:

- additional side information
- simultaneous value estimation and bidding
- equilibrium theory for multiple bidders/sellers

Thank You!