EFFICIENT ALGORITHMS FOR STOCHASTIC REPEATED SECOND-PRICE AUCTIONS

joint work with Olivier Cappé and Aurélien Garivier

SECOND PRICE AUCTIONS

Second Price Auctions

Second Price auctions *used to be* the main mechanism for Real Time Bidding. The mechanism proceeds as follows:

- An item (an ad placement linked to a user) is auctioned.
- Bidders place their bids for this specific item.
- The highest bidder wins the auction. She pays the second highest bid and observes the value of the item (possibly the occurrence of a click or a purchase).

A NICE PROPERTY OF SECOND PRICE AUCTIONS

How much should you bid in a second price auction, knowing your value ?

Notation: $v :=$ known value, $m :=$ max of the adversaries' bids.

Intuition: Bidding on the right side of m ensures the maximal utility.

A nice property of second price auctions

How much should you bid in a second price auction, knowing your value ?

Notation: $v :=$ known value, $m :=$ max of the adversaries' bids.

Intuition: Bidding on the right side of m ensures the maximal utility.

If $v < m$, the bidder needs to lose, otherwise she pays more than she wins. Losing means bidding less than m.

If $v > m$, the bidder needs to win. Any bid higher than m results in the same payment m, so that winning ensures the optimal utility whatever the bid. Winning means bidding more than m.

A nice property of second price auctions

How much should you bid in a second price auction, knowing your value ?

Notation: $v :=$ known value, $m :=$ max of the adversaries' bids.

Intuition: Bidding on the right side of m ensures the maximal utility.

If $v < m$, the bidder needs to lose, otherwise she pays more than she wins. Losing means bidding less than m.

If $v > m$, the bidder needs to win. Any bid higher than m results in the same payment m, so that winning ensures the optimal utility whatever the bid. Winning means bidding more than m.

Theorem

Second price auctions are truthful.

Stochastic Setting

Setting For t in $1, \ldots, T$,

1. the bidder submits her bid B_t for the item that is of unknown value V_t. The other players submit their bids, the maximum of which is called M_t.
2. If $M_t \leq B_t$ (which includes the case of ties), the bidder observes and receives V_t, and pays M_t. Otherwise, the bidder loses the auction and does not observe V_t.

Further assumptions $\{V_t\}_{t \geq 1}$ are iid random variables in [0,1]; their expectation is denoted by $\mathbb{E}(V_t) = v$.
The maximal bids $\{M_t\}_{t \geq 1}$ are iid random variables in $[0, 1]$; their cdf is denoted by F.
Same setting as in [Weed et al.(2016)Weed, Perchet, and Rigollet].

Evaluation

Regret

$$R_T := \max_{b \in [0,1]} \sum_{t=1}^{T} \mathbb{E}[U_t(b)] - \sum_{t=1}^{T} \mathbb{E}[U_t(B_t)].$$

where the utility is $U_t(b) = (V_t - M_t)1\{b \geq M_t\}$

A few remarks

Remarks on the setting

- The stochastic assumption is arguably a reasonable assumption for RTB auctions.
- Reserve prices are not considered, because the setting of a reserve price r is equivalent to adding a bidder who constantly bids r .
- Morally, $V_t \sim Ber(v)$ where v is small

Some intuition on the problem

A structured bandit problem.

Exploration/Exploitation Trade-Off where Exploitation consists in bidding close to v and Exploitation consists in bidding high (bidding 1 means observing everything).

UCB-TYPE ALGORITHMS

UCB algorithms

UCB algorithms are a natural solution for balancing exploration/ exploitation.

$$B_t = UCB_t(\gamma) = \begin{cases} \min\left(1, \bar{V}_{t-1} + \sqrt{\frac{\gamma \log(t)}{2N_{t-1}}}\right) \\ \text{ for UCBID [Weed et al.(2016)Weed, Perchet, and Rigollet]} \\ \inf\left\{x \in (\bar{V}_{t-1}, 1] : kl(\bar{V}_{t-1}, x) = \frac{\gamma \log(t)}{N_{t-1}}\right\} \text{ for klUCBID} \\ \min\left(1, \bar{V}_{t-1} + \sqrt{\frac{2\bar{W}_{t-1}\log(3t^{\gamma})}{N_{t-1}}} + \frac{3\log(3t^{\gamma})}{N_{t-1}}\right) \\ \text{ for BernsteinUCBID ,} \end{cases}$$

Guarantees on UCBID

Locally bounded density. There exists $\Delta > 0$ such that F admits a density f bounded on $[v, v+\Delta]$, i.e., there exists $\beta > 0$, such that $\forall x \in [v, v+\Delta],\ f(x) < \beta$.

Theorem

If $F(v) > 0$, then the regret of the UCBID algorithm with parameter $\gamma > 1$ is bounded as follows:

$$R_T \leq \frac{2\beta\gamma}{\boxed{F(v)}} \log^2 T + O(\log T).$$

Remark : $\frac{1}{\boxed{F(v)}}$ $\sim$ *average time between two successive observations under the optimal policy*

ALGORITHMS WITH TIGHTER CONFIDENCE INTERVALS

Theorem

If $F(v) \neq 0$, the kl-UCBID algorithm with parameter $\gamma > 1$ yields the following bound on the regret:

$$R_T \leq 8\gamma \boxed{v(1-v)} \frac{\beta}{\boxed{F(v)}} \log^2(T)\big(1+o(1)\big) \ .$$

whereas the Bernstein-UCBID algorithm with parameter $\gamma > 1$ yields the following bound on the regret:

$$R_T \leq 8\gamma \boxed{w} \frac{\beta}{\boxed{F(v)}} \log^2(T) + O(\log T),$$

where w is the variance.

Theorem

Without further assumption, the maximal regrets of UCBID and klUCBID are $O(\sqrt{T}\log T)$*. If F has a density that is bounded from below and above by non negative constants, the maximal regret of UCBID remains of the same order, while it is reduced to* $O(T^{\frac{1}{3}}\log^2 T)$ *for BernsteinUCBID and to* $O(\log^2 T)$ *for klUCBID.*

Simulations

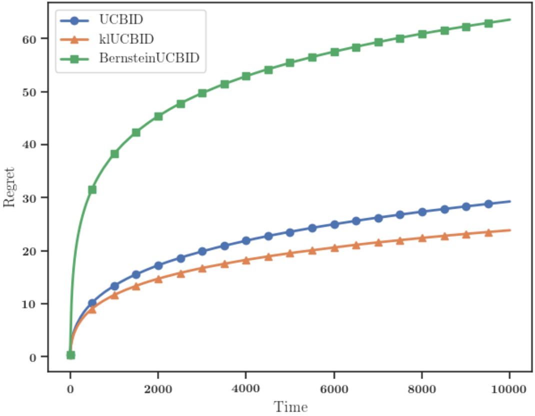

Figure: Regret plots of three UCB algorithms for values $V_t \sim \text{Ber}(0.2)$ and uniform M_t

Simulations

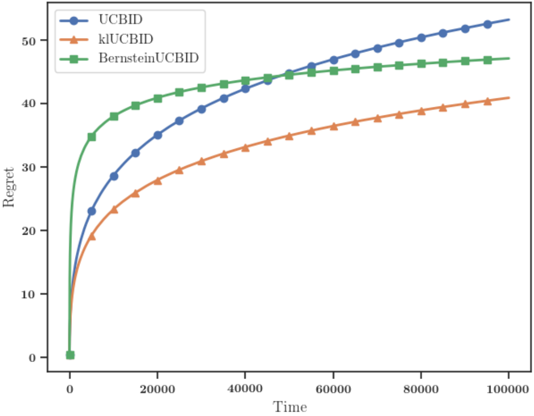

Figure: Regret plots of three UCB algorithms for V_t supported on $\{0.195, 0.205\}$ and uniform M_t

Simulations

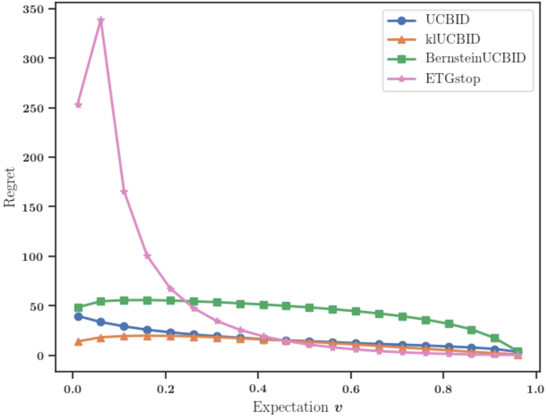

Figure: Regret at time 5000 of studied policies for uniform M_t and Bernoulli-distributed V_t of varying mean v.

PARAMETER-DEPENDENT LOWER BOUND

Lower Bound

Theorem

We consider all environments where V_t follows a Bernoulli distribution with expectation v and F admits a density f that is bounded from below and above, with $f(b) \geq \beta > 0$. If a strategy is such that, for all such environments, $R_T \leq O(T^a)$, for all $a > 0$, and if there exists $\gamma > 0$ such that for all such environments, $\mathbb{P}(B_t < v) < t^{-\gamma}$, then this strategy must satisfy:

$$\liminf_{T \to \infty} \frac{R_T}{\log T} \geq \beta \frac{v(1-v)}{16F(v)}. \tag{1}$$

The assumption is satisfied by all studied UCB algorithms

Intuition behind the proof of the lower bound

Originality : we consider a different alternative for each of the T time steps.

We fix a model in which all $(V_s)_{s=1}^T$ follow a Bernoulli distribution with expectation v, and the bids $(M_s)_{s=1}^T$ are distributed according to F. At each time t, we consider the alternative model where the values $(V_s)_{s=1}^T$ follow a Bernoulli distribution with expectation $v'_t = v + \sqrt{\frac{v(1-v)}{F(v)t}}$, and the bids M_t are distributed according to F.

Lemma

If F admits a density f, which satisfies

$$\exists\ \beta, \forall x \in [0,1], \quad \beta \leq f(x);$$

Then,

$$\frac{\beta}{2}\sum_{t=1}^{T}\mathbb{E}[(B_t - v)^2] \leq R_T.$$

PROOF OF THE LEMMA

The utility writes

$$\mathbb{E}[U_t(b)] = \int_0^b (v-m)f(m)dm$$
$$= (v-b)F(b) + \int_0^b F(m)dm,$$

The instantaneous regret writes :

$$\mathbb{E}[r_t(b)] = \mathbb{E}[U_t(v)] - \mathbb{E}[U_t(b)]$$
$$= \int_b^v F(m)dm - (v-b)F(b)$$
$$= \int_b^v (F(m) - F(b))dm$$
$$= \int_b^v \int_b^m f(u)dudm \geq \beta(v-b)^2$$

Intuition behind the proof of the lower bound

We fix a time step t. Thanks to Le Cam's method,

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}.$$

thanks to the non-underbidding assumption.

Intuition behind the proof of the lower bound

We fix a time step t. Thanks to Le Cam's method,

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}\underbrace{KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}_{kl(v,v'_t)\mathbb{E}_v[N_t]}}.$$

thanks to the non-underbidding assumption.

We fix a time step t. Thanks to Le Cam's method,

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}\underbrace{KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}_{kl(v,v'_t)\mathbb{E}_v[N_t]}}.$$

The assumption that $R_T \leq O(T^a)$ allows us to say that $\frac{\mathbb{E}_v[N_t]}{t} \to 1/F(v)$ thanks to the non-underbidding assumption.

Intuition behind the proof of the lower bound

We fix a time step t. Thanks to Le Cam's method,

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}\underbrace{KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}_{kl(v,v'_t)\mathbb{E}_v[N_t]}}.$$

The assumption that $R_T \leq O(T^a)$ allows us to say that $\frac{\mathbb{E}_v[N_t]}{t} \to 1/F(v)$

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}kl(v,v'_t)(1+\epsilon)F(v)t}.$$

thanks to the non-underbidding assumption.

We fix a time step t. Thanks to Le Cam's method,

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}\underbrace{KL(\mathbb{P}_v^{I_t}, \mathbb{P}_{v'_t}^{I_t})}_{kl(v,v'_t)\mathbb{E}_v[N_t]}}.$$

The assumption that $R_T \leq O(T^a)$ allows us to say that $\frac{\mathbb{E}_v[N_t]}{t} \to 1/F(v)$

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) + \cancel{\mathbb{P}_{v'_t}\left(B_t < \frac{v+v'_t}{2}\right)} \geq 1 - \sqrt{\frac{1}{2}kl(v,v'_t)(1+\epsilon)F(v)t - o(t^{-\gamma})}$$

thanks to the non-underbidding assumption.

Intuition behind the proof of the lower bound

We have proved

$$\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right) \geq 1 - \sqrt{\frac{1}{2}\underbrace{kl(v,v'_t)}_{\lesssim \frac{1+\epsilon}{2F(v)t}}(1+\epsilon)F(v)t - o(t^{-\gamma})}.$$

Now, $\mathbb{E}_v[(B_t - v)^2] \geq (v - \frac{v+v'_t}{2})^2\mathbb{P}_v\left(B_t > \frac{v+v'_t}{2}\right)$

This, together with the development of $kl(v, v'_t)$ yields

$$\mathbb{E}_v[(B_t - v)^2] \geq \frac{v(1-v)}{4F(v)t}\left(1 - \frac{1}{2}(1+\epsilon) - 1/t^{\gamma}\right),$$

for t large enough.

$$\liminf_{T\to\infty} \frac{\sum_{t=1}^{T}\mathbb{E}_v[(B_t - v)^2]}{\log T} \geq \frac{v(1-v)}{8F(v)}.$$

SIMPLER, NON OVERBIDDING ALGORITHMS

Explore Then Greedy

- Strategies inspired by Explore Then Commit's strategies ETG strategies.
- In the exploration phase, the maximal value of the bid ($B_t = 1$) is chosen to force observation. After a well-chosen stopping time, the bidder chooses either to abandon the bids (choosing $B_t = 0$), or to continue with the running average of observed values (greedy phase).

Motivations

- **Practical motivation**: In the context of digital advertising, simplicity is critical.
- **Other practical motivation**: ETG strategies are easy to explain and similar to truely implemented strategies.
- **Theoretical motivation**: The lower bound only works for overbidding strategies : can non-overbidding strategies work ?

ETGstop

We propose one instance of ETG, that we call ETGstop, defined by the following choice of stopping times τ_1 and τ_0:

$$\tau_1 := \inf\left\{t \in [1, T] : \exp\left(-\frac{tL_t}{8}\right) \leq \frac{1}{T^2}\right\},\ \tau_0 = \inf\left\{t \in [1, T] : U_t \leq \frac{1}{T^{\frac{1}{3}}}\right\} \tag{2}$$

where we denote by $L_t = \min\{v \in [0, \bar{V}_t[: \exp\left(-tkl(\bar{V}_t, v)\right) \leq 1/T^2\}$ and by $U_t = \max\{v \in [\bar{V}_t, 1[: \exp(-tkl(\bar{V}_t, v)) \geq 1/T^2\}$ the kl-lower and upper confidence bound for the confidence level $1/T^2$.

Idea behind the choice of stopping time

Idea : We want to ensure a minimal ratio of won auctions in this second phase.

This choice of stopping time allows to guarantee that with high probability, if τ_1 is smaller than τ_0, all bids will be larger than $\frac{v}{2}$ in the second phase. Indeed, we prove that for all n, $\mathbb{P}(\bar{V}(n) \leq \frac{v}{2}) \leq \exp(-nkl(v, \frac{v}{2})) \leq \exp(-\frac{nv}{8})$, where $\bar{V}(n)$ denotes the empirical mean of the first n observed values. With high probability, $\exp(-\frac{nv}{8}) \leq \exp(-\frac{nL_\tau}{8}) \leq \frac{1}{T^2}$, for all $n > \tau$, since L_t is a lower confidence bound of v. Therefore, the probability that there exists a time step in the second phase for which the average of the observed values is less than $v/2$ is small.

GUARANTEES

Theorem

If F admits a density f, that satisfies
$\exists\ \beta, \beta > 0, \forall x \in [0,1], \quad \beta \leq f(x) \leq \beta$, *then the regret of ETGstop satisfies :*

$$\max_{v \in [0,1]} R_T(v) \leq O(T^{\frac{1}{3}} \log^2 T),$$

and if $v > \frac{1}{T^{\frac{1}{3}}}$,

$$\textit{then} \quad R_T(v) \leq 7 + \frac{64 \log(T) + 60 T^{-1/2}}{v} + \frac{4}{F(v/2)} + \beta \frac{\log^2 T}{F(v/2)}.$$

The limitation of ETG Strategies

Theorem

If F admits a density lower-bounded by $\beta > 0$, then the regret of any ETG strategy satisfies

$$\max_{v \in [0,1]} R_T(v) \geq \frac{\beta}{4}\left(T^{\frac{1}{3}} - 1\right). \tag{3}$$

SIMULATIONS

SIMULATIONS

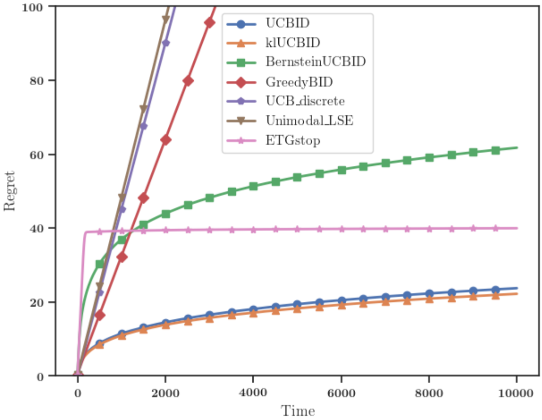

Figure: Comparison with ETGstop and other algorithms, for $V_t \sim \text{Ber}(0.3)$ and uniform M_t.

SIMULATIONS

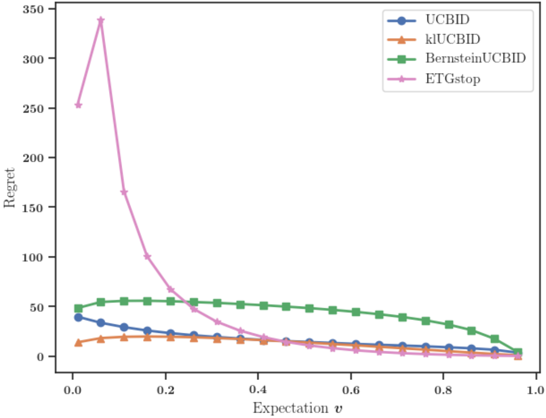

Figure: Regret at time 5000 of studied policies for uniform M_t and Bernoulli-distributed V_t of varying mean v.

J. Weed, V. Perchet, and P. Rigollet.
Online learning in repeated auctions.
In *Conference on Learning Theory*, pages 1562–1583, 2016.