Optimization of a SSP's Header Bidding Strategy using Thompson Sampling

Grégoire Jauvion
AlephD
Paris
gregoire.jauvion@gmail.com

Nicolas Grislain
AlephD
Paris
ng@alephd.com

Pascal Dkengne Sielenou
IMT
Toulouse
sielenou_alain@yahoo.fr

Aurélien Garivier
IMT
Toulouse
aurelien.garivier@math.univ-toulouse.fr

Sébastien Gerchinovitz
IMT
Toulouse
sebastien.gerchinovitz@math.univ-toulouse.fr

ABSTRACT

Over the last decade, digital media (web or app publishers) generalized the use of real time ad auctions to sell their ad spaces. Multiple auction platforms, also called Supply-Side Platforms (SSP), were created. Because of this multiplicity, publishers started to create competition between SSPs. In this setting, there are two successive auctions: a second price auction in each SSP and a secondary, first price auction, called header bidding auction, between SSPs.

In this paper, we consider an SSP competing with other SSPs for ad spaces. The SSP acts as an intermediary between an advertiser wanting to buy ad spaces and a web publisher wanting to sell its ad spaces, and needs to define a bidding strategy to be able to deliver to the advertisers as many ads as possible while spending as little as possible. The revenue optimization of this SSP can be written as a contextual bandit problem, where the context consists of the information available about the ad opportunity, such as properties of the internet user or of the ad placement.

Using classical multi-armed bandit strategies (such as the original versions of UCB and EXP3) is inefficient in this setting and yields a low convergence speed, as the arms are very correlated. In this paper we design and experiment a version of the Thompson Sampling algorithm that easily takes this correlation into account. We combine this bayesian algorithm with a particle filter, which permits to handle non-stationarity by sequentially estimating the distribution of the highest bid to beat in order to win an auction. We apply this methodology on two real auction datasets, and show that it significantly outperforms more classical approaches.

The strategy defined in this paper is being developed to be deployed on thousands of publishers worldwide.

CCS CONCEPTS

• **Mathematics of computing → Probability and statistics**; • **Applied computing → Online auctions**;

KEYWORDS

Online Learning, Multi-armed Bandit, Thompson Sampling, Bayesian Inference, Particle Filter, Sequential Monte Carlo, Big-Data, Ad-tech, Real-time, Auctions

ACM Reference Format:
Grégoire Jauvion, Nicolas Grislain, Pascal Dkengne Sielenou, Aurélien Garivier, and Sébastien Gerchinovitz. 2018. Optimization of a SSP's Header Bidding Strategy using Thompson Sampling. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3219819.3219917

1 INTRODUCTION

Real-Time Bidding (RTB) is a mechanism widely used by web publishers to sell their ad inventory through auctions happening in real time. Generally, a publisher sells its inventory through different Supply-Side Platforms (SSPs), which are intermediaries who enable advertisers to bid for ad spaces. A SSP generally runs its own auction between advertisers, and submits the result of the auction to the publisher.

There are several ways for the publisher to interact with multiple SSPs. In the typical ad selling mechanism without header bidding, called the waterfall mechanism, the SSPs sit at different priorities and are configured at different floor prices (typically the higher the priority, the higher the floor price). The ad space is sold to the SSP with the highest priority who bids a price greater than its floor price.

With header bidding, all the SSPs are called simultaneously thanks to a piece of code running in the header of the web page. Then, they compete in a first-price auction which is called the header bidding auction thereafter. In this mechanism, a SSP with a lower priority can purchase the ad if it pays more than a SSP with a higher priority. Consequently, a RTB market with header bidding is more efficient than the waterfall mechanism for the publisher.

In this paper, we take the viewpoint of a single SSP buying inventory in a RTB market with header bidding. Based on the result

KDD '18, August 19–23, 2018, London, United Kingdom
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219917

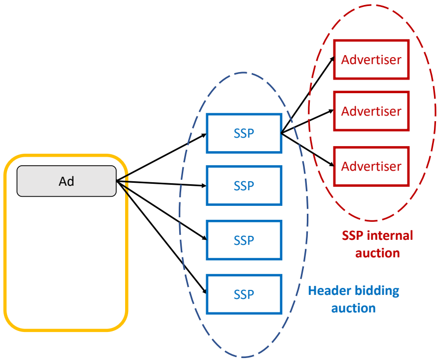

Figure 1: Ad-selling process
Typically some digital content will call a Supply-Side Platform (SSP) when loading, that will itself call many advertisers for bids in a real-time auction. To increase competition, some publishers have been calling several SSPs to introduce some competition among them. This setting is called *header bidding* because the competition among SSPs has typically been happening on the client side, in the header of HTML pages. Header bidding is, in practice, a two-staged process, with several second-price auctions happening in various SSPs, the response of which are aggregated in a final first-price auction. A SSP may be willing to adjust the bid it is responding to adapt to the first price auction context. This can be seen as an adaptative fee.

of the auction it runs internally, it submits a bid in the header bidding auction and competes with the other SSPs. This ad-selling process is summarized on Figure 1. When it wins the header bidding auction, the SSP is paid by the advertiser displaying its ad, and pays to the publisher the closing price of the header bidding auction. We study the problem of sequentially optimizing the SSP's bids in order to maximize its revenue. Quite importantly we consider a censored setting where the SSP only observes if it has won or lost once the header bidding auction has occurred. The bids of the other SSPs are not observed.

This optimization problem is formalized as a stochastic contextual bandit problem. The context is formed by the information available before the header bidding auction happens, including the result of the SSP internal auction. In each context, the highest bid among the other SSPs in the header bidding auction is modeled with a random variable and is updated in a bayesian fashion using a particle filter. Therefore, the reward (i.e., the revenue of the SSP) is stochastic. We design and experiment a version of the Thompson sampling algorithm in order to optimize bids in a sequence of auctions.

The paper is organized as follows. We discuss earlier works in Section 2. In Section 3 we formalize the optimization problem as a stochastic contextual bandit problem. We then describe our version of the Thompson sampling algorithm in Section 4. Finally, in Section 5, we present our experimental results on two real RTB auction datasets, and show that our method outperforms two more traditional bandit algorithms.

2 RELATED WORK

[31] provides a very clear introduction to the header bidding technology, and how it modifies the ad selling process in a RTB market.

Bid optimization has been much studied in the advertising literature. A lot of papers study the problem of optimizing an advertiser bidding strategy in a RTB market, where the advertiser wants to maximize the number of ads it buys as well as a set of performance goals while keeping its spend below a certain budget (see [32] and references therein). [15] and [22] state the problem as a control problem and derive methods to optimize the bid online. In [34], the authors define a functional form for the bid (as a function of the impression characteristics) and write the problem as a constrained optimization problem.

The setting of an intermediary buying a good in an auction and selling it to other buyers (which is what the SSP does in our setting) has been widely studied in the auction theory literature. In [28], the author uses the tools developed in [27] (one of the most well-known papers in auction theory) to derive an optimal auction design in this setting, and [14] and [10] analyze how the intermediary should behave to maximize its revenue.

[17] studies the optimal mechanism a SSP should employ for selling ads and analyzes optimal selling strategies. [29] analyzes the optimal behaviour of a SSP in a market with header bidding, and validates the approach on randomly generated auctions.

From an algorithmic perspective, the Thompson sampling algorithm was introduced in [30]. The papers [19, 21] studied its theoretical guarantees in parametric environments, while [23] studied it in non-parametric environments. Besides, a very clear overview of the particle filtering approach to update the posterior distribution is given in [12, 26].

Bandit algorithms were already designed and studied for repeated auctions, including RTB auctions. For instance, in repeated second-price auctions, [33] construct a bandit algorithm to optimize a given bidder's revenue, while [5] design a bandit algorithm to optimize the seller's reserve price.

In a setting very similar to ours, [18] study the situation where a given SSP competes with other SSPs in order to buy an ad space. They design an algorithm that provably enables the SSP to win most of the auctions while only paying a little more than the expected highest price of the other SSPs. Though the problem seems similar, our objective is different: we want the SSP to maximize its revenue, and not necessarily to win most auctions with a small extra-payment. In particular we cannot neglect the closing price of the SSP's internal auction in the optimization process.

We finally mention the work of [20] for the online posted-price auction: for each good in a sequence of identical goods, a seller chooses and announces a price to a new buyer, who buys the good provided the price does not exceed their private valuation (see also [24, 25] when the seller faces strategic buyers). Though their problem is different, the shape of their reward function is very similar to ours. The authors show that the classical UCB1 and Exp3 bandit algorithms applied to discretized prices are worst-case optimal under several assumptions on the sequence of the buyers' valuations. In our paper we do not tackle the worst case and instead use prior knowledge on the ad auction datasets (i.e., an empirically-validated parametric model) to better optimize the SSP's revenue.

3 PROBLEM STATEMENT

3.1 RTB market

We represent the RTB market as an infinite sequence $\mathcal{D}$ of time-ordered impressions $1, \ldots, n, \ldots$ happening at times $t_1, \ldots, t_n, \ldots$. We note $\mathcal{D}_t$ the sequence of impressions happening before time t (including t).

Impression i happening at time t_i is characterized by a context c_i, which summarizes all the information relative to impression i that is available before the header bidding auction starts. It may contain the ad placement (where it is located on the web page), some properties of the internet user (for example its operating system), or the time of the day. An important variable of the context which is specific to our setting is the closing price of the SSP internal auction, which is known before the header bidding auction happens.

We assume that the context is categorical with a finite number of categories C. A continuous variable can be discretized to meet this assumption. Without loss of generality, we assume that the categories are $1, \ldots, c, \ldots, C$. We note $\mathcal{D}_{t,c}$ the subsequence of $\mathcal{D}_t$ containing all impressions i such that $c_i = c$.

3.2 Ad selling process with header bidding

We assume that S SSPs: $\mathcal{S}_1, \ldots, \mathcal{S}_S$ compete in the header bidding auctions (possibly bidding 0 if they are not interested in purchasing the ad). We note $b_{i,s}$ the bid of $\mathcal{S}_s$ in the header bidding auction for impression i. As the header bidding auction is a first-price auction, its closing price is $\max_s(b_{i,s})$.

From now on, we consider the problem from $\mathcal{S}_1$ standpoint. We note $q_i = b_{i,1}$ the bid submitted by $\mathcal{S}_1$ in the header bidding auction, which is the variable to optimize. We also write $x_i = \max(b_{i,2}, \ldots, b_{i,S})$, which is the highest bid among the other SSPs.

In each impression i, we assume that $\mathcal{S}_1$ runs an internal auction between advertisers, whose closing price is denoted p_i. p_i is the amount paid by the advertiser winning the internal auction to $\mathcal{S}_1$ should $\mathcal{S}_1$ win the header bidding auction. Note that we do not need to know the detailed internal auction mechanism but only its closing price.

Before header-bidding, a SSP would run a second-price auction with an advertiser bidding \$10 and closing at p_i = \$8. Then the SSP would respond $q_i = p_i$ − fees = \$6 to the publisher. In this context, the advertiser pays \$8, the publisher receives \$6 and the SSP gets its fees: \$2. In a header bidding context, the SSP is in competition with other SSPs in a first price auction, it may lose an opportunity by taking too much fees or pay too much if it is sure to win and take too little fees.

3.3 Revenue function for the SSP $\mathcal{S}_1$

The revenue function $R_i(.)$ of $\mathcal{S}_1$ at impression i can be written as $R_i(q) = \mathbf{1}_{q \geq x_i}(p_i - q)$. Indeed:

- When $q \geq x_i$, $\mathcal{S}_1$ wins the header bidding auction. It is paid p_i by the advertiser winning the internal auction, and it pays q to the publisher.
- Otherwise, $\mathcal{S}_1$ does not display any ad and gets no revenue in the auction.

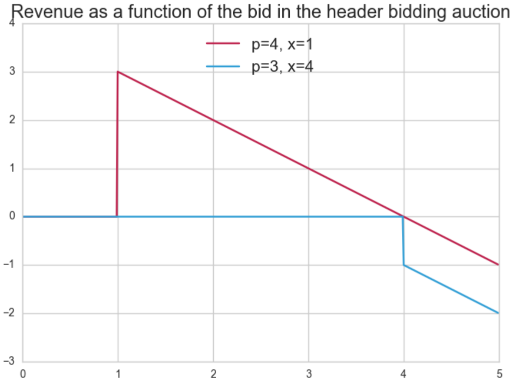

Figure 2: $\mathcal{S}_1$ revenue as a function of its bid q

In Figure 2 we plot $\mathcal{S}_1$'s revenue as a function of its bid q, for two sets of values for the closing price p of the internal auction and the highest bid x among the other SSPs.

Note that in the setting described here, we ignore some factors having an impact on $\mathcal{S}_1$'s revenue. Indeed $\mathcal{S}_1$ may charge a fee to the advertiser in addition to the closing price of the internal auction it runs. Also, the cost of running the internal auction may lower $\mathcal{S}_1$'s revenue. These factors would impact the revenue function but the strategy described in this paper would remain applicable.

3.4 Optimization problem statement

Before the header bidding auction for impression i happens, the value x_i of the highest bid among the other SSPs is unknown and is modeled with the random variable X_i. Thus, the revenue optimization problem over n impressions can be expressed as follows:

$$\begin{aligned} &\max_{q_1,\ldots,q_n} \mathbf{E}\left[\sum_{i=1}^{n} \mathbf{1}_{q_i \geq X_i}(p_i - q_i)\right] \\ &= \max_{q_1,\ldots,q_n} \sum_{c=1}^{C} \sum_{i \in \mathcal{D}_{t_n,c}} \mathbf{E}[\mathbf{1}_{q_i \geq X_i}(p_i - q_i)]\,, \end{aligned} \tag{1}$$

where the maximum is over the prices $q_1, \ldots, q_n$ that the SSP can choose as a function of the past observations.

It would be tempting to model the variables X_i as independent and identically distributed within any context c, with an unknown distribution ϕ_c. Under this assumption, the task presented here boils down to a contextual stochastic bandit problem. A closer look at the data, however, shows that there are significant non-stationarities in time. We explain below that our final model does address this issue, by the use of a particle filter within the bandit algorithm.

We emphasize that, after the header bidding auction for impression i has occurred, $\mathcal{S}_1$ does not observe the value of the bid x_i, but only observes if it has won or lost the header bidding auction, i.e., $\mathbf{1}_{q_i \geq x_i}$. This censorship issue must be tackled in the optimization methodology.

4 REVENUE OPTIMIZATION USING THOMPSON SAMPLING

In this section we present a method to sequentially optimize the bids q_i. It combines the Thompson sampling algorithm with a parametric model for the distributions ϕ_c (recall that ϕ_c is the distribution of the other SSPs' highest bids X_i within context c). Note that all the contexts c are modeled independently.

4.1 Parametric estimation of ϕ_c

We introduce f_θ a family of distributions parametrized with θ, and we note F_θ the corresponding cumulative density functions. For each context c, we assume that the distribution ϕ_c of the other SSPs' highest bids X_i belongs to the family f_θ; let θ_c be such that $\phi_c = f_{\theta_c}$.

According to the Thompson sampling method, we fix a prior distribution $\pi_{c,0}(\theta)$ over θ_c. Then, for all t, we consider the posterior distribution $\pi_{c,t}(\theta)$ given all the observations available at the end of the t-th auction, i.e., the censored observations $\mathbf{1}_{q_i \ge x_i}$ for $i \in \mathcal{D}_{t,c}$.

The Bayes rule yields the following expression for $\pi_{c,t}$:

$$\pi_{c,t}(\theta) \propto \pi_{c,0}(\theta) \prod_{i \in \mathcal{D}_{t,c}} \left[F_\theta(q_i)\mathbf{1}_{x_i \le q_i} + (1 - F_\theta(q_i))\mathbf{1}_{x_i > q_i}\right] \quad (2)$$

4.2 Overview of the methodology

In our model the Thompson sampling algorithm unfolds as follows: before any impression i,

- Sample a value θ from the posterior distribution $\pi_{c_i, t_{i-1}}$;
- Compute the bid q_i that would maximize the SSP's expected revenue if $X_i \sim f_\theta$ (see below);
- Observe the auction outcome $\mathbf{1}_{q_i \ge x_i}$ and update the posterior π_{c_i, t_i}.

As the particle filter provides a discrete approximation of the posterior distribution, the sampling step is straightforward. The optimization of the bid q_i is a one-dimensional optimization problem: when $X_i \sim f_\theta$, the maximal SSP expected revenue is

$$\max_q (p_i - q)F_\theta(q) \ .$$

There is no closed form solution in general, but this problem can be solved numerically for example by using Newton's method.

The difficult step of the algorithm is the update of the posterior, which is explained in the next section.

4.3 Updating the posterior distribution

It would be difficult to sample directly from the posterior distribution $\pi_{c,t}$, which does not have a simple or tractable form. Even the use of MCMC methods like Metropolis-Hastings would be hazardous, since computing the density of the posterior distribution has a linear cost in the number of past observations which is huge in advertising [1].

To overcome these difficulties, we approximate the posterior distribution with a *particle filter*, a powerful sequential Monte-Carlo method for Hidden Markov Models (HMM). For an introduction on HMM and particle filtering, we refer to [4]. The basic idea of a particle filter is to approximate the sequence of posterior distributions by a sequence of discrete probability distributions which are derived from one another by an *evolution* procedure (which may include a *selection* step). The posterior distribution $\pi_{c,t}$ is estimated by a discrete distribution on K points called *particles*. The particles are denoted by $(\theta_{c,1,t}, \dots, \theta_{c,K,t})$ and their respective weights by $(w_{c,1,t}, \dots, w_{c,K,t})$. The evolution procedure and the selection step we use are described below.

A very important strength of the particle filter approach is that it allows to handle *non-stationarity*: the HMM model encompasses the possibility that the hidden variable (here, the unknown parameter θ) evolves in time according to a Markovian dynamic of kernel $p(\theta'|\theta)$, thus forming an unobserved sequence $(\theta_t)_t$. We use this possibility by assuming that the parameter θ_t is equal to θ_{t-1} plus a small step in an unknown direction: this permits to handle parameter drift directly inside of the model.

The theory of particle filters for general state space HMM [8, 9, 11, 12] suggests that, in cases such as ours, the particle approximations converge to the true posterior distributions of the parameter θ when the number of particles tends to infinity.

4.3.1 Evolution: updating the distribution. Recall that we run C independent instances of Thompson Sampling, one for each context c. Next we focus on one context c and recall how to update the particle distribution in the particle filter. To simplify the notation, we write $t-1$ and t for the times of two consecutive impressions within context c, even if other contexts appeared in between.

The update consists of two steps. First the particles $\theta_{c,k,t}$ are sampled from a proposal distribution $q(\theta_{c,k,t}|\theta_{c,k,t-1}, \mathbf{1}_{q_t \ge x_t})$. We then compute new unnormalized weights $\hat{w}_{c,k,t}$ by importance sampling:

$$\hat{w}_{c,k,t} = \hat{w}_{c,k,t-1} \times \left[F_{\theta_{c,k,t}}(q_t)\mathbf{1}_{x_t \le q_t} + (1 - F_{\theta_{c,k,t}}(q_t))\mathbf{1}_{x_t > q_t}\right] \times \frac{p(\theta_{c,k,t}|\theta_{c,k,t-1})}{q(\theta_{c,k,t}|\theta_{c,k,t-1}, \mathbf{1}_{q_t \ge x_t})} \ , \quad (3)$$

where $p(\theta'|\theta)$ is the transition kernel of the hidden process. Here, we may simply take the proposal distribution $q(\theta_{c,k,t}|\theta_{c,k,t-1}, \mathbf{1}_{q_t \ge x_t})$ to be equal to the transition distribution $p(\theta_{c,k,t}|\theta_{c,k,t-1})$, which yields:

$$\hat{w}_{c,k,t} = \hat{w}_{c,k,t-1} \times \left[F_{\theta_{c,k,t}}(q_t)\mathbf{1}_{x_t \le q_t} + (1 - F_{\theta_{c,k,t}}(q_t))\mathbf{1}_{x_t > q_t}\right] . \quad (4)$$

The normalized weights $w_{c,k,t}$ can be computed as:

$$w_{c,k,t} = \frac{\hat{w}_{c,k,t}}{\sum_{k'=1}^{K} \hat{w}_{c,k',t}} \ .$$

4.3.2 Selection: resampling step. The basic update described previously generally fails after a few steps because of a well-known and general problem: weight degeneracy. Indeed, most of the particles soon get a negligible probability, and the discrete approximation becomes very poor. A standard strategy used to tackle this issue is the use of a resampling step when the degree of degeneracy is considered to be too high. We use the following methodology given in [26] for resampling:

[1] Indeed, the profile of the payoff function induces a posterior distribution that cannot be simplified. Hence, computing the posterior density exactly, cannot be done better than by computing the product of all bayesian updates, which in practice is intractable and rules-out MCMC sampling.

- Compute $S = \left(\sum_{k=1}^{K} w_{c,k,t}^{2}\right)^{-1}$ to quantify the degree of degeneracy of the particle filter
- If $S < S_{\min}$ ($S_{\min}$ is a hyperparameter of the particle filter), resample all the particles by sampling K times with replacement the current set of weighted particles $\left\{\theta_{c,1,t}, \ldots, \theta_{c,K,t}\right\}$. The result is an unweighted sample of K particles, so we set the new weights to $\hat{w}_{c,k,t} = \frac{1}{K}$.

There exist some alternative resampling schemes that could be used: see [11] for a presentation of some of them, and for a discussion on their convergence properties and computation cost.

Time and space complexities. Recall that K is the number of particles and that C is the number of contexts. After each new impression i, since i only falls into one context c, the evolution and selection steps described above need only be carried out for this particular c. This thus only requires $O(K)$ elementary operations per impression (including calls to the cumulative distribution function $F_\theta(x)$). As for space complexity, a direct upper bound is $O(CK)$ since we need to store weight vectors for each context $c = 1, \ldots, C$.

4.4 Implementation of Thompson sampling

We may now detail our modelling and algorithmic choices for the particle filter within the Thompson sampling algorithm.

Distribution of the highest bids x_i. We model the highest bids x_i among the other SSPs with a lognormal distribution, a standard choice in econometrics or finance. Lognormal distributions are parametrized by $\theta = \left(\theta^{(1)}, \theta^{(2)}\right)$, where $\theta^{(1)} = \sigma > 0$ and $\theta^{(2)} = \mu \in \mathbb{R}$. Here, μ and σ are respectively location and scale parameters for the normally distributed logarithm $\ln(x_i)$.

Particle filter. We write θ_c for the parameters of the lognormal distribution associated with context c. The particle filter for the posterior distributions works as follows:

1. In order to handle non-stationarity, we model the parameters θ_c by Markov chains $\theta_{c,t}$ such that $\log\left(\theta_{c,t}^{(1)}\right) = \log\left(\theta_{c,t-1}^{(1)}\right) + E_1$ and $\theta_{c,t}^{(2)} = \theta_{c,t-1}^{(2)} + E_2$, where $E_1, E_2 \sim \mathcal{N}(0, \epsilon = 0.005)$ are independent Gaussian variables with mean 0 and standard deviation ϵ.
2. At each time t, for each context $c \in \{1, \ldots, C = 100\}$, we use $K = 100$ particles $\theta_{c,k,t} = \left(\theta_{c,k,t}^{(1)}, \theta_{c,k,t}^{(2)}\right)$.
3. As explained above, the particles $\theta_{c,k}$ evolve at step t according to the same dynamic as the unobserved parameters $\theta_{c,t}$: $\log\left(\theta_{c,k,t}^{(1)}\right) = \log\left(\theta_{c,k,t-1}^{(1)}\right) + \mathcal{N}(0, \epsilon = 0.005)$ and $\theta_{c,k,t}^{(2)} = \theta_{c,k,t-1}^{(2)} + \mathcal{N}(0, \epsilon = 0.005)$.
4. We use a uniform distribution as prior $\pi_{c,k,0}(\theta)$ for the parameter $\theta_{c,0}$, and thus uniformly generate the components of the initial particle $\theta_{c,k,0} = \left(\theta_{c,k,0}^{(1)}, \theta_{c,k,0}^{(2)}\right)$. Because of the high number of auctions in each context, the choice of the prior distribution $\pi_{c,k,0}(\theta)$ has little impact on the result, as long as its support contains the parameter θ_c.
5. Finally, we choose $S_{min} = K/2$ as a resampling threshold criterion.

Table 1: Some properties of the datasets P_1 and P_2.

	P_1	P_2
Number of auctions	1,496,294	410,840
Number of users	875,634	269,272
Number of ad placements	3,526	31
Share of auctions where $x_i \leq p_i$	55.2%	48.4%

5 EXPERIMENTS ON RTB AUCTIONS DATASETS

5.1 Constrution of the datasets

In practice, the SSPs generally do not share their bids with one another, and we do not have a dataset with the bids from all SSPs in header bidding auctions. The datasets we have used in these experiments give, for two web publishers, the bids as well as the names of the advertisers in RTB auctions run by a particular SSP over one week, in a setting without header bidding.

For these two web publishers, a dataset giving both the bids in $\mathcal{S}_1$ internal auction and the bids from other SSPs in the header bidding auction has been artificially built the following way:

- All the advertisers competing in the RTB auctions (typically a few dozens) have been randomly assigned to one of two groups of advertisers named A and B
- In each auction, the bids coming from advertisers in the group A are supposed to be the bids of the internal auction run by the SSP $\mathcal{S}_1$, and the bids coming from advertisers in the group B are supposed to be the bids coming from the other SSPs in the header bidding auction
- Hence, in a given auction i, the closing price of the internal auction p_i is given by the second highest bid from advertisers in the group A, and the highest bid among other SSPs x_i is given by the highest bid from advertisers in the group B
- The auctions where there are less than two bids from advertisers in the group A or less than one bid from advertisers in the group B have been removed from the dataset

These two datasets are named P_1 and P_2 thereafter. A brief description is given in Table 1. We give the share of auctions where $x_i \leq p_i$, which is the share of auctions where the SSP $\mathcal{S}_1$ could have won the header bidding auction while generating a positive revenue, by choosing $q_i \in [x_i, p_i]$.

The experiments have been performed in the two following configurations:

- Stationary environment: the data is shuffled. This configuration is used to evaluate the strategy in a stationary environment
- Non-stationary environment: the data is sorted in chronological order. In this case, the data is non-stationary, as the bids highly depend on the time of the day. This configuration is used to evaluate the strategy in a non-stationary environment

Note that all the bids have been multiplied by a constant.

5.2 Definition of the contexts

In the experiments, we define the context in auction i by the closing price of $\mathcal{S}_1$ internal auction p_i. The closing price p_i is transformed into a categorical context by discretizing it into C disjoint bins.

The l-th bin contains all auctions where $p_i \in \left[q(\frac{l-1}{C}), q(\frac{l}{C})\right[$, where q is the empirical quantile function of the closing prices (p_i) estimated on the data. Consequently, each one of the C contexts contains approximately the same number of auctions.

The number of contexts should be chosen carefully. A high number of contexts enables to model more precisely the distribution of the highest bid among other SSPs, which is modeled independently on each context, at the price of a slower convergence. In the experiments, we have chosen $C = 100$ which yields a good performance on the datasets.

5.3 Baseline strategies

We define in this section the baseline strategies used to assess the quality of the Thompson sampling strategy. They correspond to the use of classical multi-armed bandit (MAB) models [6]. Each arm $j = 1, \ldots, J$ corresponds to a coefficient $\alpha^{(j)} = \frac{j}{J}$ applied to the closing price of the internal auction p_i to obtain the bid of the SSP $\mathcal{S}_1$, $q_i = \alpha^{(j)} \cdot p_i$. Note that this strategy implies that $q_i \leq p_i$, as the revenue for the SSP $\mathcal{S}_1$ can not be positive when $q_i > p_i$.

In each auction i, the SSP $\mathcal{S}_1$ chooses an arm $j(i)$ and bids $q_i = \alpha^{j(i)} \cdot p_i$. Then, it receives a reward equal to $\mathbf{1}_{q_i \geq x_i}(p_i - q_i)$, and the rewards associated to the other arms are unknown.

The goal of the SSP is to maximise their expected cumulative reward. In the MAB literature, this reward maximisation is typically defined via the minimisation of the equivalent measure of cumulative regret. The regret is the difference between the cumulative rewards of the SSP $\mathcal{S}_1$ and the one that could be acquired by a policy assumed to be optimal. In our case, the optimal policy (or the oracle strategy) consists in playing for each auction i the price $q_i = x_i \, \mathbf{1}_{\{x_i \leq p_i\}}$.

We consider two baseline strategies, corresponding to two distinct state-of-the-art policies:

- the Upper Confidence Bound (UCB) policy [1, 3]. Under the assumption that the rewards of each arm are independent, identically distributed, and bounded, the UCB policy achieves an order-optimal upper bound on the cumulative regret;
- the Exponential-weight algorithm for Exploration and Exploitation (Exp3) policy [2, 3]. Without any assumption on the possibly non-stationary sequence of rewards (except for boundedness), the Exp3 policy achieves a worst-case order-optimal upper bound on the cumulative regret.

The number of arms J has a high impact on the performance of these baseline strategies. A high number of arms makes the discretization of the coefficient applied to the bid p_i very precise, but slows the convergence as the average reward for each arm is learnt independently. We have used $J = 100$ in the experiments.

5.4 Evaluation of the Thompson sampling strategy

This section compares the performance of the Thompson sampling strategy (TS) defined in Sections 4 and 4.4 with the performance of the two baseline strategies (UCB and Exp3) introduced in Section 5.3 on the datasets P_1 and P_2.

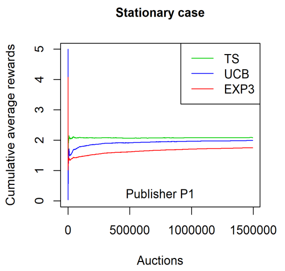

Figure 3: Evolution of the average rewards of TS, UCB, and Exp3 for dataset P1 (stationary environment).

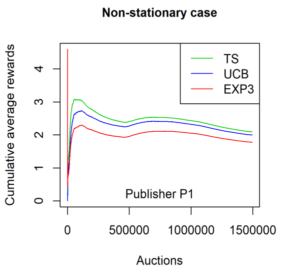

Figure 4: Evolution of the average rewards of TS, UCB, and Exp3 for dataset P1 (non-stationary environment).

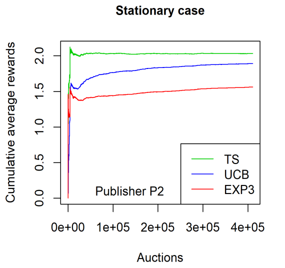

Figure 5: Evolution of the average rewards of TS, UCB, and Exp3 for dataset P2 (stationary environment).

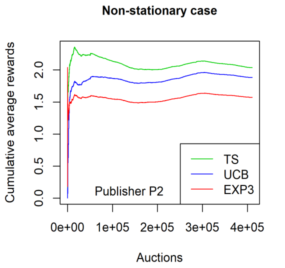

Figure 6: Evolution of the average rewards of TS, UCB, and Exp3 for dataset P2 (non-stationary environment)

The performance of a strategy after n auctions is measured with the average reward:

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{q_i \geq x_i}(p_i - q_i) \, .$$

Figures 3-4 plot the average reward as a function of n on dataset P_1, in a stationary environment (i.e. on the shuffled dataset) and in a non-stationary environment (i.e. on the ordered dataset). Figures 5-6 plot the same results on dataset P_2.

The TS strategy clearly outperforms the baseline strategies EXP3 and UCB in both stationary and non-stationary environments. Moreover, one can observe that the convergence of the TS strategy is faster than that of the EXP3 and the UCB strategies. This convergence speed is expressed in terms of the smallest number of auctions needed by the strategy to reach the overall average reward on the whole dataset.

On the dataset P_1, the average reward with TS strategy is 2.0888 for the stationary case and 2.0937 for the non-stationary case. The corresponding success rates (i.e. the share of auctions won $n^{-1} \sum_{i=1}^{n} \mathbf{1}_{q_i \geq x_i}$) are 32.19% and 32.09% respectively.

On the dataset P_2, the average reward with TS strategy is 2.0312 for the stationary case and 2.0342 for the non-stationary case. The corresponding success rates are 29.36% and 29.10% respectively.

5.5 Advantages of the Thompson sampling strategy

The main advantage of the Thompson sampling strategy introduced in this paper is that it relies on a random modeling of the highest bid among other SSPs x_i, which is the unknown variable. Then, the revenue function of the problem is introduced explicitly to determine an optimal bid in each auction. In the strategies EXP3 and UCB, the rewards corresponding to each arm are learnt independently whereas they are highly correlated because they derive from a common revenue function.

In addition, as argued above, the use of a particle filter within Thompson sampling permits to handle elegantly a parameter drift, a problem which is still under investigation for classical bandit algorithms. We ran experiments using the non-stationary bandits algorithms of [16], but the results were not better than plain UCB strategies. In contrast, the algorithm proposed above significantly outperforms the classical approaches.

The price for this improvement is an increased computational cost (proportional to the number of particles), and the presence of an additional parameter ϵ which controls the intensity of the drift. It must be chosen so as to reach a good tradeoff between accuracy of the discrete approximation and the adaptation to the parameter drift. Experiments show, however, that even a very rough choice of ϵ does lead to good performance, and that over-estimating the drift intensity has little impact.

5.6 Discussion on the parameters of the Thompson sampling strategy

5.6.1 Choice of the contexts. As precised in Section 5.2, the number of contexts has a high impact on the performance of the strategy and should be chosen carefully.

In the experiments presented in this paper, we have defined the context as the closing price of the internal auction run by the SSP $\mathcal{S}_1$. This definition of the context is intuitively a good choice, as the result of the internal auction measures the value of the ad space being sold according to the advertisers bidding in this auction. This value is probably highly correlated with the bids of other SSPs for this ad space.

The definition of the contexts could be improved by using characteristics of the ad placement or of the internet user. Experiments show that defining the context as the ad placement does not improve the results.

5.6.2 Choice of the parametric distribution. We chose the log-normal distribution to model the highest bid x_i among the other SSPs both because it is frequently used in practice for online auctions and it fitted our datasets reasonably well. However, when the number of other SSPs is sufficiently large, using the generalized extreme value distributions or the generalized Pareto distributions [7, 13] might be more relevant.

Some preliminary studies we conducted show that Fréchet distributions fit well the sample maxima of the bids x_i within each context. The reason is that such probability distributions are stable and relevant to model and to track the extreme values (sample maxima or peaks over threshold) of independent and identically distributed random variables, whatever the behavior of their tail distributions. In such situations, the associated Thompson sampling strategy could yield even higher cumulative revenues.

5.6.3 Computation time. We have measured the running time (the CPU response time) of the TS strategy using standard computer (μP 2.8GHz, RAM 8GB). Updating the full distribution model and estimating the optimal price q_i for an auction i requires about 0.14ms. This running time is below the limit of 1ms at which the optimal price must be decided.

Note that the running time is strongly related to the parametric probability distribution modeling the highest bid among other SSPs x_i and to the number of particles K used to approximate the corresponding posterior distributions.

6 CONCLUSION AND FUTURE WORK

We have formalized the problem of optimizing the sequence of bids of a given SSP as a contextual stochastic bandit problem. This problem is tackled using the Thompson sampling algorithm, which relies on a bayesian parametric estimation of the distribution of the highest bid among other SSPs. The distribution of the highest bid among other SSPs is approximated with a particle filtering approach. It provides a very efficient way to sequentially update the distribution and sample from it to apply the Thompson sampling algorithm.

The results obtained on datasets artificially built from real RTB auctions show that the Thompson sampling strategy outperforms other bandit approaches for this problem. Also, the estimation of the optimal bid for each impression is fast enough and the strategy can be used in real conditions where a bid prediction must be performed in a few milliseconds. This strategy is currently being developed to be deployed on thousands of web publishers worldwide.

The particle filtering models naturally the non-stationarity of the bid distributions through the hypothesis $p(\theta_{c,k,t}|\theta_{c,k,t-1})$. This hypothesis should be linked to the non-stationarity of the distributions, as decreasing its standard deviation (named ϵ in the paper) enables to forget past observations faster.

In the approach described here, the contexts are modeled independently. The learning speed of the algorithm could be increased by taking into account the correlations between the contexts. In particular, these correlations may be very high when the context is defined by a continuous variable. This point may lead to improvements in the strategy.

Finally, we are planning to explore further how the performance of the strategy depends on the parametric distribution used to model the highest bid among other SSPs x_i.

Aknowledgment: This work was partially funded by the French Government under the grant <ANR-13-CORD-0020> (ALICIA Project).

REFERENCES

[1] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.* 47, 2-3 (May 2002), 235–256. https://doi.org/10.1023/A:1013689704352

[2] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. 2002. The Nonstochastic Multiarmed Bandit Problem. *SIAM J. Comput.* 32, 1 (2002), 48–77. https://doi.org/10.1137/S0097539701398375 arXiv:https://doi.org/10.1137/S0097539701398375

[3] Sébastien Bubeck and Nicolò Cesa-Bianchi. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *CoRR* abs/1204.5721 (2012). http://arxiv.org/abs/1204.5721

[4] Olivier Cappé, Eric Moulines, and Tobias Ryden. 2005. *Inference in Hidden Markov Models (Springer Series in Statistics).* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[5] N. Cesa-Bianchi, C. Gentile, and Y. Mansour. 2015. Regret minimization for reserve prices in second-price auctions. *IEEE Trans. Inform. Theory* 61, 1 (2015), 549–564.

[6] Nicolo Cesa-Bianchi and Gabor Lugosi. 2006. *Prediction, Learning, and Games.* Cambridge University Press, New York, NY, USA.

[7] S. G. Coles. 2001. *An introduction to statistical modeling of extreme values.* Springer Series in Statistics.

[8] Dan Crisan and Arnaud Doucet. 2000. *Convergence of Sequential Monte Carlo Methods.* Technical Report.

[9] Dan Crisan and Arnaud Doucet. 2002. A Survey of Convergence Results on Particle Filtering Methods for Practitioners. (2002).

[10] Xiaotie Deng, Paul W. Goldberg, Bo Tang, and Jinshan Zhang. 2014. Revenue maximization in a Bayesian double auction market. *Theor. Comput. Sci.* 539 (2014), 1–12. https://doi.org/10.1016/j.tcs.2014.04.013

[11] Randal Douc, AurÃĺlien Garivier, Eric Moulines, and Jimmy Olsson. 2011. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* 21, 6 (12 2011), 2109–2145. https://doi.org/10.1214/10-AAP735

[12] Arnaud Doucet, Nando de Freitas, and Neil Gordon (Eds.). 2001. *Sequential Monte Carlo Methods in Practice.* Springer.

[13] P. Embrechts, C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

[14] Jon Feldman, Vahab Mirrokni, S. Muthukrishnan, and Mallesh M. Pai. 2010. Auctions with Intermediaries: Extended Abstract. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC '10).* ACM, New York, NY, USA, 23–32. https://doi.org/10.1145/1807342.1807346

[15] J. Fernandez-Tapia, O. Guéant, and J.-M. Lasry. 2015. Optimal Real-Time Bidding Strategies. *ArXiv e-prints* (Nov. 2015). arXiv:math.OC/1511.08409

[16] Aurélien Garivier and Eric Moulines. 2011. *On Upper-Confidence Bound Policies for Switching Bandit Problems.* Springer Berlin Heidelberg, Berlin, Heidelberg, 174–188. https://doi.org/10.1007/978-3-642-24412-4_16

[17] Renato Gomes and Vahab S. Mirrokni. 2014. Optimal revenue-sharing double auctions with applications to ad exchanges. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014.* 19–28. https://doi.org/10.1145/2566486.2568029

[18] H. Heidari, M. Mahdian, U. Syed, S. Vassilvitskii, and S. Yazdanbod. 2016. Pricing a low-regret seller, In Proceedings of The 33rd International Conference on Machine Learning (ICML 2016). *Proceedings of The 33rd International Conference on Machine Learning (ICML'16)*, 2559–2567.

[19] E. Kaufmann, N. Korda, and R. Munos. 2012. Thompson Sampling: An Asymptotically Optimal Finite Time Analysis. *ArXiv e-prints* (May 2012). arXiv:stat.ML/1205.4217

[20] R. Kleinberg and T. Leighton. 2003. The value of knowing a demand curve: bounds on regret for on-line posted-price auctions. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS 2003).* 594–605. Full version avalaible at https://www.cs.cornell.edu/ rdk/papers/oppa.pdf.

[21] N. Korda, E. Kaufmann, and R. Munos. 2013. Thompson Sampling for 1-Dimensional Exponential Family Bandits. *ArXiv e-prints* (July 2013). arXiv:stat.ML/1307.3400

[22] K.-C. Lee, A. Jalali, and A. Dasdan. 2013. Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising. *ArXiv e-prints* (May 2013). arXiv:cs.GT/1305.3011

[23] J. Leike, T. Lattimore, L. Orseau, and M. Hutter. 2016. Thompson Sampling is Asymptotically Optimal in General Environments. *ArXiv e-prints* (Feb. 2016). arXiv:cs.LG/1602.07905

[24] M. Mohri and A. M. Medina. 2014. Optimal regret minimization in posted-price auctions with strategic buyers. In *Advances in Neural Information Processing Systems (NIPS'14).*

[25] M. Mohri and A. M. Medina. 2015. Revenue optimization against strategic buyers. In *Advances in Neural Information Processing Systems (NIPS'15).*

[26] Kevin P. Murphy. 2012. *Machine Learning: A Probabilistic Perspective* (6th ed.). Cambridge, MA: MIT Press.

[27] Roger B. Myerson. 1981. Optimal Auction Design. *Math. Oper. Res.* 6, 1 (Feb. 1981), 58–73. https://doi.org/10.1287/moor.6.1.58

[28] Roger B Myerson and Mark A Satterthwaite. 1983. Efficient mechanisms for bilateral trading. *Journal of Economic Theory* 29, 2 (1983), 265 – 281. https://doi.org/10.1016/0022-0531(83)90048-0

[29] Rui Qin, Yong Yuan, and Fei-Yue Wang. 2017. Optimizing the Revenue for Ad Exchanges in Header Bidding Advertising Markets. In *IEEE International Conference on Systems, Man, and Cybernetics.*

[30] W. R. Thompson. 1933. On the Likelihood that one Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* 25 (1933), 285–294.

[31] Ratko Vidakovic. 2017. The Beginner's Guide To Header Bidding. (March 2017).

[32] Jun Wang, Weinan Zhang, and Shuai Yuan. 2017. Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting. *Foundations and Trends in Information Retrieval* 11, 4-5 (2017), 297–435. https://doi.org/10.1561/1500000049

[33] J. Weed, V. Perchet, and P. Rigollet. 2016. Online learning in repeated auctions. In *Proceedings of the 29th Annual Conference on Learning Theory (COLT'16).*

[34] Weinan Zhang, Shuai Yuan, and Jun Wang. 2014. Optimal Real-time Bidding for Display Advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14).* ACM, New York, NY, USA, 1077–1086. https://doi.org/10.1145/2623330.2623633