Real-Time Bidding Strategies with Online Learning*

Joaquin Fernandez-Tapia†, Olivier Guéant‡, Jean-Michel Lasry§

Abstract

One critical issue in the control of Markov processes is that, in order to successfully apply dynamic programming tools, the knowledge of the statistical laws governing the system is required. Sometimes, when these laws are difficult to estimate beforehand using historical data, the estimation/calibration task needs to be performed at runtime (this is known as online learning). Bayesian inference provides a useful way to address this problem by defining probability distributions for the model parameters and update them with the incoming information. It is particularly relevant in the case of most conjugate Bayesian priors as their use preserves the Markovian properties of the model, making it possible therefore to apply classical dynamic programming / stochastic optimal control tools. In this paper, we apply such a Bayesian approach for the control of a bidding algorithm participating in a high-frequency stream of (Vickrey) auctions. This is of particular interest in real-time bidding (RTB) advertising.

Keywords: Stochastic optimal control, Online learning, Exploration-Exploitation, Real-Time Bidding (RTB), Hamilton-Jacobi-Bellman (HJB) equations, Bayesian inference.

1 Introduction

The theory of Markov decision processes (MDPs) – and/or its continuous-time counterpart: stochastic optimal control theory – provides an appealing framework for the design of algorithms (i) acting on a system through a sequence of decisions whose immediate outcome is subject to randomness, and (ii) aiming at optimizing an expected payoff (depending on future states of the system). In situations where the randomness governing the system is described by statistical laws which are known to the controller, it is often possible to exactly characterize the optimal strategy by applying tools such as the dynamic programming principle (Bellman equations). One

*The authors would like to thank Dominique Delport (Havas Media), Julien Laugel (MFG Labs), Pierre-Louis Lions (Collège de France), Arnaud Parent (Havas Media), and Jiang Pu (Université Paris Diderot and Institut Louis Bachelier) for the discussions they had on the topic.

†Independent researcher. Joaquin holds a PhD in Applied Mathematics from Université Pierre et Marie Curie.

‡Full Professor at Université Paris 1 Panthéon-Sorbonne. Centre d'Economie de la Sorbonne. 106, Bd de l'Hôpital, 75013 France. Member of the Scientific Advisory Board of Havas Media. The content of this article does not reflect the official opinion or the practices of Havas Media. Corresponding author: olivier.gueant@univ-paris1.fr.

§Institut Louis Bachelier. Member of the Scientific Advisory Board of Havas Media. The content of this article does not reflect the official opinion or the practices of Havas Media.

of the main limitations of the aforementioned framework is the need to model (and estimate thereafter) the randomness. If the randomness is not exactly modeled or if the model parameters are difficult/impossible to estimate, the Bellman equation is not well defined and, from a computational perspective, the problem does not boil down, *a priori*, to the numerical approximation of the solution of a recursive system of equations (the Bellman equation) by backward induction.

Many approaches have been proposed in order to address optimization and control problems in an uncertain environment. Robust control theory is an important example. Reinforcement learning is another. When the algorithm can be played on similar data sets, reinforcement learning techniques such as Q-learning or SARSA can be used to learn the optimal policy. However, if the dataset of interest can only be played once, the problem is an exploration-exploitation one which requires to learn the environment at runtime (*i.e.*, online learning). When the optimal strategy does not depend on the state of the algorithm, bandit algorithms can often be used. However, in the more general MDP case the situation is more complex and a different treatment is needed.

A real-life instance where this kind of problems arises is in the area of digital media buying, more exactly the design of optimal real-time bidding strategies. In a nutshell, an important part of the global digital advertising inventory is purchased nowadays through auctions: each time a web page (on which an ad slot is available) is loaded, an ad exchange proposes this ad slot to potential buyers – sometimes through Demand-Side Platforms (DSPs). Buyers then have a lapse of a few milliseconds to bid a price. Millions of such auction processes are launched every day to sell ad inventory. Therefore, media buying agencies and ad trading companies providing ad-buying services to firms have to solve a complex control problem for defining how to bid on each auction in order to optimize one or several key performance indicators (KPIs).[1]

Several digital media buying problems have been addressed in the academic literature – see bibliography. In this paper, we consider Bayesian extensions of the models presented in the companion papers [7, 8]. We consider a trading desk receiving auction requests from J different sources. These sources may correspond to different ad exchanges and/or different segments (usually cookie segments). For each auction request, the algorithm bids a price and the inventory is purchased by the algorithm if and only if the bid sent is greater than the price to beat (*i.e.*, the best price proposed by the other participants in the auction). If the inventory is purchased by the algorithm, then there may be or may not be a conversion – a conversion corresponds, depending on the context, either to a click, or to a purchase, or to a subscription, or to the internet user reaching a specific page. The goal of the ad trading desk is either to maximize the expected value of a KPI (*e.g.*, the number of conversions) for a given level of spending, or – it is the dual problem – to minimizing the expected amount of cash spent subject to the constraint of reaching

[1] KPIs are usually functions of macroscopic parameters such as the total budget, the inventory purchased, the number of conversions obtained (a conversion can be a click, a purchase, a subscription, the internet user reaching a specific web page, etc.).

a minimum level of a given KPI. However, the trading desk may not know all the parameters of the model. For instance, the intensity of arrival of requests from the different sources may not be known. Similarly, the distribution of the price to beat for each of the sources may not be known. More importantly if we consider a KPI linked to the number of conversions, the conversion rate of each of the J sources may be unknown. In spite of the numerous uncertainties, the Bayesian and control framework we propose enables to address the problem faced by the trading desk in a Markovian way.

In Section 2, we present the modelling framework without uncertainty. In Section 3, we consider one of the most important uncertainties: the value of the probability of conversion associated with each source. In Section 4, we address the learning problem associated with the distribution of the price to beat for each source. We assume that the price to beat for each source of auction requests follows an exponential distribution and we show that the use of conjugate priors (here Gamma priors) allows to write the learning and control problems we consider in the form of an HJB equation. In Section 5, we focus on the uncertainty associated with the number of auction requests coming from each source. In each section, we consider both the primal problem and the dual problem.

2 The basic modelling framework without learning

2.1 Setup and notations

Let us fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ satisfying the usual conditions. We assume that all stochastic processes are defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$.

We consider a time horizon $T > 0$ and an ad trader (or an ad trading desk) buying ad inventory over a time window $[0, T]$ for a given campaign.

The ad trader receives auction requests at random times from a set of J sources. The J types of auction requests can differ in the actual ad exchange sending the auction request or in the type of population targeted (through the use of cookie segments for instance). The auction requests arising from the J sources are modeled with J marked Poisson processes $(N^1, \dots, N^J)$. The arrival of a new auction request from the source $j \in \{1, \dots, J\}$ is triggered by the jump of the Poisson process N^j. The intensity associated with N^j is denoted by λ^j. As far as the marks are concerned, we denote by p_n^j and ξ_n^j, respectively the highest bid sent by the other participants during the n^{th} auction coming from source j (*i.e.*, the price to beat), and the occurrence ($\xi_n^j \in \{0, 1\}$) of a conversion – the latter matters only in the case where the auction has been won by the ad trader.

At time t, if an ad trader receives an auction request from the source j, then we denote his bid by b_t^j. If this request turns out to be the n^{th} auction associated with the source j, then the outcome of the auction is the following:

- If $b_t^j > p_n^j$, then the ad trader wins the auction: he pays the price p_n^j and his

ad is displayed. Moreover, a conversion occurs if and only if $\xi_n^j = 1$.

- If $b_t^j \leq p_n^j$, then another trader wins the auction.

We assume that for each $j \in \{1, \ldots, J\}$, $(p_n^j)_{n\in\mathbb{N}^*}$ are *i.i.d.* random variables distributed according to an exponential distribution with parameter μ^j. We also assume that the random variables $(p_n^j)_{j\in\{1,\ldots,J\},n\in\mathbb{N}^*}$ are all independent.

As far as the variables $(\xi_n^j)_{j\in\{1,\ldots,J\},n\in\mathbb{N}^*}$ are concerned, we assume that they are all independent and independent from the variables $(p_n^j)_{j\in\{1,\ldots,J\},n\in\mathbb{N}^*}$. Moreover, we assume that for each $j \in \{1, \ldots, J\}$, $(\xi_n^j)_{n\in\mathbb{N}^*}$ are *i.i.d.* random variables distributed according to a Bernoulli distribution with parameter $\nu^j \in [0, 1]$.

The amount of cash spent is denoted by $(S_t)_t$. Its dynamics is the following:

$$dS_t = \sum_{j=1}^{J} p^j_{N_t^j} \mathbf{1}_{\{b_t^j > p^j_{N_t^j}\}} dN_t^j, \quad S_0 = 0.$$

For each $j \in \{1, \ldots, J\}$, the number of impressions associated with the auction requests coming for the source j is modeled by an inventory process $(I_t^j)_t$. For each $j \in \{1, \ldots, J\}$, the dynamics of $(I_t^j)_t$ is:

$$dI_t^j = \mathbf{1}_{\{b_t^j > p^j_{N_t^j}\}} dN_t^j, \quad I_0^j = 0.$$

For each $j \in \{1, \ldots, J\}$, the number of conversions associated with the auction requests coming for the source j is modeled by a process $(C_t^j)_t$. For each $j \in \{1, \ldots, J\}$, the dynamics of $(C_t^j)_t$ is:

$$dC_t^j = \xi^j_{N_t^j} \mathbf{1}_{\{b_t^j > p^j_{N_t^j}\}} dN_t^j, \quad C_0^j = 0.$$

2.2 The primal problem: maximizing a KPI for a given level of spending

In our framework, developed for the first time in [7], the goal of the ad trader was to minimize an objective function of the form

$$\mathbb{E}\left[-\sum_{j=1}^{J} \zeta^j I_T^j - \sum_{j=1}^{J} \upsilon^j C_T^j + K \min\left(\bar{S} - S_T, 0\right)^2 \right],$$

over $(b_t^1, \ldots, b_t^J)_t \in \mathcal{A}^J$, where $\mathcal{A}$ is the set of predictable processes with values in $\mathbb{R}_+ \cup \{+\infty\}$.

We can in fact consider a more general problem of the form

$$\inf_{(b_t^1,\ldots,b_t^J)_t \in \mathcal{A}^J} \mathbb{E}\left[\Phi(I_T^1, \ldots, I_T^J, C_T^1, \ldots, C_T^J) + K \min\left(\bar{S} - S_T, 0\right)^2 \right].$$

This objective function corresponds to a relaxed form of the problem consisting in maximizing

$$E\left[-\Phi(I_T^1, \ldots, I_T^J, C_T^1, \ldots, C_T^J) \right]$$

over the strategies $(b_t^1, \dots, b_t^J)_t \in \mathcal{A}^J$ such that the total amount spent is $\bar{S}$ (*i.e.*, $S_T = \bar{S}$). The constant K in the relaxed problem is assumed to be large enough so that S_T never becomes too much greater than $\bar{S}$.

The state of the bidding algorithm at any given time t is described by the vector (I_t, C_t, S_t) which contains information about the number of ad slots already purchased, the current number of conversions, and the amount already spent. The problem is therefore characterized by a 4-variable (or in fact a function in dimension $2J+2$) value function

$$(t, I, C, S) \mapsto u(t, I, C, S)$$

and the Hamilton-Jacobi-Bellman equation[2]

$$\partial_t u(t,I,C,S) + \sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p} \left[(1-\nu^j)(u(t, I+e^j, C, S+p) - u(t,I,C,S)) \right.$$
$$\left. + \nu^j (u(t, I+e^j, C+e^j, S+p) - u(t,I,C,S)) \right] dp = 0, \tag{1}$$

with terminal condition

$$u(T, I^1, \dots, I^J, C^1, \dots, C^J, S) = \Phi(I^1, \dots, I^J, C^1, \dots, C^J) + K \min\left(\bar{S} - S, 0\right)^2.$$

Eq. (1) is a non-standard integro-differential HJB equation in dimension $2J+2$. In the case of a linear form Φ, we have shown in [7] that the problem boils down to a problem of dimension 2 through the use of the ansatz

$$u(t, I^1, \dots, I^J, C^1, \dots, C^J, S) = \Phi(I^1, \dots, I^J, C^1, \dots, C^J) + v(t, S),$$

where v satisfies another integro-differential Hamilton-Jacobi-like equation. In the case of a general function Φ, the ad trader faces, however, a high-dimensional problem.

The numerical problem we face in the case of the primal problem is a bit involved. We can see Eq. (1) as a system (indexed by I and C) of integro-differential equations. The solution can be approximated on a grid by backward induction, but it requires to have enough computation capacity (in space and speed). It is noteworthy that, in order to approximate numerically a solution of Eq. (1), it is interesting to modify the equation by carrying out a first order expansion in S. In that case indeed, the optimal bids can easily be written as functions of u and its gradient with respect to S. In that case, the system of integro-differential equations is replaced by a system of first-order PDEs.

2.3 The dual problem: minimizing the amount spent to reach KPI thresholds

We developed in [8] a dual framework in which the goal of the ad trader is to minimize an objective function of the form

$$\mathbb{E}\left[S_T + \Phi(I_T^1, \dots, I_T^J, C_T^1, \dots, C_T^J)\right],$$

[2] We denote by $(e^1, \dots, e^J)$ the canonical basis of $\mathbb{R}^J$.

over $(b_t^1, \ldots, b_t^J)_t \in \mathcal{A}^J$, where $\mathcal{A}$ is the set of predictable processes with values in $\mathbb{R}_+ \cup \{+\infty\}$, and where Φ is a penalty function, which is for instance of the form

$$\Phi(i^1, \ldots, i^J, c^1, \ldots, c^J) = \pi_I \cdot (i^1 + \ldots + i^J - \overline{I})_- + \pi_C \cdot (c^1 + \ldots + c^J - \overline{C})_-,$$

where $\overline{I}$ and $\overline{C}$ are target levels respectively for the number of ad slots already purchased and for the number of conversions.

This objective function corresponds to a relaxed form of the problem consisting in minimizing the average amount spent $E[S_T]$ over the bidding strategies $(b_t^1, \ldots, b_t^J)_t \in \mathcal{A}^J$, such that the total number of impressions (resp. conversions) at time T is above $\overline{I}$ (resp. $\overline{C}$).[3]

The state of the bidding algorithm at any given time t is described by the vector (I_t, C_t, S_t) which contains information about the number of ad slots already purchased, the current number of conversions, and the amount already spent. The problem is therefore characterized by a 4-variable (or in fact a function in dimension $2J + 2$) value function

$$(t, I, C, S) \mapsto u(t, I, C, S)$$

and the Hamilton-Jacobi-Bellman equation

$$\partial_t u(t, I, C, S) + \sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p} \left[(1 - \nu^j)(u(t, I + e^j, C, S + p) - u(t, I, C, S)) \right.$$

$$\left. + \nu^j (u(t, I + e^j, C + e^j, S + p) - u(t, I, C, S)) \right] dp = 0, \tag{2}$$

with terminal condition

$$u(T, I^1, \ldots, I^J, C^1, \ldots, C^J, S) = S + \Phi(I^1, \ldots, I^J, C^1, \ldots, C^J).$$

Eq. (2) is a non-standard integro-differential HJB equation in dimension $2J + 2$. It can be simplified by using the ansatz

$$u(t, I^1, \ldots, I^J, C^1, \ldots, C^J, S) = S + \theta(t, I^1, \ldots, I^J, C^1, \ldots, C^J).$$

We have indeed that Eq. (2) then boils down to:

$$\partial_t \theta(t, I, C) + \sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p} \left[p + (1 - \nu^j)(\theta(t, I + e^j, C) - \theta(t, I, C)) \right.$$

$$\left. + \nu^j (\theta(t, I + e^j, C + e^j) - \theta(t, I, C)) \right] dp = 0, \tag{3}$$

with terminal condition

$$\theta(T, I^1, \ldots, I^J, C^1, \ldots, C^J) = \Phi(I^1, \ldots, I^J, C^1, \ldots, C^J).$$

[3] The higher π_I and π_C, the harder the constraints on the KPI thresholds.

It is noteworthy that the optimal bids can be computed as functions of θ by using the first order condition in Eq. (3). We obtain indeed

$$\begin{aligned} & b^{j*}_{I,C}(t) \\ = \; & \left[-((1-\nu^j)(\theta(t,I+e^j,C)-\theta(t,I,C))+\nu^j(\theta(t,I+e^j,C+e^j)-\theta(t,I,C)))\right]_+ \\ = \; & \left[\theta(t,I,C)-(\nu^j\theta(t,I+e^j,C+e^j)+(1-\nu^j)\theta(t,I+e^j,C))\right]_+ . \end{aligned}$$

Moreover, by using an integration by parts, Eq. (3) can be written as

$$\partial_t\theta(t,I,C)-\sum_{j=1}^{J}\lambda^j\int_0^{b^{j*}_{I,C}(t)}(1-e^{-\mu^j p})dp=0,$$

or equivalently

$$\partial_t\theta(t,I,C)-\sum_{j=1}^{J}\lambda^j\left(b^{j*}_{I,C}(t)-\frac{1}{\mu^j}\left(1-e^{-\mu^j b^{j*}_{I,C}(t)}\right)\right)=0, \tag{4}$$

with terminal condition

$$\theta(T,I^1,\ldots,I^J,C^1,\ldots,C^J)=\Phi(I^1,\ldots,I^J,C^1,\ldots,C^J).$$

Eq. (4) is very interesting because it shows that, unlike what happens with the primal problem, solving the dual problem boils down to solving a large but very simple system of ODEs indexed by I and C. It requires computation capacity if the number of sources J is large, but there is no mathematical difficulty!

2.4 Towards Bayesian learning

In order to write, and subsequently in order to approximate numerically the solution of Eqs. (1) and (3) / (4), we need to know the value of $3J$ parameters. First, the conversion rates (or probabilities) associated with the J different sources of auction requests are characterized by a set of J parameters: $\nu^1,\ldots,\nu^J$. The exponential distributions of the price to beat associated with the J sources of inventory are also characterized by J parameters: $\mu^1,\ldots,\mu^J$. Finally, the J Poisson processes $N^1,\ldots,N^J$ are characterized by J intensities $\lambda^1,\ldots,\lambda^J$. These $3J$ parameters can be estimated using past campaigns, but most of the time it would be worth updating the estimates over the course of a campaign. For that purpose, we propose to mix Bayesian learning and optimal control.

3 Online learning of conversion rates

3.1 Bayesian learning

In the previous framework, the J different sources of auction requests are associated with J probabilities of conversion: $\nu^1,\ldots,\nu^J$. In practice, assuming that the value of these probabilities is known is a strong assumption and it is instead reasonable to

assume that we have a prior distribution on each of these J parameters.

In what follows we assume that for each $j \in \{1, \dots, J\}$, we have at time $t = 0$ a Beta prior distribution on ν^j, *i.e.*,

$$\nu^j \sim B(\alpha_0^j, \beta_0^j).$$

After the n^{th} auction from the source j, if the ad slot has been purchased by the ad trader, we know whether or not a conversion occurred ($\xi_n^j = 1$ or $\xi_n^j = 0$). This enables to update the prior distribution of ν^j through the use of the Bayes rule. If the n^{th} auction from the source j occurs at time t then, assuming that

$$\nu^j | \mathcal{F}_{t-} \sim B(\alpha_{t-}^j, \beta_{t-}^j),$$

we obtain the following:

- If a conversion took place ($\xi_n^j = 1$), then

$$\mathcal{L}(\nu^j | \mathcal{F}_t) \propto \nu^j \cdot {\nu^j}^{\alpha_{t-}^j - 1} (1 - \nu^j)^{\beta_{t-}^j - 1} \propto {\nu^j}^{\alpha_{t-}^j} (1 - \nu^j)^{\beta_{t-}^j - 1}.$$

- If the purchased ad slot did not yield to a conversion ($\xi_n^j = 1$), then

$$\mathcal{L}(\nu^j | \mathcal{F}_t) \propto (1 - \nu^j) \cdot {\nu^j}^{\alpha_{t-}^j - 1} (1 - \nu^j)^{\beta_{t-}^j - 1} \propto {\nu^j}^{\alpha_{t-}^j - 1} (1 - \nu^j)^{\beta_{t-}^j}.$$

In other words,

$$\nu^j | \mathcal{F}_t \sim B(\alpha_t^j, \beta_t^j),$$

where

$$d\alpha_t^j = \xi_{N_t^j}^j \mathbf{1}_{\{b_t^j > p_{N_t^j}^j\}} dN_t^j = dC_t^j$$

and

$$d\beta_t^j = (1 - \xi_{N_t^j}^j) \mathbf{1}_{\{b_t^j > p_{N_t^j}^j\}} dN_t^j = dI_t^j - dC_t^j.$$

We obtain therefore straightforwardly that

$$\nu^j | \mathcal{F}_t \sim B(\alpha_0^j + C_t^j, \beta_0^j + I_t^j - C_t^j).$$

In particular,

$$\mathbb{E}[\nu^j | \mathcal{F}_t] = \frac{\alpha_0^j + C_t^j}{\alpha_0^j + \beta_0^j + I_t^j}.$$

3.2 A new Hamilton-Jacobi-Belmann equation for the primal problem

In the above paragraph we derived the expected value of the ν^js conditionally on the information available at a given point in time. One could therefore replace at any time the ν^js by their expected value and recompute the optimal strategy – this may be called myopic learning. However, this would be a time inconsistent approach (it is used in many situations though), because we know the mechanism by which we

update the prior distributions of the ν^js.[4]

With the above Bayesian framework, the optimal control problem

$$\inf_{(b_t^1,\ldots,b_t^J)_t\in\mathcal{A}^J} \mathbb{E}\left[\Phi(I_T^1,\ldots,I_T^J,C_T^1,\ldots,C_T^J)+K\min\left(\bar{S}-S_T,0\right)^2\right].$$

is in fact characterized by the following HJB equation:

$$\begin{aligned}0 &= \partial_t u(t,I,C,S)\\ &+\sum_{j=1}^J \lambda^j \inf_{b^j\in\mathbb{R}_+}\int_0^{b^j}\mu^j e^{-\mu^j p}\left[\frac{\beta_0^j+I^j-C^j}{\alpha_0^j+\beta_0^j+I^j}(u(t,I+e^j,C,S+p)-u(t,I,C,S))\right.\\ &\left.+\frac{\alpha_0^j+C^j}{\alpha_0^j+\beta_0^j+I^j}(u(t,I+e^j,C+e^j,S+p)-u(t,I,C,S))\right]dp, \qquad (5)\end{aligned}$$

with terminal condition

$$u(T,I^1,\ldots,I^J,C^1,\ldots,C^J,S)=\Phi(I^1,\ldots,I^J,C^1,\ldots,C^J)+K\min\left(\bar{S}-S,0\right)^2.$$

Eq. (5) is a non-standard integro-differential HJB equation in dimension $2J+2$ which can be seen as a system (indexed by I and C) of integro-differential equations (or a system of first-order PDEs if we consider the same approximation as the one we proposed for Eq. (1)). For approximating numerically the solution of Eq. (5), the same methods as for Eq. (1) can be employed. In particular, taking account of the Bayesian learning on the conversion rates does not cost anything – in addition to the case of constant ν^js – in terms of computation requirements and computation time.

3.3 New equations for the dual problem

In the case of the dual problem, the optimal control problem

$$\inf_{(b_t^1,\ldots,b_t^J)_t\in\mathcal{A}^J} \mathbb{E}\left[S_T+\Phi(I_T^1,\ldots,I_T^J,C_T^1,\ldots,C_T^J)\right].$$

is in fact characterized by the following HJB equation:

$$\begin{aligned}0 &= \partial_t u(t,I,C,S)\\ &+\sum_{j=1}^J \lambda^j \inf_{b^j\in\mathbb{R}_+}\int_0^{b^j}\mu^j e^{-\mu^j p}\left[\frac{\beta_0^j+I^j-C^j}{\alpha_0^j+\beta_0^j+I^j}(u(t,I+e^j,C,S+p)-u(t,I,C,S))\right.\\ &\left.+\frac{\alpha_0^j+C^j}{\alpha_0^j+\beta_0^j+I^j}(u(t,I+e^j,C+e^j,S+p)-u(t,I,C,S))\right]dp, \qquad (6)\end{aligned}$$

with terminal condition

$$u(T,I^1,\ldots,I^J,C^1,\ldots,C^J,S)=S+\Phi(I^1,\ldots,I^J,C^1,\ldots,C^J).$$

[4] See [9] for a similar framework in finance developed by one of the authors.

As in the non-Bayesian case, this equation can be simplified by using the ansatz

$$u(t, I^1, \dots, I^J, C^1, \dots, C^J, S) = S + \theta(t, I^1, \dots, I^J, C^1, \dots, C^J).$$

The Bayesian counterpart of Eq. (3) is indeed:

$$\begin{aligned} 0 = \partial_t \theta(t, I, C) \\ + \sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p} \left[p + \frac{\beta_0^j + I^j - C^j}{\alpha_0^j + \beta_0^j + I^j} (\theta(t, I + e^j, C) - \theta(t, I, C)) \right. \\ \left. + \frac{\alpha_0^j + C^j}{\alpha_0^j + \beta_0^j + I^j} (\theta(t, I + e^j, C + e^j) - \theta(t, I, C)) \right] dp, \end{aligned} \tag{7}$$

with terminal condition

$$\theta(T, I^1, \dots, I^J, C^1, \dots, C^J) = \Phi(I^1, \dots, I^J, C^1, \dots, C^J).$$

The optimal bids can be computed as above by:

$$b^{j*}_{I,C}(t) = \left[\theta(t, I, C) - \left(\frac{\alpha_0^j + C^j}{\alpha_0^j + \beta_0^j + I^j} \theta(t, I + e^j, C + e^j) + \frac{\beta_0^j + I^j - C^j}{\alpha_0^j + \beta_0^j + I^j} \theta(t, I + e^j, C) \right) \right]_+ .$$

And we have

$$\partial_t \theta(t, I, C) - \sum_{j=1}^{J} \lambda^j \int_0^{b^{j*}_{I,C}(t)} (1 - e^{-\mu^j p}) dp = 0,$$

or equivalently

$$\partial_t \theta(t, I, C) - \sum_{j=1}^{J} \lambda^j \left(b^{j*}_{I,C}(t) - \frac{1}{\mu^j} \left(1 - e^{-\mu^j b^{j*}_{I,C}(t)} \right) \right) = 0, \tag{8}$$

with terminal condition

$$\theta(T, I^1, \dots, I^J, C^1, \dots, C^J) = \Phi(I^1, \dots, I^J, C^1, \dots, C^J).$$

As in the non-Bayesian case, the problem boils down to solving a large but very simple system of ODEs. Computation capacity is needed to numerically solve this equation, but there is no mathematical difficulty, and most importantly, no additional difficulty associated with the consideration of Bayesian learning.

It is noteworthy that for both the primal problem and the dual problem, adding Bayesian learning of the conversion rates in the initial setting does not make the problem more complex to solve. This is largely linked to the fact that the Bayesian learning procedure for the conversion rates does not involve any new state variable (as α_t and β_t are simple functions of I_t and C_t).

4 Online learning of the distributions of the prices to beat

4.1 Bayesian learning

In the above models, the J different sources are characterized by J different distributions for the price to beat. In this article, we assume that these distributions are exponential and consequently characterized by the J parameters $\mu^1, \dots, \mu^J$. In practice, we may not assume that the value of these parameter is known, and instead assume that we have a prior distribution on each of these J parameters.

In what follows we assume that for each $j \in \{1, \dots, J\}$, we have at time $t = 0$ a Gamma prior distribution on μ^j, *i.e.*,

$$\mu^j \sim \Gamma(k_0^j, \vartheta_0^j).$$

After the n^{th} auction from the source j, we know whether or not the ad trader won that auction. Moreover, because the auction is of the Vickrey type, being the best bidder at an auction enables to know the value of the price to beat (*i.e.*, the second best price). We can therefore update the prior distribution of μ^j through the use of the Bayes rule. If the n^{th} auction from the source j occurs at time t then, assuming that

$$\mu^j | \mathcal{F}_{t-} \sim \Gamma(k_{t-}^j, \vartheta_{t-}^j),$$

we obtain the following:

- If the ad trader was the best bidder ($b_t^j > p_n^j$), then

$$\mathcal{L}(\mu^j | \mathcal{F}_t) \propto \mu^j e^{-\mu^j p_n^j} \cdot {\mu^j}^{k_{t-}^j - 1} e^{-\vartheta_{t-}^j \mu^j} \propto {\mu^j}^{k_{t-}^j} e^{-(\vartheta_{t-}^j + p_n^j)\mu^j}.$$

- If the ad trader was not the best bidder ($b_t^j \leq p_n^j$), then

$$\mathcal{L}(\mu^j | \mathcal{F}_t) \propto e^{-\mu^j b_t^j} \cdot {\mu^j}^{k_{t-}^j - 1} e^{-\vartheta_{t-}^j \mu^j} \propto {\mu^j}^{k_{t-}^j - 1} e^{-(\vartheta_{t-}^j + b_t^j)\mu^j}.$$

In other words,

$$\mu^j | \mathcal{F}_t \sim \Gamma(k_t^j, \vartheta_t^j),$$

where

$$dk_t^j = \mathbf{1}_{\{b_t^j > p^j_{N_t^j}\}} dN_t^j = dI_t^j$$

and

$$d\vartheta_t^j = \left(p^j_{N_t^j} \mathbf{1}_{\{b_t^j > p^j_{N_t^j}\}} + b_t^j \mathbf{1}_{\{b_t^j \leq p^j_{N_t^j}\}} \right) dN_t^j = \inf\left(b_t^j, p^j_{N_t^j}\right) dN_t^j.$$

In particular, $k_t^j = k_0^j + I_t^j$. The variable k is not therefore a new state variable. However, ϑ is a new state variable!

The terms in μ^j involved in the HJB equations are of the form $\mu^j e^{-\mu^j p}$. Therefore, we need to compute $\mathbb{E}[\mu^j e^{-\mu^j p} | \mathcal{F}_t]$. We have:

$$\begin{aligned}\mathbb{E}[\mu^j e^{-\mu^j p}|\mathcal{F}_t] &= \frac{1}{\Gamma(k_t^j)}\int_0^{+\infty} x e^{-xp} x^{k_t^j-1}\vartheta_t^{j\,k_t^j} e^{-\vartheta_t^j x}dx \\ &= \frac{\Gamma(k_t^j+1)}{\Gamma(k_t^j)}\frac{\vartheta_t^{j\,k_t^j}}{\left(\vartheta_t^j+p\right)^{k_t^j+1}} \\ &= k_t^j\frac{\vartheta_t^{j\,k_t^j}}{\left(\vartheta_t^j+p\right)^{k_t^j+1}}.\end{aligned}$$

4.2 A new Hamilton-Jacobi-Belmann equation for the primal problem

With the above Bayesian framework, the value function associated with the optimal control problem

$$\inf_{(b_t^1,\ldots,b_t^J)_t\in\mathcal{A}^J}\mathbb{E}\left[\Phi(I_T^1,\ldots,I_T^J,C_T^1,\ldots,C_T^J)+K\min\left(\bar{S}-S_T,0\right)^2\right].$$

is a 5-variable function (or in fact a function in dimension $3J+2$)

$$(t,I,C,\vartheta,S)\mapsto u(t,I,C,\vartheta,S).$$

The associated HJB equation is:

$$\begin{aligned}0 &= \partial_t u(t,I,C,\vartheta,S) \\ &+\sum_{j=1}^J\lambda^j\inf_{b^j\in\mathbb{R}_+}\left[\int_0^{b^j}(k_0^j+I^j)\frac{\vartheta^{j\,k_0^j+I^j}}{\left(\vartheta^j+p\right)^{k_0^j+I^j+1}}\left[(1-\nu^j)(u(t,I+e^j,C,\vartheta+pe^j,S+p)-u(t,I,C,\vartheta,S))\right.\right. \\ &\qquad +\nu^j(u(t,I+e^j,C+e^j,\vartheta+pe^j,S+p)-u(t,I,C,\vartheta,S))\Big]dp \\ &\qquad \left.\left.+\left(\frac{\vartheta^j}{\vartheta^j+b^j}\right)^{k_0^j+I^j}(u(t,I,C,\vartheta+b^je^j,S)-u(t,I,C,\vartheta,S))\right]\right., \qquad (9)\end{aligned}$$

with terminal condition

$$u(T,I^1,\ldots,I^J,C^1,\ldots,C^J,\vartheta^1,\ldots,\vartheta^J,S)=\Phi(I^1,\ldots,I^J,C^1,\ldots,C^J)+K\min\left(\bar{S}-S,0\right)^2.$$

Eq. (9) is very complex because it involves two continuous variables: a scalar variable S and a d-dimensional variable ϑ. In particular, the optimal bids are characterized by complex implicit equations involving the gradient of the value function with respect to ϑ. Approximating numerically the solution to Eq. (9) seems to be very complex, except maybe for some specific choices of the function Φ.

4.3 New equations for the dual problem

In the case of the dual problem, the optimal control problem

$$\inf_{(b_t^1,\ldots,b_t^J)_t\in\mathcal{A}^J}\mathbb{E}\left[S_T+\Phi(I_T^1,\ldots,I_T^J,C_T^1,\ldots,C_T^J)\right].$$

is characterized by the following HJB equation:

$$0 = \partial_t u(t, I, C, \vartheta, S)$$
$$+\sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \left[\int_0^{b^j} (k_0^j + I^j) \frac{{\vartheta^j}^{k_0^j + I^j}}{(\vartheta^j + p)^{k_0^j + I^j + 1}} \Big[(1-\nu^j)(u(t, I + e^j, C, \vartheta + pe^j, S + p) - u(t, I, C, \vartheta, S)) \right.$$
$$+\nu^j (u(t, I + e^j, C + e^j, \vartheta + pe^j, S + p) - u(t, I, C, \vartheta, S)) \Big] \, dp$$
$$\left. + \left(\frac{\vartheta^j}{\vartheta^j + b^j} \right)^{k_0^j + I^j} (u(t, I, C, \vartheta + b^j e^j, S) - u(t, I, C, \vartheta, S)) \right], \tag{10}$$

with terminal condition

$$u(T, I^1, \ldots, I^J, C^1, \ldots, C^J, \vartheta^1, \ldots, \vartheta^J, S) = S + \Phi(I^1, \ldots, I^J, C^1, \ldots, C^J).$$

As in the non-Bayesian case, this equation can be simplified by using the ansatz

$$u(t, I^1, \ldots, I^J, C^1, \ldots, C^J, \vartheta^1, \ldots, \vartheta^J, S) = S + \theta(t, I^1, \ldots, I^J, C^1, \ldots, C^J, \vartheta^1, \ldots, \vartheta^J).$$

The Bayesian counterpart of Eq. (3) is then:

$$0 = \partial_t \theta(t, I, C, \vartheta)$$
$$+\sum_{j=1}^{J} \lambda^j \inf_{b^j \in \mathbb{R}_+} \left[\int_0^{b^j} (k_0^j + I^j) \frac{{\vartheta^j}^{k_0^j + I^j}}{(\vartheta^j + p)^{k_0^j + I^j + 1}} \Big[p + (1-\nu^j)(\theta(t, I + e^j, C, \vartheta + pe^j) - \theta(t, I, C, \vartheta)) \right.$$
$$+\nu^j (\theta(t, I + e^j, C + e^j, \vartheta + pe^j) - \theta(t, I, C, \vartheta)) \Big] \, dp$$
$$\left. + \left(\frac{\vartheta^j}{\vartheta^j + b^j} \right)^{k_0^j + I^j} (\theta(t, I, C, \vartheta + b^j e^j) - \theta(t, I, C, \vartheta)) \right], \tag{11}$$

with terminal condition

$$\theta(T, I^1, \ldots, I^J, C^1, \ldots, C^J, \vartheta^1, \ldots, \vartheta^J) = \Phi(I^1, \ldots, I^J, C^1, \ldots, C^J).$$

Eq. (11) is very complex because it involves the d-dimensional continuous variable ϑ. In particular, the optimal bids are characterized by complex implicit equations involving the gradient of the function θ with respect to ϑ. Approximating numerically the solution to Eq. (10) / (11) seems to be very complex, except maybe for some specific choices of the function Φ.

It is noteworthy that for both the primal problem and the dual problem, adding Bayesian learning of the parameters μ^js in the initial setting does make the problem far more complex to solve. This is largely linked to the fact that the Bayesian learning procedure for the distribution of the price to beat involves a new state variable ϑ which is (i) continuous and not discrete, and (ii) related to the control variables b^js.

In practice, it seems that the only realistic option for using Bayesian learning of the distributions of the prices to beat is through myopic learning.

5 Online learning of the auction frequencies

5.1 Bayesian learning

As for the conversion rates and the distributions of the prices to beat, the J parameters $\lambda^1, \dots, \lambda^J$ characterizing the frequency at which each of the J sources sends auction requests can be learnt on the run. In what follows we assume that for each $j \in \{1, \dots, J\}$, we have at time $t = 0$ a Gamma prior distribution on λ^j, *i.e.*,

$$\lambda^j \sim \Gamma(\kappa_0^j, \tau_0^j).$$

These prior distributions can be updated with the number of auction requests received from the J different sources. We have indeed:

$$\mathcal{L}(\lambda^j | N_t^j) \propto (\lambda^j t)^{N_t^j} e^{-\lambda^j t} \cdot \lambda^{j\kappa_0^j - 1} e^{-\tau_0^j \lambda^j} \propto \lambda^{j k_0^j + N_t^j - 1} e^{-(\tau_0^j + t)\lambda^j}.$$

In other words,

$$\lambda^j | \mathcal{F}_t \sim \Gamma(k_0^j + N_t^j, \tau_0^j + t).$$

In particular,

$$\mathbb{E}[\lambda^j | \mathcal{F}_t] = \frac{k_0^j + N_t^j}{\tau_0^j + t}.$$

5.2 A new Hamilton-Jacobi-Belmann equation for the primal problem

With the above Bayesian framework, the value function associated with the optimal control problem

$$\inf_{(b_t^1, \dots, b_t^J)_t \in \mathcal{A}^J} \mathbb{E}\left[\Phi(I_T^1, \dots, I_T^J, C_T^1, \dots, C_T^J) + K \min\left(\bar{S} - S_T, 0\right)^2\right].$$

is a 5-variable function (or in fact a function in dimension $3J + 2$)

$$(t, I, C, N, S) \mapsto u(t, I, C, N, S).$$

The associated HJB equation is:

$$0 = \partial_t u(t, I, C, N, S) + \sum_{j=1}^{J} \frac{k_0^j + N^j}{\tau_0^j + t} \Bigg[(u(t, I, C, N + e^j, S) - u(t, I, C, N, S))$$

$$+ \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p} \left[(1 - \nu^j)(u(t, I + e^j, C, N + e^j, S + p) - u(t, I, C, N + e^j, S)) \right.$$

$$\left. + \nu^j (u(t, I + e^j, C + e^j, N + e^j, S + p) - u(t, I, C, N + e^j, S)) \right] dp \Bigg], \tag{12}$$

with terminal condition

$$u(T, I^1, \dots, I^J, C^1, \dots, C^J, N^1, \dots, N^J, S) = \Phi(I^1, \dots, I^J, C^1, \dots, C^J) + K \min\left(\bar{S} - S, 0\right)^2.$$

Eq. (12) is a non-standard integro-differential HJB equation in dimension $3J+2$ which can be seen as a system (indexed by I, C, and N) of integro-differential equations (or a system of first-order PDEs if we consider the same approximation as the one we proposed for Eq. (1)). For approximating numerically the solution of Eq. (12), the same methods as for Eq. (1) can be employed. The only difference is that the system of equations is far larger because of the additional variable N.

5.3 New equations for the dual problem

In the case of the dual problem, the optimal control problem

$$\inf_{(b_t^1,\dots,b_t^J)_t \in \mathcal{A}^J} \mathbb{E}\left[S_T + \Phi(I_T^1,\dots,I_T^J,C_T^1,\dots,C_T^J)\right].$$

is characterized by the following HJB equation:

$$0 = \partial_t u(t,I,C,N,S) + \sum_{j=1}^{J} \frac{k_0^j + N^j}{\tau_0^j + t}\Bigg[(u(t,I,C,N+e^j,S) - u(t,I,C,N,S))$$

$$+ \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p}\left[(1-\nu^j)(u(t,I+e^j,C,N+e^j,S+p) - u(t,I,C,N+e^j,S))\right.$$

$$\left. + \nu^j (u(t,I+e^j,C+e^j,N+e^j,S+p) - u(t,I,C,N+e^j,S))\right] dp\Bigg], \qquad (13)$$

with terminal condition

$$u(T,I^1,\dots,I^J,C^1,\dots,C^J,N^1,\dots,N^J,S) = S + \Phi(I^1,\dots,I^J,C^1,\dots,C^J).$$

As in the non-Bayesian case, this equation can be simplified by using the ansatz

$$u(t,I^1,\dots,I^J,C^1,\dots,C^J,N^1,\dots,N^J,S) = S + \theta(t,I^1,\dots,I^J,C^1,\dots,C^J,N^1,\dots,N^J).$$

The Bayesian counterpart of Eq. (3) is then:

$$0 = \partial_t \theta(t,I,C,N) + \sum_{j=1}^{J} \frac{k_0^j + N^j}{\tau_0^j + t}\Bigg[(\theta(t,I,C,N+e^j) - \theta(t,I,C,N))$$

$$+ \inf_{b^j \in \mathbb{R}_+} \int_0^{b^j} \mu^j e^{-\mu^j p}\left[p + (1-\nu^j)(\theta(t,I+e^j,C,N+e^j) - \theta(t,I,C,N+e^j))\right.$$

$$\left. + \nu^j (\theta(t,I+e^j,C+e^j,N+e^j) - \theta(t,I,C,N+e^j))\right] dp\Bigg], \qquad (14)$$

with terminal condition

$$\theta(T,I^1,\dots,I^J,C^1,\dots,C^J,N^1,\dots,N^J,S) = \Phi(I^1,\dots,I^J,C^1,\dots,C^J).$$

As in the non-Bayesian case, the problem boils down to solving a large, but structurally simple, system of ODEs. Computation capacity is needed to numerically solve this equation, but there is no mathematical difficulty. It is noteworthy

that the consideration of Bayesian learning only adds here a dimension (the variable N) for the indices of the equations in the system.

It is noteworthy that for both the primal problem and the dual problem, adding Bayesian learning of the frequencies in the initial setting makes the problem more time-consuming as far as numerical approximations are concerned, but there is no additional mathematical difficulty. This is related to the fact that the Bayesian learning procedure for the frequencies involves a new state variable, but a discrete one and not a continuous one.

Conclusion

In this paper, we consider two classical problems faced by ad trading desks – in their relaxed forms. The first (primal) problem consists in buying inventories so as to maximize expected KPIs, with the constraint of spending no more than a given amount. The second (dual) problem consists in minimizing the average amount spent, with the constraint of reaching thresholds for different KPIs. Our goal was to understand how to incorporate (Bayesian) learning in these two optimal control problems. We have seen that learning conversion rates online is very easy and does not make the problem more complex than the initial one without learning. As far as learning the distributions of the best prices bid by other market participants, we have seen that only myopic learning is reasonable in terms of mathematical complexity. Finally, we have seen that learning on the fly the frequency at which each source sends auction requests makes the problem slightly more involved in terms of computation time, but not more complicated mathematically speaking.

It is noteworthy that we have considered learning procedures for each type of unknown parameters independently. However, it is easy to generalize our equations in order to learn on the fly both the conversion rates and the frequencies for instance.

References

[1] Amin, K., Kearns, M., Key, P., & Schwaighofer, A. (2012). Budget optimization for sponsored search: Censored learning in mdps. arXiv preprint arXiv:1210.4847.

[2] Baradel, N., Bouchard, B., & Dang, N. M. (2016). Optimal control under uncertainty and Bayesian parameters adjustments. arXiv preprint arXiv:1604.06340.

[3] Baradel, N., Bouchard, B., & Dang, N. M. (2016). Optimal trading with online parameters revisions. arXiv preprint arXiv:1604.06342.

[4] Badanidiyuru, A., Kleinberg, R., & Slivkins, A. (2013, October). Bandits with knapsacks. In Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on (pp. 207-216). IEEE.

[5] Fernandez-Tapia, J. (2016). Statistical modeling of Vickrey auctions and applications to automated bidding strategies. Optimization Letters.

[6] Fernandez-Tapia, J. (2015). An analytical solution to the budget-pacing problem in programmatic advertising. Technical report.

[7] Fernandez-Tapia, J., Guéant, O., & Lasry, J. M. (2015). Optimal Real-Time Bidding Strategies. To appear in Applied Mathematics Research eXpress.

[8] Fernandez-Tapia, J., Guéant, O., & Lasry, J. M. (2016). On the pricing of performance-based programmatic ad buying contracts.

[9] Guéant, O., & Pu, J. (2016). Portfolio choice under drift uncertainty: a Bayesian learning and stochastic optimal control approach.

[10] Khaledi, M., & Abouzeid, A. (2016). Optimal Bidding in Repeated Wireless Spectrum Auctions with Budget Constraints. arXiv preprint arXiv:1608.07357.

[11] Mohri, M., & Medina, A. M. (2016). Learning Algorithms for Second-Price Auctions with Reserve. Journal of Machine Learning Research, 17(74), 1-25.

[12] Perchet, V., Rigollet, P., & Weed, J. (2015). Online learning in repeated auctions.

[13] Sani, A., Neu, G., & Lazaric, A. (2014). Exploiting easy data in online optimization. In Advances in Neural Information Processing Systems (pp. 810-818).

[14] Tran-Thanh, L., Stavrogiannis, L. C., Naroditskiy, V., Robu, V., Jennings, N. R., & Key, P. (2014). Efficient regret bounds for online bid optimisation in budget-limited sponsored search auctions.

[15] Tran-Thanh, L., & Yu, J. Y. (2014). Functional Bandits. arXiv preprint arXiv:1405.2432.

[16] Wang, J., Yuan, S., & Zhang, W. (2016, March). Real-Time Bidding Based Display Advertising: Mechanisms and Algorithms. In European Conference on Information Retrieval (pp. 897-901). Springer International Publishing.