POLITECNICO
MILANO 1863

Online Learning Applications

Project assignment

Authors:
Balestrieri Niccolò, Bertogalli Andrea, Cavalieri Francesco, Tombini Nicolò

Professors:
Matteo Castiglioni, Gianmarco Genalti

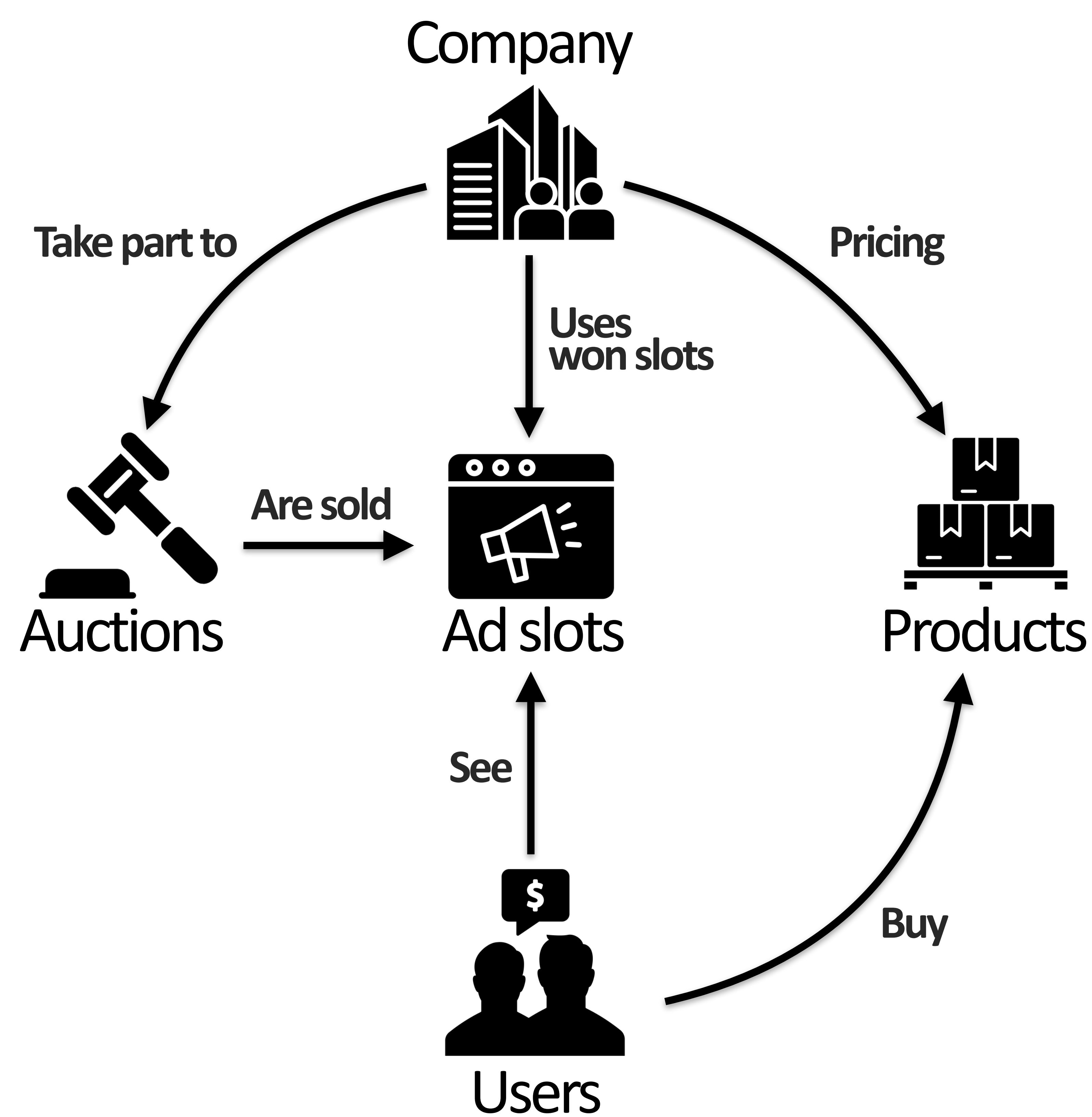
Accademic year 2024/2025

Introduction - The task

The goal of the project is to design online learning algorithms to handle a marketing campaign to sell products, including an advertising campaign and a pricing problem.

The setting is then:

- For each day the company chooses a price p.
- The company faces a sequence of auctions. For each auction:
    - The company chooses a bid b.
    - A slot is (possibly) assigned to the company depending on b, the competing bids, and the auction format.
    - If the ad is clicked, an user visit the company web page.
    - The user buy the company product with a probability that depends on the price p.



REFERENCES:
Project assignment, M. Castiglioni & G. Genalti, 2024.

Stochastic environment - The setting

Here we consider a stochastic environment, this includes a distribution over the bids of the other agents ($\mathscr{U}(0,1)$) and a function specifying the probability with which an user buys for every price ($D(p)$). Here (but also in adversarial environments) we consider the probability λ_s of a slot s to be observed as constant. In particular considering a **continuous** set of prices $p \in [0,\ 1]$ we employed **GP-UCB** for pricing, while for the bidding problem we considered a sequence of **second price auction** and employed a **Multiplicative Pacing Strategy** and a **UCB-Like** approach.

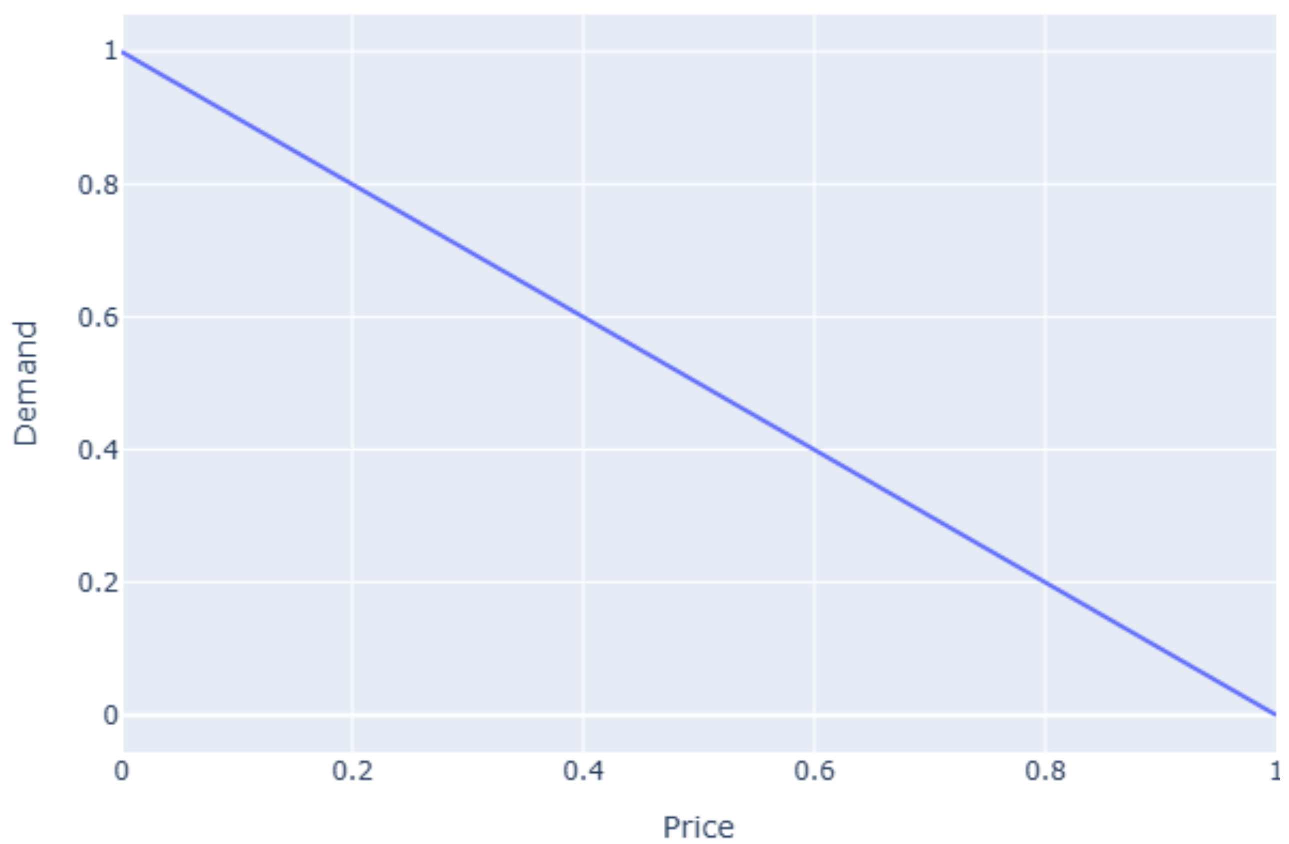


Fig. 1 - The demand curve which represent the conversion probability, in this case this function is modeled simply as $1-p$.

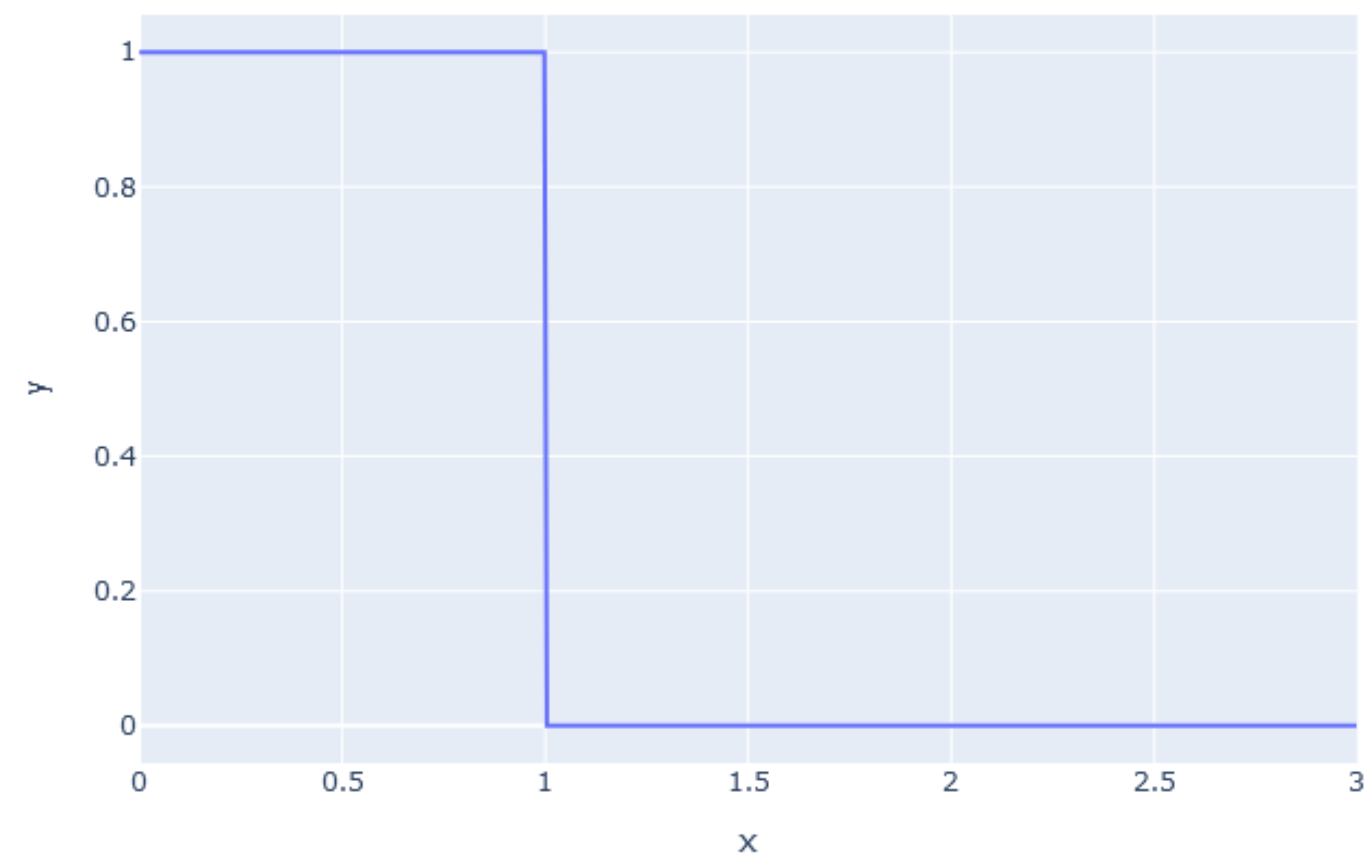


Fig. 2 - The distribution over competing bids, in a stochastic environment is modeled as a uniform distribution $\mathscr{U}(0,1)$.

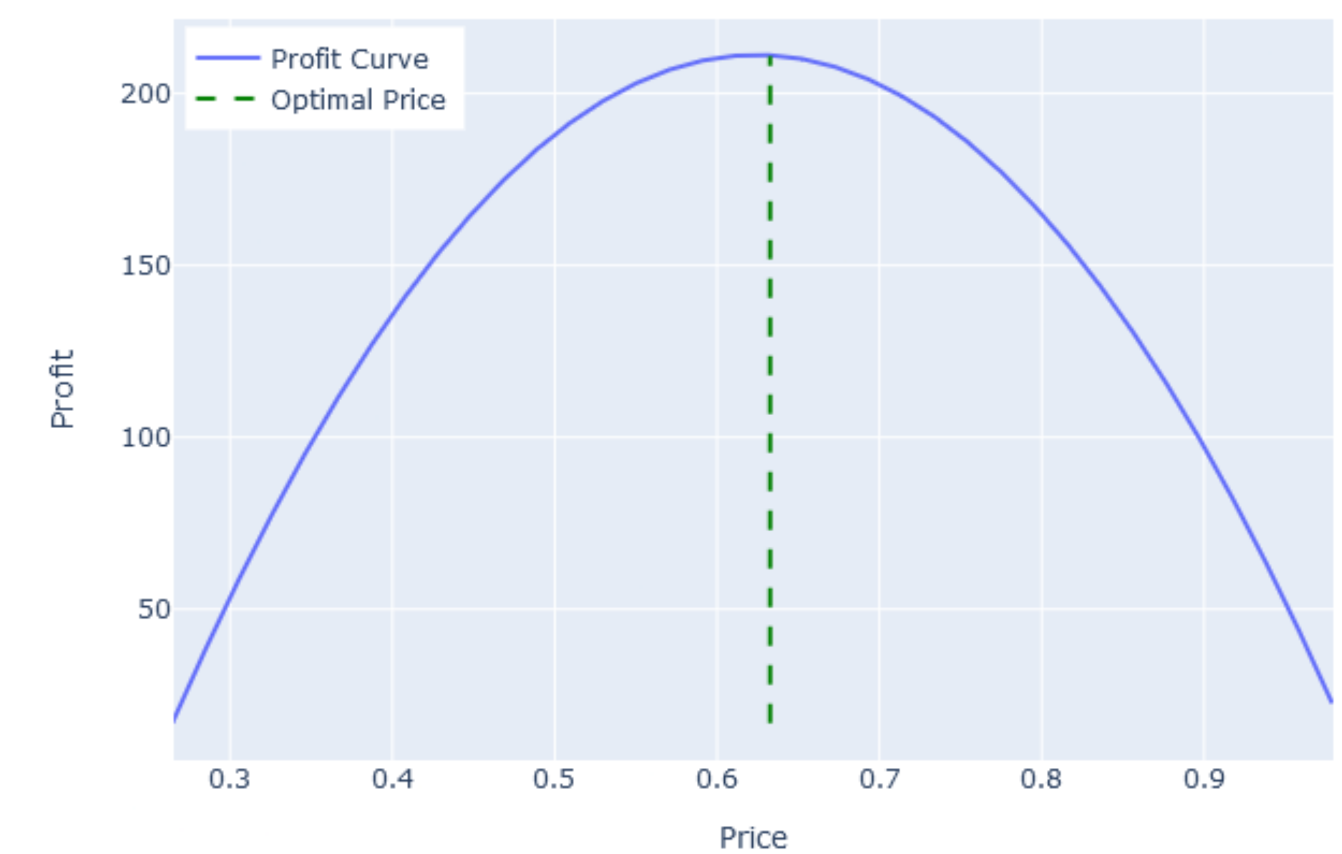


Fig. 3 - Clairvoyant profit curve representing the profit for each price, the profit is compute as $P = (p-c) \cdot n_t$ where $n_t = D(p) \cdot n_c$.

Stochastic environment - Pricing GP-UCB

Here we employ **GP-UCB** over a continuous set of prices $p \in [0,\ 1]$ considering the following parameters: the cost $c = 0.25$, the number of rounds $T = 250$, $K = 50$, customers = 1500. The execution have been made over 10 trials.

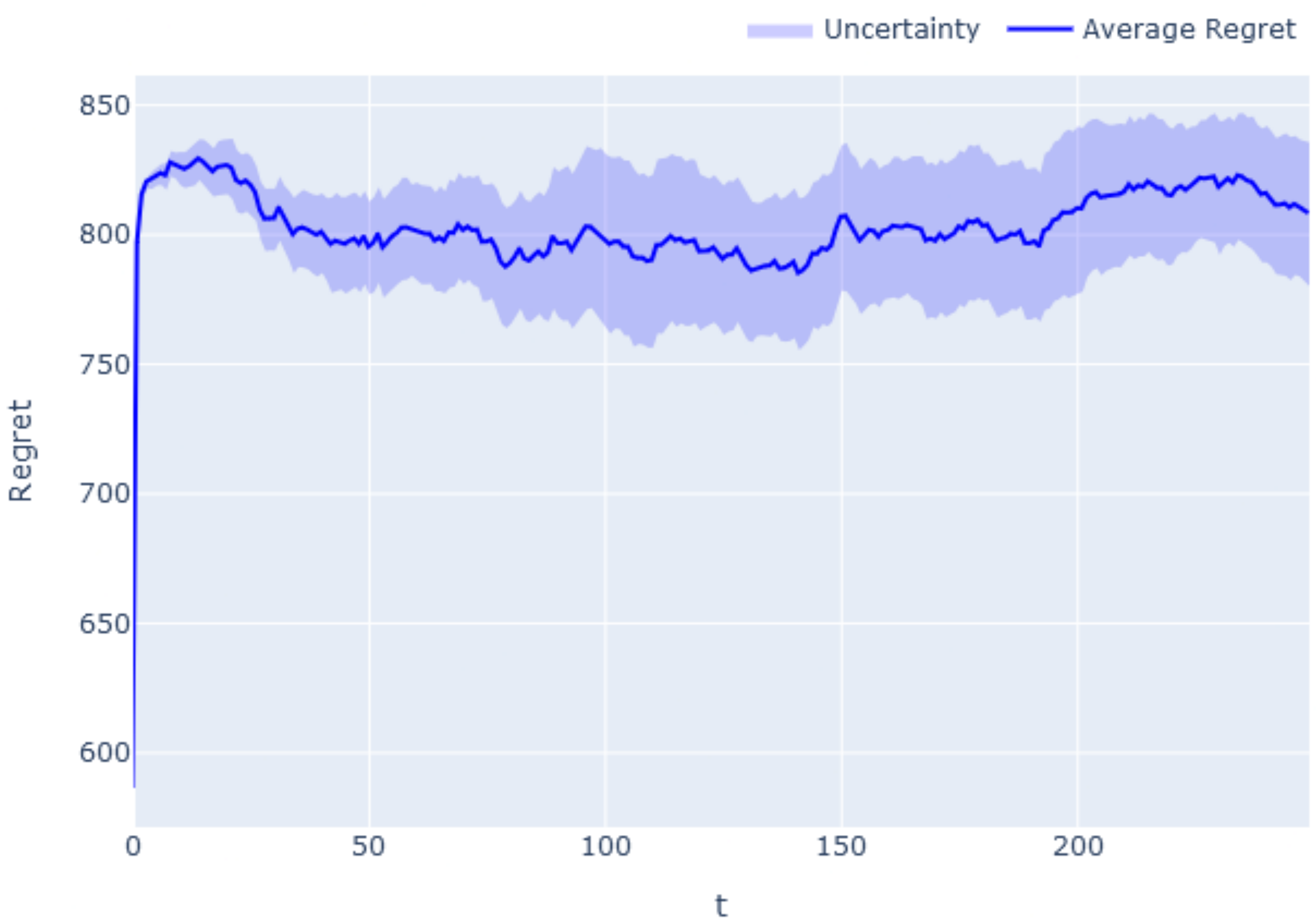


Fig. 4 - The average regret achieved over 10 trials by GP-UCB.

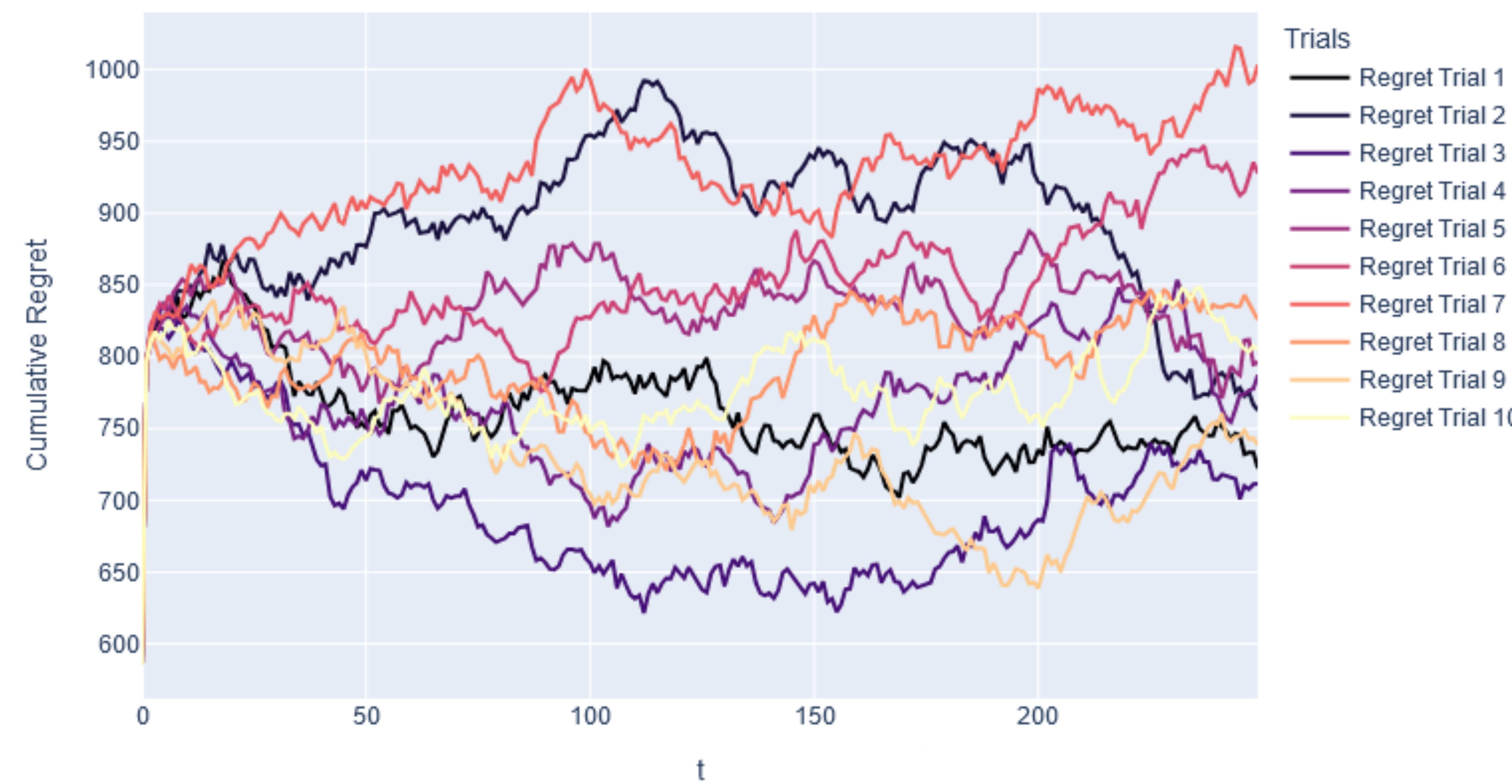


Fig. 5 - The regret for every trial archived by GP-UCB.

Stochastic environment - Pricing GP-UCB

Here we inspect the chosen prices, as we can see the algorithm perform well (almost) converging to the optimal price. Notice that the heat map that we show is discretised for visualisation purposes.

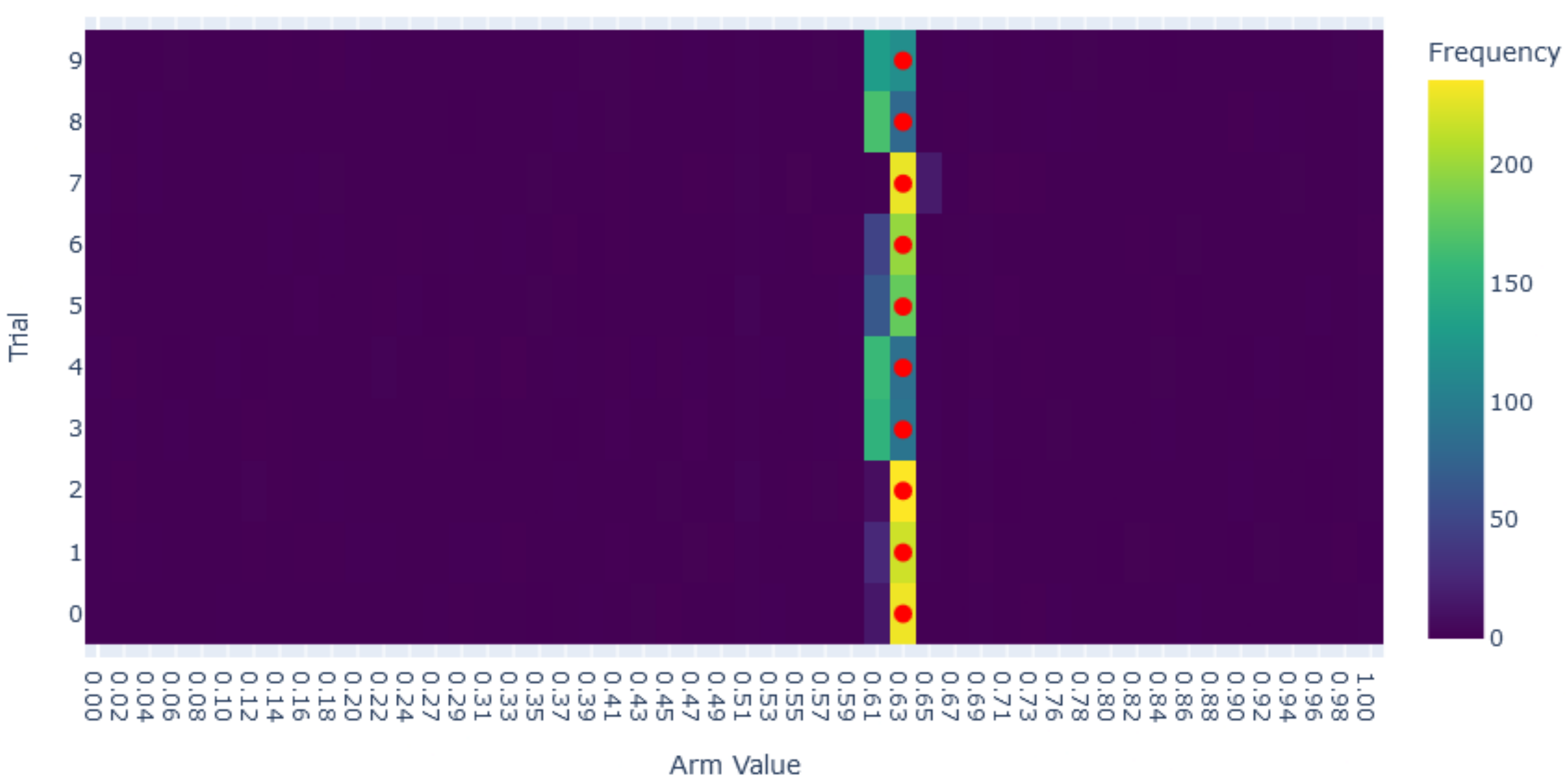


Fig. 6 - The chosen prices frequency ($N(t)$) for each trial - GP-UCB.

Fig. 7 - Here we plot the final estimated profit curve (normalized), along with the chosen prices samples.

Stochastic environment - Bidding multiplicative pacing

Here we employ a **Multiplicative Pacing Strategy** on the bidding problem in stochastic environment (shown in the previous slide). Here we consider the following parameters: the budget $B = 135$, 3 advertisers, 1500 customers ($n_c = 1500$), these results were obtained over 10 trials.

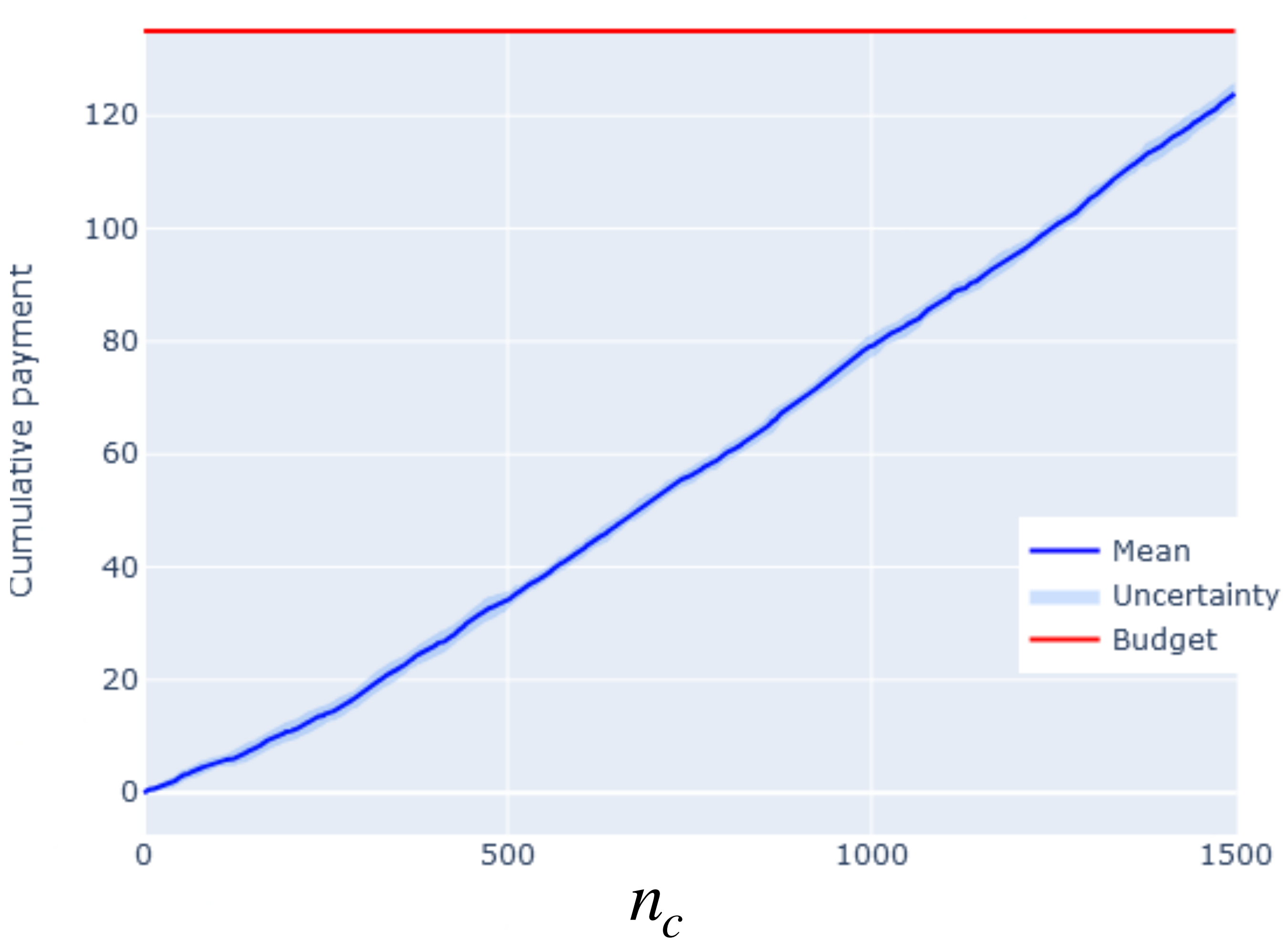


Fig. 8 - The cumulative payments over the users in red we have the budget.

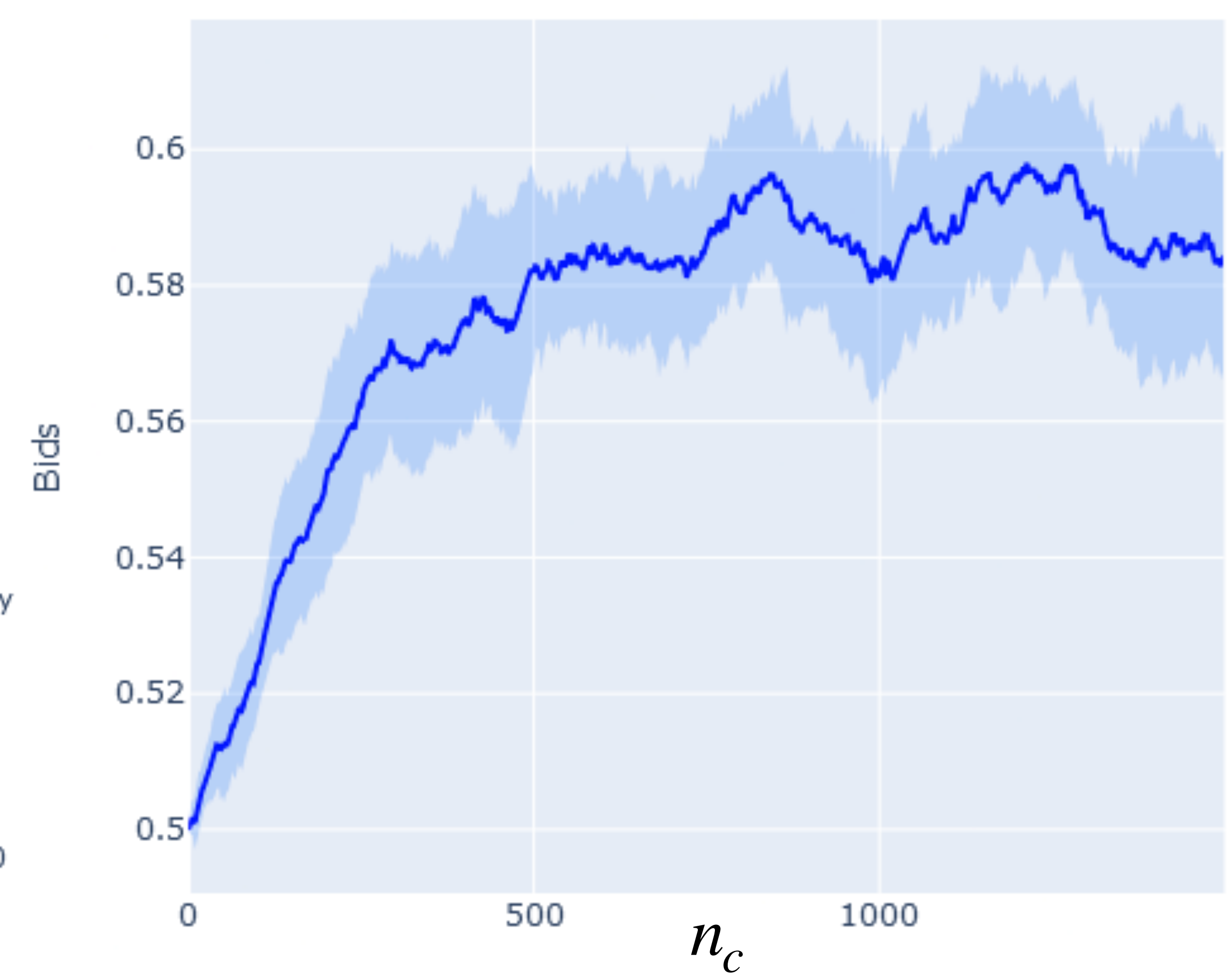


Fig. 9 - The chosen bids curve.

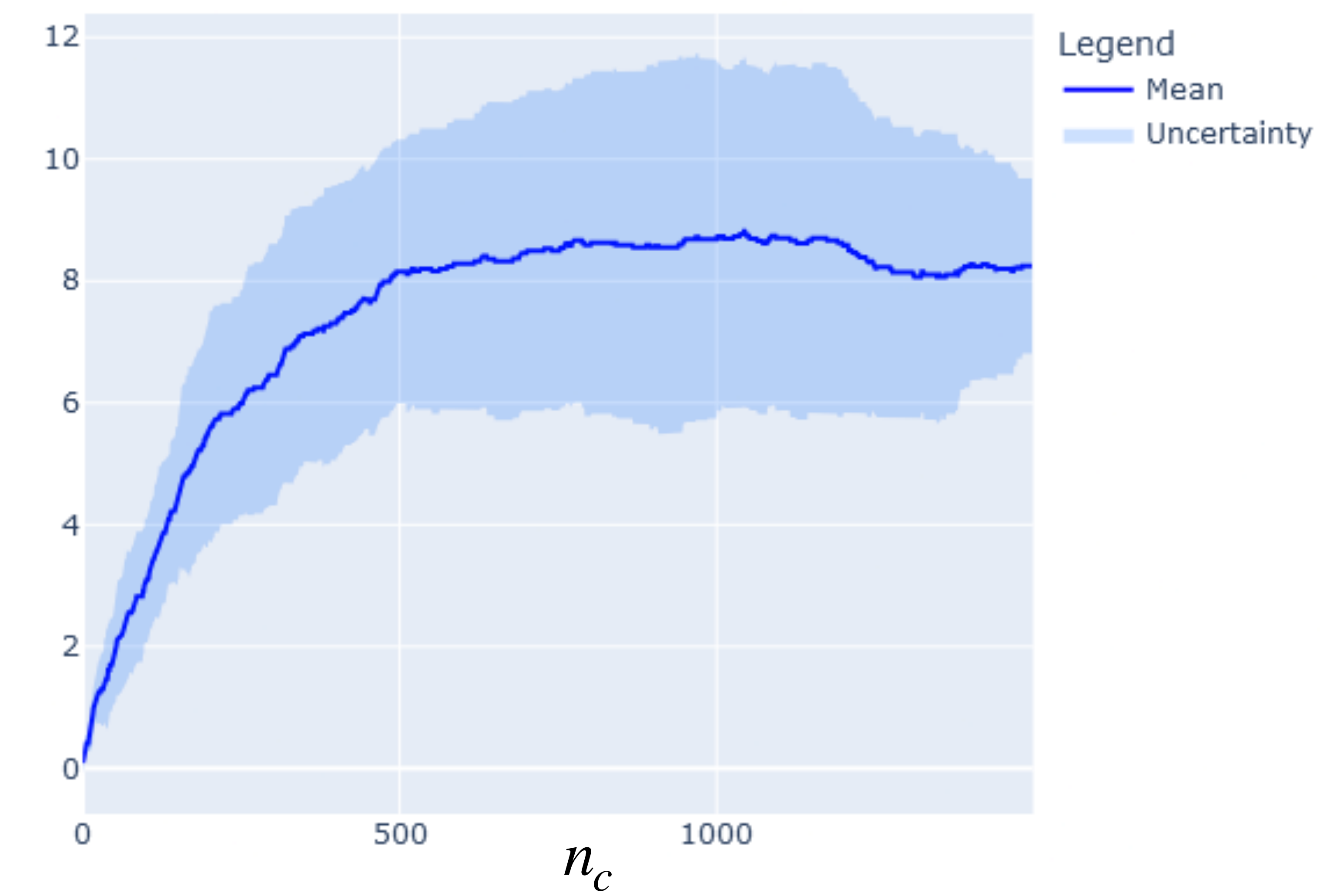


Fig. 10 - The cumulative regret archived by the multiplicative pacing strategy.

N. Balestrieri, A. Bertogalli, F. Cavalieri, N. Tombini

Stochastic environment - Bidding UCB-Like

Here we employ a **UCB-Like approach** on the bidding problem in stochastic environment (shown in the previous slide). Here we consider the following parameters: the budget $B = 135$, 3 advertisers, 1500 customers ($n_c = 1500$), these results were obtained over 10 trials.

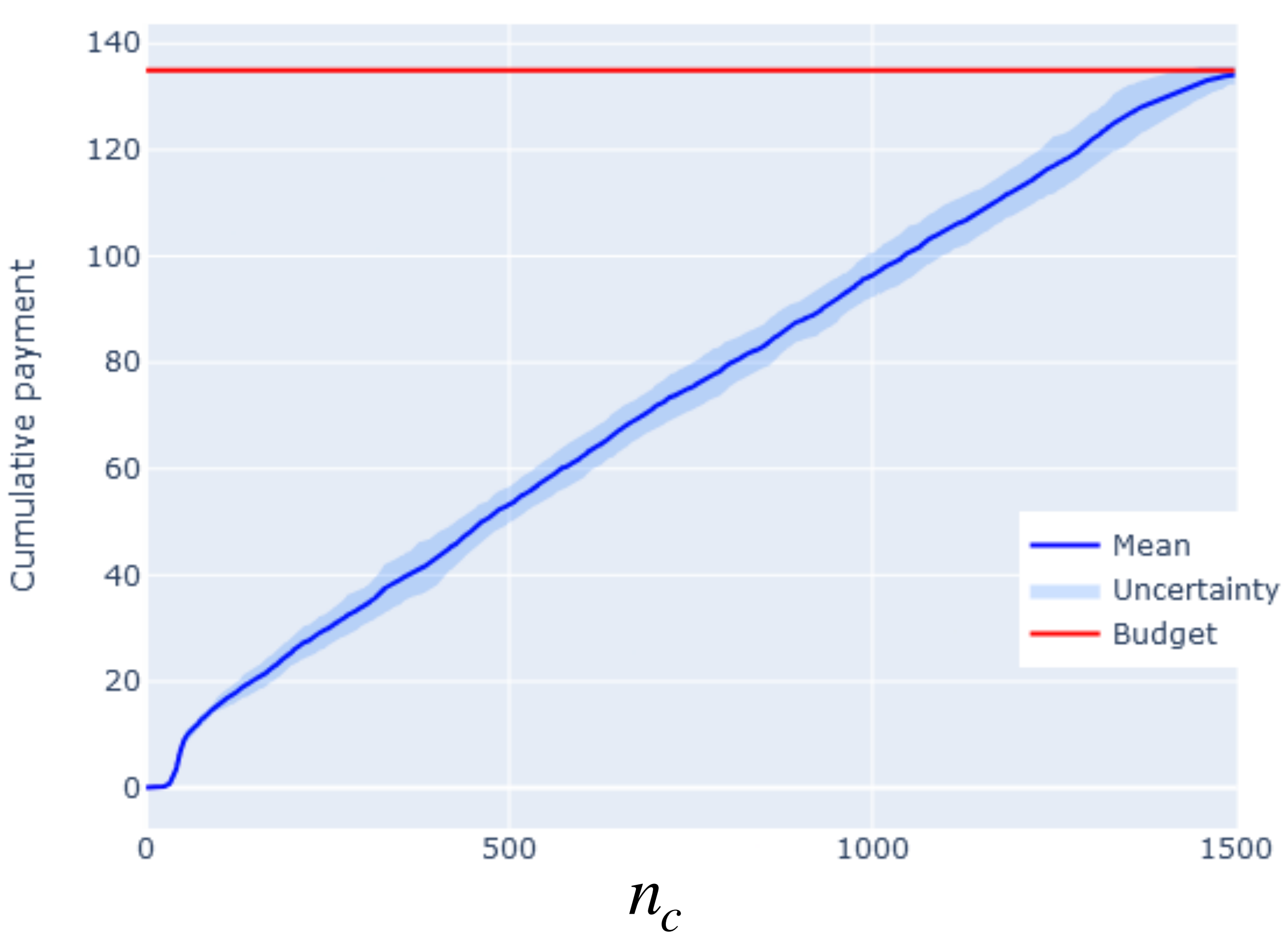


Fig. 11 - The cumulative payments over the users in red we have the budget.

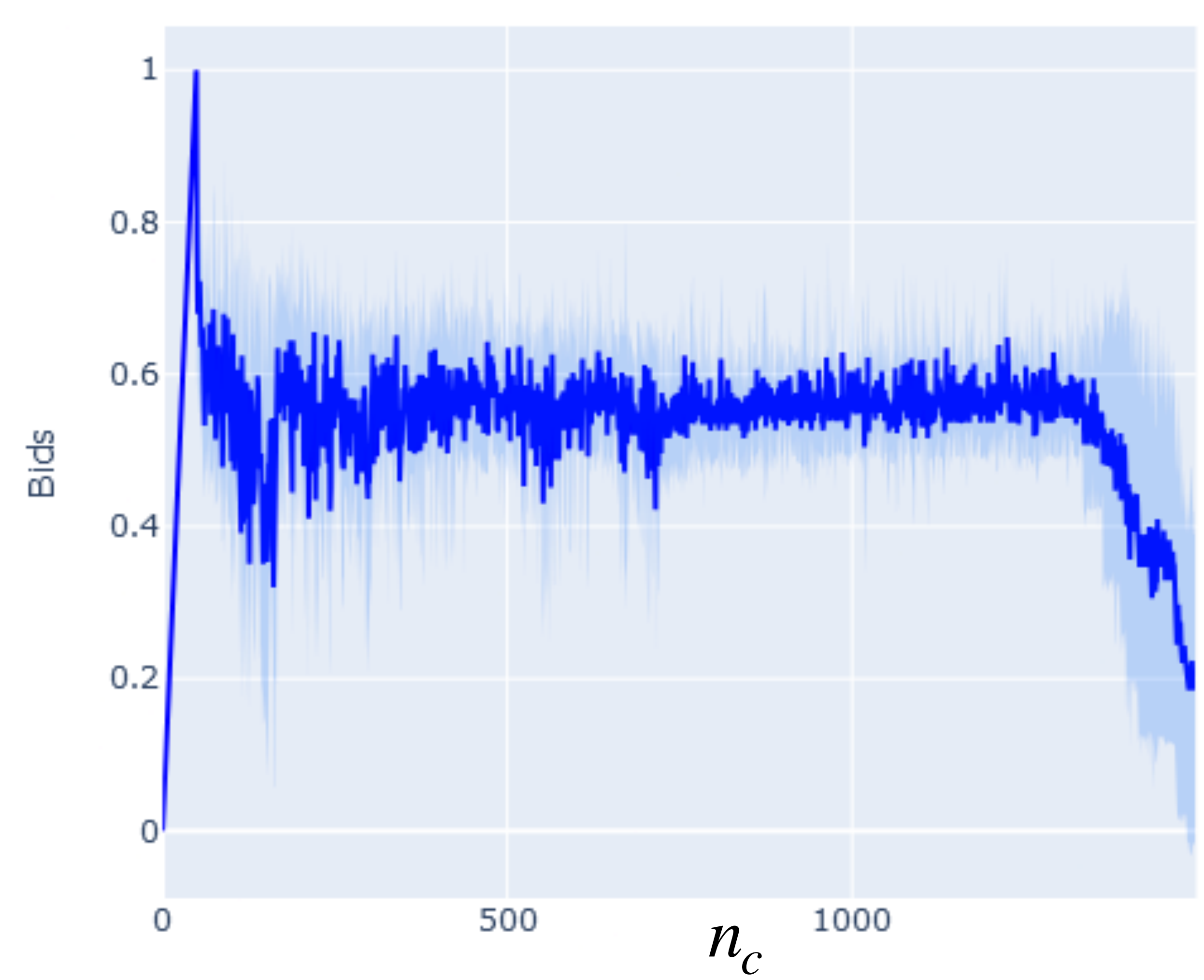


Fig. 12 - The chosen bids curve.

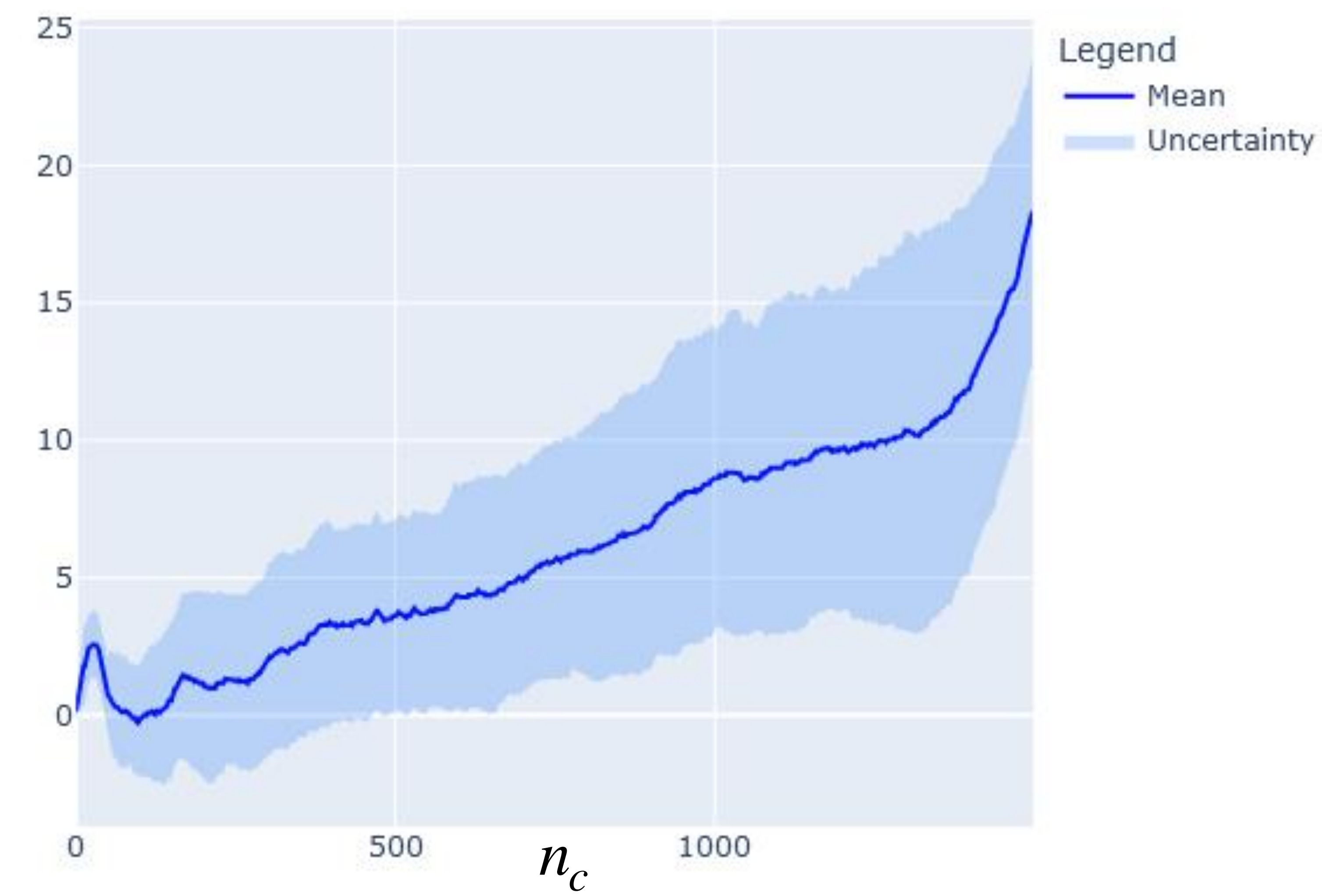


Fig. 13 - The cumulative regret archived by the UCB-Like.

Stochastic environment - Pricing + Bidding (MPS)

Here we evaluate **Multiplicative Pacing Strategy** for the bidding and **GP-UCB** for the pricing, with parameters: auctions = 1500, advertisers = 3, budget = 135, days = 10, trials = 10, cost = 0.25.

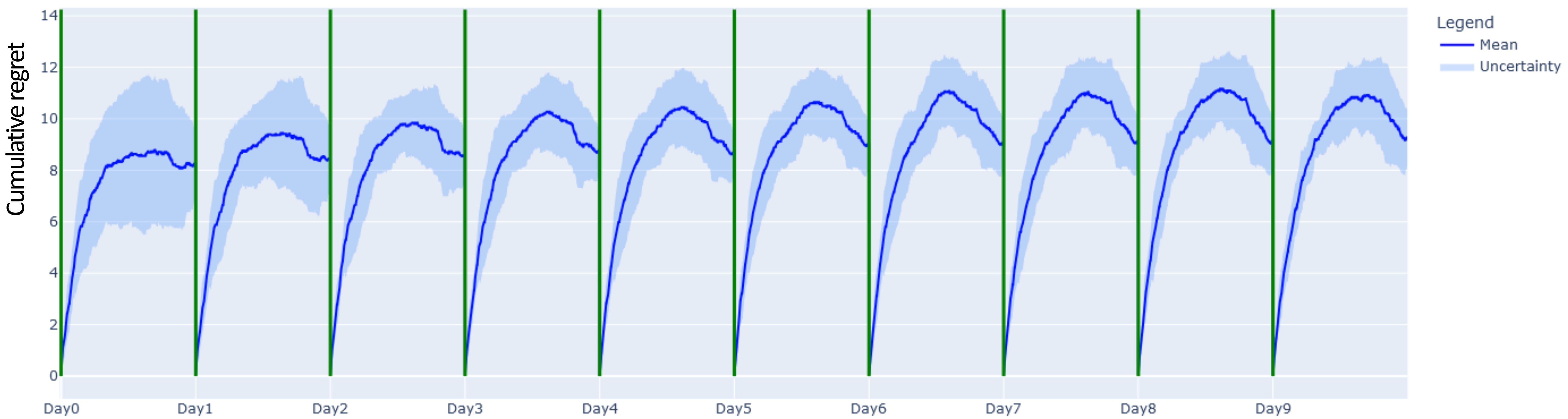


Fig. 14 - The cumulative regret in 10 days of pricing + bidding interaction.

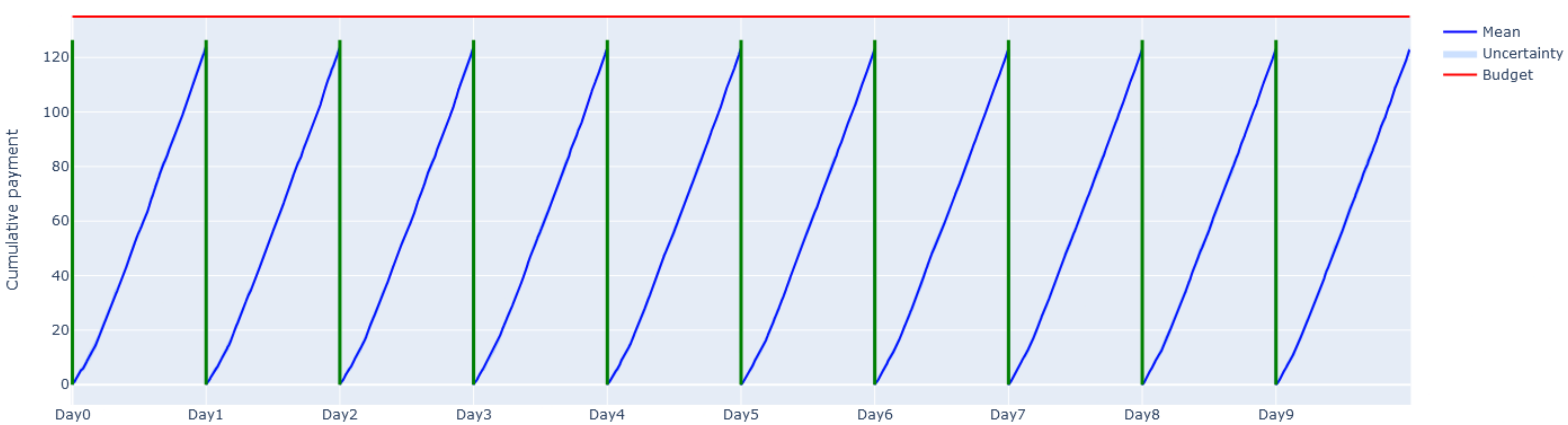


Fig. 15 - The cumulative payments in 10 days of pricing + bidding interaction.

Stochastic environment - Pricing + Bidding (UCB-Like)

Here we evaluate **UCB-Like algorithm** for the bidding and **GP-UCB** for the pricing, with parameters: auctions = 1500, advertisers = 3, budget = 135, days = 10, trials = 10, cost = 0.25.

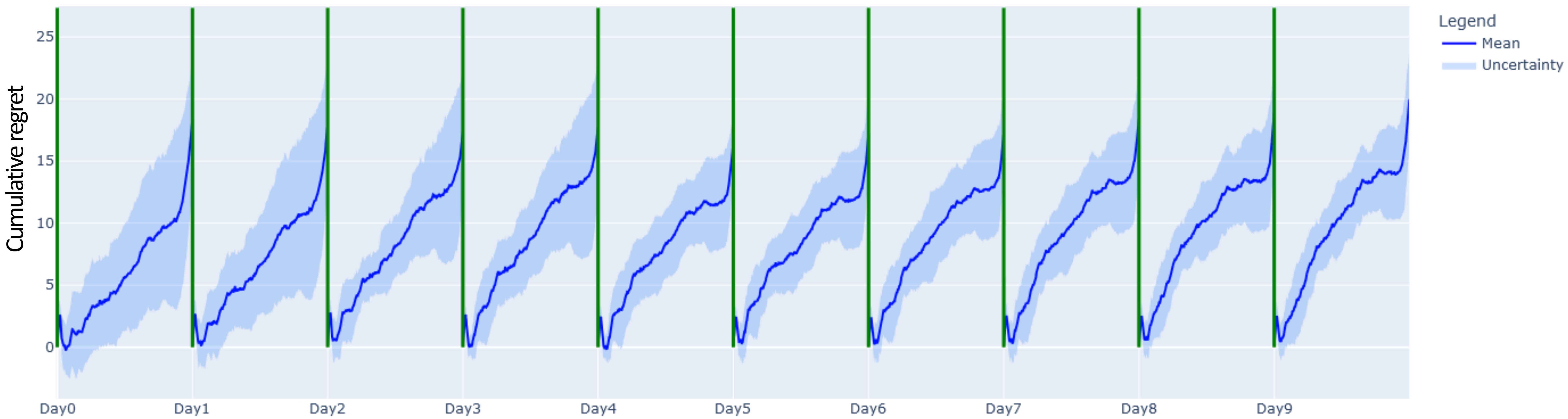


Fig. 16 - The cumulative regret in 10 days of pricing + bidding interaction.

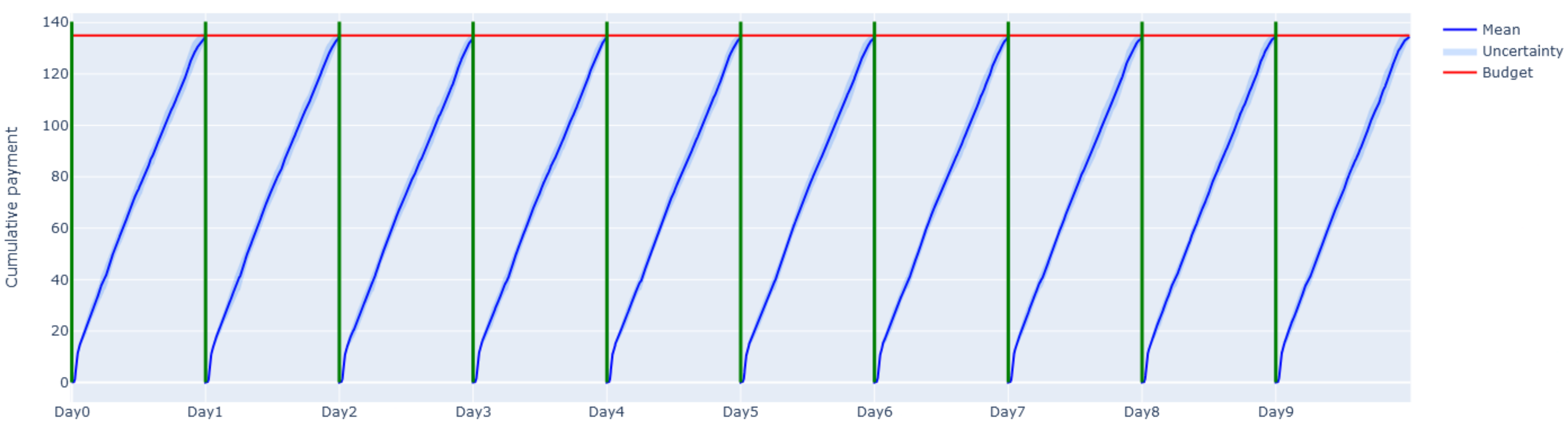


Fig. 17 - The cumulative payments in 10 days of pricing + bidding interaction.

Adversarial environment - The setting

Here we are considering an **highly non stationary** environment (modeled through a **Gaussian Drift**). In this setting we have a time-varying probability distribution over bids and a time-varying function which gives the user's buying probability.

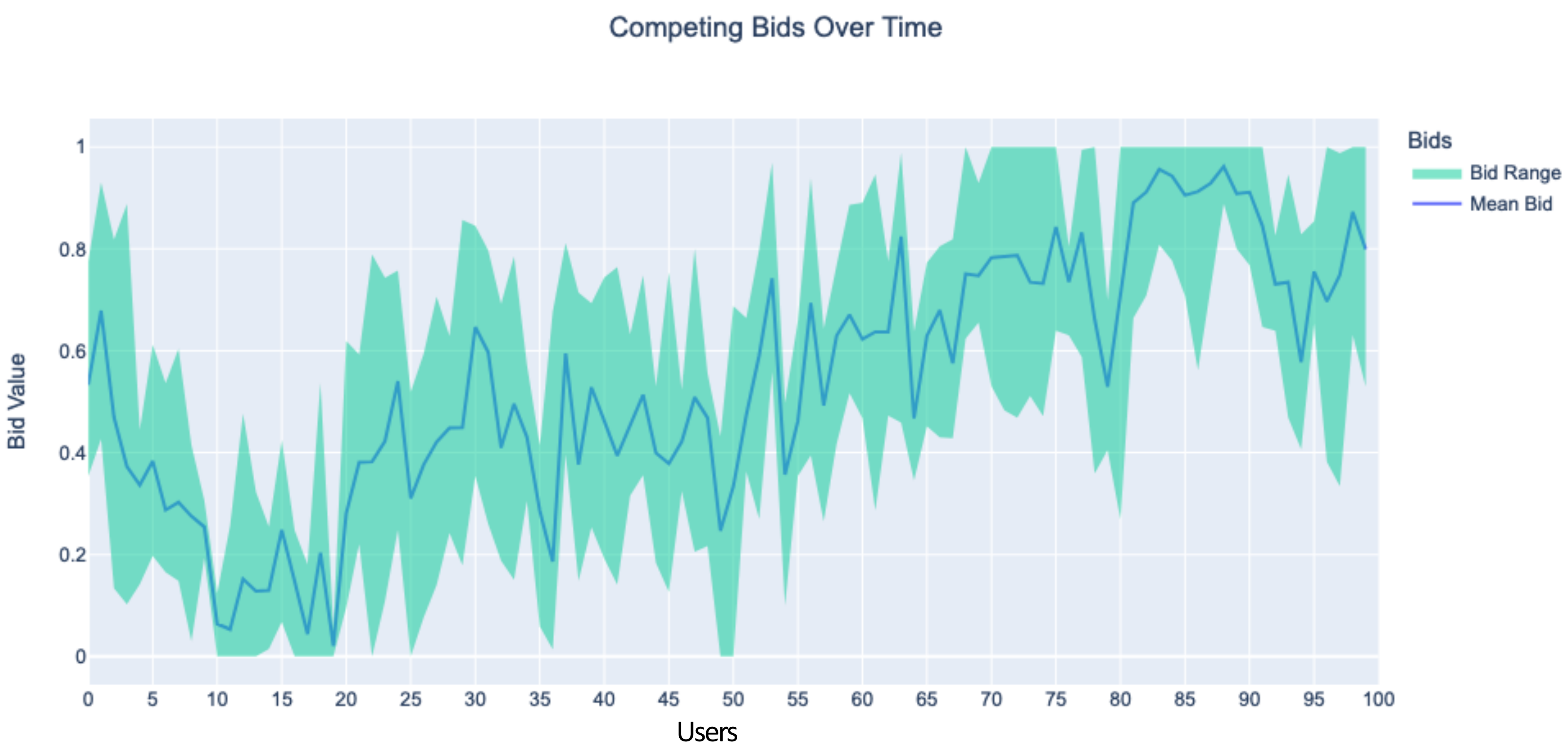


Fig. 18 - The competing bids over time of the competing advertiser present in the auction $B(t) = \mathcal{N}(\mu_t, 0.2)$ where $\mu_t = \mu_{t-1} + \mathcal{U}(i, j)$ and $i, j = \pm 0.1$.

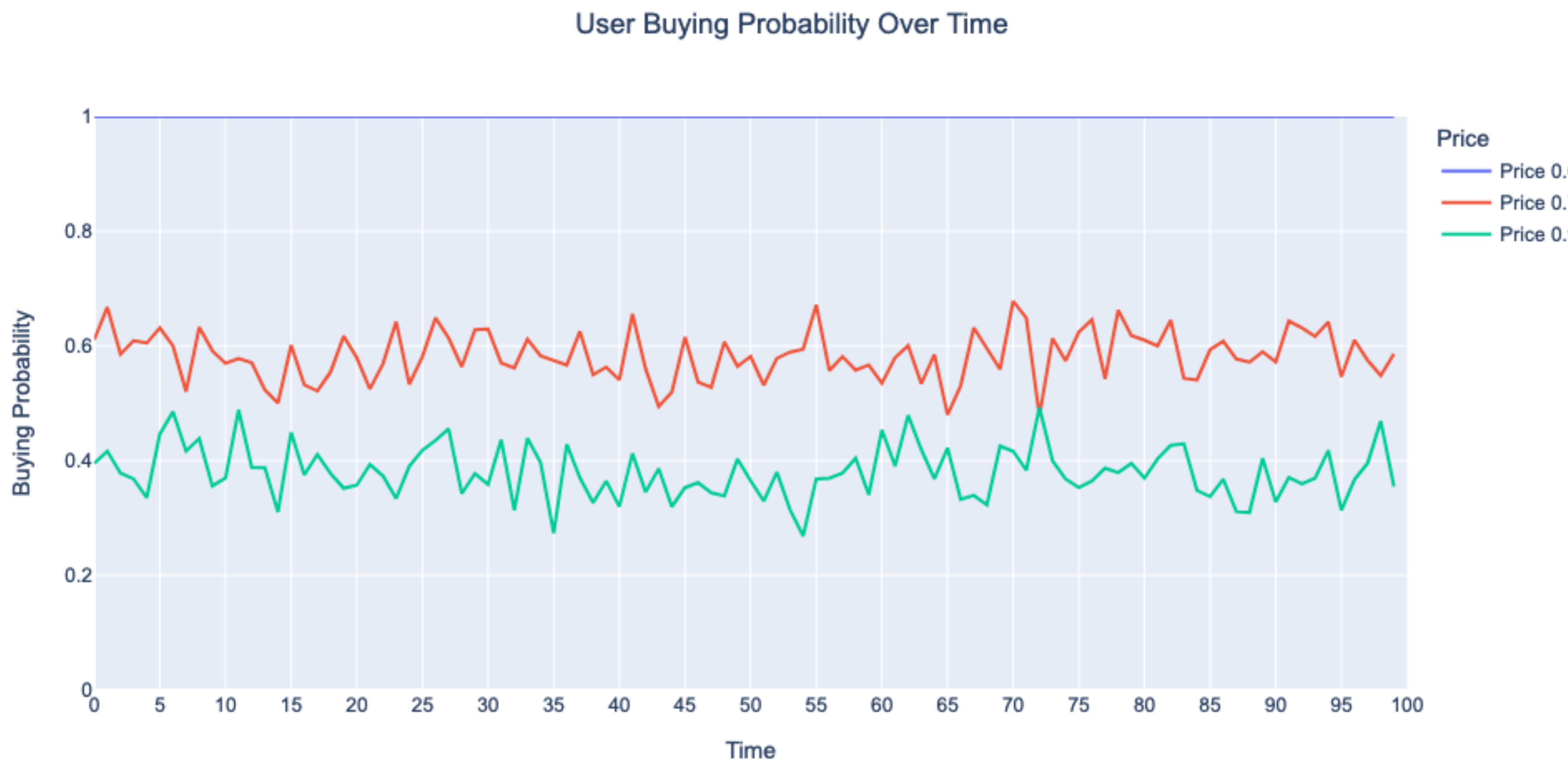


Fig. 19 - The buying probability over time for 3 fixed prices, as we can see the curve is oscillating, this is due to the change in the noise mean. This curve is defined as $D(t) = (1 - p) \ + \ \mathcal{N}(\mu_t, 0.05)$ where $\mu_t = \mu_{t-1} + \mathcal{U}(i, j)$ and $i, j = \pm 0.005$.

Adversarial environment - Pricing EXP3

Here we employ **EXP3** on discretized set of prices $p \in [0,1]$ considering the following parameters: the cost c = 0.3, the number of rounds T = 7000, K = 20, the number of customer = 30. The execution have been made over 8 trials.

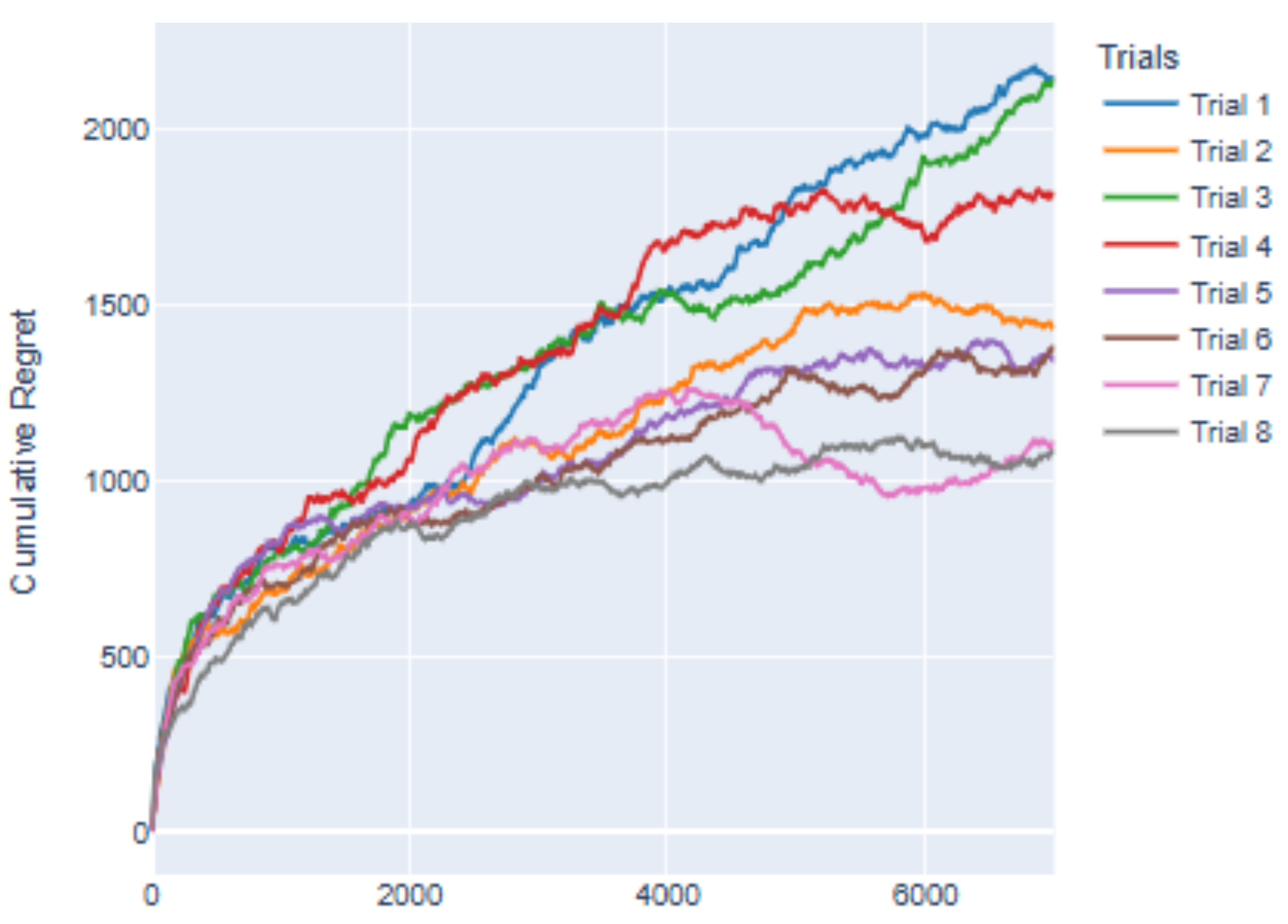


Fig. 20 - The regret for every trial archived by EXP3.

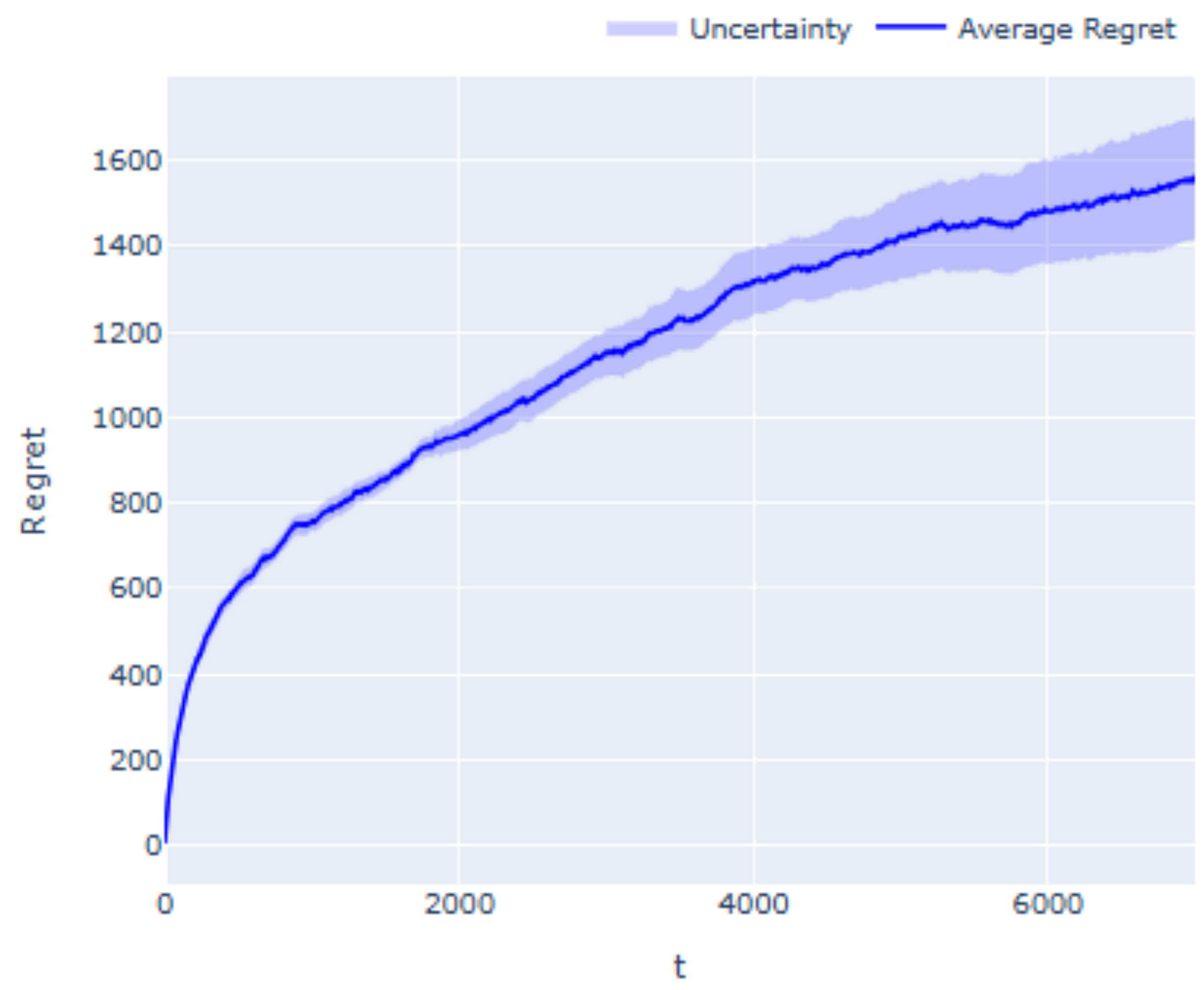


Fig. 21 - The average regret archived over 8 trials by EXP3.

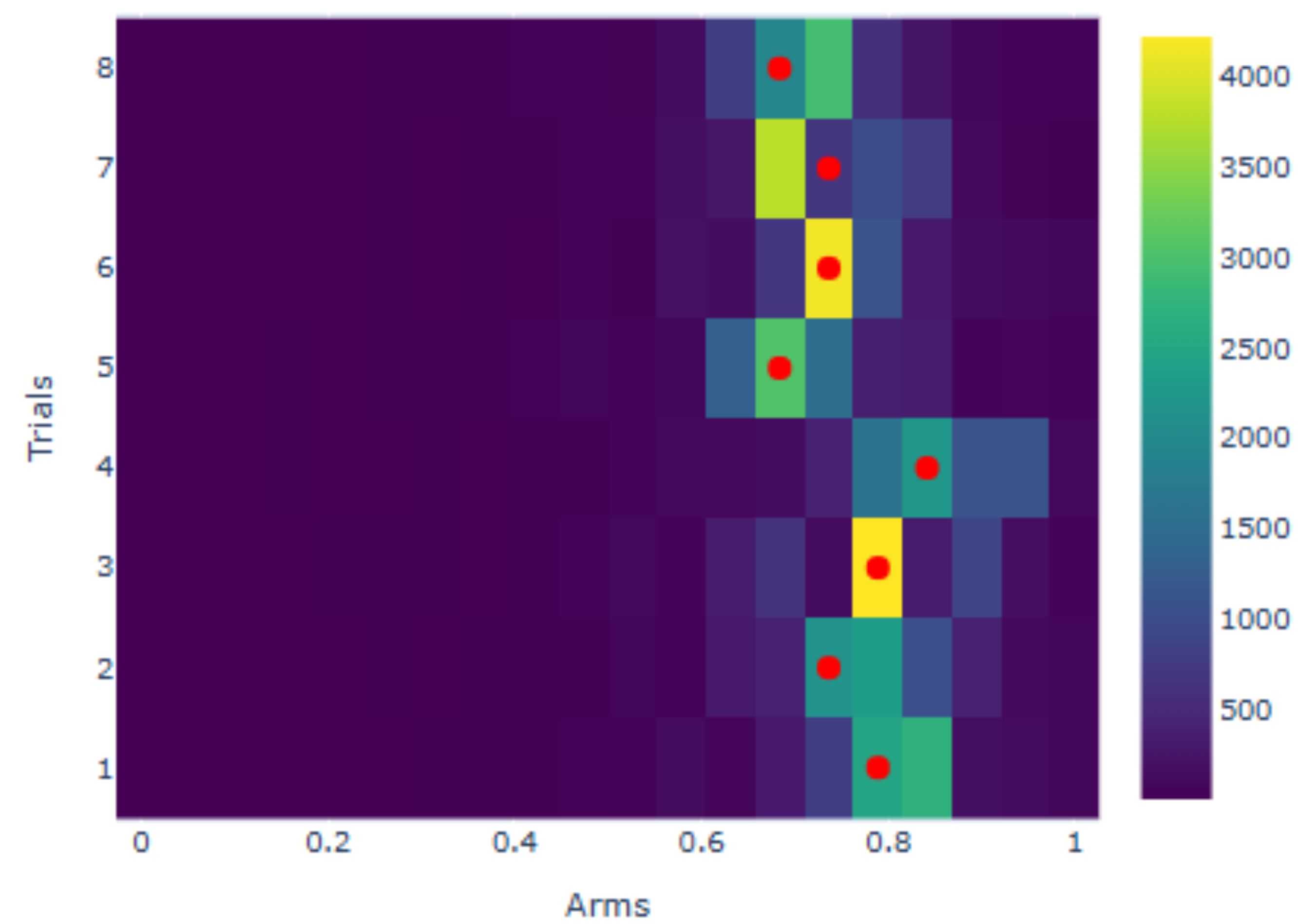


Fig. 22 - The chosen prices for each trials by EXP3.

Adversarial environment - Pricing Hedge

Here we employ **Hedge** on discretized set of prices $p \in [0,1]$ considering the following parameters: the cost c = 0.3, the number of rounds T = 7000, K = 20, the number of customer = 30. The execution have been made over 8 Trials.

Fig. 23 - The regret for every trial archived by Hedge

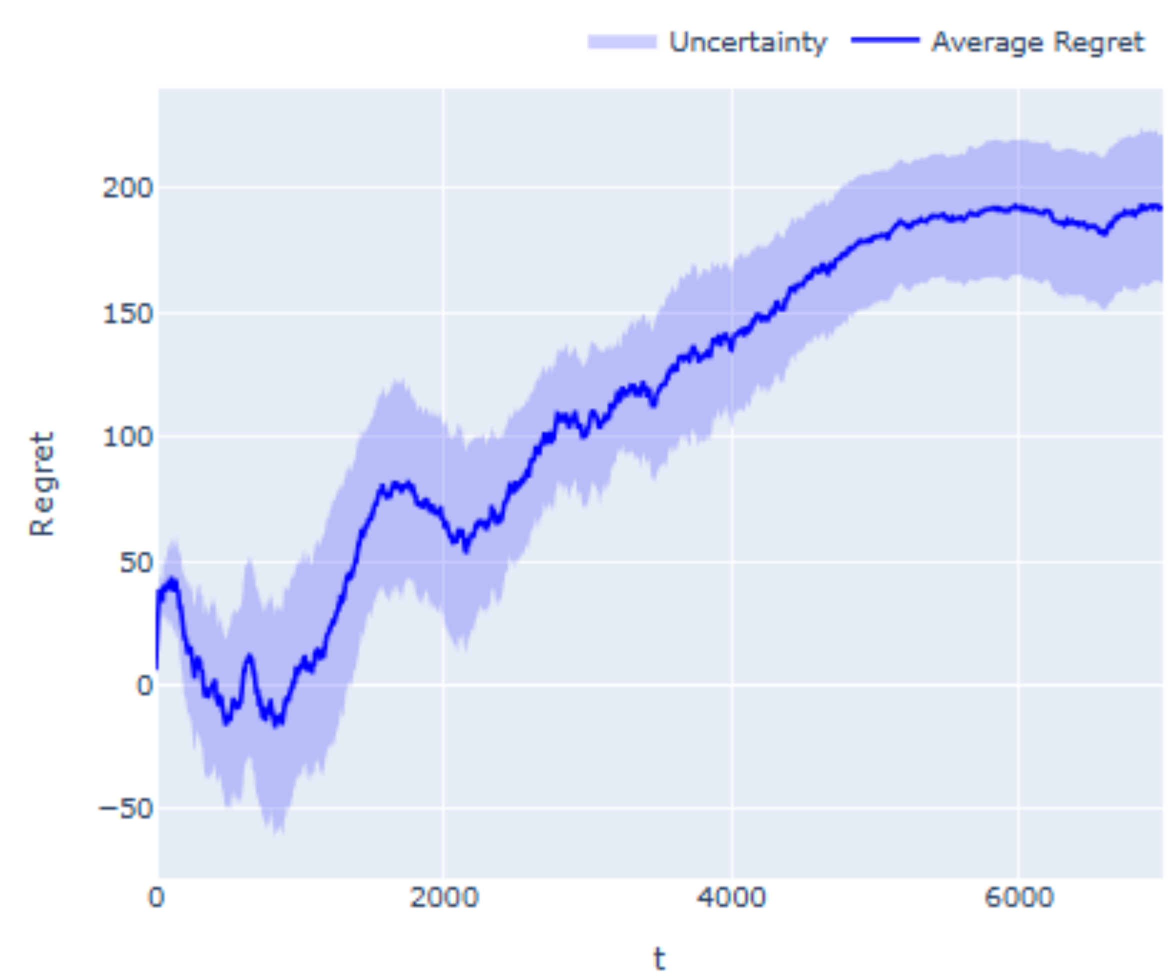


Fig. 24 - The average regret archived over 8 trials by Hedge.

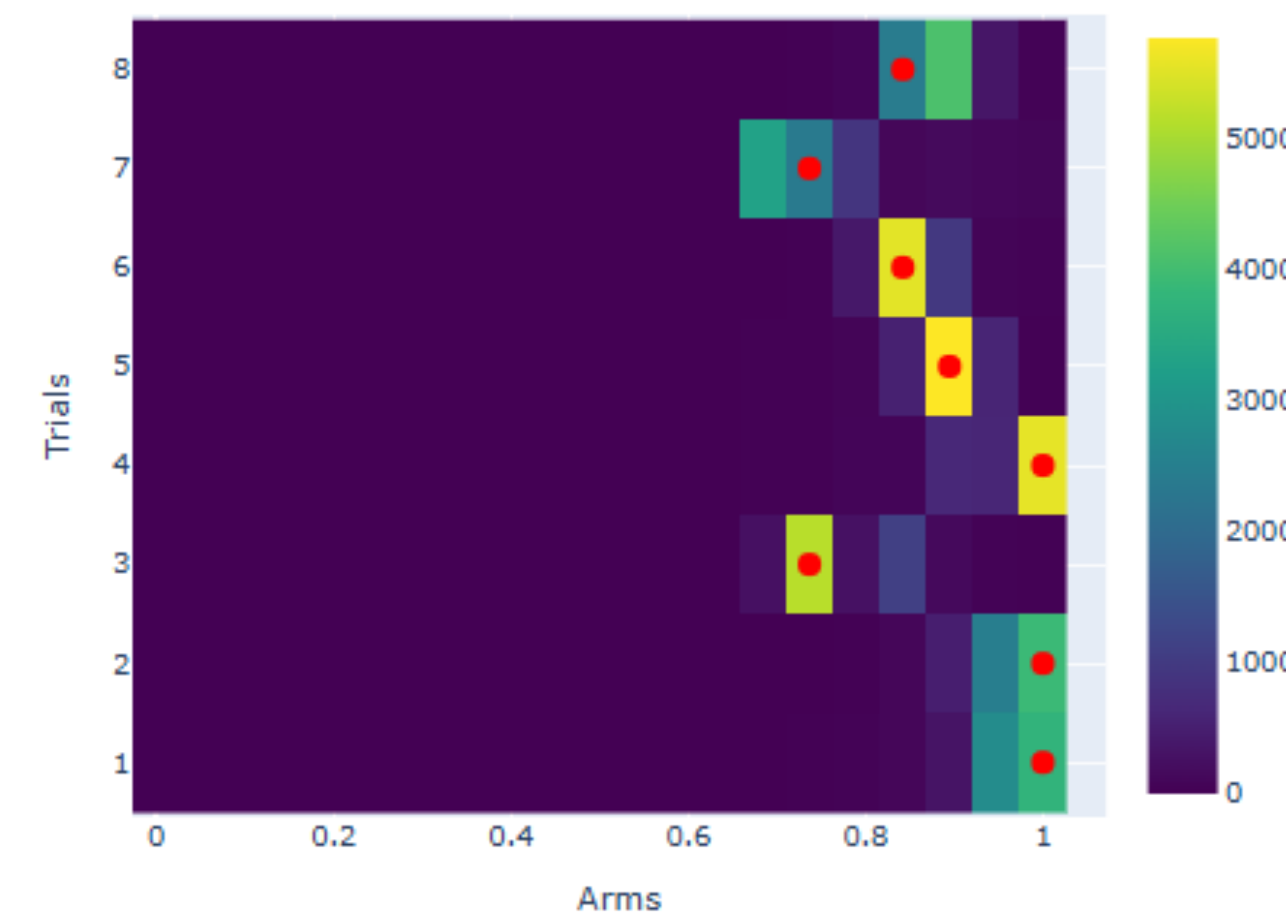


Fig. 25 - The choosen prices for each trials by Hedge.

OLA - Project Assignment
N. Balestrieri, A. Bertogalli, F. Cavalieri, N. Tombini

Adversarial environment - Bidding Hedge (MPS)

Here we use an Hedge Multiplicative Pacing Strategy on the bidding problem in adversarial environment (shown in the previous slide). Here we consider the following parameters: the budget B = 25, 5 advertisers, 150 users, value = 0.4, slots = 3.

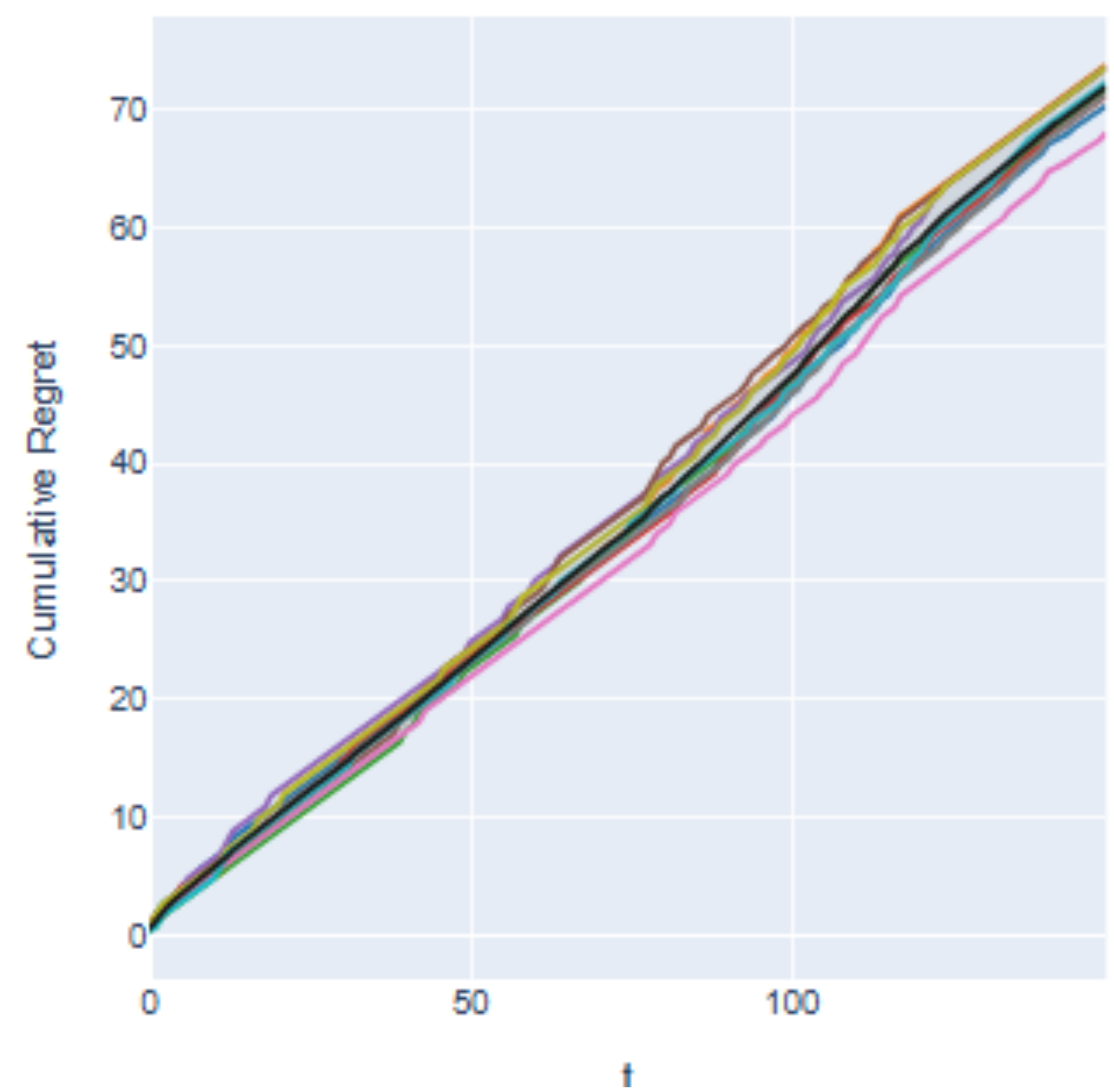


Fig. 26 - The regret for every trial archived by Hedge.

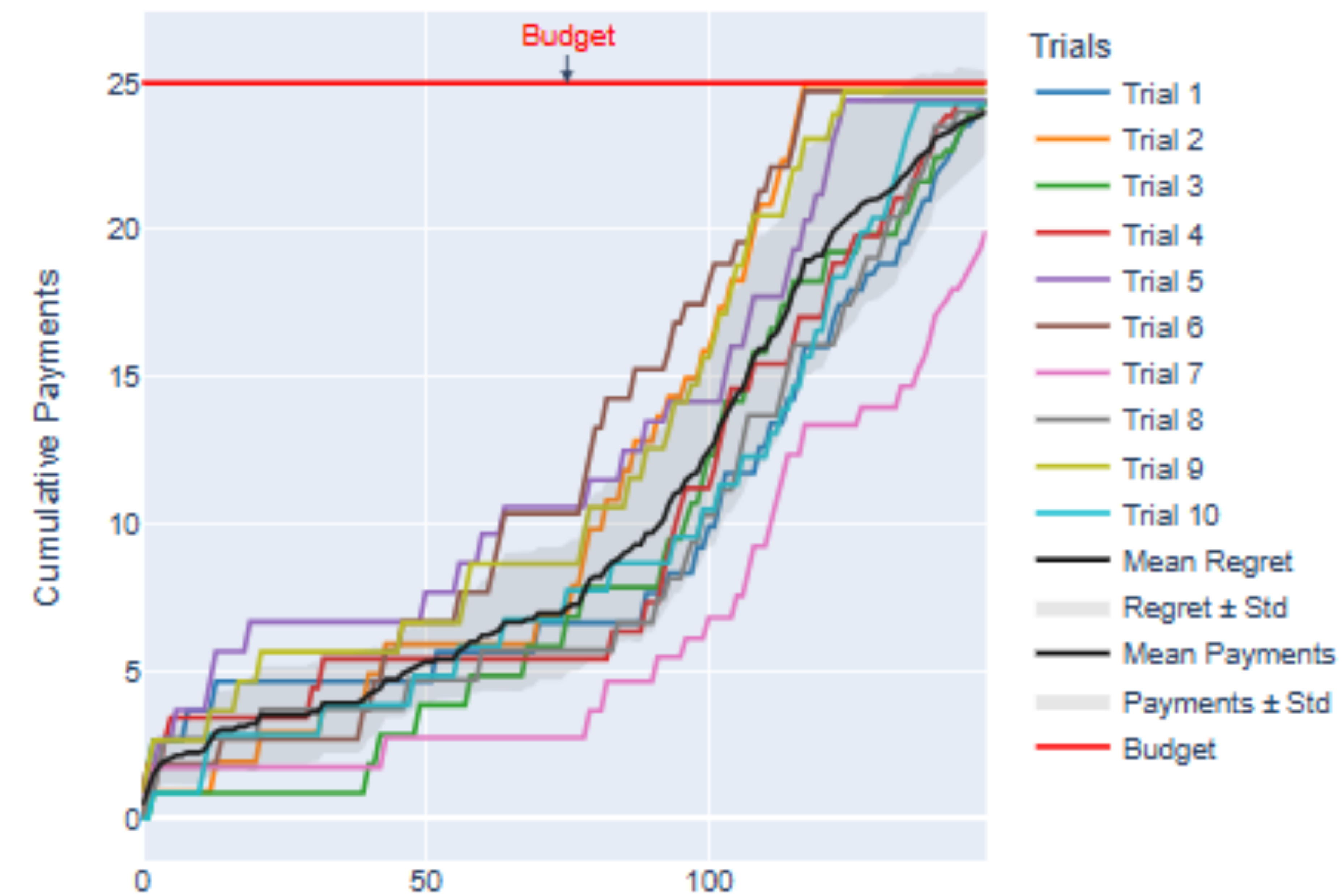


Fig. 27- The cumulative payments for each trial.

Adversarial environment - Pricing + Bidding

Here we evaluate the interaction with parameters: advertisers = 3, slot = 3, users = 500 (auction per day), value=0.6, budget=120, days=365, trials=2, K=20, cost=0.4, using EXP3 for the pricing and Hedge for the bidding.

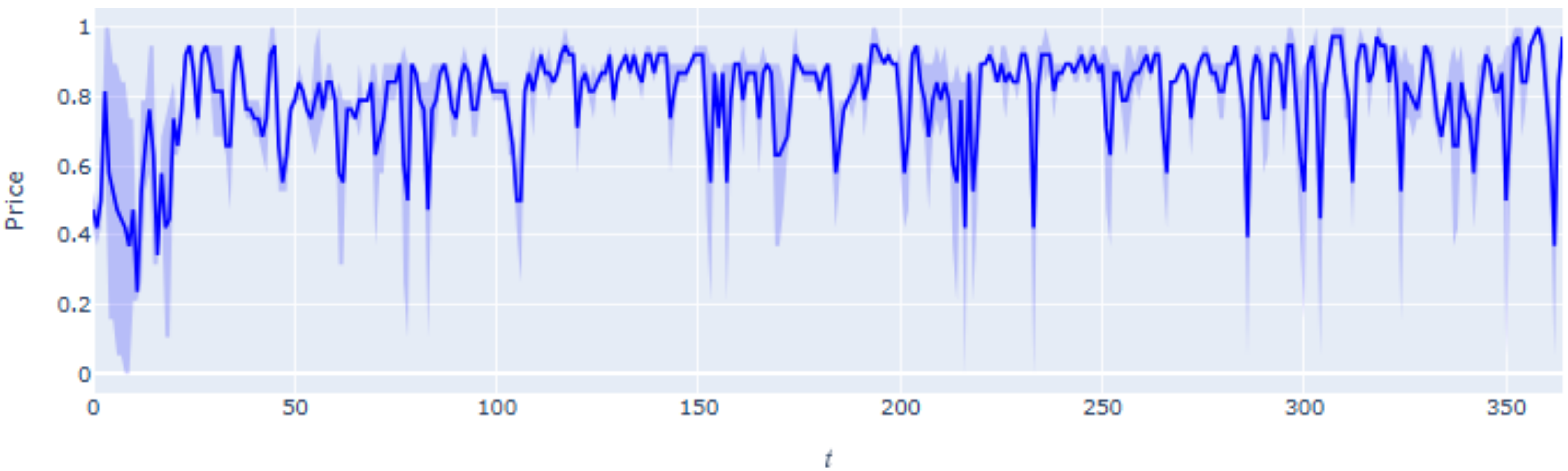


Fig. 28 - Here are represented the prices chosen by agent EXP3 during the campaign days.

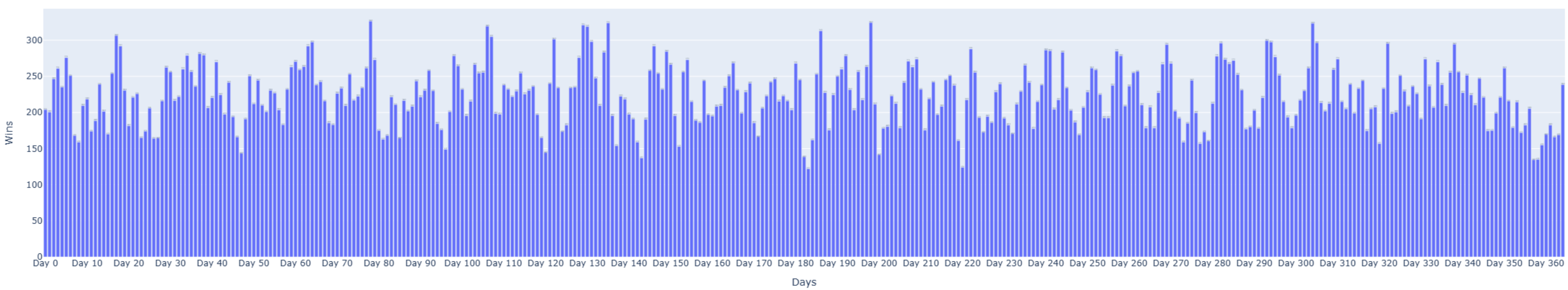


Fig. 29 - Here are represented the wins per day obtained during the entire duration of the campaign.

Adversarial environment - Pricing + Bidding

Here we evaluate the interaction with parameters: advertisers = 3, slot = 3, users = 500 (auction per day), value=0.6, budget=120, days=365, trials=2, K=20, cost=0.4.

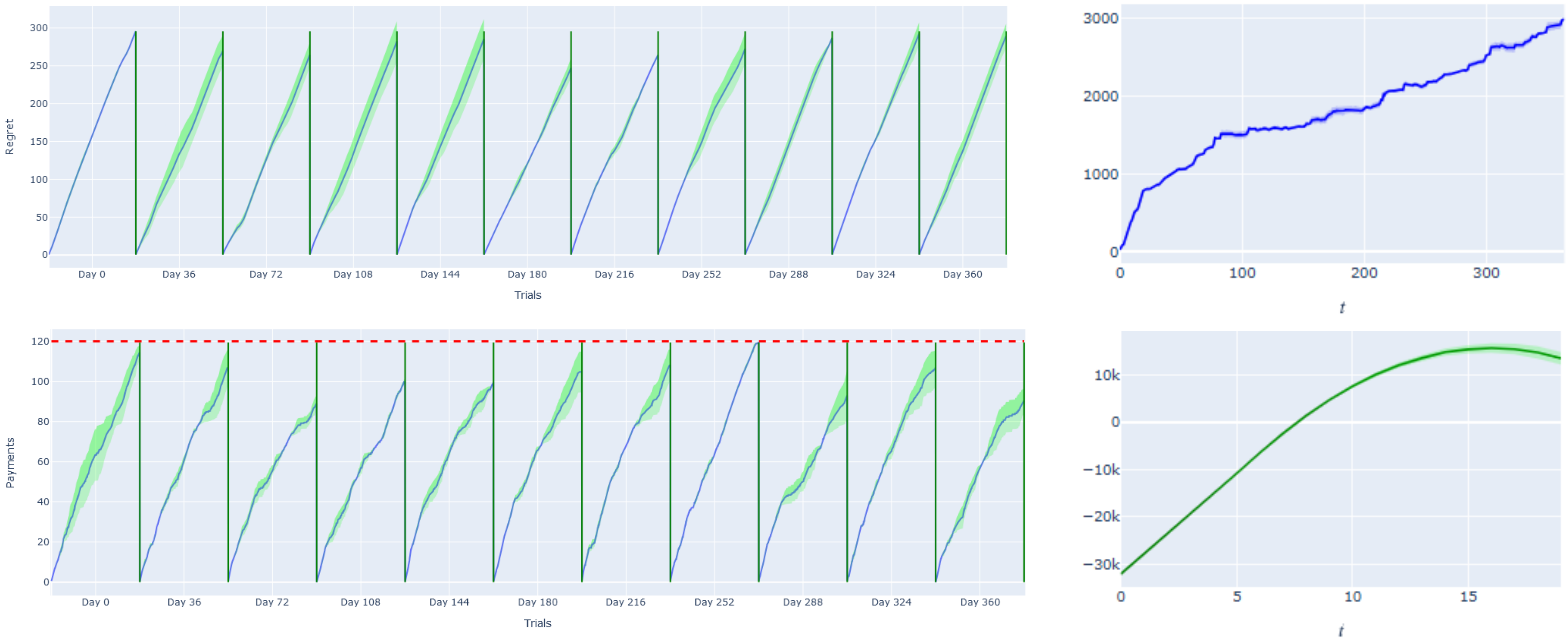


Fig. 30 - Here we show the bidding performances in 10 days in the years in terms of regret and payments.

Non-stationary environment - The setting

Here we are considering a non-stationary environment for the pricing problem. In particular days are partitioned in intervals, in each interval the demand curve is different. The demand curve with noise specifies how many buyers will buy for every price depending on the current interval. We focus on building a pricing strategy using the discretization of the prices $p \in [0, 1]$. To handle non-stationarity we employ **Sliding-Window** and **CUSUM**.

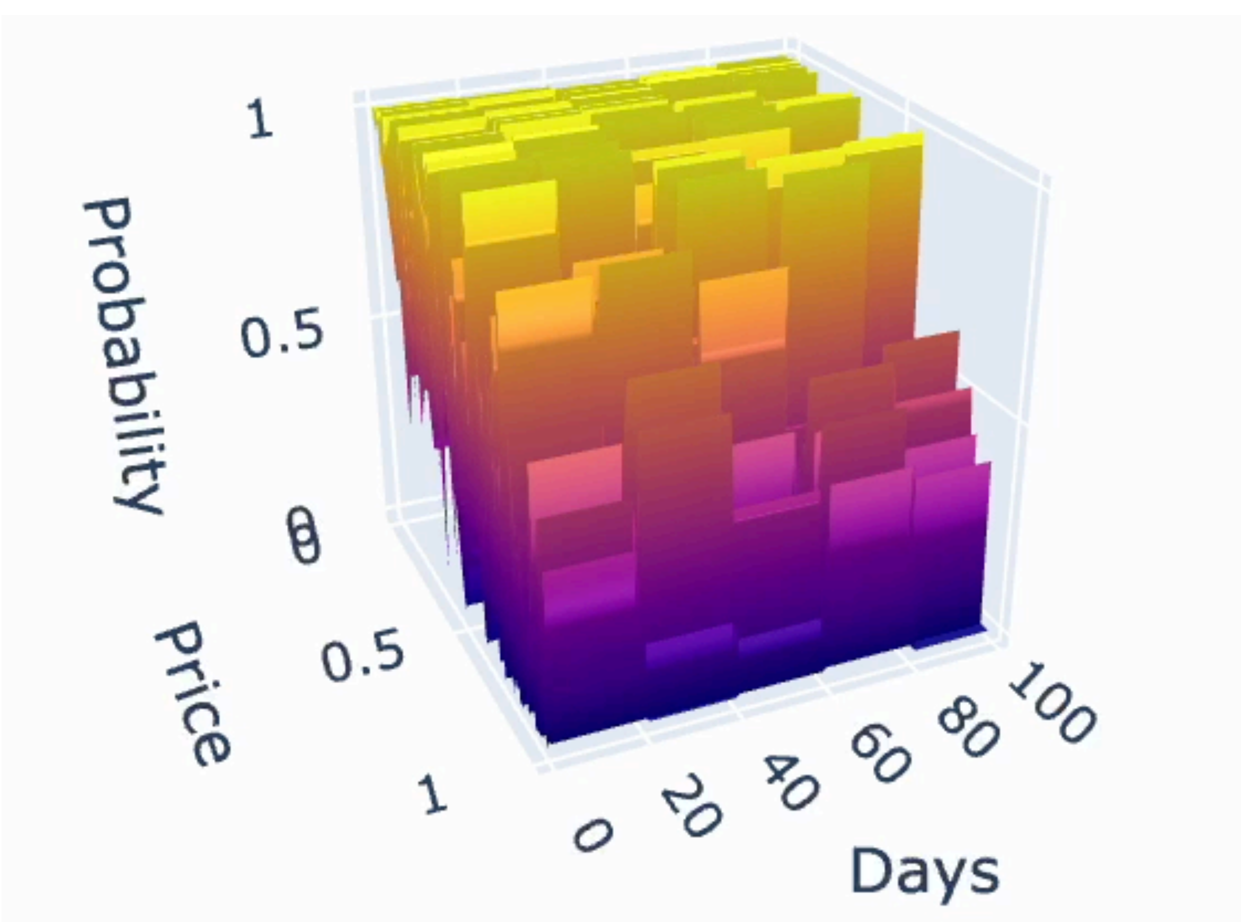


Fig. 31 - The demand curve expression: $D(t) = (1 - p) + \mathcal{N}(0, 0.30)$ where the noise is sampled every 20 rounds.

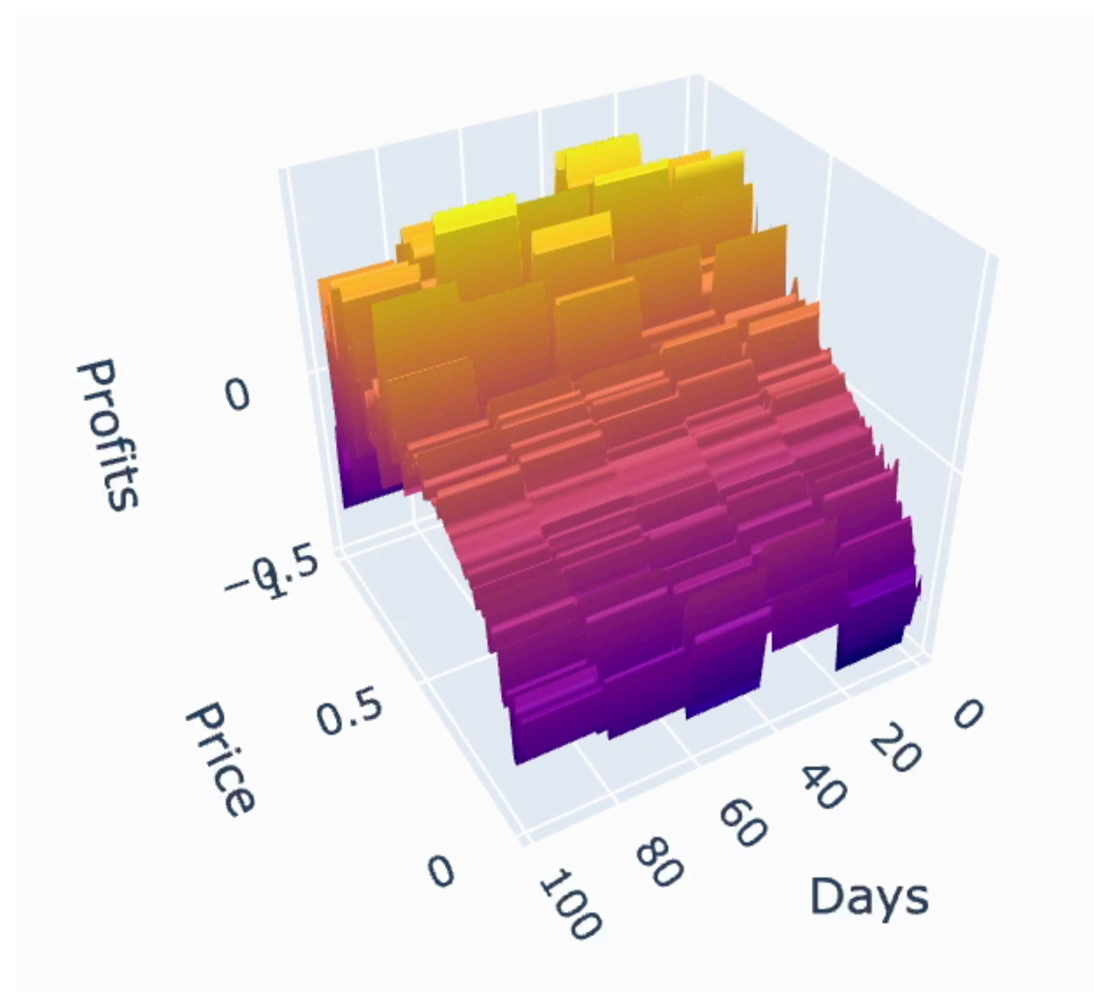


Fig. 32 - The profit curve

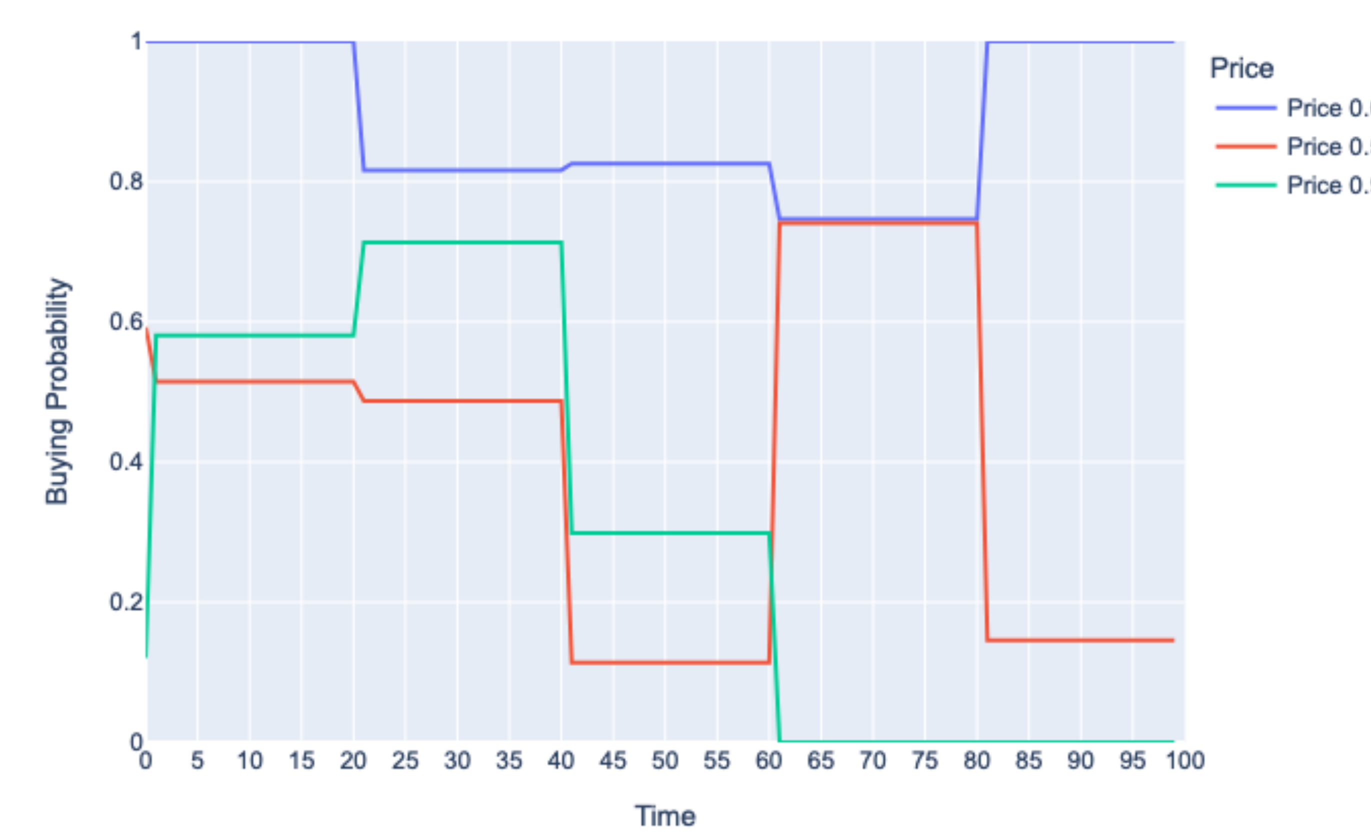


Fig. 33 - User's buying probability over time considering three possible prices

Non-stationary environment - CUSUM

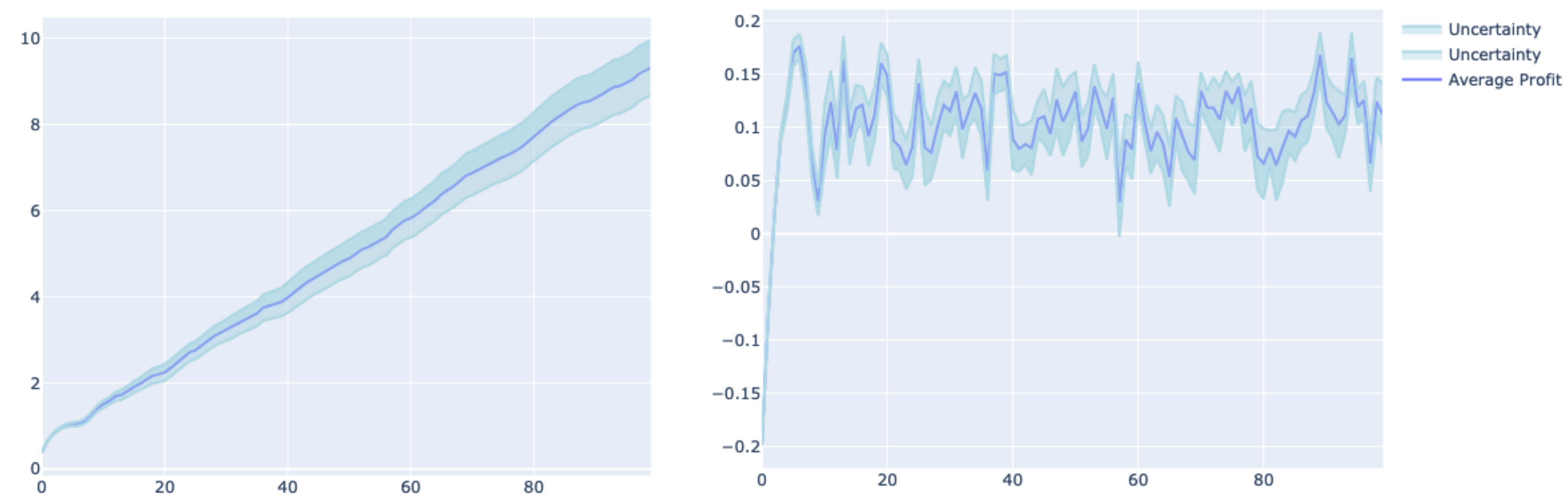


Fig. 34 - $H = 2.41, \alpha = 0.6, M = 1.$

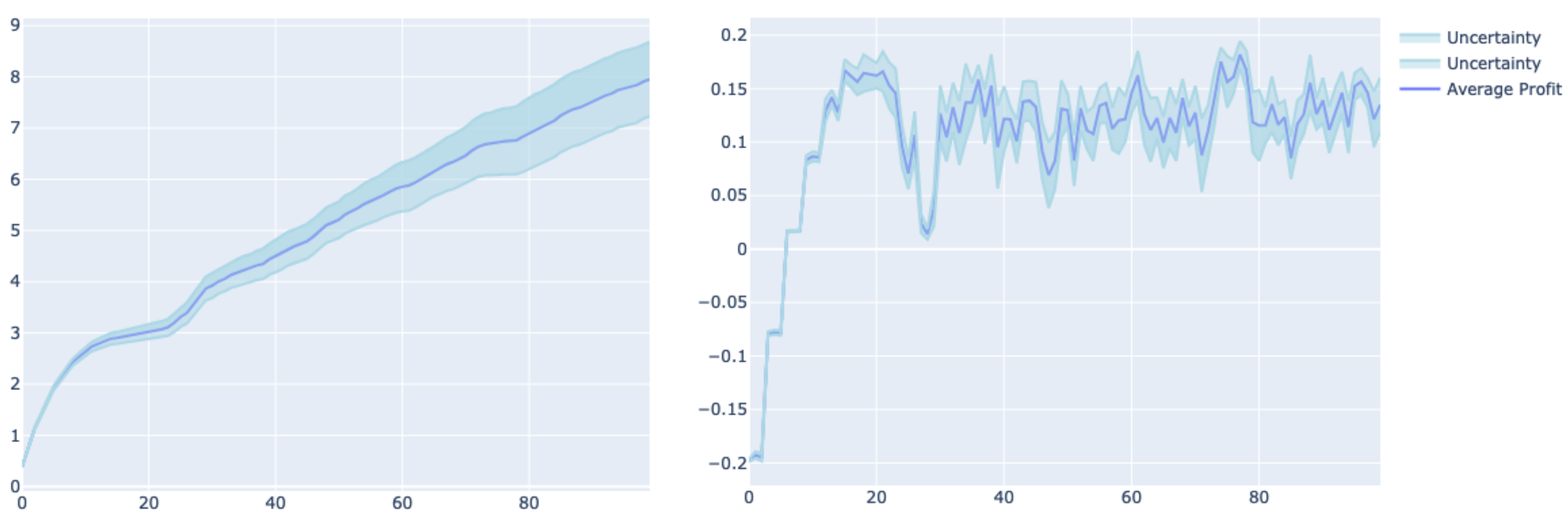


Fig. 35 - $H = 5.99, \alpha = 0.39, M = 3.$

Here we are considering the CUSUM-UCB Algorithm. We used the same non-stationary environment setting but changed the price discretization by using a linear-space with $K = 10$, we considered two different cases:

- One overestimating the number of abrupt changes ($U_t = 30$)
- Uno sapendo a priori il numero di abrupt changes ($U_t = 5$)

$$H = 2 * log(\frac{T}{U_T}) \qquad \alpha = \sqrt{U_T \cdot \frac{log(\frac{T}{U_T})}{T}}$$

$$M = log(\frac{T}{U_T})$$

Non-stationary environment -Sliding Window

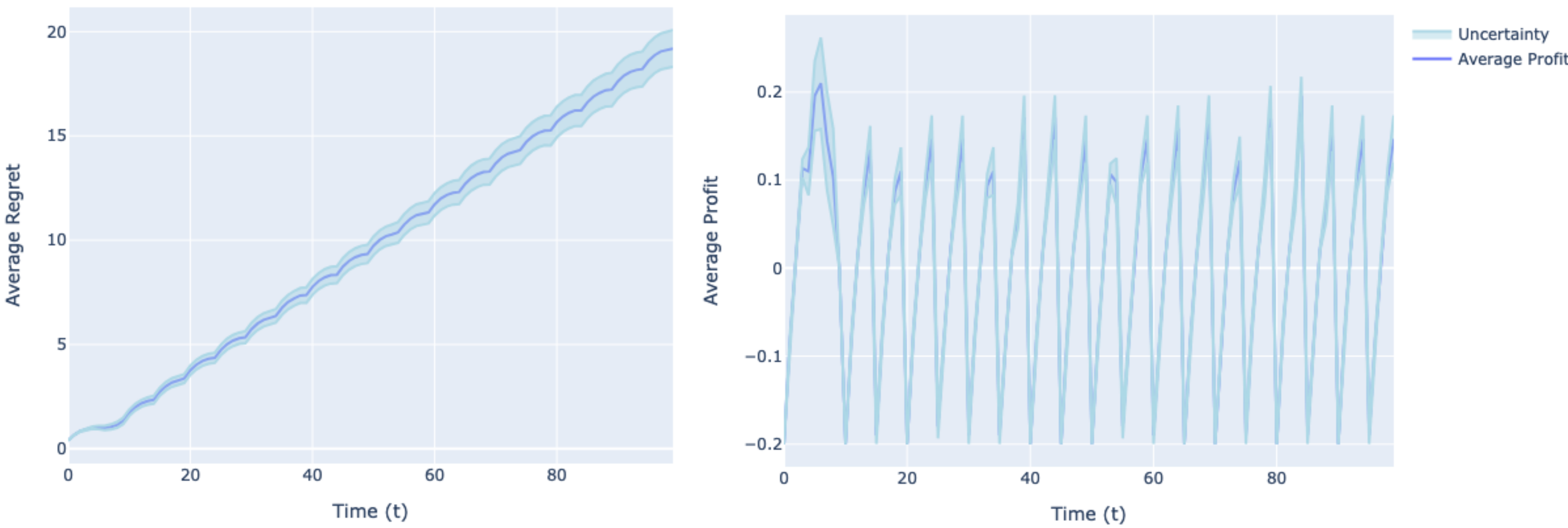


Fig. 36 - $W = 4, \quad U_T = T$

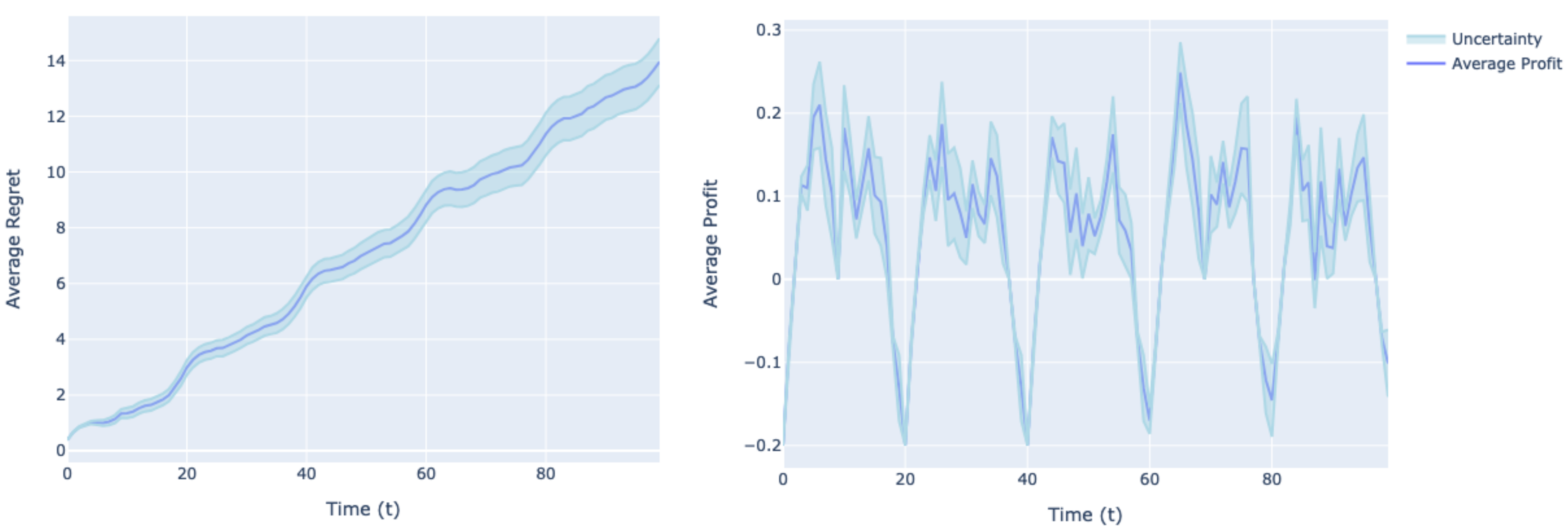


Fig. 37 - $W = 19, \quad U_T = 5$

Here we are considering the Sliding Window - UCB Algorithm, using the same non-stationary setting but changing the price discretization by using a linear-space with $K = 10$ we considered two different cases:

- Knowing a priori the number of abrupt changes

$$W = \left\lceil 2B\sqrt{\frac{T \log T}{\Upsilon_T}} \right\rceil, U_T = 5$$

- no prior information

$$W = \left\lceil 2B\sqrt{\frac{T \log T}{\Upsilon_T}} \right\rceil, U_T = T$$

Non-stationary environment - Two products

Here we are considering a two-item stochastic pricing environment. In particular, the seller proposes two items i_1 and i_2 the demand curve $D(p_1, p_2)$ is a two variable function + noise that specifies how many buyers will buy product i_1 and how many will buy product i_2 depending on price p_1 of product i_1 and price p_2 of product i_2.

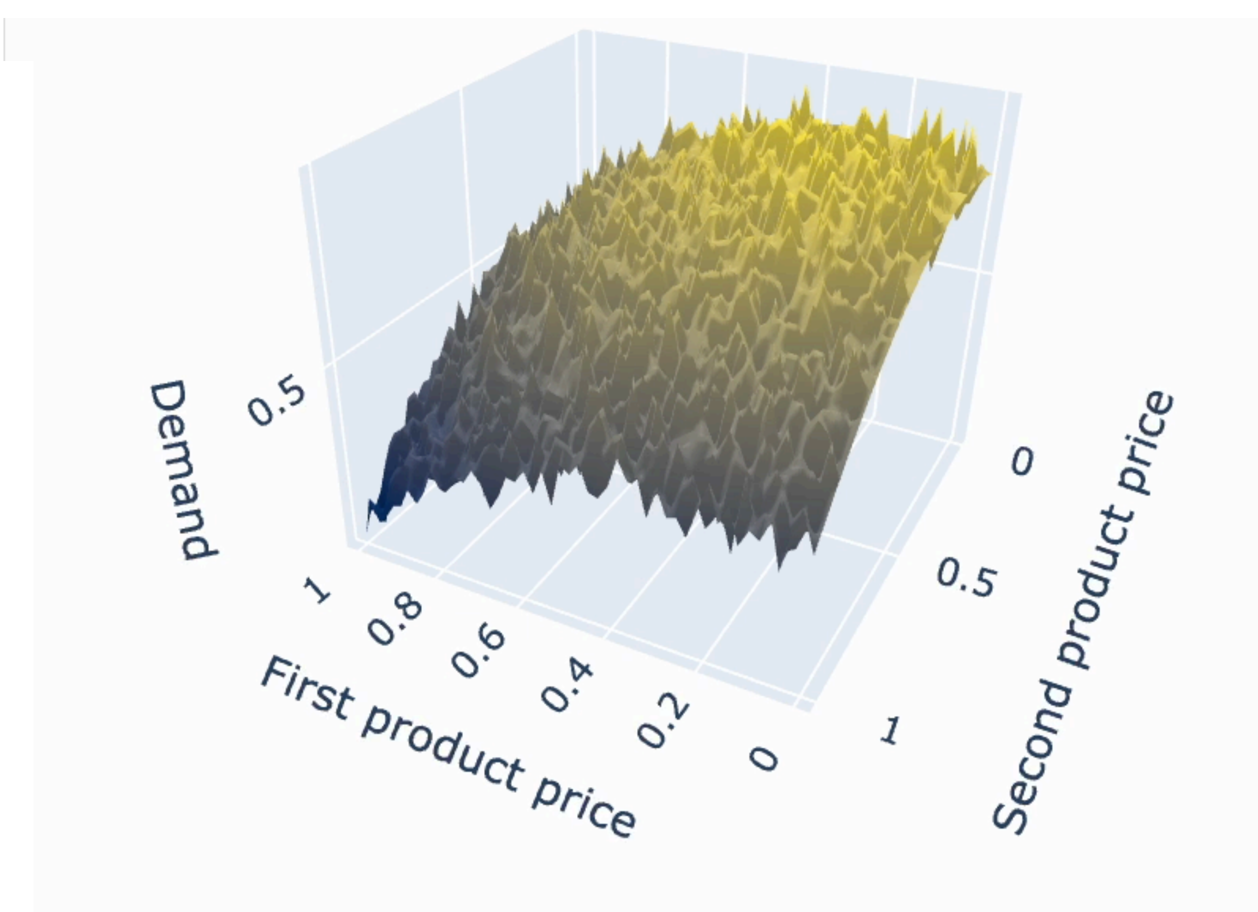


Fig. 38 - $D(t) = -0.8 \cdot \left((p_1 - 0.4)^2 + (p_2 - 0.6)^2\right) + 1.0 + \mathcal{N}(0,\ 0.04)$ the demand curve for the product 1.

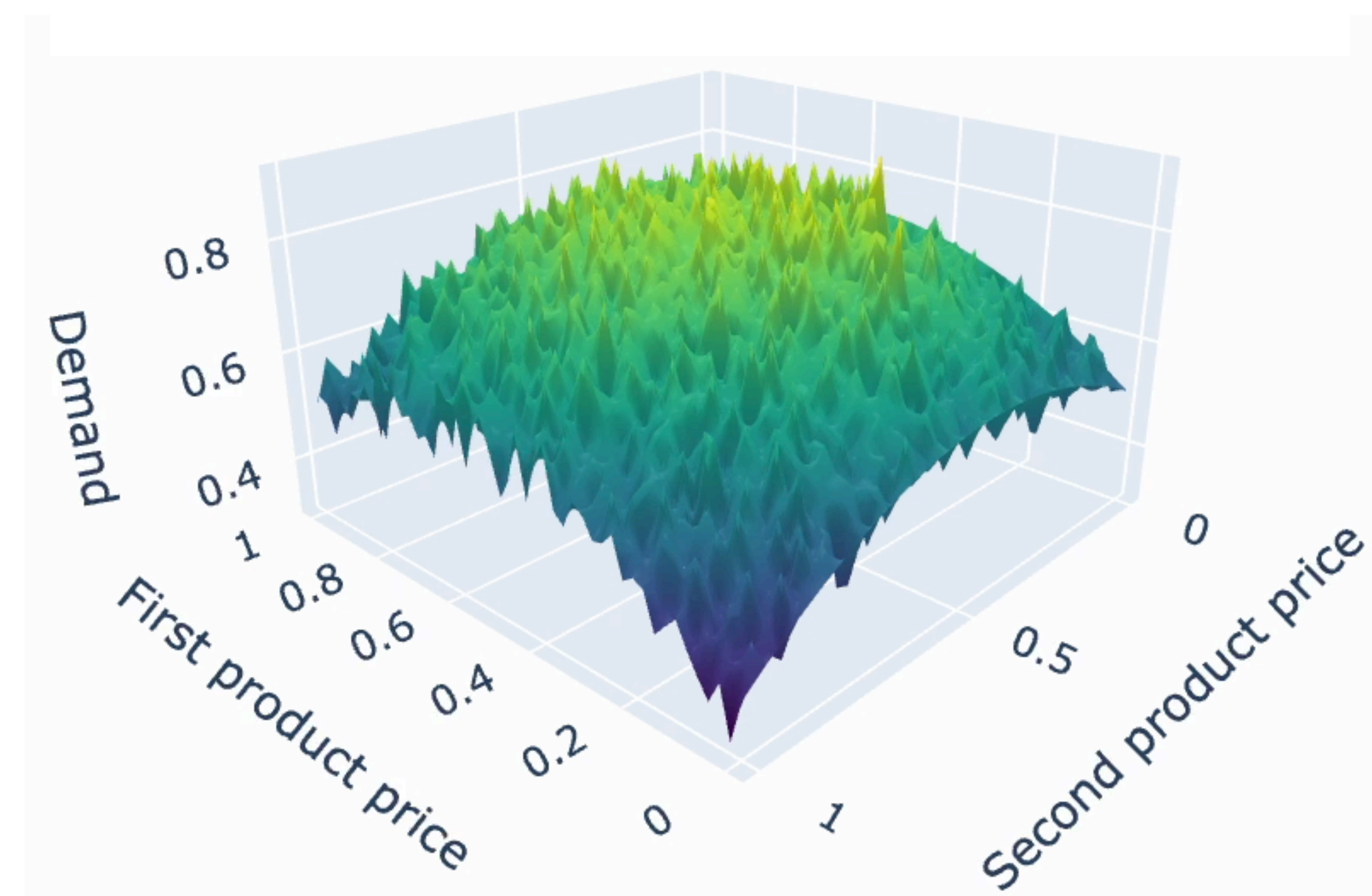


Fig. 39 - $D(t) = -0.8 \cdot \left((p_1 - 0.2)^2 + (p_2 - 0.3)^2\right) + 1.0 + \mathcal{N}(0,\ 0.04)$ the demand curve for the product 2.

Non-stationary environment - Two products

Here we are considering a two-item stochastic pricing environment. In particular, the seller proposes two items i_1 and i_2 the demand curve $D(p_1, p_2)$ is a two variable function + noise that specifies how many buyers will buy product i_1 and how many will buy product i_2 depending on price p_1 of product i_1 and price p_2 of product i_2.

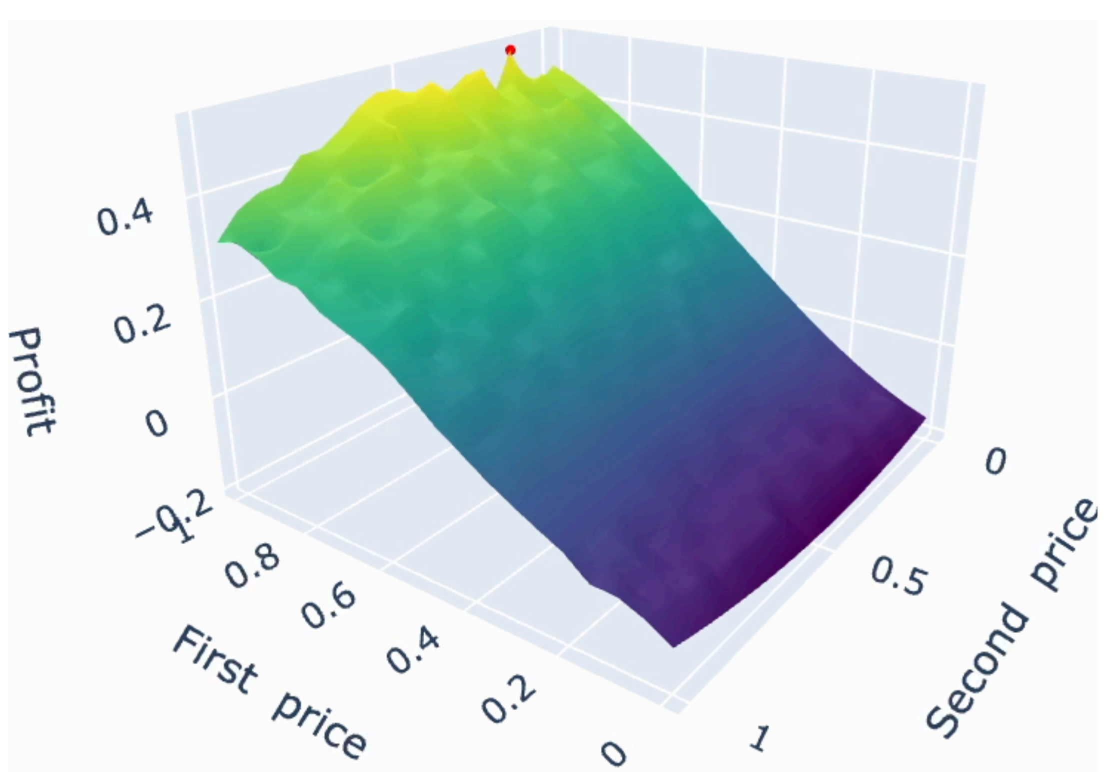


Fig. 40 - The profit curve for the product 1.

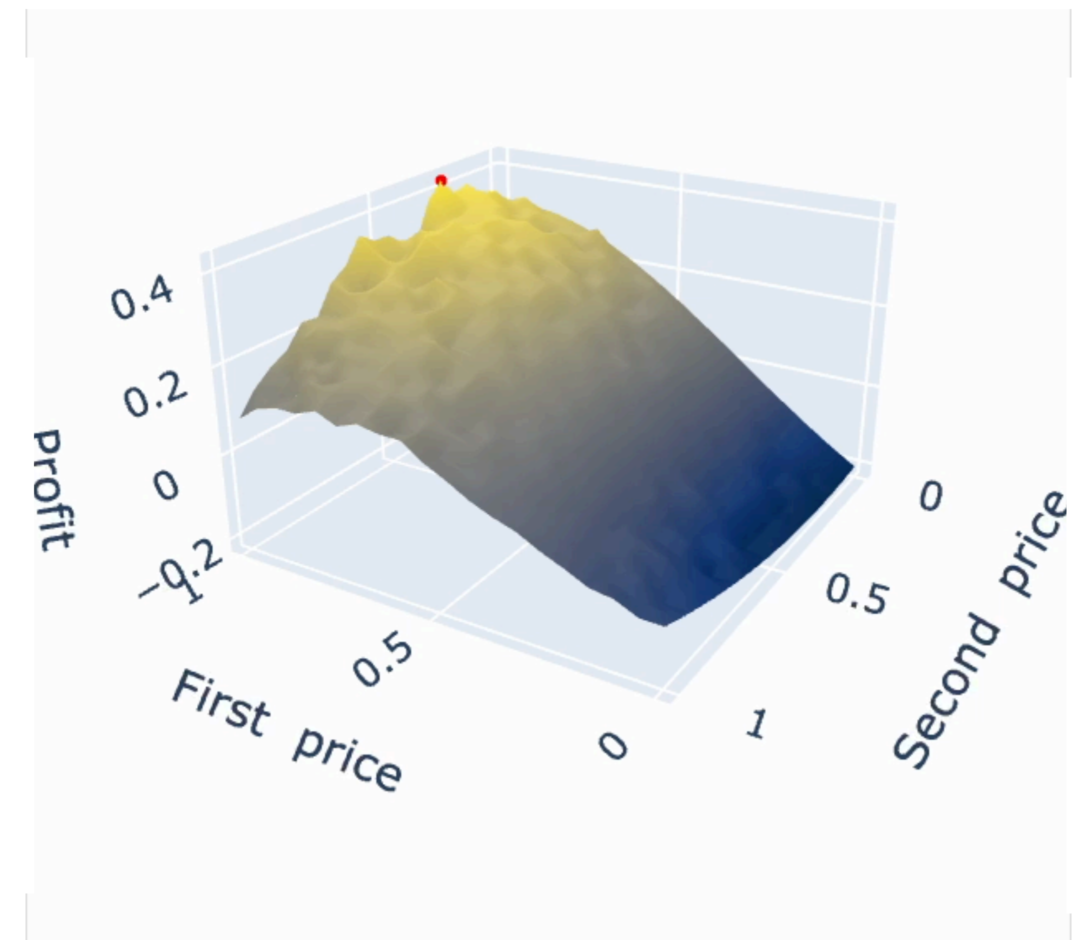


Fig. 41 - The profit curve for the product 2.

Non-stationary environment - Two products results

To model this type of enviroment we used a Gaussian UCB approach, using a Gaussian Regressor to estimate the profit curve, optimizing it along the couple of price $p_1,\ p_2$that were be used as arms. This type of approach permitted us to use the same pipeline used at with one product even if the problem has 2 different products.

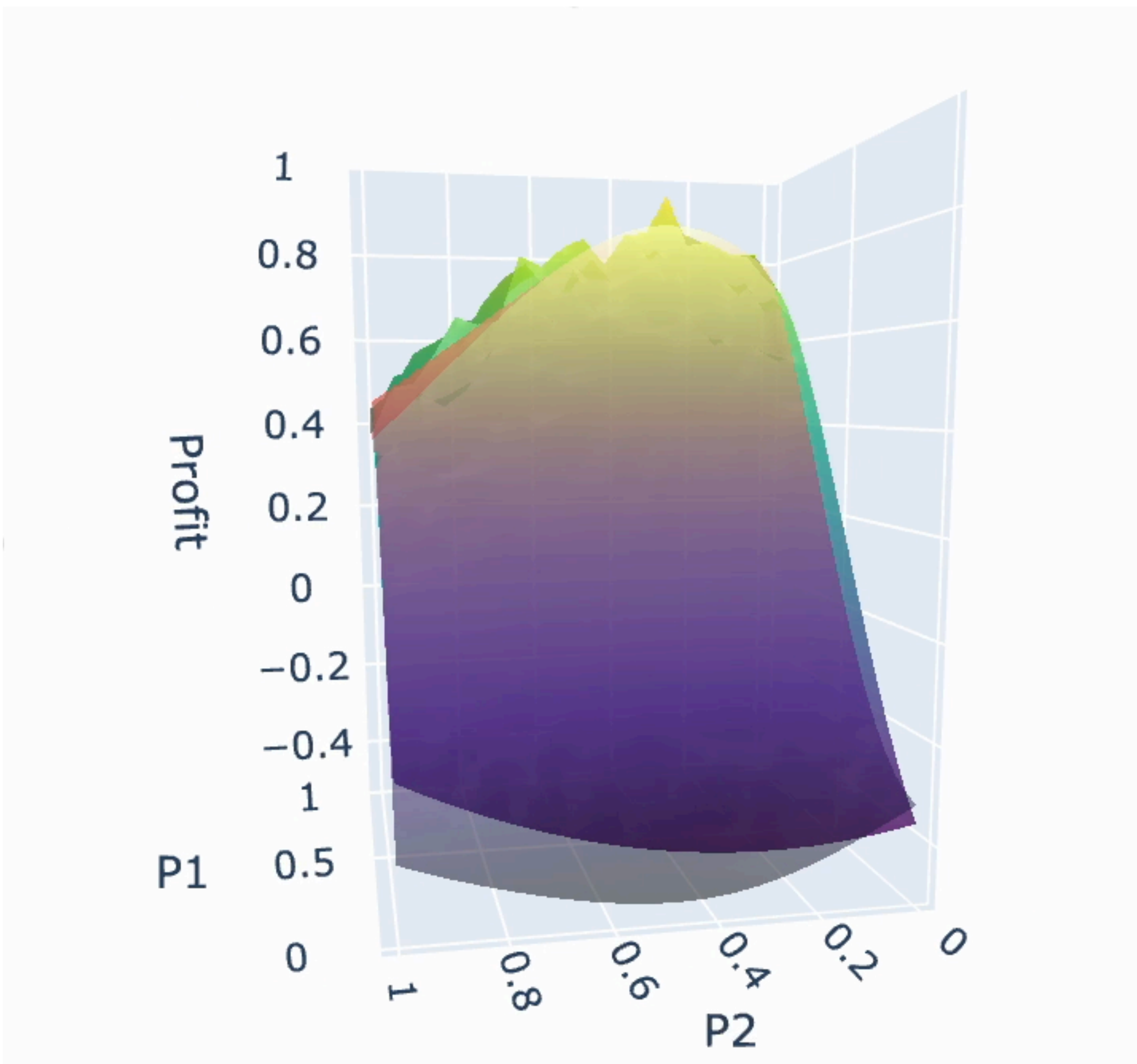


Fig. 42 - Estimated profit curve.

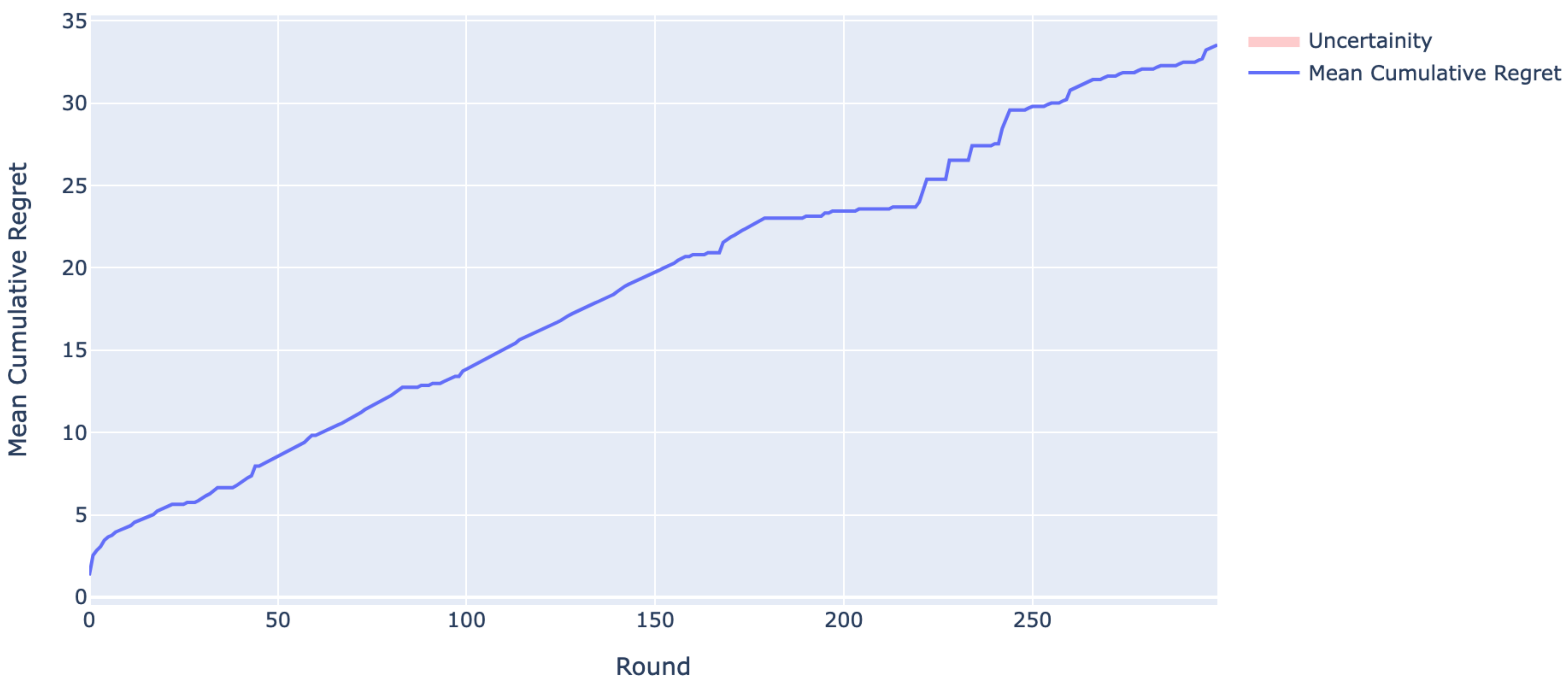


Fig. 43 - GP-UCB cumulated regret.

Competition among algorithms - The setting

The goal here is to compare the performances of different bidding algorithms, competing over a **Generalized First Price Auction**, in particular we considered 3 algorithms:

- A primal-dual algorithm for truthful auctions (Multiplicative Pacing Strategy)
- A primal-dual algorithm for non-truthful auctions (Hedge Multiplicative Pacing)
- A UCB-like approach (UCB-Like Multiplicative Pacing)

We also evaluated different parameters for the algorithms i.e. we tried different valuations for each algorithm. The parameters considered are:

SLOTS	TRIALS	BUDGET	VALUATIONS	K	USERS
2	5	100	$v \in \{0.3, 0.5\}$	10	600

Fig. 44 - The considered parameters.

Competition among algorithms - Comparison

Here we compare the results of the competition among algorithms:

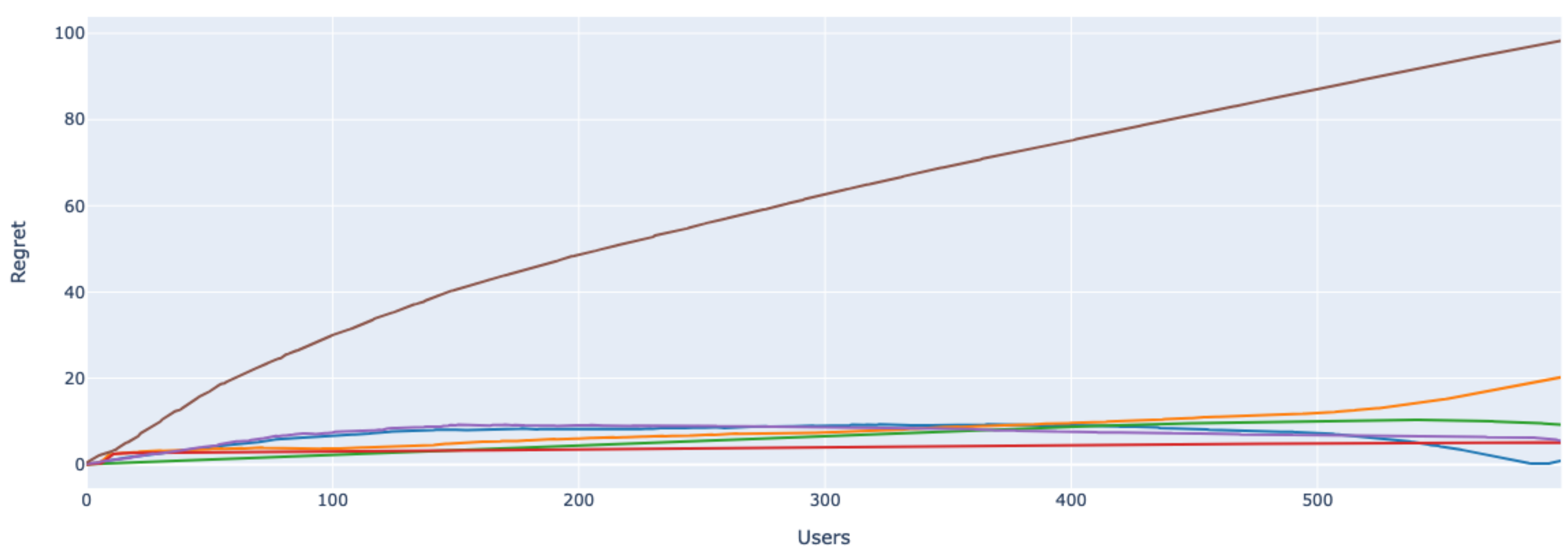


Fig. 45 - Cumulative regret per algorithm.

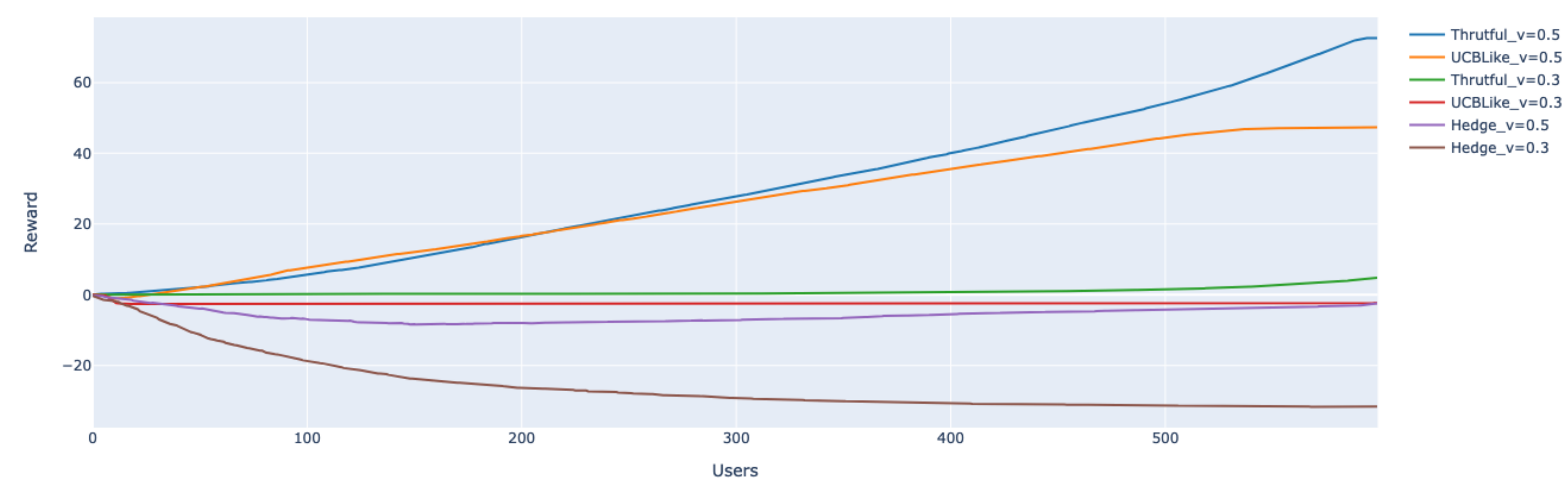


Fig. 46 - Cumulative reward per algorithm.

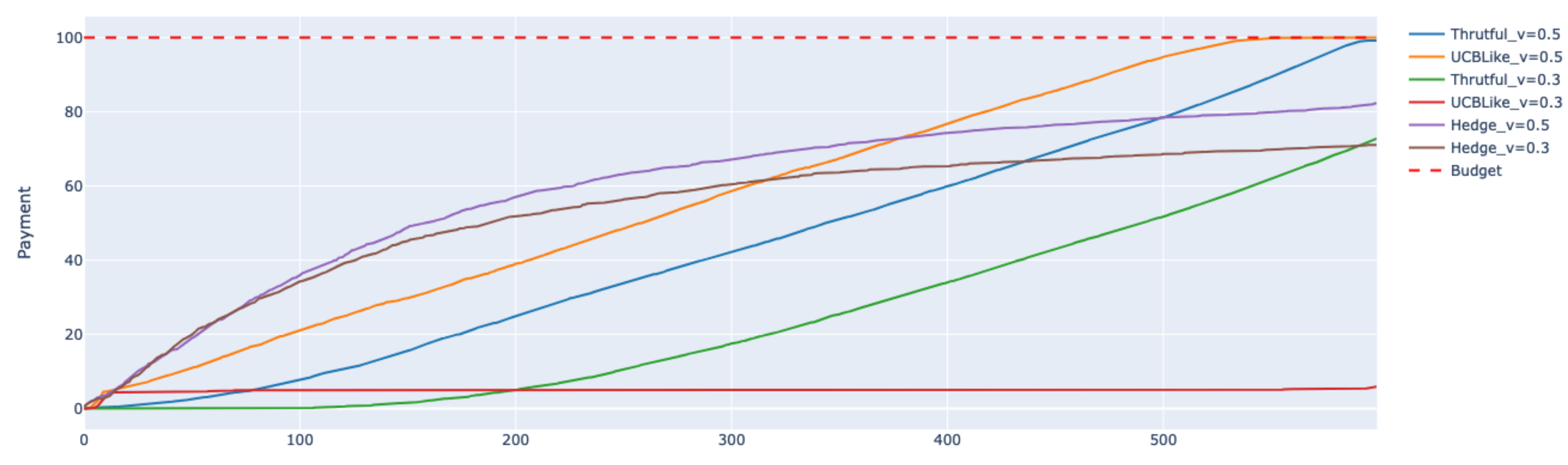


Fig. 47 - Cumulative payments per algorithm.

POLITECNICO
MILANO 1863

Thanks

Authors:
Balestrieri Niccolò, Bertogalli Andrea, Cavalieri Francesco, Tombini Nicolò

Professors:
Matteo Castiglioni, Gianmarco Genalti

Accademic year 2024/2025