Paper Report

Dynamic Online Pricing with Incomplete Information Using Multiarmed Bandit Experiments

Kanishka Misra,
Eric M. Schwartz,
Jacob Abernethy

CONTENTS

Introduction

- Propose an alternative dynamic price experimentation policy for online shopping
- Extends the problem of multiarmed bandit (MAB)
- The tweaked algorithm is proved to be analytically asymptotically optimal
- The methods are then tested under simulations and field experiments and are proved to be raising long term profits

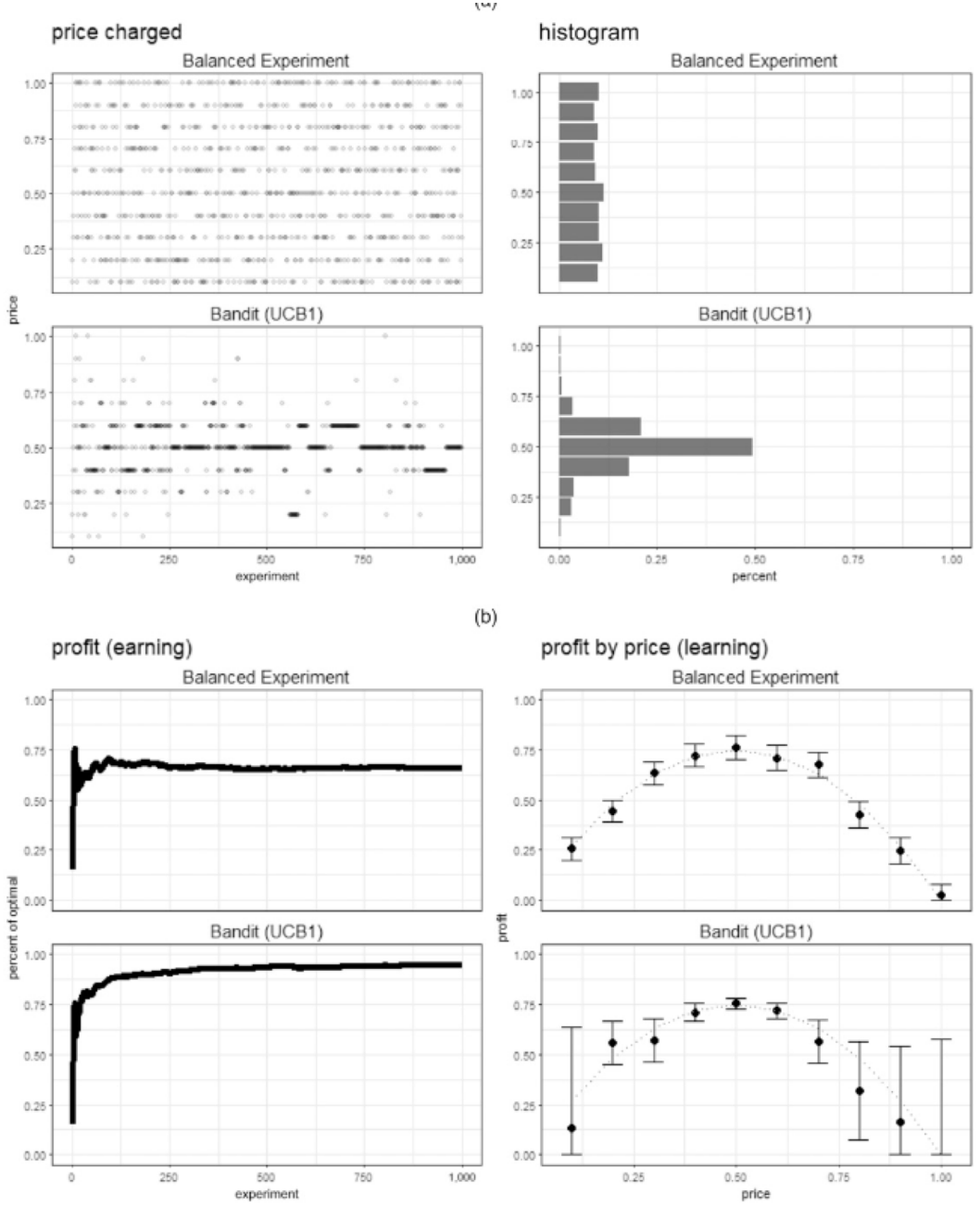
Keypoints

E-commerce vs Physical Stores

- The amounts of products selled
- The accessibility of price change (Menu Cost)

Keypoints

Traditional Field Experiments
vs
Multiarmed Bandit Experiments



Literature Reviews - Pricing

- Pricing are usually considered under parametric distribution
- Firms get information about demands by changing prices and see the correspondent demand
- Here, the goal is to minimize the maximum regret, I,e,, the distance to the optimized price cause by pricing strategies
- Price changes can only happen across time but not consumers.
- The relationship between utility and preference is decided by

  $u = f(Z) + v - p$, where Z is the customer data a firm holds

Literature Reviews - MAB

The key components of MAB problems

- Output -> Demand (D(Pk))
- Arms -> Different Prices (Pk)
- Times -> Rounds of experiment (T)
- Total Return -> Final profits (sum of $\pi(p)$)
- Regret -> loss due to strategies

$$\begin{aligned}\mathrm{Regret}(\Psi, \{\pi(\mathrm{p_k})\}, t) &= \mathbb{E}\left[\sum_{\tau=1}^{t} \pi^* - \pi_{\mathrm{p}_\tau}\right] \\ &= \sum_{\tau=1}^{t} (\pi^* - \pi(\mathrm{p}_\tau)) \\ &= \pi^* t - \sum_{k=1}^{K} \pi(p_k)\mathbb{E}\left[n_{kt}\right],\end{aligned}$$

The index rules do not provide the exactly optimal solution under finite horizons. However, the expected sum of return is maximized as T approaches infinite.

Assumptions

- Stable preference for a single customers
- Stable budget
- Stable outside option
- Choice structure under WARP

- The segment s customer I belongs to is known through the observable data

- The profit outcomes in any two actions are independent

- Fully nonparametrc heterogeneity across segments

Notations

Notation	Description
p	Price, with $p \in \{p_1, \ldots, p_K\}$
$D(p)$	Demand at price p
$\pi(p)$	Profit at price p (i.e., $\pi(p) = pD(p)$))
v_{is}	Valuation of individual consumer i in segment s
v_s	Midpoint of valuations in segment s, with its estimated value $\hat{v}_{st}$ after t rounds
δ	Estimated range of valuations across with $v_{is} \in [v_s - \delta, v_s + \delta_s]$, $\forall s$
$\hat{\delta_{st}}$	Lowest value of δ that can rationalize all data for segment s after t rounds
$\hat{\delta}_t$	Estimated range of valuations across all segments at time t
Ψ	Policy for dynamic pricing (i.e., data-driven decision rule $p_{t+1} = \Psi(\{p_1, \pi_{p_1}, \ldots, p_t, \pi_{p_t}\})$)
n_{kt}	Number of times price p_k was tested through time t
n_{st}	Number of times any price was tested with segment s through time t
ψ_s	Percentage of consumers in segment s
$LB_t(x)$, $U, B_t(x)$	Estimated lower and upper bounds through t of a parameter x (e.g., either v_{is}, $D(p)$, or $\pi(p)$)
$H_t(x)$	Partially identified set through t of a parameter x (i.e., $H_t(x) \equiv [LB_t(x), U, B_t(x)]$)
UCB	Upper confidence bound, with its original implementations **UCB1**
UCB-PI	Upper confidence bound with partial identification (proposed in this paper)

Proposed Methods

$$\text{UCB1}_{kt} = \bar{\pi}_{kt} + \sqrt{\frac{2\log t}{n_{kt}}}.$$

$$\text{V}_{kt} = \left(\frac{1}{n_{kt}}\sum_{\tau=1}^{n_{kt}} \pi_{k\tau}^2\right) - \bar{\pi}_{kt}^2 + \sqrt{\frac{2\log t}{n_{kt}}}$$

$$\text{UCB-tuned}_{kt} = \bar{\pi}_{kt} + \sqrt{\frac{\log t}{n_{kt}}\min\left(\frac{1}{4}, \text{V}_{kt}\right)}.$$

$$\textbf{UCB-PI-untuned}_{kt} = \begin{cases} \bar{\pi}_{kt} + p_k\sqrt{\frac{2\log t}{n_{kt}}} & \text{if } UB_t(\pi(p_k)) > \max_l LB_t(\pi(p_l)), \\ 0 & \text{if } UB_t(\pi(p_k)) \leq \max_l LB_t(\pi(p_l)). \end{cases}$$

$$\text{V}_{kt} = \left(\frac{1}{n_{kt}}\sum_{\tau=1}^{n_{kt}} \pi_{k\tau}^2\right) - \bar{\pi}_{kt}^2 + \sqrt{\frac{2\log t}{n_{kt}}}.$$

$$\textbf{UCB-PI-tuned}_{kt} = \begin{cases} \bar{\pi}_{kt} + 2p_k\hat{\delta}\sqrt{\frac{\log t}{n_{kt}}\min\left(\frac{1}{4}, \text{V}_{kt}\right)} \\ \qquad \text{if } UB_t(\pi(p_k)) > \max_l LB_t(\pi(p_l)), \\ 0 \qquad \text{if } UB_t(\pi(p_k)) \leq \max_l LB_t(\pi(p_l)). \end{cases}$$

- Scaled by price
- Tuned by considering variance of observed outcomes and the size of the bound

Experiments1 - Simulation

Setup

Price changes every 10 customers

Price = $0.01, $0.02, …. , $1,00, K=100

Segments, S = 1000

delta = 0.1

Valuation generating distribution:

1. right-skewed Beta distribution given by Beta(2, 9),
2. symmetric Beta distribution given by Beta(2, 2),
3. left-skewed Beta distribution given by Beta(9, 2),
4. bimodal continuous given by Beta(0.2, 0.3),
5. discontinuous finite mixture model with each vs equal to either $0.2 (with 70% chance) or $0.9 (30%).

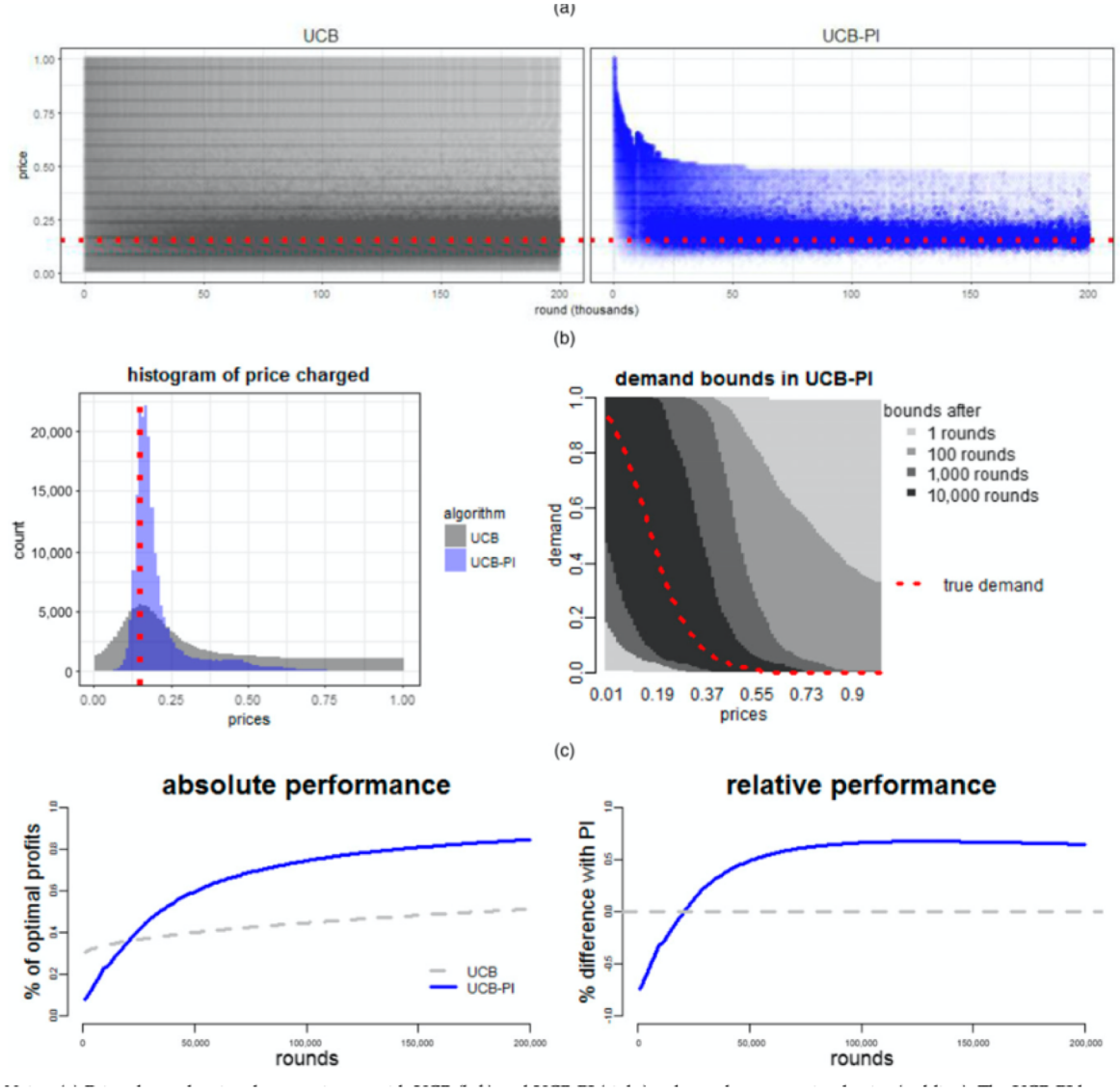
Experiments1 - Simulation

Competitors

1. UCB1
2. UCB1 Tuned
3. UCB PI
4. UCB PI Tuned
5. Learn-then-earn with different hyperparameters:

0.5%, 1%, 5%, 10%, 25%, 100%

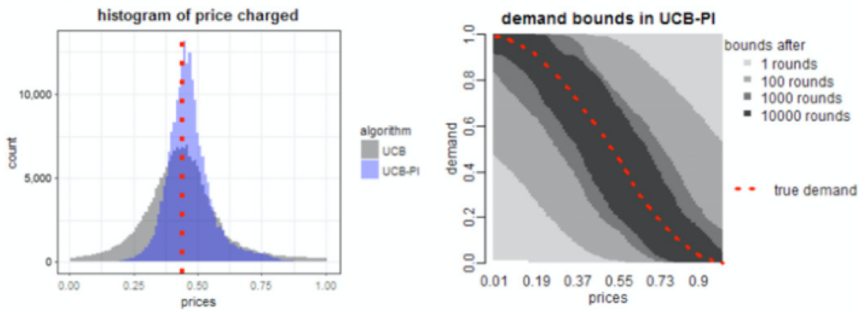
Experiments1 - Simulation



Experiments1 - Simulation

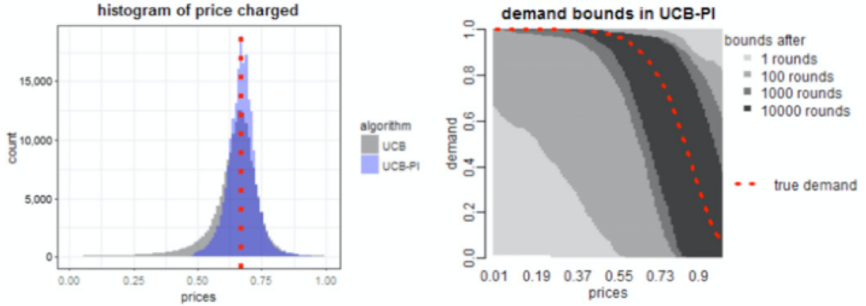
Setting: Symmetric

(a)



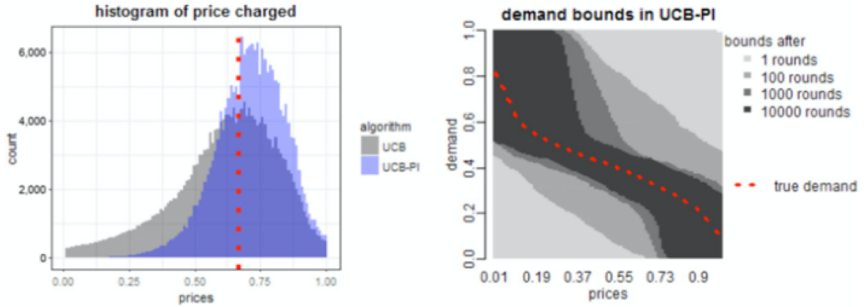
Setting: Left-Skewed

(b)



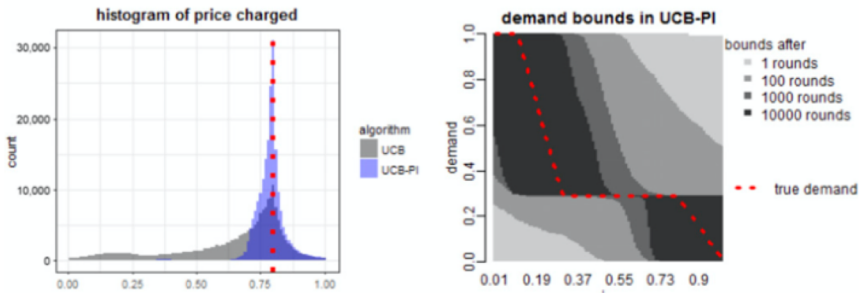
Setting: Bimodal Continuous

(c)



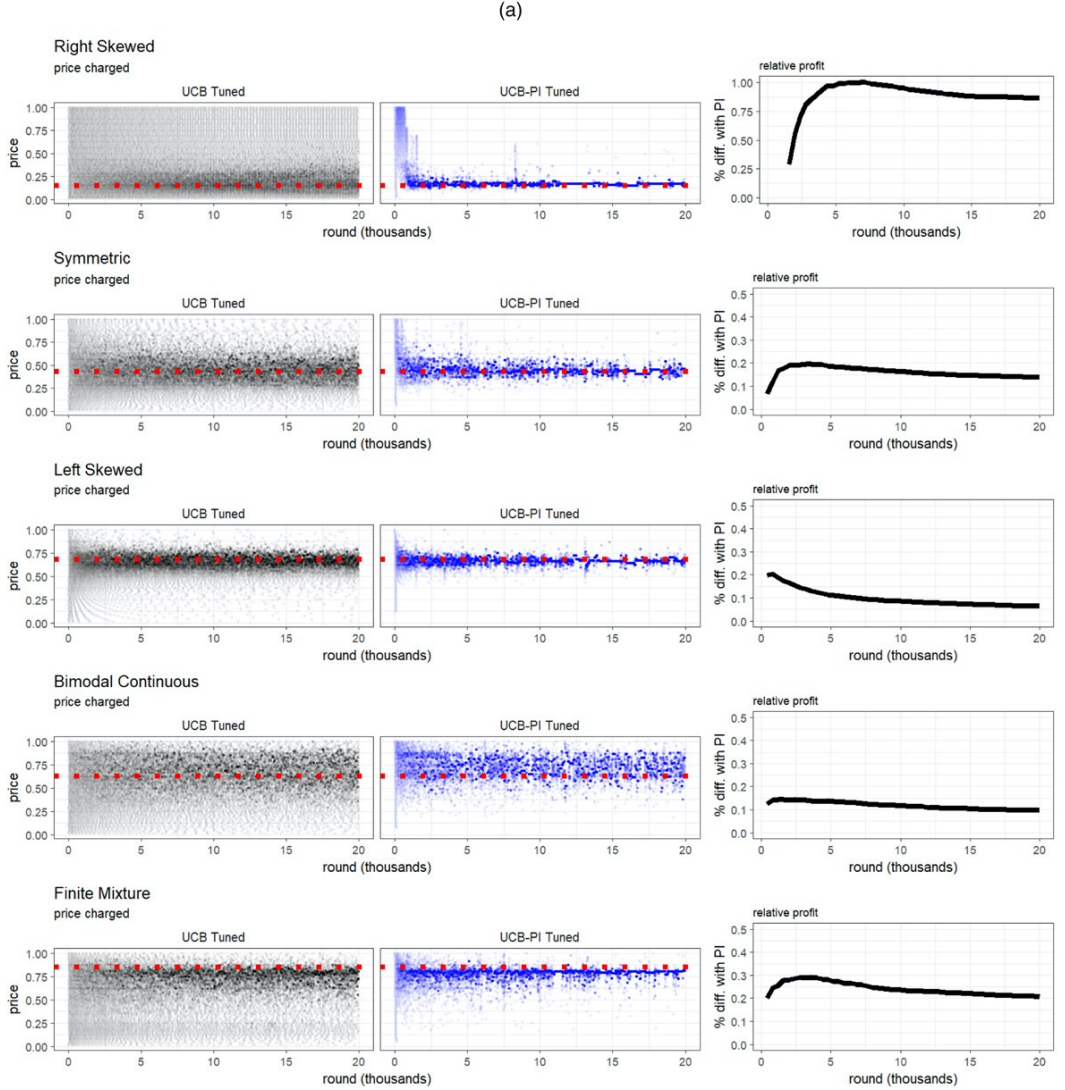
Setting: Finite Mixture

(d)

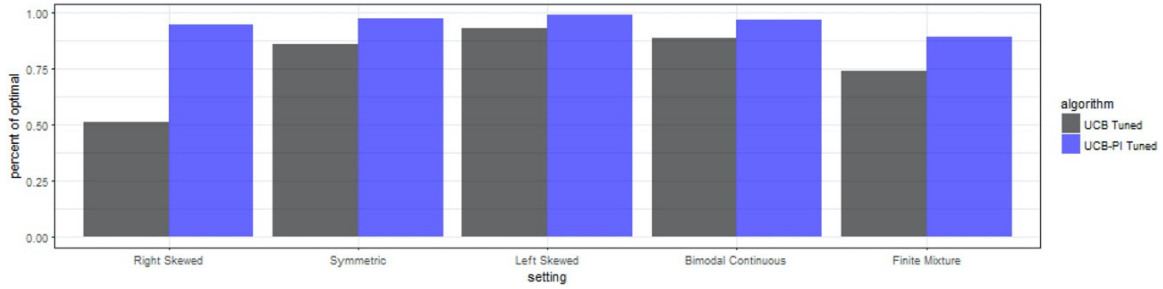


Experiments1 - Simulation

(a)

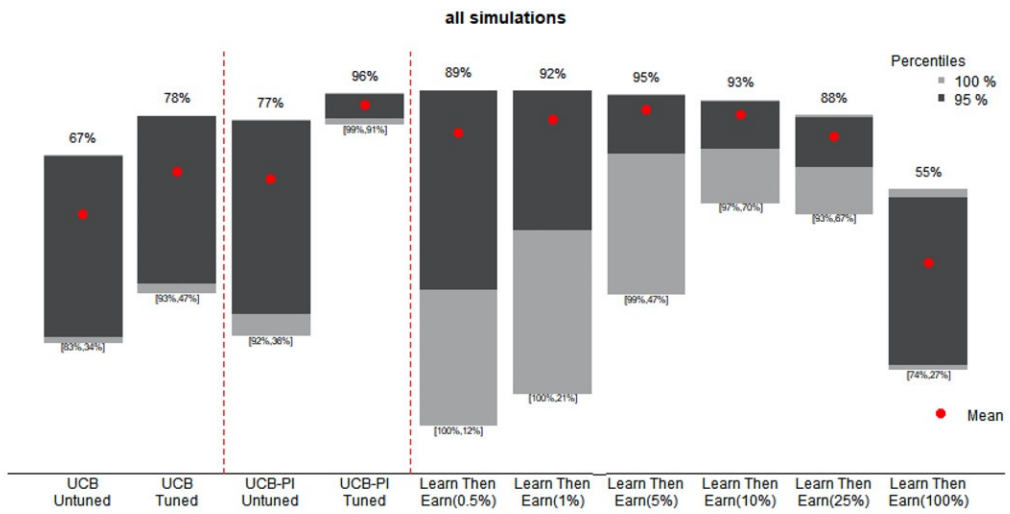


(b)

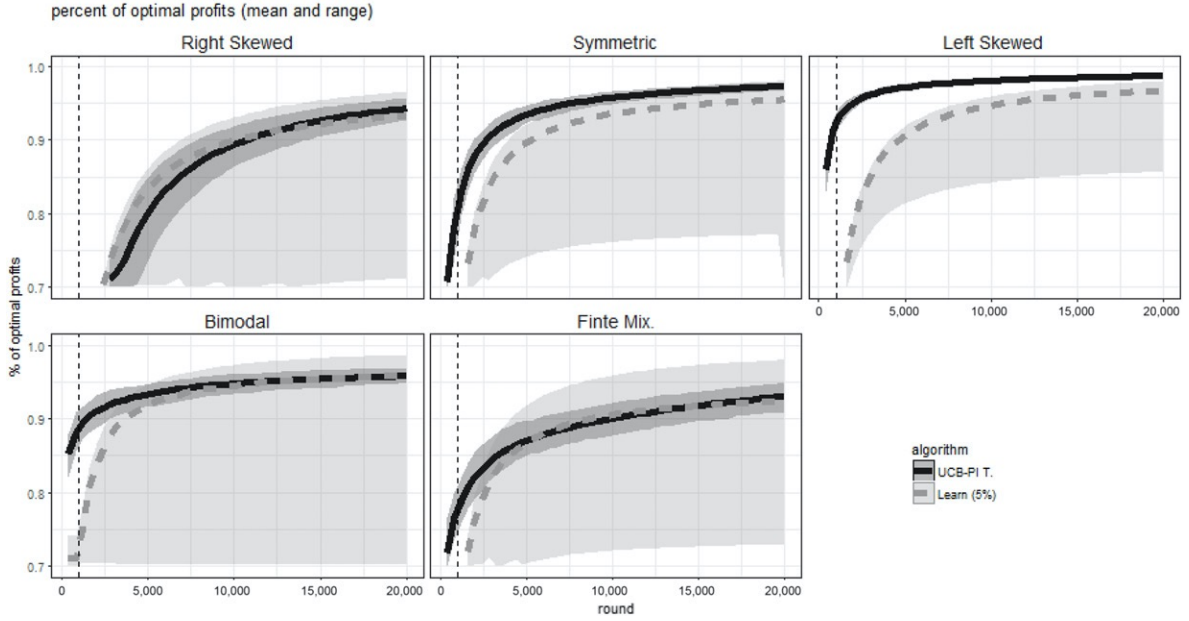


Experiments1 - Simulation

(a)



(b)



Experiments2 – Field Experiment

Setup

Based on a pricing field experiment by Dube and Misra

They ran an randomized price exp on Online recruiting firm with price ranging from $19 to $399, while the control and the real price is $99

Then a demand model is constructed based on these collected data.

S = 1000, equally sized based on descriptive variables.

Assume the valuations means to be uniform distributed, while the demand percentage is based on real-world data collected from the exp.

Delta = $5

Price, K = 10, $19, $39, …, $399, same as DM

Experiments2 – Field Experiment

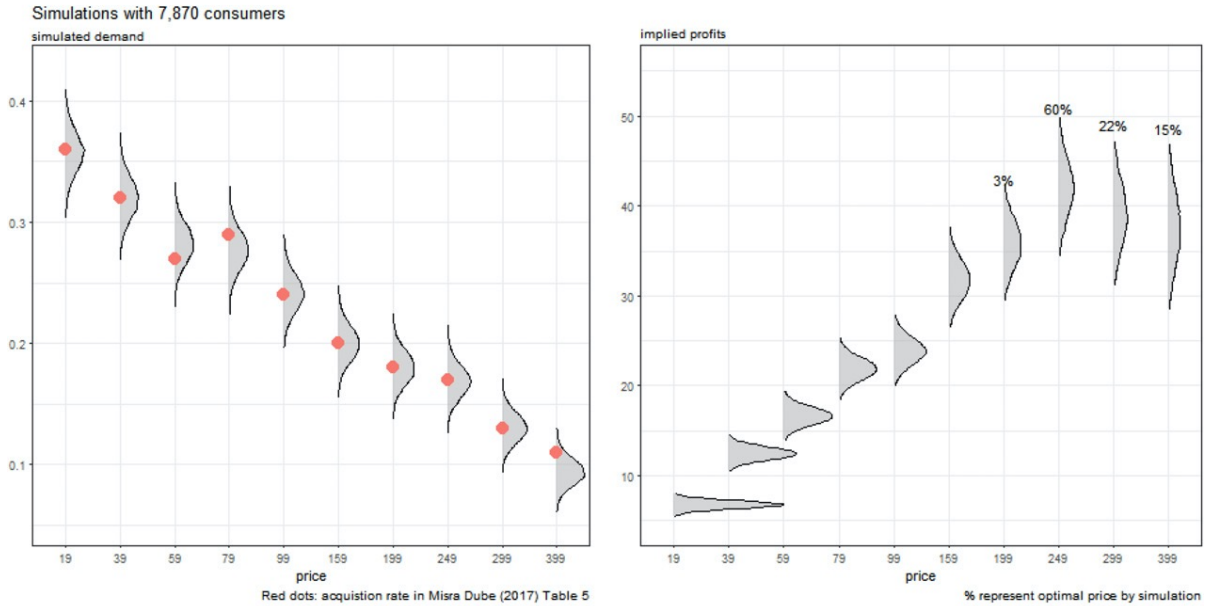
Competitors

1. UCB1 Tuned
2. UCB PI Tuned
3. Learn-then-earn with different hyperparameters:

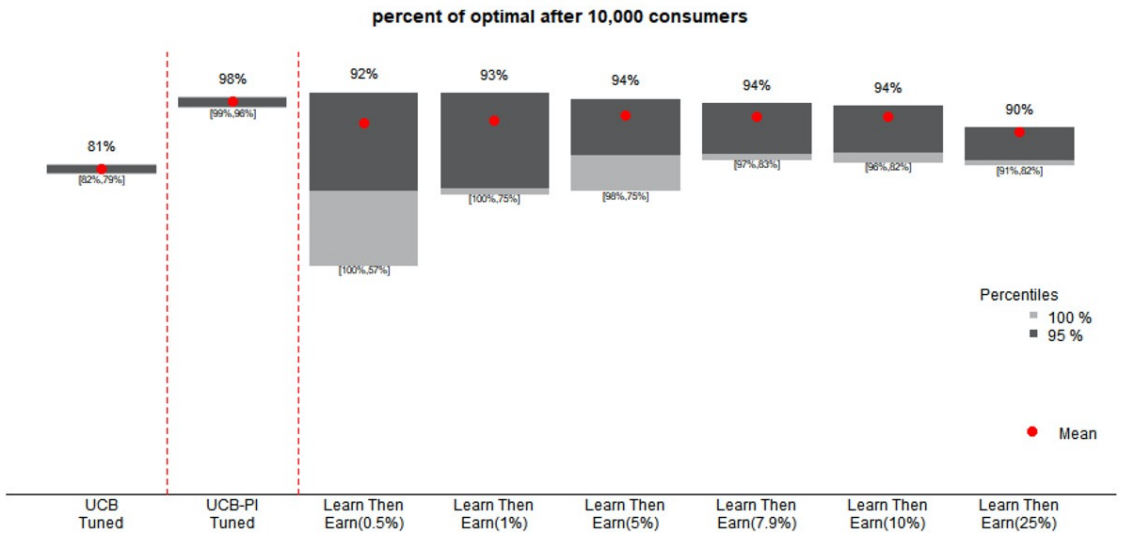
0.5%, 1%, 5%, 7.9%,10%, 25%

Experiments2 – Field Experiment

(a)

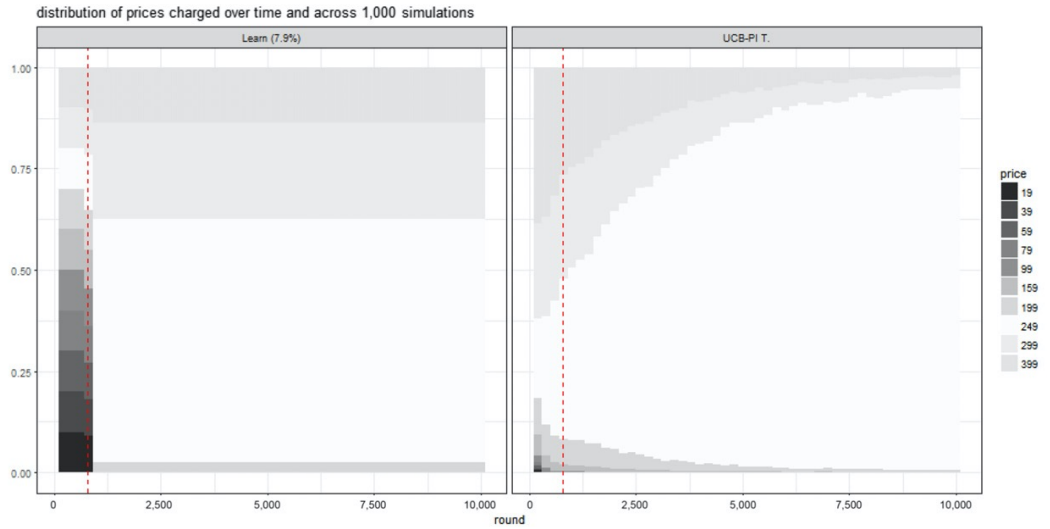


(b)

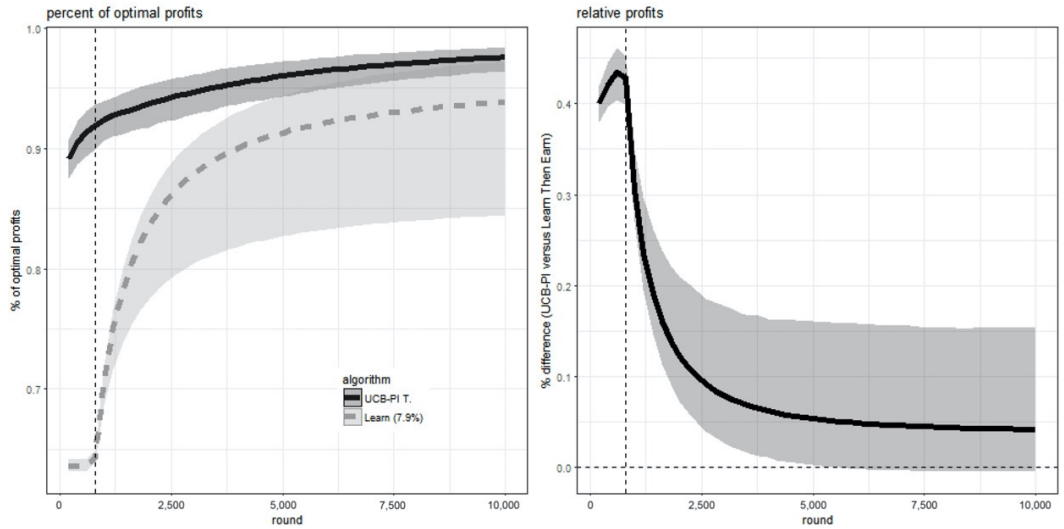


Experiments2 – Field Experiment

(a)



(b)



Conclusion

- UCBPI tend to not charge low prices due to PI
- Fast, Efficient, Optimal Profit
- Weaker assumptions, distribution-free
- Combine reinforcement learning and microeconimic theory to solve a realistic problem in marketing.

- Consider cases with dynamic preferences
- More than one products