Technical Note—Dynamic Pricing and Demand Learning with Limited Price Experimentation

Dynamic Pricing and Demand Learning with Limited Price Experimentation

Wang Chi Cheung
Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, wangchi@mit.edu

David Simchi-Levi
Engineering Systems Division, Department of Civil and Environmental Engineering, Institute for Data, Systems, and Society and Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, dslevi@mit.edu

He Wang
Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, wanghe@mit.edu

In a dynamic pricing problem where the demand function is not known a priori, price experimentation can be used as a demand learning tool. Existing literature usually assumes no constraint on price changes, but in practice sellers often face business constraints that prevent them from conducting extensive experimentation. We consider a dynamic pricing model where the demand function is unknown but belongs to a known finite set. The seller is allowed to make at most m price changes during T periods. The objective is to minimize the worst case *regret*, i.e. the expected total revenue loss compared to a clairvoyant who knows the demand distribution in advance. We demonstrate a pricing policy that incurs a regret of $O(\log^{(m)} T)$, or m iterations of the logarithm. We further show that this regret is the smallest possible up to a constant factor. Our analysis provides important structural insights into optimal pricing strategies. Finally, we describe an implementation at Groupon, a large e-commerce marketplace for daily deals. The field study shows significant impact on revenue and market share.

Key words: revenue management, dynamic pricing, learning-earning trade-off, price experimentation

1. Introduction

Groupon is a large e-commerce marketplace where customers can purchase discount deals from local merchants such as restaurants, spas and house cleaning services. The revenue from selling deals is split between local merchants and Groupon. Every day, thousands of new deals are launched on Groupon's website globally. Groupon is faced with high level of demand uncertainty mainly because newly launched deals have no previous sales data that can be used for demand forecasting.

This challenge presents an opportunity for Groupon to learn about customer demand using real time sales data after deals have been launched so as to obtain more accurate demand estimation.

In general, when the underlying relationship between demand and price is unknown a priori, price experimentation can be used for demand learning. In this paper, we consider a dynamic pricing model where the exact demand function is unknown but belongs to a finite set of possible demand functions, or demand hypotheses. The seller faces an exploration-exploitation tradeoff between actively adjusting price to gather demand information and optimizing price for revenue maximization.

Dynamic pricing under a finite set of demand hypotheses has been previously considered by Rothschild (1974) and Harrison et al. (2012). But unlike the current paper, both of these papers focus on customized pricing, where price is changed for every arriving customer. For example, the motivation of Harrison et al. (2012) is the pricing of financial services such as consumer and auto loans, where sellers can quote a different interest rate for each customer. However, for many e-commerce sellers like Groupon, charging a different price for each arriving customer is impossible either because of implementation constraints, negative customer response, or for fear of confusing their customers.

For example, Groupon expressed a preference to use as few price changes as possible for each deal. Motivated by this practical business constraint on price experimentation, the model in this paper includes an explicit constraint on the number of price changes during the sales horizon. We quantify the impact of this constraint on the seller's revenue using *regret*, defined as the gap between the revenue of a clairvoyant who has full information on the demand function and the revenue achieved by a seller facing unknown demand.

Our main finding is a characterization of regret as a function of the number of price changes. When there are T periods in the sales horizon, we propose a pricing policy with at most m price changes, whose regret is bounded by $O(\log^{(m)} T)$, or m iterations of the logarithm. Furthermore, we prove that the regret of any *non-anticipating* pricing policy, a policy where current price does

not depend on future demands, is lower bounded by $\Omega(\log^{(m)} T)$. Thus, the regret bound achieved by the proposed pricing policy is tight up to a constant factor.

A natural question is how frequently one needs to change price to achieve a constant regret. Harrison et al. (2012) shows that a semi-myopic policy can achieve a constant regret, but the policy requires changing price for every time period. To answer this question, we show that a modified version of our algorithm with no more than $O(\log^* T)$ price changes, where $\log^* T$ is the smallest number m such that $\log^{(m)} T \leq 1$, achieves a constant regret. Interestingly, while the value $\log^* T$ is unbounded when T increases, the growth rate is extremely slow. For example, if the number of time periods T is less than $3,000,000$, we have $\log^* T \leq 3$.

This characterization of the regret bound has two important implications. First, imposing a price change constraint always incurs a cost on revenue, since the seller cannot achieve a constant regret using any *finite* number of price changes. Second, the incremental effect of price changes decreases quickly. The first price change reduces regret from $O(T)$ to $O(\log T)$; each additional price change thereafter compounds a logarithm to the order of regret. As a result, the first few price changes generate most of the benefit of dynamic pricing.

Motivated by these results, we implemented a pricing strategy at Groupon where each deal can have at most one price change. The results from a field experiment show significant improvement in revenue.

The remainder of this paper is organized as follows. In Section 2, we review related literature. In Section 3, we define the mathematical model of the learning and pricing problem. The main theoretical results on the regret bound as a function of the number of price changes are presented in Section 4. Section 5 reports the implementation results at Groupon. Finally, we summarize in Section 6 with some concluding remarks.

2. Related Literature

Joint learning-and-pricing problems have received extensive research attention over the last decade. Recent surveys by Aviv and Vulcano (2012) and den Boer (2015) provide a comprehensive overview

of this area. Some papers that consider price experimentation for learning demand curves include Besbes and Zeevi (2009), Boyacı and Özer (2010), Wang et al. (2014) and Besbes and Zeevi (2015). These problems typically focus on the tradeoffs between learning and earning, which is closely related to the multi-armed bandit literature (e.g. Kleinberg and Leighton 2003, Mersereau et al. 2009, Rusmevichientong and Tsitsiklis 2010).

Recently, a stream of papers focuses on semi-myopic pricing policies using various learning methods. Examples include maximum likelihood estimation (Broder and Rusmevichientong 2012), Bayesian methods (Harrison et al. 2012), maximum quasi-likelihood estimation (den Boer and Zwart 2014, den Boer 2014) and iterative least-squares estimation (Keskin and Zeevi 2014).

Unlike the current paper, all the literature mentioned above does not assume any constraint on price experimentation. In the dynamic pricing literature with complete demand information, several papers consider limitation on price changes, e.g. Feng and Gallego (1995), Bitran and Mondschein (1997), Netessine (2006) and Chen et al. (2015). Caro and Gallien (2012) reports that the fashion retailer Zara uses a clearance pricing policy with a pre-determined price set, which essentially allows for only a limited number of mark-down prices. Finally, Zbaracki et al. (2004) provide empirical results on the cost of price changes.

To the best of our knowledge, the only work that considers price-changing constraints in an unknown demand setting is Broder (2011). The author assumes that the demand function belongs to a known family, e.g. linear, but has unknown parameters. He shows that in order to achieve the optimal regret, a pricing policy needs at least $\Theta(\log T)$ price changes. However, the result only applies to a restricted class of policies where the seller cannot use any knowledge of T.

Our model is different than the model by Broder (2011) in the following aspects. First, we assume finite demand hypotheses, while Broder (2011) assumes a parametric family of demand functions. This is a fundamental difference because the optimal regret in Broder's case is $\Theta(\sqrt{T})$, while in our case the regret can be bounded by a constant. Second, we do not assume a restricted class of policies as in Broder (2011), and our results hold for any pricing policies. Last but not least,

unlike Broder (2011) where the number of price changes is an output from the model, we design a pricing algorithm that accepts the number of price changes as an input and achieves the best possible regret bound, under that constraint.

3. Problem Formulation

We consider a seller offering a single product with unlimited supply for T periods. The set of allowable prices is denoted by $\mathcal{P}$. In the t^{th} period ($t=1,\ldots,T$), the seller offers a unit price $P_t \in \mathcal{P}$, and observes a random customer demand X_t, i.e. the number of units purchased by customers. Given $P_t = p$, the distribution of X_t is solely determined by price p and is independent of previous sales history $X_1,\ldots,X_{t-1}$. We use $D(p) \sim X_t$ to denote a random variable distributed as X_t given $P_t = p$. The corresponding *mean demand function* $d:\mathcal{P}\to\mathbb{R}_+$ is defined as $d(p)=\mathbb{E}[D(p)]$.

The distribution of $D(p)$ is unknown to the seller. However, the seller knows that the distribution belongs to a finite set of *demand models*, or demand distributions as a function of p. The demand models are indexed by $i=1,\ldots,K$. Let $\mathbb{P}_i(\cdot)$ and $\mathbb{E}_i(\cdot)$ be the probability measure and expectation under demand model i. In particular, the mean demand function $d(p)$ belongs to a finite set of K demand functions, denoted by $\Phi=\{d_1(p),\ldots,d_K(p)\}$, where $d_i(p)=\mathbb{E}_i[D(p)]$. For each demand function $d_i\in\Phi$, ($i=1,\ldots,K$), the expected revenue per period is $r_i(p)=pd_i(p)$. We also denote the optimal revenue for demand function d_i by $r_i^*=\max_{p\in\mathcal{P}} r_i(p)$ and an optimal price by $p_i^*\in \arg\max_{p\in\mathcal{P}} r_i(p)$. The seller does not necessarily know the distribution of demand model i apart from the mean $d_i(p)$.

For all $p\in\mathcal{P}$ and $i=1,\ldots,K$, the probability distribution of $D(p)$ is assumed to be *light-tailed* with parameters (σ,b), where $\sigma,b>0$. That is, we have $\mathbb{E}_i[e^{\lambda(D(p)-d_i(p))}]\le\exp(\lambda^2\sigma^2/2)$ for all $|\lambda|<1/b$. Note that the class of light-tailed distributions includes sub-Gaussian distributions. Some common light-tailed distributions include normal, Poisson and Gamma distributions, as well as all distributions with bounded support, such as binomial and uniform distributions.

3.1. Pricing Policies

We say π is a non-anticipating pricing policy if the price P_t offered at period t is determined by the realized demand $(X_1,\ldots,X_{t-1})$ and previous prices $(P_1,\ldots,P_{t-1})$, but does not depend on future

demand. For $i = 1, \ldots, K$, let $\mathbb{P}_i^\pi(\cdot)$ and $\mathbb{E}_i^\pi(\cdot)$ be the probability measure and expectation induced by policy π if the underlying demand model is i. In this case, the seller's expected revenue in T periods under policy π is given by

$$R_i^\pi(T) = \mathbb{E}_i^\pi \left[\sum_{t=1}^T P_t X_t \right] = \mathbb{E}_i^\pi \left[\sum_{t=1}^T P_t \mathbb{E}_i^\pi [X_t \mid P_t] \right] = \mathbb{E}_i^\pi \left[\sum_{t=1}^T r_i(P_t) \right]. \tag{1}$$

As discussed in the introduction, the seller faces a constraint on the number of price changes in many pricing applications. Specifically, we assume that the seller can make at most m changes to the price over the course of the sales event, where m is a fixed integer. So a feasible policy π should satisfy the following condition:

$$\mathbb{P}_i^\pi \left(\sum_{t=2}^T I(P_t \neq P_{t-1}) \leq m \right) = 1, \quad \forall i = 1, \ldots, K,$$

where $I(\cdot)$ is the indicator function. We refer to a policy with at most m price changes as an *m-change policy*.

The performance of pricing policies is measured against the optimal policy in the full information case. If the true demand is d_i, then a clairvoyant with full knowledge of the demand function would offer price p_i^* and obtain expected revenue r_i^* for every period. The *regret* with respect to demand d_i is defined as the gap between the expected revenue achieved by the clairvoyant and the one achieved by policy π, namely

$$\text{Regret}_i^\pi(T) = T r_i^* - R_i^\pi(T) = \mathbb{E}_i^\pi \left[\sum_{t=1}^T (r_i^* - r_i(P_t)) \right]. \tag{2}$$

Finally, we define the (minimax) regret for the demand set, $\Phi = \{d_1, \ldots, d_K\}$, as

$$\text{Regret}_\Phi^\pi(T) = \max_{i=1,\ldots,K} \text{Regret}_i^\pi(T).$$

When there is no ambiguity of which policy we are referring to, we suppress the superscript "π" in the notation for clarity, so $\mathbb{E}_1 := \mathbb{E}_1^\pi$, $\mathbb{P}_1 := \mathbb{P}_1^\pi$.

3.2. Notations

We use $\log^{(m)} T$ to represent m iterations of the logarithm, $\log(\log(\dots\log(T)))$, where m is the number of price changes. For convenience, we let $\log(x) = 0$ for all $0 \le x < 1$, so the value of $\log^{(m)} T$ is defined for all $T \ge 1$. Similarly, we define $e^{(0)} := 1$ and $e^{(\ell)} := \exp(e^{(\ell-1)})$ for $\ell \ge 1$. As mentioned in the introduction, $\log^* T$ denotes the smallest nonnegative integer m such that $\log^{(m)} T \le 1$. For any real number x, $\lceil x \rceil$ denotes the minimum integer greater than or equal to x, and for any finite set S, $|S|$ is the cardinality of S. We sometimes use the abbreviations $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$.

4. Main Results: Upper and Lower Bounds on Regret

In this section we prove the main results of the paper: an upper bound and a lower bound on regret as a function of the number of price changes. We first design a non-anticipating pricing policy that changes price m times and achieves a regret of $O(\log^{(m)} T)$. Then, we show that the regret of any non-anticipating policy with at most m price changes is at least $\Omega(\log^{(m)} T)$. Thus, our proposed pricing policy achieves the optimal regret bounds up to a constant factor.

4.1. Upper Bound

We propose a policy mPC (which stands for "m-price change") that achieves a regret of $O(\log^{(m)} T)$ with at most m price changes. An important feature of policy mPC is that it applies a *discriminative* price for every period. A price p is *discriminative* if the values $d_1(p), \cdots, d_K(p)$ are mutually distinct.

We make the following assumption on the set of demand functions Φ:

Assumption 1. *For all $d_i \in \Phi = \{d_1, \cdots, d_K\}$, there exists a corresponding revenue-optimal price $p_i^* \in argmax_{p \in \mathcal{P}} r_i(p)$ such that p_i^* is a discriminative price for Φ, that is, $d_1(p_i^*), \cdots, d_K(p_i^*)$ are distinct. Moreover, such price p_i^* can be efficiently computed.*

Assumption 1 ensures that the seller is able to learn the underlying demand curve while maximizing its revenue for any given demand function $d_i \in \Phi$. In fact, as we demonstrate in Section 4.4,

Algorithm 1 m-change policy mPC

INPUT:

- A set of demand functions $\Phi = \{d_1, \cdots, d_K\}$.
- A discriminative price P_0^*.

(Learning) Set $\tau_0 = 0$.

for $\ell = 0, \cdots, m-1$ **do**

  if $\log^{(m-\ell)} T = 0$ **then**

    Set $\tau_{l+1} = 0$ and $P_{\ell+1}^* = P_\ell^*$.

  else

    From period $\tau_\ell + 1$ to $\tau_{\ell+1} := \tau_\ell + \left\lceil M_\Phi(P_\ell^*) \log^{(m-\ell)} T \right\rceil$, set the offered price as P_ℓ^*.

    At the end of period $\tau_{\ell+1}$, compute the sample mean $\bar{X}^\ell$ from period $\tau_\ell + 1$ to $\tau_{\ell+1}$:

$$\bar{X}^\ell := \frac{\sum_{j=\tau_\ell+1}^{\tau_{\ell+1}} X_j}{\tau_{\ell+1} - \tau_\ell}, \text{ where } X_j = \text{Number of items sold in period } j.$$

    Choose an index $i_\ell \in \{1, \cdots, K\}$ which solves

$$\min_{i \in \{1,\cdots,K\}} \left| \bar{X}^\ell - d_i(P_\ell^*) \right|.$$

    Set the next offered price as $P_{\ell+1}^* = p_{i_\ell}^*$, where $p_{i_\ell}^*$ is the optimal price for demand d_{i_ℓ}.

  end if

end for

(Earning) From period $\tau_m + 1$ to period $\tau_{m+1} = T$, set the selling price as P_m.

this condition turns out to be both sufficient and necessary for achieving regret bound better than $o(\log T)$.

Algorithm 1 describes our mPC policy. The policy partitions the finite time horizon $1, \cdots, T$ into $m+1$ phases. For each $0 \le \ell \le m$, a single price P_ℓ^* is offered through Phase ℓ, which starts at period $\tau_\ell + 1$ and ends at $\tau_{\ell+1}$. Phase 0 to Phase $m-1$ are called the *learning phases*, and Phase m is referred to as the *earning phase*. Except for a constant factor $M_\Phi(P_\ell^*)$, which is to be defined

later, the lengths of phases are exponentially increasing, which ensures an optimal balance between exploration and exploitation.

At the end of learning phase ℓ, policy mPC computes the sample mean $\bar{X}^\ell$ of the sales under price P_ℓ^* (in line 8 of the algorithm). Since price P_ℓ^* is discriminative, the seller gains new information about the underlying demand in this learning phase. She then updates her belief on the true demand distribution to be $d_{i_{\ell+1}}$ (in line 9), and sets the offered price $P_{\ell+1}^*$ to be $p_{i_{\ell+1}}^*$ in the next phase. In going through all the learning phases, the seller progressively refines her estimate on the optimal price, which enables her to establish the choice of optimal price in the earning phase.

The function $M_\Phi(P)$ in line (7) of the mPC algorithm is defined as follows.

DEFINITION 1. Let $p \in \mathcal{P}$ be a discriminative price. We define $M_\Phi(p)$ as

$$M_\Phi(p) := \frac{16\sigma^2}{\min_{i\neq j}\left(d_i(p)-d_j(p)\right)^2} \vee \frac{8b}{\min_{i\neq j}|d_i(p)-d_j(p)|}, \tag{3}$$

where the minimum is taken over distinct pairs of indices $i, j \in \{1, \cdots, K\}$.

Since we assume that p is a discriminative price, $M_\Phi(p)$ is well defined. The function $M_\Phi(p)$ measures the distinguishability of the demand functions $d_1, \cdots, d_k$ under the discriminative price p. We explain the definition of $M_\Phi(p)$ further in the analysis of mPC.

Define $M_\Phi^* = \max_{i\in\{1,\cdots,K\}} M_\Phi(p_i^*)$ and $r^* = \max_{i\in\{1,\cdots,K\}} r_i^*$. The following result shows that the regret of mPC is bounded by $O(\log^{(m)} T)$.

THEOREM 1. *Suppose the demand set Φ satisfies Assumption 1. For all $T \geq 1$, the regret of mPC is bounded by*

$$\text{Regret}_\Phi^{mPC}(T) \leq C_\Phi(P_0^*)\max\{\log^{(m)} T, 1\} + 4(M_\Phi^* + 1)r^*,$$

where $C_\Phi(P_0^*) = \max_{i\in\{1,\cdots,K\}}\{M_\Phi(P_0^*)(r_i^* - r_i(P_0^*))\}$.

Proof Idea of Theorem 1. In the proof, we establish that the regret incurred in Phase 0 is $O(\log^{(m)} T)$, and the cumulative regret incurred in the remaining phases is $O(1)$. At the beginning of Phase 0, which is also the beginning of the sale horizon, the seller has no information on the

optimal price. Thus, the regret during Phase 0 is proportional to the length of Phase 0. However, in each of the subsequent phases, the seller can choose a price based on the previous sale history. By choosing the lengths of the subsequent phases appropriately, we ensure that the total regret in these phases is $O(1)$.

Proof of Theorem 1. Suppose d_1 is the underlying demand function. The regret under demand d_1 can be decomposed as

$$\text{Regret}_1^{\mathsf{mPC}}(T) = \mathbb{E}_1\left[\sum_{t=1}^{T}(r_1^* - r_1(P_t))\right] = \sum_{\ell=0}^{m}\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right].$$

We first consider the case where $\log^{(m)} T > 0$. By definition, $\tau_1 = \left\lceil M_\Phi(P_0^*)\log^{(m)} T\right\rceil$, so the regret during Phase 0 is equal to

$$\mathbb{E}_1\left[\sum_{t=1}^{\tau_1}(r_1^* - r_1(P_t))\right] = \left\lceil M_\Phi(P_0^*)\log^{(m)} T\right\rceil (r_1^* - r_1(P_0^*)). \tag{4}$$

Next, we show that for each $1 \le \ell \le m$, the regret during Phase ℓ is bounded by

$$\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right] \le \frac{2M_\Phi^* r_1^*}{\log^{(m-\ell)} T} + \frac{2r_1^*}{(\log^{m-\ell} T)^2}, \tag{5}$$

where $M_\Phi^* = \max_{i\in\{1,\cdots,K\}} M_\Phi(p_i^*)$.

For $1 \le \ell \le m$, the regret during Phase ℓ satisfies the following bound:

$$\begin{aligned}
&\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right] \\
=&\mathbb{E}_1\left[(\tau_{\ell+1} - \tau_\ell)\times(r_1^* - r_1(P_\ell^*))\right] \\
\le&\mathbb{E}_1\left[\left(M_\Phi(P_\ell^*)\log^{(m-\ell)} T + 1\right)\times(r_1^* - r_1(P_\ell^*))\right] \\
\le&\left(M_\Phi^*\log^{(m-\ell)} T + 1\right)\sum_{i=1}^{K}(r_1^* - r_1(p_i^*))\times\mathbb{P}_1(P_\ell^* = p_i^*) \qquad (6)\\
\le&\left(M_\Phi^*\log^{(m-\ell)} T + 1\right) r_1^* \times \sum_{i=2}^{K}\mathbb{P}_1(P_\ell^* = p_i^*).
\end{aligned}$$

In the above calculation, the expectation is taken on the price offered in Phase ℓ, P_ℓ^*, which is a random variable depending on the realized demand in phases $0,\cdots,\ell-1$. In equation (6), we use the fact that for all $\ell = 1,\cdots,m$, the offered price $P_\ell^* \in \{p_1^*,\cdots,p_K^*\}$ (see line 10 of mPC).

To complete the proof of inequality (5), we prove the following inequality:

$$\sum_{i=2}^{K} \mathbb{P}_1(P_\ell^* = p_i^*) \le \frac{2}{(\log^{(m-\ell)} T)^2}. \tag{7}$$

By the definition of mPC, the choice of price P_ℓ^* is determined by the sample mean $\bar{X}_{\ell-1}$ in Phase $\ell-1$, so we have

$$\sum_{i=2}^{K} \mathbb{P}_1(P_\ell^* = p_i^*) = \mathbb{P}_1\left(|\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)| \ge |\bar{X}^{\ell-1} - d_i(P_{\ell-1}^*)| \text{ for some } i \ne 1\right).$$

Now, if $|\bar{X}^{\ell-1} - d_1(P_\ell^*)| \ge |\bar{X}^{\ell-1} - d_i(P_\ell^*)|$ for some $i \ne 1$, we have

$$|\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)| \ge \frac{1}{2}\left(|\bar{X}^{\ell-1} - d_i(P_{\ell-1}^*)| + |\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)|\right) \ge \frac{1}{2}|d_i(P_{\ell-1}^*) - d_1(P_{\ell-1}^*)|,$$

where the last step uses the triangle inequality. This results in the following bound:

$$\sum_{i=2}^{K} \mathbb{P}_1(P_\ell^* = p_i^*) \le \mathbb{P}_1\left(\left|\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)\right| \ge \frac{1}{2}\min_{i\ne 1}\left|d_1(P_{\ell-1}^*) - d_i(P_{\ell-1}^*)\right|\right). \tag{8}$$

Given price $P_{\ell-1}^*$, sample mean $\bar{X}_{\ell-1}$ is the average of i.i.d. random variables with mean $d_1(P_{\ell-1})$. Because demand in each period is lighted-tailed with parameters (σ, b), we can apply the Chernoff inequality: conditioning on $P_{\ell-1}^*$, for any $\epsilon > 0$, it holds that

$$\mathbb{P}_1\left(|\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)| \ge \epsilon \middle| P_{\ell-1}^*\right) \le 2\exp\left(-(\tau_\ell - \tau_{\ell-1})\left(\frac{\epsilon^2}{2\sigma^2} \wedge \frac{\epsilon}{2b}\right)\right).$$

Let $\epsilon = \frac{1}{2}\min_{i\ne 1}\left|d_1(P_{\ell-1}^*) - d_i(P_{\ell-1}^*)\right|$. Because $\tau_\ell - \tau_{\ell-1} = \left\lceil M_\Phi(P_{\ell-1}^*)\log^{(m-\ell+1)} T\right\rceil$, we have

$$\begin{aligned}
&\mathbb{P}_1\left(\left|\bar{X}^{\ell-1} - d_1(P_{\ell-1}^*)\right| \ge \frac{1}{2}\min_{i\ne 1}\left|d_1(P_{\ell-1}^*) - d_i(P_{\ell-1}^*)\right| \middle| P_{\ell-1}^*\right) \\
\le& 2\mathbb{E}_1\left[\exp\left(-\left\lceil M_\Phi(P_{\ell-1}^*)\log^{(m-\ell+1)} T\right\rceil\left(\frac{\epsilon^2}{2\sigma^2} \wedge \frac{\epsilon}{2b}\right)\right) \middle| P_{\ell-1}^*\right] \\
\le& 2\mathbb{E}_1\left[\exp\left(-M_\Phi(P_{\ell-1}^*)\log^{(m-\ell+1)} T\left(\frac{\epsilon^2}{2\sigma^2} \wedge \frac{\epsilon}{2b}\right)\right) \middle| P_{\ell-1}^*\right] \\
\le& 2\mathbb{E}_1\left[\exp\left(-2\log^{(m-\ell+1)} T\right) \middle| P_{\ell-1}^*\right] \\
=& \frac{2}{\left(\log^{(m-\ell)} T\right)^2},
\end{aligned} \tag{9}$$

where step (9) uses the definition

$$M_\Phi(P^*_{\ell-1}) = 2 \times \left(\frac{2\sigma^2}{\frac{1}{4}\min_{i\neq j}\left(d_i(P^*_{\ell-1}) - d_j(P^*_{\ell-1})\right)^2} \vee \frac{2b}{\frac{1}{2}\min_{i\neq j}\left|d_i(P^*_{\ell-1}) - d_j(P^*_{\ell-1})\right|} \right).$$

By integrating over the realizations of $P^*_{\ell-1}$ in the above bound, we have established inequality (7), which in turn proves (5).

Combining equations (4) and (5), we can prove the regret bound on mPC under demand d_1 as follows:

$$\begin{aligned}
\text{Regret}_1^{\mathsf{mPC}}(T) &= \mathbb{E}_1\left[\sum_{t=1}^{\tau_1}(r_1^* - r_1(P_t))\right] + \sum_{\ell=1}^{m}\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right] \\
&\leq \left(M_\Phi(P_0^*)\log^{(m)}T + 1\right)(r_1^* - r_1(P_0^*)) + \sum_{\ell=1}^{m}\left(\frac{2M_\Phi^* r_1^*}{\log^{(m-\ell)}T} + \frac{2r_1^*}{(\log^{(m-\ell)}T)^2}\right).
\end{aligned}$$

Since $\log^{(m)}T > 0$, it is easily verified that $\log^{(m-\ell)}T \geq e^{\ell-1}$ for all $\ell \geq 1$, so

$$\sum_{\ell=1}^{m}\frac{1}{\log^{(m-\ell)}T} \leq \sum_{\ell=1}^{\infty}\frac{1}{e^{\ell-1}} \leq 2, \quad \sum_{\ell=1}^{m}\frac{1}{(\log^{(m-\ell)}T)^2} \leq \sum_{\ell=1}^{\infty}\frac{1}{e^{2\ell-2}} \leq \frac{3}{2}.$$

Therefore,

$$\begin{aligned}
\text{Regret}_1^{\mathsf{mPC}}(T) &\leq \left(M_\Phi(P_0^*)\log^{(m)}T + 1\right)(r_1^* - r_1(P_0^*)) + 4M_\Phi^* r_1^* + 3r_1^* \\
&\leq M_\Phi(P_0^*)(r_1^* - r_1(P_0^*))\log^{(m)}T + 4M_\Phi^* r_1^* + 4r_1^*.
\end{aligned}$$

The minimax regret of demand set Φ is bounded by

$$\text{Regret}_\Phi^{\mathsf{mPC}}(T) = \max_{i=1,\ldots,K}\text{Regret}_i^{\mathsf{mPC}}(T) \leq C_\Phi(P_0^*)\log^{(m)}T + 4M_\Phi^* r^* + 4r^*,$$

where $C_\Phi(P_0^*) = \max_{i\in\{1,\cdots,K\}}\{M_\Phi(P_0^*)(r_i^* - r_i(P_0^*))\}$ and $r^* = \max_{i\in\{1,\cdots,K\}} r_i^*$.

If $\log^{(m)}T = 0$, let $m' \leq m$ be the largest integer such that $\log^{(m')}T > 0$. Clearly, $\log^{(m')}T \leq 1$. In this case, policy mPC applied to T periods uses only m' price changes, so

$$\text{Regret}_\Phi^{\mathsf{mPC}}(T) \leq C_\Phi(P_0^*)\log^{(m')}T + 4M_\Phi^* r^* + 4r^* \leq C_\Phi(P_0^*) + 4M_\Phi^* r^* + 4r^*.$$

Combining both cases for $\log^{(m)}T > 0$ and $\log^{(m)}T = 0$, we have

$$\text{Regret}_\Phi^{\mathsf{mPC}}(T) \leq C_\Phi(P_0^*)\max\{\log^{(m)}T, 1\} + 4M_\Phi^* r^* + 4r^*.$$

□

REMARK 1. In Phase 0, the discriminative price P_0^* is given as a input. One can further reduce the regret bound by choosing a discriminative price P_0^* which minimizes the regret during Phase 0, namely $C_\Phi(P_0^*)$.

REMARK 2. In line (9) of Algorithm 1, the test to select a demand function i_ℓ is a simple comparison between the sample mean $\bar{X}^\ell$ and the mean demand function value $d_i(P_\ell^*)$. Therefore, *the algorithm does not require the seller to know the demand distributions for each demand model.* Nevertheless, if the seller does know the demand distribution, line (9) can be replaced by other selection criteria, such as a likelihood ratio test.

REMARK 3. The proof shows that in the special case of $m = 1$, Assumption 1 is not required for Theorem 1. We only require that the initial price P_0^* is discriminative.

4.2. Lower Bound

We show next that for a family of problem instances, any m-change policy incurs a regret of $\Omega(\log^{(m)} T)$. Thus, the regret achieved by the m-change policy mPC is optimal up to a constant factor.

Consider a problem instance (Γ) that satisfies the following conditions:

1. There exists a constant $Q_\Gamma > 0$, such that $\sum_{i=1}^K (r_i^* - r_i(p)) \geq Q_\Gamma$ for all $p \in \mathcal{P}$.
2. The demand $D(p) \in \mathbb{N}$ for any price $p \in \mathcal{P}$.
3. Given $p \in \mathcal{P}$, there exists a subset $\mathcal{B}_p \subset \mathbb{N}$, such that for all i, $\mathbb{P}_i(D(p) = d) > 0$ if and only if $d \in \mathcal{B}_p$.
4. There exists a constant $0 < \kappa_\Gamma < 1$, such that $\mathbb{P}_i(D(p) = d)/\mathbb{P}_j(D(p) = d) \geq \kappa_\Gamma$ for all $i, j \in \{1, \ldots, K\}$, $p \in \mathcal{P}, d \in \mathcal{B}_p$.

The first condition states that there is no price $p \in \mathcal{P}$ that simultaneously maximizes the revenue of all demand functions in Φ. This ensures that the problem instance is nontrivial and a learning process is necessary for maximizing the revenue when the demand function is unknown. The second condition is that demand must be integers. The third condition states that all demand functions

have the same support for a given price. The fourth condition states that the ratios of probability mass functions of different demand models are bounded.

The key step in the proof of the lower bound theorem is to quantify the performance of a pricing policy under different demand functions. This is made precise by following lemma.

LEMMA 1 **(Change-of-Measure Lemma)**. *Let $H_t = (P_1, X_1, \cdots, P_t, X_t)$ be the history observed by the end of period t, and let h_t be a realization of H_t. For any non-anticipating pricing policy π, we have*

$$\mathbb{P}_i^\pi(H_t = h_t) \geq \kappa_\Gamma^t \mathbb{P}_{i'}^\pi(H_t = h_t),$$

for all $i, i' \in \{1, \ldots, K\}$. The constant κ_Γ is defined in the condition (Γ).

The proof of Lemma 1 can be found in Appendix A.1.

The regret lower bound of any m-change policy is formally stated in the following.

THEOREM 2 **(Lower Bound Theorem)**. *For any m-change policy π on problem instance Γ, there exists a constant $\theta_m > 0$ such that for any $T > \theta_m$, we have*

$$\text{Regret}_\Phi^\pi(T) \geq \frac{1}{K} C_\Gamma Q_\Gamma \log^{(m)} T,$$

where $C_\Gamma := (-8 \log \kappa_\Gamma)^{-1} \wedge 1$ and Q_Γ is given by the first condition of (Γ).

Proof Idea of Theorem 2. We consider the time period τ when the first price change occurs, and compare it with $C_\Gamma \log^{(m)} T$. If $\tau > C_\Gamma \log^{(m)} T$, the seller spends at least $C_\Gamma \log^{(m)} T$ periods on learning with price P_1, which is determined without any observation. This implies that the seller must incur a regret of at least $\Omega(\log^{(m)} T)$. Otherwise, if we have $\tau \leq C_\Gamma \log^{(m)} T$, we argue that the seller has not extracted enough information about the underlying demand function, using the Change-of-Measure Lemma. In addition, the seller can perform at most $m - 1$ price changes after $C_\Gamma \log^{(m)} T$ periods. It turns out that these two facts cause the seller to incur a regret of at least $\Omega(\log^{(m)} T)$.

Proof of Theorem 2. Without loss of generality, we restrict π to be a deterministic policy, since the regret of a randomized policy is the expectation of the regret of corresponding deterministic policies. In other words, we restrict price P_t to be a deterministic function of the history $(P_1, X_1, \ldots, P_{t-1}, X_{t-1})$.

We prove the theorem by establishing the following induction claim.

Induction Claim-(m) There exists $\theta_m > 0$ such that for any m-change policy π and any $T > \theta_m$, we have

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi}(T) \geq C_\Gamma Q_\Gamma \log^{(m)} T,$$

where $C_\Gamma := (-8 \log \kappa_\Gamma)^{-1} \wedge 1$ and Q_Γ is given by the first condition of (Γ).

Note that the constants C_Γ, Q_Γ are independent of the time horizon T, the number of price changes m, and the choice of pricing policy π. If the induction claim is established for all $m \geq 0$, the theorem is easily proved since we have

$$\text{Regret}_\Phi^{\pi}(T) = \max_{i=1,\ldots,K} \text{Regret}_i^{\pi}(T) \geq \frac{1}{K} \sum_{i=1}^{K} \text{Regret}_i^{\pi}(T) \geq \frac{1}{K} C_\Gamma Q_\Gamma \log^{(m)} T.$$

Basic induction hypothesis $m = 0$. In this case, the seller must use a fixed price throughout the sales horizon, i.e., $P_t = P_1$ for all $t = 1, \ldots, T$. By the first condition of (Γ), the regret of any 0-change policy π is at least

$$\sum_{i=1}^{K} \text{Regret}_i^{\pi}(T) = \sum_{i=1}^{K} \sum_{t=1}^{T} \Delta_i(P_1) = \sum_{t=1}^{T} \left(\sum_{i=1}^{K} \Delta_i(P_1) \right) \geq Q_\Gamma T \geq C_\Gamma Q_\Gamma T,$$

where $\Delta_i(P_1) = r_i^* - r_i(P_1)$. This proves the case for $m = 0$ by setting $\theta_0 = 0$.

Induction step. For some $m > 0$, suppose the induction claim is true for $m - 1$. We prove that the induction claim is also true for m. Without loss of generality, we assume $\log^{(m)} T > 0$, otherwise the induction claim trivially holds. For a given m-change policy π, let τ be the time period when the first price change occurs, i.e., $\tau = \min_{1 \leq t \leq T} \{t : P_t \neq P_{t-1}\}$. Let $T_m = \lceil C_\Gamma \log^{(m)} T \rceil$, where $C_\Gamma = (-8 \log \kappa_\Gamma)^{-1} \wedge 1$. Note that the constant $\kappa_\Gamma \in (0, 1)$, so $C_\Gamma > 0$. We use $\mathcal{L}$ to denote the event $\mathcal{L} = \{\tau > T_m\}$.

We decompose the regret $\text{Regret}^{\pi}_{\Phi}(T)$ and bound it from below as follows:

$$\begin{aligned}
&\sum_{i=1}^{K} \text{Regret}^{\pi}_{i}(T) = \sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=1}^{T} \Delta_i(P_t) \right] \\
=& \sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=1}^{T} \Delta_i(P_t) \middle| \mathcal{L} \right] \mathbb{P}_i(\mathcal{L}) + \sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i(\mathcal{L}^C) \\
\geq& \underbrace{\sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=1}^{T_m} \Delta_i(P_t) \middle| \mathcal{L} \right] \mathbb{P}_i(\mathcal{L})}_{(\dagger)} + \underbrace{\sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{T_m+1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i(\mathcal{L}^C)}_{(\ddagger)}.
\end{aligned}$$

Consider the regret term (†). Conditioned on the event $\mathcal{L}$, the first price change only occurs after the T_m^{th} period. Thus, we have $P_t = P_1$ for all $1 \leq t \leq T_m$, and the term (†) can bounded by

$$(\dagger) \geq \sum_{i=1}^{K} C_\Gamma \log^{(m)}(T) \Delta_i(P_1) \mathbb{P}_i(\mathcal{L}). \tag{10}$$

Next, we analyze the regret term (‡). Recall that $H_t = (P_1, X_1, \cdots, P_t, X_t)$ is the history observed by the seller at the end of period t, and let h_t be a specific realization of H_t. We define the set

$$\mathcal{H}_m^{\Delta} = \{h_{T_m} = (P_1, X_1, \cdots, P_{T_m}, X_{T_m}) : P_s \neq P_{s+1} \text{ for some } 1 \leq s \leq T_m - 1\}$$

as the set of history for which a price change occurs before period T_m. By the definition, we have $\mathbb{P}_i(\mathcal{L}^C) = \mathbb{P}_i(H_{T_m} \in \mathcal{H}_m^{\Delta})$.

Thus, term (‡) is bounded by

$$\begin{aligned}
(\ddagger) =& \sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i(\mathcal{L}^C) \\
=& \sum_{i=1}^{K} \mathbb{E}_i \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| \mathcal{L}^C \right] \mathbb{P}_i(H_{T_m} \in \mathcal{H}_m^{\Delta}) \\
=& \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \mathbb{E}_i \left[\sum_{t=T_m+1}^{T} \Delta_i(P_t) \middle| H_{T_m} = h_{T_m} \right] \mathbb{P}_i \left(H_{T_m} = h_{T_m} \right) && (11) \\
=& \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \text{Regret}_i^{\pi(h_{T_m})}(T - T_m) \mathbb{P}_i \left(H_{T_m} = h_{T_m} \right) && (12) \\
\geq& \sum_{i=1}^{K} \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \text{Regret}_i^{\pi(h_{T_m})}(T - T_m) \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota \left(H_{T_m} = h_{T_m} \right) \right) && (13) \\
=& \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\ldots,K\}} \mathbb{P}_\iota \left(H_{T_m} = h_{T_m} \right) \right) \sum_{i=1}^{K} \text{Regret}_i^{\pi(h_{T_m})}(T - T_m)
\end{aligned}$$

In step (11), we decompose the previous expression into a summation of conditional expectations over realized history $h_{T_m} \in \mathcal{H}_m^{\Delta}$. Note that the set $\mathcal{H}_m^{\Delta}$ is countable, since the demand X_t is integer and the price P_t is completely determined by the previous history. In step (12), the pricing policy $\pi(h_{T_m})$ denotes the policy adopted by the seller from period T_m+1 to period T, after she observes the history h_{T_m} from period 1 to period T_m. Note that the policy $\pi(h_{T_m})$ is determined after the history h_{T_m} is realized. Thus, the expression in (12) is a weighted sum of regret under strategies $\{\pi(h_{T_m}), h_{T_m} \in \mathcal{H}_m^{\Delta}\}$, where each regret term is weighted by the probability of the corresponding history. Step (13) applies the Change-of-Measure Lemma.

Let θ'_{m-1} the threshold such that $\left\lceil C_\Gamma \log^{(m)} T \right\rceil \le 2C_\Gamma \log^{(m)} T$ for all $T > \theta'_{m-1}$. So we have

$$\kappa_\Gamma^{T_m} \ge \kappa_\Gamma^{2C_\Gamma \log^{(m)} T} \ge \exp(-\frac{1}{4}\log^{(m)} T) \ge \exp\left(-\Big(1 - \frac{\log(2\log^{(m)} T)}{\log^{(m)} T}\Big)\log^{(m)} T\right) = \frac{2\log^{(m)} T}{\log^{(m-1)} T}. \tag{14}$$

The first inequality uses the definition of T_m, the second inequality applies the definition of C_Γ, and the third inequality uses the fact that $\log(2x)/x < 3/4$ for all $x > 0$.

For all $h_{T_m} \in \mathcal{H}_m^{\Delta}$, the policy $\pi(h_{T_m})$ changes price no more than $m-1$ times during period T_m to period T, because at least one price change is exhausted before period T_m. Applying the induction claim for $(m-1)$, we know that for all $h_{T_m} \in \mathcal{H}_m^{\Delta}$, we have

$$\sum_{i=1}^{K} \mathrm{Regret}_i^{\pi(h_{T_m})}(T - T_m) \ge C_\Gamma Q_\Gamma \log^{(m-1)}(T - T_m)$$

for T such that $T - T_m \ge \theta_{m-1}$. Furthermore, let $\theta''_{m-1} > 0$ be a threshold such that $T > T_m + \theta_{m-1}$ and $\log^{(m-1)}(T - \left\lceil C_\Gamma \log^{(m)} T \right\rceil) \ge \frac{1}{2}\log^{(m-1)} T$ for all $T \ge \theta''_{m-1}$. Then, for $T \ge \theta''_{m-1}$, we have:

$$\sum_{i=1}^{K} \mathrm{Regret}_i^{\pi(h_{T_m})}(T - T_m) \ge \frac{1}{2} C_\Gamma Q_\Gamma \log^{(m-1)} T. \tag{15}$$

Combining (14) and (15), we have the following:

$$\begin{aligned}
(\ddagger) &\ge \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \left(\kappa_\Gamma^{T_m} \max_{\iota \in \{1,\dots,K\}} \mathbb{P}_\iota\left(H_{T_m} = h_{T_m}\right)\right) \sum_{i=1}^{K} \mathrm{Regret}_i^{\pi(h_{T_m})}(T - T_m) \\
&\ge \sum_{h_{T_m} \in \mathcal{H}_m^{\Delta}} \left(\frac{2\log^{(m)} T}{\log^{(m-1)} T} \max_{\iota \in \{1,\dots,K\}} \mathbb{P}_\iota\left(H_{T_m} = h_{T_m}\right)\right) \frac{1}{2} C_\Gamma Q_\Gamma \log^{(m-1)} T
\end{aligned}$$

$$
\begin{aligned}
=&C_\Gamma Q_\Gamma \log^{(m)} T \sum_{h_{T_m}\in\mathcal{H}_m^\Delta} \left(\max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota \left(H_{T_m}=h_{T_m}\right)\right)\\
\geq&C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\dots,K\}} \sum_{h_{T_m}\in\mathcal{H}_m^\Delta} \mathbb{P}_\iota \left(H_{T_m}=h_{T_m}\right)\\
=&C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C). \qquad (16)
\end{aligned}
$$

Altogether, by both (10) and (16), we have

$$
\begin{aligned}
&\sum_{i=1}^K \mathrm{Regret}_i^\pi(T) \geq (\dagger)+(\ddagger)\\
\geq&\sum_{i=1}^K C_\Gamma \log^{(m)}(T)\Delta_i(P_1)\mathbb{P}_i(\mathcal{L}) + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C)\\
\geq&C_\Gamma \log^{(m)}(T)\left(1-\max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C)\right)\sum_{i=1}^K \Delta_i(P_1) + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C)\\
\geq&C_\Gamma \log^{(m)}(T)\left(1-\max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C)\right) Q_\Gamma + C_\Gamma Q_\Gamma \log^{(m)} T \max_{\iota\in\{1,\dots,K\}} \mathbb{P}_\iota(\mathcal{L}^C)\\
=&C_\Gamma Q_\Gamma \log^{(m)} T,
\end{aligned}
$$

for all $T \geq \max\{\theta'_{m-1}, \theta''_{m-1}\}$. By setting $\theta_m := \max\{\theta'_{m-1}, \theta''_{m-1}\}$, the induction step is established. This completes the proof. □

Taken together, the proofs of the upper and lower bounds provide important insights into the structure of any optimal m-change policy. With high probability, an optimal m-change policy has $m-1$ learning phases of lengths $\Theta(\log^{(m)} T), \cdots, \Theta(\log T)$. They are followed by the last phase, which is the earning phase on the last $T - \Theta(\log T)$ time periods, see Fig 1.

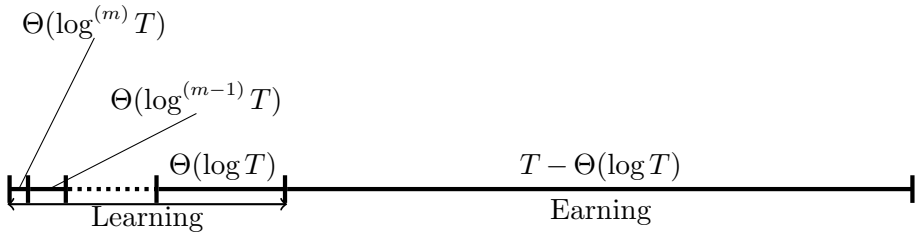

Figure 1 The structure of an optimal m-change policy.

The lengths of the learning phases are set in a way to ensure an optimal balance between learning and earning. If any of the learning phases is shortened significantly, such lack of learning will incur

a large regret in the subsequent phases. In general, for each $\ell \in \{1, \cdots, m\}$, if the ℓ^{th} learning phase is of length $o(\log^{(m-\ell+1)} T)$, then a regret of $\Omega(\log^{(m-\ell)} T)$ is incurred in the subsequent phases. This quantifies the value of learning in any m-change policy.

4.3. Unbounded but Infrequent Price Experiments

Policy mPC defines m learning phases with exponentially increasing lengths. This motivates us to consider a modification of mPC, which improves the regret bound to a constant. We call this modified policy uPC (which standing for "unbounded price changes"), see Algorithm 2. Although the number of price changes under this policy is not bounded by any finite number as T increases, it grows extremely slowly with order $O(\log^* T)$, where $\log^* T = \min\{m \in \mathbb{Z}^+ : \log^{(m)} T \leq 1\}$. For example, for $T \leq 3,000,000$, we have $\log^* T \leq 3$. According to the Lower Bound Theorem in Section 4.2, this is the minimum growth rate possible.

PROPOSITION 1. *Suppose Assumption 1 holds. For all $T \geq 1$, the pricing policy* uPC *has regret*

$$\text{Regret}^{\mathsf{uPC}}(T) \leq C_\Phi(P_0^*) + 2(M_\Phi^* + 1)r^*,$$

where $C_\Phi(P_0^*) = \max_{i \in \{1,\cdots,K\}}\{M_\Phi(P_0^*)(r_i^* - r_i(P_0^*))\}$.

The proof of Proposition 1 is in Appendix A.2.

Furthermore, uPC is an *anytime* policy, meaning that the seller can apply uPC algorithm without any knowledge of T. Anytime policies can be used for customized pricing. In customized pricing, each customer arrival is modeled as a single time period, so T is the total number of customers arrivals (see Harrison et al. 2012, Broder and Rusmevichientong 2012). Since uPC is an anytime policy, the seller is not required to know the total number of customers arrivals.

In comparison, if the seller is only allowed to change price m times, it is impossible to achieve the optimal regret bound $O(\log^{(m)} T)$ if the seller does not know T. The lower bound theorem shows that in order to achieve the optimal regret bound, the ℓ^{th} price change must happen at $\Theta(log^{(m-\ell+1)}T)$, so it is impossible to determine when to change price without knowing T.

Algorithm 2 Policy uPC

INPUT:

- A set of demand functions $\Phi = \{d_1, \cdots, d_K\}$.
- A discriminative price P_0^*.

Set $\tau_0 = 0$.

for $\ell = 0, 1, \cdots$ **do**

From period $\tau_\ell + 1$ to $\tau_{\ell+1} := \tau_\ell + \left\lceil M_\Phi(P_\ell^*) e^{(\ell)} \right\rceil$, set the offered price as P_ℓ^*.

if $T \le \tau_{\ell+1}$ **then** stop the algorithm at period T.

else

At the end of period $\tau_{\ell+1}$, compute the sample mean $\bar{X}^\ell$ from period $\tau_\ell + 1$ to $\tau_{\ell+1}$:

$$\bar{X}^\ell := \frac{\sum_{j=\tau_\ell+1}^{\tau_{\ell+1}} X_j}{\tau_{\ell+1} - \tau_\ell}, \text{ where } X_j = \text{Number of items sold in period } j.$$

Choose an index $i_\ell \in \{1, \cdots, K\}$, which solves

$$\min_{i \in \{1,\cdots,K\}} \left| \bar{X}^\ell - d_i(P_\ell^*) \right|.$$

Set the next offered price as $P_{\ell+1}^* = p_{i_\ell}^*$, where $p_{i_\ell}^*$ is an optimal price for demand d_{i_ℓ}.

end if

end for

4.4. Discussion on the Discriminative Price Assumption

The $O(\log^{(m)} T)$ regret of mPC and the $O(1)$ regret of uPC hold under the assumption that there exists an optimal discriminative price for each demand function (Assumption 1). In fact, one can show that this assumption is necessary for any non-anticipating policy to achieve a regret better than $o(\log T)$.

PROPOSITION 2. *If Assumption 1 is violated, then there exists a price set $\mathcal{P}$ and a demand set Φ such that any non-anticipating pricing policy incurs a regret of $\Omega(\log T)$, even if that policy is allowed to change price for infinitely many times.*

The proof of Proposition 2 is in Appendix A.3. It implies that the best possible regret bound without Assumption 1 is $O(\log T)$. The question remaining is what is the best regret upper bound we can have when Assumption 1 does not hold. Below we show that for any set of K demand functions, policy kPC (see Algorithm 3) achieves regret bound of $O(\log T)$ with at most $K-1$ price changes.

For this purpose, we need the following definition:

DEFINITION 2. For any nonempty subset of demand functions $A \subset \{d_1, \ldots, d_K\}$, let

$$\tilde{p}_A := \arg\max_{p \in \mathcal{P}} |\{d_i(p) \mid d_i \in A\}|,$$

i.e., $\tilde{p}_A$ is the price that maximizes the number of distinct values of $d_i(p)$ for all $d_i \in A$.

Furthermore, define

$$\tilde{M}_A(p) := \frac{8\sigma^2}{\min_{(i,j):d_i(p)\neq d_j(p)} \left(d_i(p) - d_j(p)\right)^2} \vee \frac{4b}{\min_{(i,j):d_i(p)\neq d_j(p)} |d_i(p) - d_j(p)|}, \tag{17}$$

where the minimum is taken over all pairs of demand functions $d_i, d_j \in A$ such that $d_i(p) \neq d_j(p)$.

Note that if $|A| \geq 2$, for any pair of demand functions d_i and d_j in A, we can always find a price p such that $d_i(p) \neq d_j(p)$, because otherwise the two demand functions are identical. So the value $\tilde{M}_A(\tilde{p}_A)$ in line (4) of Algorithm 3 is well defined for any $|A| \geq 2$.

PROPOSITION 3. *For all $T \geq 1$, the regret of kPC is bounded by*

$$\text{Regret}_{\Phi}^{kPC}(T) \leq (K-1)(\tilde{M}_{\Phi} r^* \log T + 3r^*),$$

where $\tilde{M}_{\Phi} = \max_{A \subset \{1,\ldots,K\}} \tilde{M}_A(\tilde{p}_A)$.

Proof Idea of Proposition 3. In each of the learning phases, the definition of the algorithm (line 6) guarantees that at least one demand function is eliminated. So the number of iterations in the while loop is at most $K-1$, and the regret of the learning phases is $O((K-1)\log T)$. Then, we show that with high probability, the single demand function remained in the earning phase is the true demand function. The complete proof is in Appendix A.4.

Algorithm 3 Policy kPC.

1: INPUT: A set of demand functions $\Phi = \{d_1, \cdots, d_K\}$.

2: (Learning) Set $A \leftarrow \Phi$. Set $\ell = 0$, $\tau_0 = 0$.

3: **while** $|A| \neq 1$ **do**:

4: Set the price as $P_\ell^* = \tilde{p}_A$ from period $\tau_\ell + 1$ to $\tau_{\ell+1} := \tau_\ell + \left\lceil \tilde{M}_A(P_\ell^*) \log T \right\rceil$.

5: At the end of period $\tau_{\ell+1}$, compute the sample mean $\bar{X}^\ell$ from period $\tau_\ell + 1$ to $\tau_{\ell+1}$:

$$\bar{X}^\ell := \frac{\sum_{j=\tau_\ell+1}^{\tau_{\ell+1}} X_j}{\tau_{\ell+1} - \tau_\ell}, \text{ where } X_j = \text{Number of items sold in period } j.$$

6: Update A: keep all d_i in set A if it is a minimizer of $\min_{d_i \in A} \left| \bar{X}^\ell - d_i(P_\ell^*) \right|$. Eliminate other demand functions from A. If there are two minimizers d_i and d_j such that $d_i(P_\ell^*) < \bar{X}^\ell < d_j(P_\ell^*)$, remove d_j and only keep d_i.

7: Set $\ell \leftarrow \ell + 1$.

8: **end while**

9: (Earning) Suppose $A = \{d_i\}$. From period $\tau_\ell + 1$ to period $\tau_{\ell+1} = T$, set the selling price as $P_\ell^* = p_i^*$.

5. Field Experiment at Groupon

We collaborated with Groupon, a large e-commerce marketplace for daily deals, to implement the dynamic pricing strategies proposed in the previous section. Groupon offers subscribed customers discount deals from local merchants. By the second quarter of 2015, Groupon served more than 500 cities worldwide, had nearly 49 million active customers and featured more than 510,000 active deals globally.

As an example, Figure 2 shows a local restaurant deal on Grouopon's website. The deal can be purchased through Groupon at \$17 and redeemed at the local restaurant for \$30. The amount paid by customer (\$17) is called "booking". The booking is then split between Groupon and the local merchant. For example, in a 50/50 split, the local business gets \$8.5 and Groupon keeps \$8.5 as its revenue. In most cases, a deal is only available for a limited time, ranging from several weeks to several months.

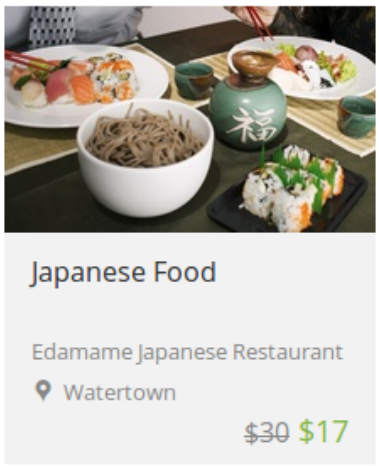

Figure 2 Screenshot of a restaurant deal on the Groupon website.

Prior to our collaboration, Groupon applied a fixed price strategy for each deal until it expires. Our initial analysis suggested that Groupon could benefit from the dynamic learning and pricing algorithm that we proposed in this paper for the following reasons:

- Groupon launches thousands of new deals everyday across its global markets, and most of these deals are offered on its website for the first time. There is not enough historical data to predict demand before the new deals are launched. So there is an opportunity to learn demand using real time sales data.
- Most deals are offered for a limited time, so there is a time tradeoff between price experimentation and revenue maximization, a tradeoff addressed by our pricing algorithm.
- Groupon prefers to use as few price changes as possible for each deal. Since the result from Section 4 shows that most benefit of dynamic pricing is captured by the first price change, we apply a single price change in our implementation. More specifically, we use the mPC algorithm of Section 4.1 with $m = 1$.
- Each deal has a monthly cap that specifies the maximum quantity that can be sold within a month. But historical data show that only a small fraction of deals have actually reached their monthly caps. So the unlimited inventory assumption is a reasonable approximation of reality.

The pricing algorithm mPC requires two inputs: a set of demand functions, Φ, and an initial discriminative price, P_1. We discuss how they were generated in the following two subsections.

5.1. Generating the Demand Function Set

Although it is unlikely that we can find a finite set Φ such that the underlying demand function belongs to it, our objective is to generate a set Φ so that the true demand function is well approximated by at least one of the functions in it.

We suppose the price set is a continuous interval $\mathcal{P} = [p_l, p_u]$, and each demand function in Φ is linear. Every time period is one day.

We first collect sales data of previous deals that have been tested for dynamic pricing. Note that all deals in this dataset have been offered under more than one price. In the preprocessing step, the demand data are normalized to remove the time effect (e.g. holiday/weekend effect). Then, given a new deal, we generate a set of linear demand functions using the following process:

1. Select a subset of deals from the historical data that share similar features with the new deal (e.g. initial price, category/subcategory, discount rate).
2. Since any given deal in this subset has been offered under more than one price, we can fit a linear demand function to it by least squares method. The linear demand function is then mapped to a point on a plane, where the y-coordinate is the slope, and the x-coordinate is the demand function valued at the initial price of the new deal. For example, suppose we fit a demand function $d_i(p) = a - bp$ for an old deal, and the new deal has an initial price P_1, then the demand function is mapped to the point $(a - bP_1, b)$, see Figure 3. Every deal in the subset is now represented by a point on the plane.
3. Apply K-means clustering to group the points into K clusters. For example, Figure 3 shows $K = 3$ clusters. Note that the center of each cluster also represents a linear demand function. In particular, if the center is located at (x_i, y_i), it corresponds to linear function $d_i(p) = x_i + y_i(P_1 - p)$. So the set of centers contains exactly K linear demand functions, which forms the demand set Φ.

To determine the best value of K, we apply cross-validation. The previous deals are randomly split into training and testing sets. For each deal in the testing set with two prices p_1, p_2, we treat

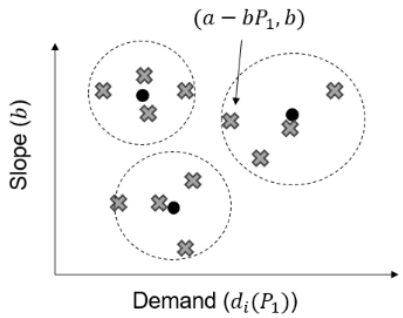

Figure 3 Applying K-means clustering to generate K linear demand functions.

it as a new deal with initial price p_1. For different values of K, we generate K demand functions using the training set following the process described above. Then, we select one function among the K functions whose value at p_1 is the closest to the actual demand of the new deal at p_1, since this is the function that would have been chosen by our learning algorithm. Next, we compare the realized demand under price p_2 to the mean demand predicted by the selected function at p_2. The difference between these two values can be interpreted as the prediction error of our learning algorithm.

In Figure 4, we plot the mean squared error of demand prediction and bookings prediction for different values of K. The error is large for small values of K, and then decreases as K increases. This implies that for small values of K, none of the demand functions in set Φ is close to the true demand function, so the prediction error is large. Therefore, it is important to choose a K large enough so that at least one of the demand function in Φ is close to the true demand function. Notice that what is not shown in the figure is that the error will eventually go up due to over-fitting when K becomes sufficiently large. We didn't have enough data points in this example to demonstrate over-fitting.

5.2. Choosing the Initial Price and the Time of Price Change

The initial price P_1 is negotiated by the local merchant and Groupon. Basically, this is the price that Groupon would have used in its fixed pricing policy.

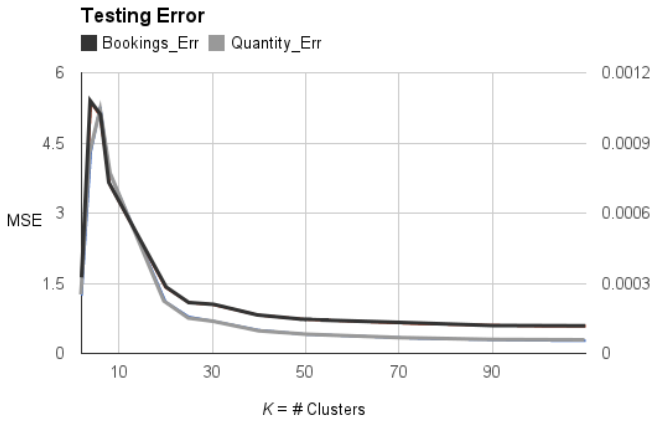

Figure 4 Prediction error of re for different values of K.

Since a finite set of linear demand functions has only finite non-discriminative prices in the price interval $[p_l, p_u]$, it is unlikely that the initial price is non-discriminative. In fact, in all the examples that we tested, the initial price P_1 is discriminative with respect to demand set Φ.

In the definition of mPC for $m = 1$, the price is changed at period $\lceil M_\Phi(P_0) \log T \rceil$, where the constant $M_\Phi(P_0)$ is given by Equation (3) in Section 4.1. However, this constant is mainly designed to prove the theoretical regret bound, and may not be a good choice for implementation. In practice, we tested several price switching times (between 1 to 7 days), and the value with the best performance was chosen.

When changing price, Groupon has a constraint that price can only be decreased between 5% to 30%. If the output of the algorithm either decreases price by less than 5% or increases price, then no price change is made. If the output decreases price by more than 30%, then we decrease price by only 30%. More importantly, the merchant's share of bookings is unchanged after price decrease. For example, if a deal has initial price \$20 and the merchant get \$10 from each purchase, it would still get \$10 after price decrease. Therefore, the merchant is never worse off after price change, and hopefully is more willing to accept the dynamic pricing policy that Groupon proposed. However, for those merchants who prefer fixed pricing, there is always an option to keep using fixed pricing.

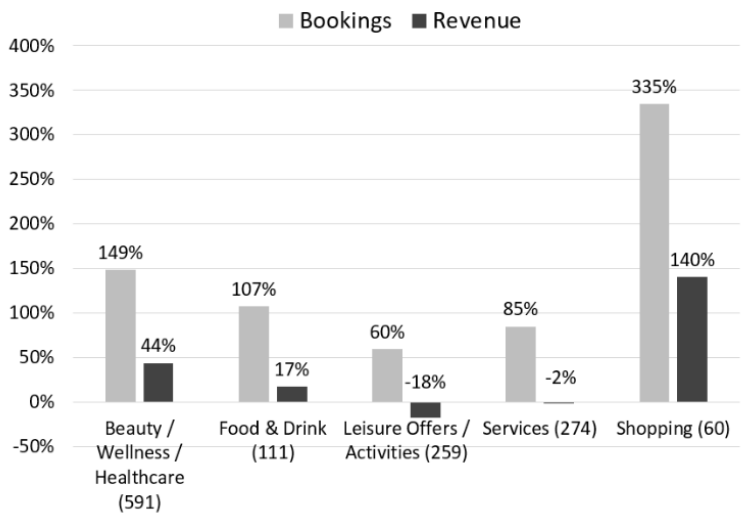

Figure 5 Bookings and revenue increase by deal category.

5.3. Field Experiment Results

In the field experiment, we included 1,295 deals that span five product categories: Beauty, Food & Drink, Activities, Services, and Shopping. We focused on two performance measurements. One is the total amount money paid by the customers to Groupon, referred to as *bookings*, which is directly related to Groupon's market share; the other is the portion of money that Groupon keeps after paying local merchants, referred to as *revenue*. For each product category, we compare the average bookings and revenue before and after price change. Since the initial price of dynamic pricing is determined in the same way as in fixed pricing, the bookings and revenue before the price change represent the performance for fixed pricing strategy. Note that if a deal is tested using our pricing algorithm but the algorithm does not recommend price decrease, then this deal is not included in the 1,295 selected deals.

Figure 5 shows the average increase in bookings and revenue of after price changes by category. The numbers in parentheses are the quantity of deals tested in each category. Among the five categories, Beauty, Food & Drink, and Shopping have significant revenue increase, Services category has almost no revenue change but significant bookings increase, and Activities category has a decrease in revenue. Overall, bookings is increased by 116%, and revenue is increased by 21.7%. The revenue improvement may also be partly attributed to increased exposure, because deals with

price decrease are featured on a separate webpage labeled as "clearance", so they may attract more customer visits.

Further analysis of the field experiment result shows that reducing price has a much bigger impact on deals that have fewer bookings per day. For deals with bookings per day less than the median (across all product categories), the average increase in revenue is 116%, while the increase is only 14% for deals with bookings per day more than the median. This explains the big increase in bookings and revenue for the Shopping category, because the average daily bookings of the Shopping category is only around 1/10 of the average daily bookings the Food & Drink category.

Our pricing algorithm has a poor performance for the Activities category, despite the fact that this categories has almost the same level of average daily bookings as the Beauty category. We suspect that some information of customer demand for Activities is not included in our demand model. For example, it might be that the weekend/holiday effect is much more significant for this category than we estimated, or perhaps the holiday effect happens a few days before the actual holiday. Further work is needed to improve the demand prediction method for the Activities category.

6. Conclusion

We consider a dynamic pricing problem where the latent demand model is unknown but belongs to a finite set of demand functions. The seller faces a constraint that price can be changed at most m times. We propose a pricing policy that incurs a regret of $O(\log^{(m)} T)$, where T is the length of the sales horizon. In addition, we show that this regret bound is the best possible, up to a constant factor.

We then implement this pricing algorithm at Groupon, a website that sells deals from local merchants. We design a process to generate a set of linear demand functions from historical data, and use it as an input to our pricing algorithm. The algorithm allows for at most one price change per deal and price decrease only. Field experiment shows that the algorithm has a significant improvement on revenue and bookings.

Acknowledgments

We gratefully acknowledge assistance of data analysis provided by Alex Weinstein of MIT Operations Research Center and the Data Science team of Groupon. We thank the Associate Editor and three anonymous referees for comments that have greatly improved the manuscript.

Appendix A: Additional Proofs of the Results in Section 4

A.1. Proof of Lemma 1

Proof of Lemma 1. Let $h_t = (p_1, x_1, \ldots, p_t, x_t)$ be a realization of $H_t = (P_1, X_1, \cdots, P_t, X_t)$. We first assume $\mathbb{P}_i^\pi(H_t = h_t) > 0$, so we have

$$\begin{aligned}
\mathbb{P}_i^\pi(H_t = h_t) =& \prod_{s=1}^{t} \mathbb{P}_i^\pi(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \\
=& \prod_{s=1}^{t} \left(\mathbb{P}_{i'}^\pi(D(p_s) = x_s) \cdot \frac{\mathbb{P}_i^\pi(D(p_s) = x_s)}{\mathbb{P}_{i'}^\pi(D(p_s) = x_s)} \right) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad (18) \\
\geq& \prod_{s=1}^{t} (\mathbb{P}_{i'}^\pi(D(p_s) = x_s) \cdot \kappa_\Gamma) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad (19) \\
=& \kappa_\Gamma^t \prod_{s=1}^{t} \mathbb{P}_{i'}^\pi(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \\
=& \kappa_\Gamma^t \prod_{s=1}^{t} \mathbb{P}_{i'}^\pi(D(p_s) = x_s) \prod_{s=1}^{t-1} \mathbb{P}_{i'}^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) \qquad (20) \\
=& \kappa_\Gamma^t \mathbb{P}_{i'}^\pi(H_t = h_t).
\end{aligned}$$

Step (18) uses the third condition of (Γ), which states that all demand functions have the same support under a given price, so $\mathbb{P}_{i'}^\pi(D(p_s) = x_s) \neq 0$. Step (19) uses the fourth condition of (Γ). Step (20) holds because price P_{s+1} is determined by policy π and realized history h_s, and is independent of the underlying demand model. Note that if π is a deterministic policy, we always have $\mathbb{P}_i^\pi(P_{s+1} = p_{s+1} \mid H_s = h_s) = 1$ for all i.

Finally, if $\mathbb{P}_i^\pi(H_t = h_t) = 0$, we have $\mathbb{P}_{i'}^\pi(H_t = h_t) = 0$, too. This is again due to the third condition of (Γ), which states that all demand functions have the same support under a given price. □

A.2. Proof of Proposition 1

Proof of Proposition 1. Let m be the integer such that $\tau_m < T \leq \tau_{m+1}$. Suppose d_1 is the underlying demand function. The regret under demand d_1 can be composed as

$$\mathrm{Regret}_1^{\mathsf{uPC}}(T) = \mathbb{E}_1 \left[\sum_{t=1}^{T} (r_1^* - r_1(P_t)) \right] \leq \sum_{\ell=0}^{m} \mathbb{E}_1 \left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}} (r_1^* - r_1(P_t)) \right].$$

The regret during Phase 0 is equal to

$$\mathbb{E}_1\left[\sum_{t=1}^{\tau_1}(r_1^*-r_1(P_t))\right]=\lceil M_\Phi(P_0^*)\rceil(r_1^*-r_1(P_0^*)).$$

For $1\le\ell\le m$, the offered price $P_\ell^*\in\{p_1^*,\cdots,p_K^*\}$, so the regret during Phase ℓ is bounded by:

$$\begin{aligned}
&\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^*-r_1(P_t))\right]\\
=&\mathbb{E}_1\left[(\tau_{\ell+1}-\tau_\ell)\times(r_1^*-r_1(P_\ell^*))\right]\\
\le&\mathbb{E}_1\left[\left(M_\Phi(P_\ell^*)e^{(\ell)}+1\right)\times(r_1^*-r_1(P_\ell^*))\right]\\
\le&\left(M_\Phi^*e^{(\ell)}+1\right)\sum_{i=1}^{K}(r_1^*-r_1(p_i^*))\times\mathbb{P}_1(P_\ell^*=p_i^*)\\
\le&\left(M_\Phi^*e^{(\ell)}+1\right)r_1^*\times\sum_{i=2}^{K}\mathbb{P}_1(P_\ell^*=p_i^*).
\end{aligned}$$

By the definition of the uPC policy, the choice of price P_ℓ^* is determined by the sample mean $\bar{X}_{\ell-1}$ in Phase $\ell-1$. Similar to the proof of Theorem 1, letting $\epsilon=\frac{1}{2}\min_{i\neq1}\left|d_1(P_{\ell-1}^*)-d_i(P_{\ell-1}^*)\right|$, we have

$$\begin{aligned}
&\sum_{i=2}^{K}\mathbb{P}_1(P_\ell^*=p_i^*)\\
\le&\mathbb{P}_1\left(\left|\bar{X}^{\ell-1}-d_1(P_{\ell-1}^*)\right|\ge\epsilon\right)\\
=&\mathbb{E}_1\left[\mathbb{P}_1\left(\left|\bar{X}^{\ell-1}-d_1(P_{\ell-1}^*)\right|\ge\epsilon\middle|P_{\ell-1}^*\right)\right]\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-(\tau_\ell-\tau_{\ell-1})\left(\frac{\epsilon^2}{2\sigma^2}\wedge\frac{\epsilon}{2b}\right)\right)\middle|P_{\ell-1}^*\right]\right]\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-M_\Phi(P_{\ell-1}^*)e^{(\ell-1)}\left(\frac{\epsilon^2}{2\sigma^2}\wedge\frac{\epsilon}{2b}\right)\right)\middle|P_{\ell-1}^*\right]\right]\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-2e^{(\ell-1)}\right)\middle|P_{\ell-1}^*\right]\right]\\
=&2/(e^{(\ell)})^2.
\end{aligned}$$

So

$$\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^*-r_1(P_t))\right]\le\left(M_\Phi^*e^{(\ell)}+1\right)r_1^*\cdot\frac{2}{(e^{(\ell)})^2}=\frac{2M_\Phi^*r_1^*}{e^{(\ell)}}+\frac{2r_1^*}{(e^{(\ell)})^2}.$$

In sum, the regret of uPC under demand d_1 is bounded by

$$\begin{aligned}
&\mathrm{Regret}_1^{\mathsf{uPC}}(T)=\mathbb{E}_1\left[\sum_{t=1}^{\tau_1}(r_1^*-r_1(P_t))\right]+\sum_{\ell=1}^{m}\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^*-r_1(P_t))\right]\\
\le&(M_\Phi(P_0^*)+1)(r_1^*-r_1(P_0^*))+\sum_{\ell=1}^{m}\left(\frac{2M_\Phi^*r_1^*}{e^{(\ell)}}+\frac{2r_1^*}{(e^{(\ell)})^2}\right)\\
\le&M_\Phi(P_0^*)(r_1^*-r_1(P_0^*))+r_1^*+(2M_\Phi^*r_1^*+r_1^*)\\
=&M_\Phi(P_0^*)(r_1^*-r_1(P_0^*))+2M_\Phi^*r_1^*+2r_1^*.
\end{aligned}$$

The minimax regret of demand set Φ is given by

$$\text{Regret}_{\Phi}^{\mathsf{uPC}}(T) = \max_{i=1,\dots,K} \text{Regret}_{i}^{\mathsf{uPC}}(T) \le C_{\Phi}(P_0^*) + 2M_{\Phi}^* r^* + 2r^*,$$

where $C_{\Phi}(P_0^*) = \max_{i\in\{1,\cdots,K\}}\{M_{\Phi}(P_0^*)(r_i^* - r_i(P_0^*))\}$ and $r^* = \max_{i\in\{1,\cdots,K\}} r_i^*$. □

A.3. Proof of Proposition 2

Proof of Proposition 2. Consider a price set $\mathcal{P} = \{1,2\}$ and two demand functions $d_1(1) = 0.6, d_1(2) = 0.25; d_2(1) = 0.4, d_2(2) = 0.25$. Demand per period has a Bernoulli distribution. It is clear that the optimal prices are $p_1^* = 1, p_2^* = 2$. This demand model violates Assumption 1, because $p_2^* = 2$ is not a discriminative price. We show that for this model, any non-anticipating policy must have a regret of $\Omega(\log T)$.

The one period regret for not using the optimal price is $a = 0.1$ under either demand function. For any policy, we let T_1 be the number of the times that $p = 1$ is used.

We prove the result by contradiction. Suppose $\text{Regret}_2(T) = a \cdot \mathbb{E}_2[T_1] = o(1) \cdot \log T$ and $\text{Regret}_1(T) = a(\mathbb{E}_1[T - T_1]) = o(1) \cdot \log T$. The change-of-measure inequality (see proof of Lemma 1) implies that for any event A,

$$\mathbb{P}_2(A) \le \mathbb{E}_1[1_A \exp(bT_1)].$$

where $b = \log(0.6/0.4)$.

Consider the event: $A = \{T_1 \le \log T/(2b)\}$, then we have

$$\mathbb{P}_2(A) \le \mathbb{P}_1(A) \exp(b \cdot \log T/(2b)) = \mathbb{P}_1(A)\sqrt{T}.$$

By Markov's inequality,

$$\mathbb{P}_1(A) = \mathbb{P}_1(T - T_1 \ge T - \log T/(2b)) \le \frac{\mathbb{E}_1[T - T_1]}{T - \log T/(2b)} = \frac{o(1)\log T}{T - \log T/(2b)}.$$

Thus, we have

$$\mathbb{P}_2(A) \le \frac{o(1)\sqrt{T}\log T}{T - \log T/(2b)} = o(1).$$

Using Markov's inequality again, we get

$$\mathbb{E}_2[T_1] \ge \frac{\log T}{2b}\mathbb{P}_2(T_1 \ge \frac{\log T}{2b}) = \frac{\log T}{2b}(1 - \mathbb{P}_2(A)) = \frac{\log T}{2b}(1 - o(1)).$$

This contradicts the assumption that $\mathbb{E}_2[T_1] = o(1) \cdot \log T$. □

A.4. Proof of Proposition 3

Proof of Proposition 3. Suppose d_1 is the underlying demand function. Let $k \le K-1$ be the number of iterations in the while loop.

The regret under demand d_1 can be composed as

$$\text{Regret}_1^{\mathsf{kPC}}(T) = \mathbb{E}_1\left[\sum_{t=1}^{T}(r_1^* - r_1(P_t))\right] \le \mathbb{E}_1\left[\sum_{\ell=0}^{k}\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right].$$

Let $\epsilon = \frac{1}{2}\min_{i:d_1(P^*_{\ell-1})\neq d_i(P^*_{\ell-1})}\left|d_1(P^*_{\ell-1}) - d_i(P^*_{\ell-1})\right|$. The probability that demand d_1 is eliminated in phase $\ell < k$ is bounded by

$$\begin{aligned}
&\mathbb{P}_1\left(|\bar{X}^{\ell-1} - d_1(P^*_{\ell-1})| \ge |\bar{X}^{\ell-1} - d_i(P^*_{\ell-1})| \text{ for some } i \neq 1\right)\\
\le&\mathbb{P}_1\left(\left|\bar{X}^{\ell-1} - d_1(P^*_{\ell-1})\right| \ge \epsilon\right) \qquad (21)\\
=&\mathbb{E}_1\left[\mathbb{P}_1\left(\left|\bar{X}^{\ell-1} - d_1(P^*_{\ell-1})\right| \ge \epsilon \middle| P^*_{\ell-1}\right)\right]\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-(\tau_\ell - \tau_{\ell-1})\left(\frac{\epsilon^2}{2\sigma^2}\wedge\frac{\epsilon}{2b}\right)\right)\middle| P^*_{\ell-1}\right]\right] \qquad (22)\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-\tilde{M}_\Phi(P^*_{\ell-1})\log T\left(\frac{\epsilon^2}{2\sigma^2}\wedge\frac{\epsilon}{2b}\right)\right)\middle| P^*_{\ell-1}\right]\right]\\
\le&\mathbb{E}_1\left[2\mathbb{E}_1\left[\exp\left(-\log T\right)\middle| P^*_{\ell-1}\right]\right]\\
=&2/T.
\end{aligned}$$

Inequality (21) is proved in Theorem 1, and (22) uses the Chernoff bound. Since $k \le K-1$, we have

$$\mathbb{P}_1\left(|\bar{X}^{\ell-1} - d_1(P^*_{\ell-1})| \ge |\bar{X}^{\ell-1} - d_i(P^*_{\ell-1})| \text{ for some } i \neq 1, 0 \le \ell < k\right) \le \frac{2(K-1)}{T}.$$

For each of the learning phase ($0 \le \ell \le k-1$), the regret is bounded by

$$\mathbb{E}_1\left[\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right] = \mathbb{E}_1\left[\left\lceil\tilde{M}_A(P^*_\ell)\log T\right\rceil(r_1^* - r_1(P^*_\ell))\right] \le \tilde{M}_\Phi r_1^*\log T + r_1^*.$$

The regret in the earning phase ($\ell = k$) is bounded by

$$\mathbb{E}_1\left[\sum_{t=\tau_k+1}^{T}(r_1^* - r_1(P_t))\right] \le T r_1^*\mathbb{P}_1(P_k \neq p_i^*).$$

So the regret of kPC under demand d_1 is bounded by

$$\begin{aligned}
&\text{Regret}_1^{\mathsf{kPC}}(T) = \mathbb{E}_1\left[\sum_{\ell=0}^{k-1}\sum_{t=\tau_\ell+1}^{\tau_{\ell+1}}(r_1^* - r_1(P_t))\right] + \mathbb{E}_1\left[\sum_{t=\tau_k+1}^{T}(r_1^* - r_1(P_t))\right]\\
\le&(K-1)(\tilde{M}_\Phi r_1^*\log T + r_1^*) + T r_1^*\frac{2(K-1)}{T}\\
=&(K-1)\tilde{M}_\Phi r_1^*\log T + 3(K-1)r_1^*.
\end{aligned}$$

The minimax regret of demand set Φ is given by

$$\text{Regret}_{\Phi}^{\mathsf{kPC}}(T) = \max_{i=1,\ldots,K} \text{Regret}_{i}^{\mathsf{kPC}}(T) \leq (K-1)\tilde{M}_{\Phi} r^* \log T + 3(K-1)r^*.$$

□

References

Aviv, Y. and Vulcano, G. (2012). Dynamic list pricing. In Özer, Ö. and Phillips, R., editors, *The Oxford Handbook of Pricing Management*, pages 522–584. Oxford University Press, Oxford.

Besbes, O. and Zeevi, A. (2009). Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420.

Besbes, O. and Zeevi, A. (2015). On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739.

Bitran, G. R. and Mondschein, S. V. (1997). Periodic pricing of seasonal products in retailing. *Management Science*, 43(1):64–79.

Boyacı, T. and Özer, Ö. (2010). Information acquisition for capacity planning via pricing and advance selling: When to stop and act? *Operations Research*, 58(5):1328–1349.

Broder, J. (2011). *Online Algorithms For Revenue Management*. PhD thesis, Cornell University.

Broder, J. and Rusmevichientong, P. (2012). Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980.

Caro, F. and Gallien, J. (2012). Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, 60(6):1404–1422.

Chen, Q., Jasin, S., and Duenyas, I. (2015). Real-time dynamic pricing with minimal and flexible price adjustments. *Management Science*. To appear.

den Boer, A. and Zwart, B. (2014). Simultaneously learning and optimizing using controlled variance pricing. *Management Science*, 60(3):770–783.

den Boer, A. V. (2014). Dynamic pricing with multiple products and partially specified demand distribution. *Mathematics of operations research*, 39(3):863–888.

den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1):1–18.

Feng, Y. and Gallego, G. (1995). Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, 41(8):1371–1391.

Harrison, J., Keskin, N., and Zeevi, A. (2012). Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58(3):570–586.

Keskin, N. B. and Zeevi, A. (2014). Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations Research*, 62(5):1142–1167.

Kleinberg, R. and Leighton, T. (2003). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 594–605.

Mersereau, A. J., Rusmevichientong, P., and Tsitsiklis, J. N. (2009). A structured multiarmed bandit problem and the greedy policy. *IEEE Transactions on Automatic Control*, 54(12):2787–2802.

Netessine, S. (2006). Dynamic pricing of inventory/capacity with infrequent price changes. *European Journal of Operational Research*, 174(1):553–580.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202.

Rusmevichientong, P. and Tsitsiklis, J. N. (2010). Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411.

Wang, Z., Deng, S., and Ye, Y. (2014). Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operations Research*, 62(2):318–331.

Zbaracki, M. J., Ritson, M., Levy, D., Dutta, S., and Bergen, M. (2004). Managerial and customer costs of price adjustment: direct evidence from industrial markets. *Review of Economics and Statistics*, 86(2):514–533.