Samuel Oshay
CIS 498: Thesis
Advisor: Michael Kearns
April 2021

A low-regret, differentially-private price individualization framework

ABSTRACT

In this thesis, we develop an $O(1/\varepsilon)$-regret price individualization model for any ε-level of differential privacy.

INTRODUCTION

"No man ever came to market with less seductive goods,
and no man ever got a better price for what he had to offer." [1]

For most of economic history, prices were dynamic, with specialized peddlers in ports and markets becoming expert negotiators in their limited catalog. Over time, as catalogs grew and distribution scaled, the ability for individual sellers to optimize their revenue withered. By the Industrial Revolution, the price tag became the norm, and every buyer received the same price for each unit of sale of a good.[2] A century and a half later, e-commerce and machine learning have enabled the scaling of the older, dynamic model. Now, for the first time in centuries, different buyers at different times are receiving radically different prices for nearly every good, with price variations determined by the endless array of personally-identifiable information that has been collected about each of us, with and without our consent.

The use of user data in price discrimination algorithms brings up a tough question: is it *fair* to use someone's personal data to alter the prices they are offered? In other domains, the use of user data has had clearly undesirable effects when it comes to common notions of fairness: for example, criminal sentencing algorithms have been shown to have racial bias, and job advertisement algorithms have been shown to have gender bias.[3] In pricing, though, the use of some personal data is already commonplace. Take, for example, the senior discount, where a younger person will be up-charged for their youth, or Uber's surge pricing. In another academic department, entire theses could be written dissecting the moral arguments behind a slew of pricing case studies. For our purposes though, it suffices to say that the use of personal data in pricing is a breeding ground for moral dilemma; it is an arena with many competing imperatives continually at play.

---

[1] An old joke by H.L. Mencken, though I can't find the source or the subject.

[2] Goldstein, Jacob and Jess Jiang, hosts. "The Birth And Death Of The Price Tag." *Planet Money*, NPR, 17 Jun. 2015.

[3] Friedler, Sorelle A., Carlos Scheidegger, Suresh Venkatasubramanian. "The (Im)possibility of Fairness: Different Value Systems Require Different Mechanisms For Fair Decision Making," *Communications of the ACM*, April 2021, Vol. 64 No. 4, Pages 136-143. 10.1145/3433949.

Before we philosophize too deeply, let us ask if there is an algorithmic shortcut around this dilemma: is there a way to enable traditional price individualization and optimization models to operate *fairly*, perhaps by offering a probabilistic guarantee that a user's data will not be used to up-charge them? In this thesis, we will show that the answer is yes. This thesis combines revenue management algorithms with differential privacy to produce a novel pricing framework with both privacy and profitability guarantees. In the following pages, we will produce an $O(1/\varepsilon)$-regret price individualization model for any ε-level of differential privacy.

The significance of pricing algorithms is hard to overstate. In the current discourse, income and wealth inequality are issues of top interest. This interest, of course, rests upon the idea that a dollar for each person is equally valuable; that is, a dollar in one pocket has the ability to buy the same goods as in another. In light of dynamic pricing, this ideal is simply not the case. Two people can be charged wildly different prices for the same good, often at the same time. Adding fairness to pricing through privacy creates a safeguard against many undesirable effects.

Dynamic pricing and its early practitioners

Markets have existed for so long and in so many lands, that they are likely an intrinsic property of human cooperation. One might daydream of the oldest of markets: a *bazaar* or a *souq*, the central square on a caravanserai along an early trade route. There, one might imagine trade's steady hum, of buyers and sellers coming to terms on an exchange. Anthropological research has shown that as early as 3000 BCE, these canonical bazaars were located near citadels and palaces, placed with the express intention of selling to rulers. Of course, the sellers in these bazaars knew that the ruler's wealth implied a grand willingness-to-pay, and therefore hand-negotiating even routine deals gave sellers the power to extract wealth from buyers.[4]

This haggling, of course, was the earliest form of dynamic, or individualized, pricing: the setting and resetting of prices to optimize revenue or profits over time. The market mechanisms that lend themselves to dynamic pricing are twofold. First, if the product expires at some point in time, an urgency of sale is created that drives sellers to change prices for a certain number of units. Second, if there is a limited inventory that can only be expanded at great marginal cost or with long time delays, the value of the n-th remaining unit of a good is the willingness-to-pay of the n-th highest bidder under unit demand; this dynamic creates an ever-increasing market-clearing price, which, of course, the seller would be happy to exploit.[5] The interaction effect of these two

---

[4] Mohammadreza Pourjafar, Masoome Amini, Elham Hatami Varzaneh, Mohammadjavad Mahdavinejad, "Role of bazaars as a unifying factor in traditional cities of Iran: The Isfahan bazaar," *Frontiers of Architectural Research*, Volume 3, Issue 1, 2014, Pages 10-19, ISSN 2095-2635, https://doi.org/10.1016/j.foar.2013.11.001.

[5] McAfee, R. Preston and Vera te Velde, "Dynamic Pricing in the Airline Industry." *Handbook on Economics and Information Systems*, Ed: T.J. Hendershott, *Elsevier Handbooks in Information Systems*, Volume 1; ISBN 0444517715, 2007.

conditions is that the market for a good becomes a time-dependent system where the value of the marginal unit fluctuates dramatically based on the time to expiry, number of bidders, remaining inventory, and other pieces of the extended context. Over the following pages, we will work to mathematically define and refine these forces.

Given the immense value that can be unlocked by dynamic pricing, the practice is found in a variety of modern-day industries where these two conditions are met. Consider the average vacation:

1. You book a hotel through a travel website, such as Expedia or Priceline, where the price of the room changes day-to-day. Of course, hotel rooms on a particular night are an expiring good: if they go unbooked, no revenue is collected and the opportunity is lost. In addition, hotel rooms are a constrained good: if the hotel is full, there is no way to easily add an additional room on demand.
2. You book a flight, where the price of your seat changes day-to-day. Just as with the hotel, flight seats are both expiring and constrained.
3. You grab an Uber to the airport, where you encounter surge pricing. This surge pricing, as many know, is driven by the supply of drivers and demand of riders in your area.[6]
4. Your travel companion, though, arrives separately in their own car, having stopped at the gas station and paid slightly more per-gallon than they did yesterday.
5. You get to your hotel and order fish in the lobby restaurant, where the price is listed as market price.

While dynamic pricing is everywhere, there are various forms.

The three degrees of price discrimination

To economists, dynamic pricing is also known as price discrimination: the use of different market constructions to apply differential pricing regimes and capture *alpha*. To understand alpha, let us consider a single-price seller subject to a demand curve $D(p)$, where p is the single-price they are able to set. To optimize revenue, all this seller must do is optimize $D(p)p$, the maximum value of which we can define as R^*. Now, what if this seller could set multiple prices, maybe through a senior discount? We might represent this revenue as $R^* + \alpha$. As such, we can think of alpha as the marginal value of price individualization above and beyond single-price selling.[7]

In economic theory, there are three degrees of price discrimination, each focusing on a different subset of information to generate alpha. In price discrimination of the first degree, each

[6] "How Surge Pricing Works," *Drive for Uber*. Uber. 2021.
[7] Note that a bad pricing model might have negative alpha.

buyer is charged *exactly* their willingness-to-pay, thereby maximizing alpha. As the entire willingness-to-pay curve is captured for the seller in this case, this degree is also known as perfect price discrimination, and it is the ultimate goal for every multi-price-setting seller. As we will see, most pricing algorithms from both computer science and revenue management research are *price individualization* models, themselves just approximations of perfect pricing. In the wild, perfect price discrimination is most commonly seen in airline and entertainment tickets, multi-buyer auctions, hotels, gasoline prices, and–everyone's favorite–Uber's surge pricing.

The second degree of price discrimination is product bundling, in which the product is bundled in varied manners, leading to different pricing regimes. The most popular form of bundling is volume discounts, wherein a certain price is applied for, say, the first k units, and a lower price is applied for all units thereafter. Another common example of second-degree price discrimination is Amazon Prime, where the buyer is given a different price depending on whether or not they bought a membership. (The same designation applies to Costco memberships.) Price discrimination of the second degree has gained particular importance in software-as-a-service, where a buyer might pay a fixed setup fee plus a variable usage-based fee, such as for API access.[8]

The third degree of price discrimination is segmented pricing, wherein separate groups of individuals are given different prices. Here, we find varied examples from a senior discount to in-state and out-of-state tuition differences.

While not traditionally defined as a degree of price discrimination, the term "dynamic pricing" itself emphasizes the time of the transaction and the current supply and demand for the good being transacted. Similarly, cost management models emphasize fluctuations in the cost-basis of a good, as well as whether or not cost changes can be passed on to buyers.[9]

---

[8] McKenzie, Patrick. "A Guide to Software-as-a-Service Pricing: Pricing low-touch SaaS," *Stripe Atlas*.
[9] Acemoglu, Daron, et al. *Economics*. Pearson, 2019.

Toward a mathematical model for pricing

From the degrees of price discrimination above, we see that a great number of data points might be considered in a pricing decision. Intuitively, we might think that buyer information, time, inventory level and supply constraints, demand estimates, and cost might all be used in a pricing decision, as well as many other parameters both known and estimated. Mathematically, let us define the following ingredients for a pricing model:

1. Let $\Gamma(p)$ be our aggregate demand function and let $\Gamma_i(p)$ be our individual demand function for buyer i. In traditional economic theory, Γ is treated as a deterministic function of price. For a given price p, the market will, in aggregate, demand $\Gamma(p) = \sum_i \Gamma_i(p)$ units. In this thesis, $\widehat{\Gamma}$ will refer to our estimate of Γ, conditional on some informational state.
2. Let w be an arbitrary context vector containing buyer, seller, inventory, cost, and product information, as well as any other contextual information about the transaction.
3. Let t be time and $I(t)$ be the inventory level at time t.

Thus, we will define an ideal pricing model M as a price p-producing function that maximizes revenue while inventories are available:

$$p_i = M(w_i, t) \text{ that maximizes } \sum_i \Gamma_i(p_i|w_i) * p_i \text{ while } I(t) > 0$$

In English, we see that M produces prices that maximize revenue across all buyers, so long as goods remain in inventory.

PRICING MODELS

In practice, pricing algorithms from the field of revenue management focus on data that is most easily available to sellers: cost, supply, and demand. Cost, for the most part, is the easiest, as sellers can directly observe it during the purchasing process.[10] Similarly, when deliveries are received and shipments are sent, sellers have a real-time view into their inventory: a directly observable proxy for market supply. Demand, on the other hand, is forward-looking and cannot be observed. As such, many pricing models assume or require a demand forecasting model, which itself is a field of computational research and is outside the scope of this thesis.

Cost-plus

Cost-plus is the simplest and most commonly-used pricing rule; it gives a good's unit price by its cost c and a profit markup m:

$$p = c * (1 + m)$$

The simplicity and widespread use of cost-plus speaks to two real-life concerns of price strategy. First, sellers prefer a justifiable pricing model. Pricing, crucially, is a repeated game, in which revenue retention is far more valuable than a few percentage points here or there on any one purchase. Later on, we will quantify this "reputation premium" as the expected value of customer retention. Second, most sellers prefer a simple model, as they are often time-constrained and without the know-how or resources to invest in a more sophisticated pricing model. Thus, for millions of businesses all over the world, someone sits down, looks at their costs, picks a profit margin they like, and sets a static price. Of course, we might recognize a fatal flaw in this scenario: cost-plus has no clear mechanism for optimizing the price or the markup. How does a seller know when the markup they have chosen is right?

Optimization using the price elasticity of demand

Price elasticity of demand z measures the relationship between p and Γ:

$$z = - [\partial\Gamma / \Gamma] / [\partial p / p]$$

In theory, the value of z can say a lot about the market a seller is operating in:

---

[10] It is worth mentioning that the specifics of cost accounting get very complicated, and many businesses struggle with accurately and meaningfully calculating their cost of goods sold.

- If $|z| = \infty$, the market is perfectly elastic, meaning that any small change in price will garner a complete loss in demand. In reality, such a market is present for commodities, where a price is set and everyone must accept it.
- If $|z| = 0$, the market is perfectly inelastic, meaning that a change in price will garner precisely no movement in demand. In reality, such a market cannot exist, as everyone runs out of buying power eventually. Approximations of such a market include healthcare, wherein the cost of an operation or procedure is nearly irrelevant when death is the alternative.
- If $|z| = 1$, the market is unit elastic, meaning that a small change in price is perfectly offset by a small change in quantity, such that the overall revenue remains unchanged. As a consequence, when $|z| = 1$, revenue is maximized, as proved below.

The proof is rather straightforward. For our single-price seller, the revenue effect of a price increase is as follows:

$$
\begin{aligned}
&\frac{\partial R}{\partial p} \\
&= \frac{\partial}{\partial p}\Gamma(p) * p \\
&= \Gamma(p) + p * \frac{\partial \Gamma}{\partial p} \\
&= \Gamma(p) * [\ \Gamma(p) + p * \frac{\partial \Gamma}{\partial p}\ ] \ / \ \Gamma(p) \\
&= \Gamma(p) * [\ 1 + \frac{\partial \Gamma \ / \ \Gamma(p)}{\partial p \ / \ p}\ ] \\
&= \Gamma(p) * (1 - z)
\end{aligned}
$$

With our revenue effect established, we now have three cases:

1. If $(1 - z) > 0$, then the revenue effect is positive, as we know that $\Gamma(p)$ is definitionally positive. Thus, the seller makes more money than before and would be correct to continue to raise prices until $\Gamma(p) * (1 - z) = 0$ and maximum revenue has been reached.
2. Similarly, if $(1 - z) < 0$, then the revenue effect is negative. Thus, the seller makes less money than before and would be correct to lower prices until $\Gamma(p) * (1 - z) = 0$ and the maximum revenue has been reached.
3. If $(1 - z) = 0$, then the revenue effect is zero, and the revenue itself is already maximized. This maximization, as desired, occurs when $z = 1$.[11]

[11] Acemoglu, Daron, et al. *Economics*. Pearson, 2019.

Shortcomings of the baselines

For the economist, price optimization by the elasticity of demand is a tried-and-true strategy, though the more statistically-minded will observe a weakness: it treats preferences (so far denoted by the demand curve Γ) as a deterministic function. A traditional economists' approach to elasticity would outline a few determinants of the value, among them:

- The quality and availability of substitutes. If a substitute good, or a good that fulfills the same "job to be done," is available at better value-for-quality, the original good tends to be more elastic, as demand can shift over to substitutes in response to a price change.
- The purchasing power of the buyer. If the buyer has exceptional purchasing power relative to the cost of the good, the good tends to be inelastic, as the willingness-to-pay is not constrained.
- Necessity and addiction. In the real world, there are no perfectly inelastic markets, because everyone runs out of money eventually. Even so, we might state that the demand is (nearly) perfectly inelastic in the case of necessity and addiction.
- Time. The longer the time horizon of a market, i.e. the longer a price will be set for, the more elastic it tends to be, as buyers have more time to find substitutes or change their needs and preferences in response to price changes.[12]

Considering these "traditional" factors of elasticity, we might recognize that they are all stochastic: substitutes disappear and emerge constantly; every day fortunes rise and fall; priorities and needs change. In addition, we might consider another "factor" affecting elasticity: that multiple Nobel prizes have been awarded to the experimenters who have measured and validated just how fickle, irrational, and without complete information the average buyer is.[13] Indeed, we must think of Γ as a density function for our demand estimate: given a price, what do we believe the distribution of demand to be? How do we observe and sample it? With this framing, we must now find a model that both samples demand, as well as leverages our knowledge of it over time to truly optimize revenue.

Setting random prices

Imagine a seller who sets random prices. Say the seller gets n customers per day, who individually have demand Γ_i , where $i \in [1, n]$. Let us make some assumptions about Γ that we will maintain for the remainder of this work:

1. p will always be real and positive.

[12] Ibid.

[13] Consider the work of Daniel Kahneman, Amos Tversky, and Richard Thaler, among many others.

2. $\Gamma_i(p)$ monotonically decreases as p increases for all i. This assumption is a standard in economic theory, known as a downward-sloping demand curve.[14]
3. $\Gamma_i(p)$ is finite and non-negative for all i and all values of p.
4. There exists a finite price υ such that $\Gamma_i(p) = 0$ for all i and all $p > \upsilon$. Here, we can think of υ as one cent more than anyone would pay and υ_i as one cent more than buyer i would pay.
5. If the quote price $p_i > \upsilon_i$, the buyer finds a new seller and never returns ("churns"). In this case, the lost customer lifetime value can be stated as simply $\Gamma_i(p^*)p^*L_i$, where L_i is the length of the customer lifetime of the buyer in periods and p^* is the revenue-maximizing price for our single-price seller.[15]

With these conditions laid out, let us specify the nature of our sellers' random pricing: for each buyer, the seller picks a price p_r uniformly across $[0,\ \upsilon]$ and observes revenue $p_r\Gamma_i(p_r)$. Here we assume that we know υ, though this is largely not the case in real life. In the next section, we will relax this assumption. If we then plot $\Gamma_i(p_r)$ by p_r, we will begin to plot out our aggregate demand function Γ. As $n \to \infty$, this resultant revenue graph would seem to converge to our aggregate demand density, sampling each p_r multiple times for each buyer. With this knowledge, it is trivial to find p^*, the price that maximizes revenue in expectation.

Overall, random pricing is a robust method of price discovery, although it is very expensive, with the lost revenue or "cost of information" as follows:

$$\sum_i \Gamma_i(p^*)p^* - \sum_i \Gamma_i(p_i)p_i + \sum_i P(p_i > \upsilon_i)\Gamma_i(p^*)p^*L_i$$

In plain English, the "cost of information" is ideal revenue minus collected revenue plus the cost of lost buyers. To the astute reader, one term jumps out: there are no guarantees whatsoever about $P(p_i > \upsilon_i)$. It just might so happen that our pricing upper-limit of υ is very, very high, and we randomly quote very high prices to a great number, all of whom churn. As such, we need a better manner of price searching, perhaps one that can narrow our search range.

---

[14] Acemoglu, Daron, et al. Economics. Pearson, 2019.

[15] Ghose, A., P. Ipeirotis and A. Sundararajan. Reputation Premiums in Electronic Peer to Peer Markets: Analyzing Textual Feedback and Network Structure. Proceedings of the *ACM SIGCOMM Workshop on Economics of P2P*, Philadelphia, August 2005.

Setting prices with Binary Price Search

The good news is that we can do better than random sampling by using a binary search. Here, we can improve upon uniform random price sampling in two directions. First, we reduce $P(p_i > \upsilon_i)$ across transactions. Second, we reduce the number of observations required to find certainty, thereby reducing the maximum number of sub-optimal transactions and alienated buyers. To achieve these reductions, we must first take a page from Arthur Laffer, the economist who famously observed that tax rates of 0% and 100% both garner zero governmental revenue. Only in the interval is income tax elastic. In our pricing construction, we see that $p_i = 0$ gains no revenue, as no money is collected for each unit sold. In addition, we see that $p_i = \upsilon_i$ garners no revenue, as there is no demand. As we know that υ is finite, what if we just binary-searched this interval to find a single optimal price p^*?[16] We could do just that, recursing onto the higher-revenue quarter points on either side of our interval's midpoint, starting with $[0, \upsilon]$.

What would the cost of this search be? Let us recall that random search required n observations to move our estimate of p^* within some provable bounds of optimal. Consider the following proof that only $log(n)$ observations are needed for convergence to those same bounds with random pricing. In random pricing, we sampled an interval of length υ a total number of n times. Thus, in expectation, the distance between each observation point (and therefore the granularity of our price estimate) is $\frac{\upsilon}{n}$. Now, let the price discovered by binary search be p_b. After halving our search interval $log(n)$ times, we know that $p^* - p_b < \frac{\upsilon}{n}$.

In plain English, the cost of information under this search is ideal revenue minus revenue collected over only $log(n)$ transactions minus revenue collected over $n - log(n)$ transactions with price p_b plus the cost of lost buyers over only $log(n)$ transactions:

$$R^* - \sum_{i \in [0,\, log(n)]} \Gamma_i(p_i)p_i - \sum_{i \notin [0,\, log(n)]} \Gamma_i(p_b)p_b + \sum_{i \in [0,\, log(n)]} P(p_i > \upsilon_i)\, \Gamma_i(p^*)p^* L_i$$

Asymptotically, p_b converges to p^*, such that we can rewrite this cost as:

$$\sum_{i \in [\, 0,\, log(n)\,]} [\Gamma_i(p^*)p^* - \Gamma_i(p^*)p^* - P(p_i > \upsilon_i)\Gamma_i(p^*)p^* L_i]$$

Here, we make two observations. First, the number of sub-optimal transactions is $O(log(n))$, a different order than in random pricing, where it was $O(n)$. Second, $P(p_i > \upsilon_i)$ is less under this search than uniform for the average buyer, as we have "learned" from prior observations to set

[16] This, of course, only works if the revenue curve is convex, which we cannot prove without making assumptions about the preferences of individuals.

prices into a reduced segment of the pricing interval. Formal proof for this statement would require us to make assumptions about the demand functions Γ_i, but we can intuit a proof using the Central Limit Theorem. If indeed all deterministic demand functions Γ_i converge to normal, we know that $P(p_i > \upsilon_i)$ strictly declines relative to uniform sampling, as the probability of selecting an outlier from a normal distribution across an interval is lower than doing the same over a uniform distribution.

As we can see, better price sampling leads to strictly better outcomes: we understand what our demand curves look like while incurring fewer costs to obtain information. One must ask, though, how much better can we do? Is there a better price sampling algorithm than binary search?

BANDITS[17]

The multi-armed bandit problem is a resource allocation problem wherein an agent must allocate constrained resources to various alternatives, the payoffs of which are all uncertain. The problem derives its unique name from the "one-armed" bandit: a slot machine. Consider a gambler who approaches a slot machine M. Say that the gambler has allocated enough money for n pulls. With this decision made, the cost of play is fixed; it is a sunk cost. His interest lies solely in maximizing his payout. Consider the "win" random variable $W(M, i \in [1, n])$, parameterized by the machine and the pull number, that returns a scalar payoff, the higher the better. With these definitions, we see that our gambler's expected payoff is simply $E[W(M, i)] * n$, as he has no choice but to play machine M.

Now consider a row of k slot machines, $M_1, \ldots, M_k$. Although our gambler doesn't know the exact behavior of W, he knows that the machine itself is a parameter and that some machines pay out better than others. Now our gambler faces a resource allocation problem: how does he allocate his n pulls across the k machines to maximize his expected winnings?

Intuitively, we see that the gambler is split between two imperatives. First, the gambler must *explore* W, sampling each machine to try to understand its payoffs. Second, the gambler must *exploit* W, allocating pulls to the best machine to maximize his earnings. Solving this trade-off, of course, is the multi-armed bandit problem.

If we leave Las Vegas and return to our market, we see that our seller faces the same problems as the gambler:

- The seller sees n buyers per period sequentially, just as the gambler has n pulls.
- The seller has k pricing options, just as the gambler has k slot machine options.
- The seller faces an uncertain demand curve $\Gamma_i(p)$, just as the gambler faces an uncertain payoff function $W(M, i)$.

One nuance to be considered is the nature of continuity in pricing. In all models considered this far, the price interval has been treated as continuous and differentiable, though the mapping of pricing onto the bandit problem necessitates a finite number of prices, a finite number of "slot machines." The question arises: are prices continuous or discrete? Answer: prices are continuous but can be bucketed into a discrete number of prices over $[0, \upsilon]$. The granularity of rounding is immaterial for our analysis, whether it be millions (in the case of planes), dollars (in the case of dinners), or mille (in the case of gasoline).

With our pricing problem mapped to a well-researched domain, what type of strategies can we use to simultaneously perform price discovery (exploration) and dynamic pricing (exploitation)?

[17] Ryzhov, Ilya O. and Warren B. Powell. "Bandit Problems," *Optimal Learning*.

The Epsilon Greedy Strategy

In the epsilon-greedy bandit strategy, our seller's n pricing decisions are partitioned into an explore segment and an exploit segment across the discrete prices p_1 through p_s. To run this strategy, our seller must do the following:

1. Select a probability ε.
2. Set an initial optimal price estimate $\hat{p}$ to a random price.
3. For each buyer $i \in [1,\ n]$, select a random real $r \in [0,1]$.
    a. If $r > \varepsilon$, exploit and offer the good for sale at $\hat{p}$. Here, indicator $I = 0$.
    b. If $r < \varepsilon$, explore and offer the good for sale at a price selected at random p_r. Let the mean quantity demanded be $m(\Gamma(p_r))$, over all buyers to whom p_r was offered. If $m(\Gamma(p_r))p_r > m(\Gamma(\hat{p}))\hat{p}$, update $\hat{p} = p_r$. Here, indicator $I = 1$.

With this strategy laid out, let us notice that the random pricing model is a degenerate case of the epsilon-greedy strategy, where $\varepsilon = 1$. As we did above, let us quantify the cost of this strategy. Now, we must recognize that $\hat{p}$ is not a theoretical constant; rather it is a function of time $\hat{p}(t)$ that evolves as we make observations and learn about demand. The cost of this overall strategy has a few components:

1. $R^* = \sum_i \Gamma_i(p^*)p^*$, the theoretical maximum revenue attainable with perfect play.
2. $\sum_i I(\varepsilon, i)\Gamma_i(p_r)p_r$, the revenue gained through exploration.
3. $\sum_i [1 - I(\varepsilon, i)] * \Gamma_i(\hat{p}(i))\hat{p}(i)$, the revenue gained through exploitation.
4. $\sum_i P(p_i > \upsilon_i) * \Gamma_i(p^*)p^* L_i$, the "reputation premium" or cost of customer churn.

Let us now consider the behavior of this bandit's total cost as $n \to \infty$. Let's make the optimistic assumption that our true perfect price p^* equals our estimate of perfect price $\hat{p}(t)$ at all times. In other words, assume our initial guess was perfect and that we never change it. For all pulls with $r \in [1 - \varepsilon, 1]$, we will collect perfect revenue. For all pulls otherwise, we will incur a loss of $(L_i + 1)\Gamma_i(p^*)p^*$, assuming the worst case where we select a p_r such that $\Gamma_i(p_r) = 0$ and the customer churns. At asymptote, then, our expected cost is at most $n\varepsilon(L_i + 1)\Gamma_i(p^*)p^*$. If ε is set to some constant, as well, then we know that this strategy will have some asymptotic cost of order $O(n)$. In bandit literature, the cost of some strategy at asymptote is known as *regret*, and as such an epsilon-greedy strategy with constant ε has linear regret.

Let us now observe that the structural framework of this strategy has two major components we might be able to toy with. First, there is a partition heuristic, or a mechanism by which the agent decides whether to explore or exploit a particular pull. Second, there is an exploration heuristic, which the agent uses to decide which price to explore next. Are there ways to alter either of these components to improve performance?

In short, yes, and we don't even have to go very far. Consider any ε function where the $\lim_{n \to \infty} \varepsilon = 0$, such as $\varepsilon = 1/n$. Mathematically, it is clear that such a partition heuristic would lead to a zero-regret strategy, as despite constant costs of exploration, we do so increasingly infrequently. In reinforcement learning theory, a strategy satisfying this constraint is GLIE: greedy in the limit of infinite exploration.[18]

Assuming a random starting state[19]

The intuitive explanation above lacks a critical aspect of bandit theory: the notion of information state. How do our asymptotic costs change if we aren't gifted the perfect price as our starting position? Imagine the seller is in some information state S^k at time i where our estimated perfect price is $\hat{p}(t)$. For the following iterations, the seller can either exploit and receive some reward $\lambda = \Gamma(\hat{p}(k))\hat{p}(k)$ or explore and receive some reward $\varkappa = \Gamma(p_r)p_r$. However, by exploring, the seller progresses to state S^{k+1}, where they have more information about $\hat{p}(t)$, now knowing a better estimate of $\hat{p}(k+1)$. If the seller exploits, then they gain no information about the best alternatives and stay in the state S^k. Thus, the value of being in state S^k at time i is:

$$V(S^k, i) = max \; \{ \; \Gamma(\hat{p}(k))\hat{p}(k) + V(S^k, \; i+1) \; , \; E[\Gamma(p_r) * p_r \mid S^k] \; + \; V(S^{k+1}, \; i+1) \; \}$$

Or, more simply:

$$V(S^k, i) = max \; \{ \; \lambda \; + V(S^k, \; i+1) \; , \; E[\varkappa \mid S^k] \; + \; V(S^{k+1}, \; i+1) \; \}$$

Now, if exploitation ever becomes the dominant strategy in S^k, the seller will continue to exploit all the remaining buyers, as they never learn any information about alternatives that would lead them to deviate their strategy. Thus, once all n iterations of this strategy are complete:

$$V(S^k, i) = max \; \{ \; (n - i + 1) * \lambda \; , \; E[\varkappa \mid S^k] \; + \; V(S^{k+1}, \; i+1) \; \}$$

Let us now consider an λ that solves the following equation:

[18] Shimkin, Nahum. "Efficient Exploration," *Learning in Complex Systems*. 2011.
[19] Ryzhov, Ilya O. and Warren B. Powell. "Gittins Index," *Optimal Learning*.

$$\lambda = \Gamma(\hat{p}(k))\hat{p}(k) = \frac{1}{(n-i+1)} * [\ E[\varkappa \mid S^k]\ +\ V(S^{k+1},\ i+1)\ ]$$

If the seller ever observes that their exploitation revenue is greater than the right-hand side above, then our partition function can switch us to exploit forever. Let us call this value the trigger revenue. At this trigger value, we find ourselves close enough to p^* that additional information is not as useful as the cost of a suboptimal guess. As we know that our search converges to p^* asymptotically, we see the right-hand-side above converges to zero, and our strategy is therefore zero-regret.

Contextual Thompson Sampling[20]

Let us revisit the critical failure of the first epsilon-greedy strategy that led to linear regret: that over time, the seller using this strategy does not explore less as they gain more information. Indeed, as the construction above shows, there is a point where marginal knowledge is just not worth it, and we need to exploit forever. The upper-confidence bound strategy (UCB) solves this problem. With UCB, we build a confidence interval for each price's performance using our limited data, and we optimistically assume that the revenue gained will be at the upper limit of our interval. This strategy lends itself to more efficient exploration, as the size of the confidence interval is $O(\sqrt{n})$.[21] Moreover, by choosing the size of the confidence interval, whether 99% or 90% or otherwise, the seller gets to choose the relative ratio of exploitation and exploration.

Let us dive more deeply into a related framework: Contextual Thompson Sampling (CTS). CTS is a derivative of a UCB strategy that leads the seller to act "optimistically optimal," with respect to a random belief about the optimal price. As Thompson himself wrote in 1933, one should "randomly take action according to the probability you believe it is the optimal action."[22] We can mathematically formulate Thompson's heuristic for our seller as follows. First, let us define the following values:

1. A revenue distribution function $P(R \mid \theta,\ p,\ w)$, for learned state θ, price p, and buyer information vector w.
2. A prior on our revenue distribution function's parameters $P(\theta)$.
3. An informational state S, containing all past observations.

[20] Russo, Daniel, et al. "A Tutorial on Thompson Sampling," *Foundations and Trends in Machine Learning*, Vol. 11, No. 1, pp. 1-96, 2018.
[21] Ryzhov, Ilya O. and Warren B. Powell. "Bandit Problems," *Optimal Learning*.
[22] Thompson, William R. "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples," *Biometrika*, Vol. 25, No. 3/4 (Dec., 1933), pp. 285-294. Oxford University Press.

With these inputs, CTS leads us to select a price $\hat{p}$ with probability as follows, where I is again an indicator function:

$$\int I[\, E[\, R \mid \theta,\, \hat{p}^*,\, w\,] = max_p\, E[\, R \mid \theta,\, p,\, w\,]\,] \ * P(\theta \mid S)\, d\theta$$

Augmented for use in our streamwise price-setting environment, we might consider θ to be information about Γ_i, our particular buyer's demand curve:

1. For each time step i, sample randomly from the distribution of individual demand curves Γ to find Γ_i.
2. Quote price $p_i = \ argmax_i\, \Gamma_i(p_i \mid w_i)p_i$ and observe revenue R_i.
3. Update our understanding of Γ using the maximum likelihood estimator given our observed revenue R_i.

Earlier, we noted there exists a class of universal pricing functions that allowed not just for price optimization for a single-price seller, but rather price individualization of the first degree:

$$p_i = M(w_i, t) \text{ that maximizes } \sum_i \Gamma_i(p_i|w_i) * p_i \ \text{ while } \ I(t) \ > 0$$

This CTS model is the first bandit we have encountered that accepts a buyer state w_i. Despite accepting additional information above and beyond state, Contextual Thompson Sampling has non-zero regret; as shown by researchers, the regret bound for Contextual Thompson Sampling is: $O(\sqrt{n})$, as with UCB more generally.[23] As such, let us use CTS as an example of a low-regret model that is preferable in practice to a zero-regret model because of its other properties, such as computational efficiency.

In the next section, we will deliver on the promise of this thesis' title: we will construct a low-regret pricing model that offers a privacy guarantee as compensation for having non-zero regret.

[23] Agrawal, Shipra and Navin Goyal. "Thompson Sampling for Contextual Bandits with Linear Payoffs," Microsoft Research.

ADDING PRIVACY TO PRICING

Differential privacy[24]

A randomized algorithm f is ε-differentially private if for all input datasets A and B differing by exactly one observation and all $S \subseteq Range(f)$:

$$P(f(A) \in S) \leq exp(\varepsilon) * P(f(B) \in S)$$

In plain English, a differentially private algorithm is such that the inclusion or exclusion of a particular piece of information, say, a user's data, is nearly impossible to distinguish based on the output.

Theoretical privacy upper bound on Thompson Sampling

The probability density of a price being quoted in a Thompson Sampling model is simply:

$$E_i [\ \Gamma_i(p_i \mid w_i) * p_i\ ] \ / \int_{p=0}^{p=\upsilon} E_i [\ \Gamma_i(p_i \mid w_i) * p_i\ ]\, dp$$
$$\leq R_i / [\ \upsilon * max_i\, R_i\ ]$$
$$\leq max_i\, R_i / [\ \upsilon * max_i\, R_i\ ]$$
$$= \frac{1}{\upsilon}$$

Thus, vanilla Thompson Sampling is at most $\frac{1}{\upsilon}$-private. Intuitively, this result makes sense: if the maximum price anyone is willing to pay is very large, then deviations in price can be "hidden" by the scale of the price. As $\upsilon \to 0$, our pricing model becomes less and less private, as any small deviation is a larger percentage of the overall market.

Of course, for the purposes of controlling privacy in a pricing model, this result shows us nothing, as a seller cannot control the maximum willingness-to-pay in their market. Now, let us consider a model that gives us a tunable privacy parameter ε.

[24] Dwork, Cynthia and Aaron Roth. *The Algorithmic Foundations of Differential Privacy.* 17.

The Fair Pricing Model[25]

Let us now insert a *price transformation* that adds differential privacy to a CTS-inspired epsilon-greedy individualized pricing strategy. Consider the following Fair Pricing Model (FPM):

1. Select a probability β representing the percent of observations to be explored.[26]
2. Set an initial optimal price $\hat{p}(0)$ to a random price.
3. For each buyer $i \in [1, n]$, select a random real $r \in [0, 1]$:
    a. If r > β, exploit and offer the good for sale at a private price φ_i randomly selected according to $exp\left\{ \varepsilon * E_i[\Gamma_i(\hat{p}^*(i) \mid w_i) * \hat{p}^*(i)] \right\}$.
    b. If r < β, explore and offer the good for sale at a price selected at random p_r.
    c. Update $\hat{p}(t)$ and Γ_i simply using Bayes' rule.

Here, we have created a pricing individualization model that has both low-regret and a guarantee of differential privacy. The intuition here is as follows:

1. If using a perfectly individualized price over-leverages the user's data, let's quote a *roughly* individualized price, such that it is impossible to tell whether or not the user's data was even used in the personalization process.
2. In order to maintain revenue maximization targets, we should choose the prices that are more likely to optimize revenue over time, with a parameter that allows the "widening" and "thinning" of the probability distribution to enable the choice of more private (and therefore unoptimized, options). Notice that our density definition exponentially weights higher revenue options.

The FPM exhibits an $\varepsilon(1 - z)$-level of differential privacy

The probability density of φ_i is:

$$P(\varphi_i \mid w_i)$$

$$= \left[exp\{ \varepsilon * E_i[\Gamma_i(\hat{p}(i) \mid w_i)\hat{p}(i)] \}\right] \bigg/ \int_{p=0}^{p=\upsilon} exp\{ \varepsilon * E_i[\Gamma_i(\hat{p}(i) \mid w_i)\hat{p}(i)] \}\, dp$$

$$= \left[exp\{ \varepsilon * R_i(\hat{p}(i) \mid w_i) \}\right] \bigg/ \int_{p=0}^{p=\upsilon} exp\{ \varepsilon * R_i(\hat{p}(i) \mid w_i) \}\, dp$$

[25] McSherry, Frank and Kunal Talwar, "Mechanism Design via Differential Privacy," Microsoft Research.
[26] Used instead of epsilon to avoid collision with the privacy parameter.

Now, let $\gamma = max_i \, R_i(\hat{p}(i)| \, w_i)$ and $k = argmax_i \, R_i(\hat{p}(i) \mid w_i)$:

$$\leq [exp\{ \varepsilon * \gamma\}] \; / \int_{p=0}^{p=\upsilon} exp\{ \varepsilon * \gamma(p) \} \, dp$$
$$= exp \; \{\varepsilon * \gamma\} * [\, \varepsilon * \gamma'(p) \,] \; / \, exp \; \{\varepsilon * \gamma\}$$
$$= \varepsilon * \gamma'(p)$$
$$= \varepsilon * \Gamma_k * (1 - z)$$

If we then normalize the quantity purchased by each buyer, to say, a percentage of the total demand, we see that the final privacy parameter is at most $\varepsilon(1 - z)$, where z is the price elasticity of demand. Using the proof established earlier on the price optimization properties of unit price elasticity, we see that a model that optimizes prices over time will converge toward a 2ε-level of price elasticity.

Thus, we find ourselves with a tunably private model that depends only on the price elasticity of demand. Intuitively, this makes sense, as the more price-sensitive a market, the less likely it is that an individual will receive a price individualization without context about their specific willingness-to-pay.

The FPM is $O(1/\varepsilon)$-regret

Let us build our asymptotic regret claim by case.

Exploration case

If r < β, all we must do is show that exploration happens asymptotically infrequently, as described earlier. Setting $\beta = 1/n$ satisfies this case, causing the exploration case of the FPM to be greedy in the limit of infinite exploration. Thus, our explore case is zero-regret.

Exploitation case

If r > β, let us recognize that our privacy parameter will lead us to select non-optimal prices at times in order to maintain privacy. Let us construct a bound on the regret. If regret is the size of the loss l multiplied by the probability of the loss, we can write our regret using the density of our price transformation function, assuming we alter ε to be a normalized value:

$$
\begin{aligned}
regret &= \int_{l=R^*}^{l=0} l * P(l)\, dl \\
&= \int_{l=R^*}^{l=0} l * exp\{\, \varepsilon * (E_i[\Gamma_i(\hat{p}(i) \mid w_i)\hat{p}(i) - l)\,]\,\}\ dl \\
&\leq \int_{l=R^*}^{l=0} l * exp\{\, \varepsilon * (R^* - l)\,]\,\}\ dl \\
&= e^{\varepsilon R^*} \int_{l=R^*}^{l=0} l * e^{-\varepsilon l}\, dl \\
&= e^{\varepsilon R^*} * [e^{-\varepsilon R^*} * (\varepsilon R^* + 1) - 1\,] / \varepsilon^2 \\
&\leq e^{\varepsilon R^*} * [e^{-\varepsilon R^*} * (\varepsilon R^* + 1) - e^{-\varepsilon R^*}\,] / \varepsilon^2 \\
&= e^{\varepsilon R^*} * [e^{-\varepsilon R^*} * \varepsilon * R^*\,] / \varepsilon^2 \\
&= R^* / \varepsilon
\end{aligned}
$$

As such, our exploitation branch regret is $O(1/\varepsilon)$. Notice two things about our proof. First, we integrate from R^* to 0, which might seem unintuitive. However, we do so based on our framing of revenue: we integrate from the least revenue (when the loss is maximized) to the most revenue (when the loss is minimized). Second, we rely on the fact that $e^{-\varepsilon R^*} < 1$. To prove this, we rely on the facts that $\varepsilon > 0$ and $R^* > 0$ by definition:

$$e^{-\varepsilon R^*} < 1 \Longrightarrow e^{\varepsilon R^*} > 1 \Longrightarrow ln(e^{\varepsilon R^*}) > ln(1) \Longrightarrow \varepsilon R^* > 0$$

Putting cases together

As the sum of both cases is still $O(1/\varepsilon)$, we have produced a price model with both bounded regret and a guarantee of differential privacy. Intuitively, the relationship between these two bounds makes sense. Imagine a very large ε such that differential privacy is not maintained in any meaningful sense. In this scenario, we would have very low regret, as we are able to individualize prices to our heart's content. Now, imagine a very small ε such that differential privacy is strictly maintained. In this scenario, we would miss a lot of revenue, as we must meaningfully pick non-optimal (or randomly non-optimal) prices in order to maintain privacy.

OUTLOOK

The general principles of pricing

In both theory and practice, ideal pricing models share a few key principles. From practice, it is important that models require shallow data and are robust to incomplete or inexact data. For most businesses, recording and maintaining the data necessary to use any number of pricing algorithms is a time- and capital-intensive endeavor. There exist entire technology businesses dedicated to this proposition. Thus, from the real-world perspective, a model that cannot serve the average business, with its average-business data and average-business problems, is rather useless. Theoretically, the points above are largely served by bandits, as we have seen. Intuitively, the "error bars" and treatment of every input as a probability distribution make bandits extraordinarily tolerant of incomplete and inexact data. In addition, bandits always "do their best" with the data one has; if you give a contextual bandit whatever you got, something always comes back to you.

In addition to the principles above, theory offers one more: profit guarantees. We have discussed the meaning of a zero-regret strategy, and as a seller, it does not make a lot of sense to choose a regretful strategy, unless there is some meaningful external payoff to doing so. It remains with the seller whether or not fairness is a payoff worthy of employing a regretful strategy.

Reflecting on fairness

Let us reflect on the pricing principle of fairness. No matter what the alpha promise of a bandit may be, the fact remains that the lifetime value of a customer is typically much higher than any single transaction's marginal revenue. This "reputation premium" is the central issue for most businesses exploring dynamic pricing. As in cost-plus, sellers with strong preferences for fairness will opt for a poorly-optimized or single-price strategy, often at great financial cost. It seems, then, that there is a fundamental tradeoff between alpha and fairness: how do we extract marginal value from a customer without them feeling like they have been treated unfairly and churning?

As the name implies, the ultimate promise of our Fair Pricing Model is that sellers do not need to navigate such a tradeoff: they can gain alpha without appearing unfair to any individual buyer through the beauty of differential privacy. A final question remains, though: is differential privacy really a proxy for fairness? While tightly defining "fair pricing" is difficult, the following definition approximates fairness well: when the price quoted to you through a randomized individualization mechanism is nearly indistinguishable from the one that would have been quoted to you had your data been excluded from the model. In practice and in theory, this definition of fairness directly relates to differential privacy. The other easy-to-recite definition of fairness in pricing makes no room for any of this: just put a price tag on it. Indeed, the ultimate barrier to variable pricing in the real world is not data nor economic understanding nor computational theory. It is executives afraid of upsetting a fickle public, not knowing which definition of fairness to appeal to.

Making pricing accessible

Today, only the largest, wealthiest, and most powerful corporations in the world use dynamic pricing to its fullest potential. As discussed, it is a Herculean task to convert historical and current sales data into an optimal pricing strategy. Because of the disproportionate value a company can gain through better pricing, I believe there is a meaningful business opportunity to create a series of dynamic pricing APIs that allow users to submit data and retrieve prices programmatically. This API could be sold to e-commerce sellers who control their own source code: those who can use this pricing API to directly track users and lost sales, as well as train optimization models. The playbook for bringing such an API to market is well-worn by unicorn startups such as Twilio and Stripe. By offering fast-deployment services, our API company could market itself to startups and early-stage marketplaces, who could then embed our pricing API in their service. Our revenue model, of course, should leverage price discrimination of the second degree and charge a variable fee based on the gross merchandise value of the marketplace. Thus, as our customers grow, our revenue grows as well. If we grow our customer base only linearly and our customers grow exponentially, we grow exponentially by extension. Thus, by marrying innovations in economic and statistical theory with the execution of a startup, we can make dynamic pricing scalable and accessible.

Future research

The central point of this thesis was to offer a price selection heuristic with certain privacy and profit guarantees, namely regret $O(1/\varepsilon)$. In this upper bound lies quite a complicated constant factor, and this thesis has not examined what the actual regret value is: we just gave a bound. Given the setup of this thesis' objective, we might close by proposing a "holy grail": what selection heuristic gives the minimum regret for any ε-level of differential privacy?