The Promise and Perils of Myopia in Dynamic Pricing With Censored Information

Meenal Chhabra[1], **Sanmay Das**[2], **and Ilya Ryzhov**[3]
[1] Virginia Tech (now at Square Inc.)
[2] Washington University in St. Louis
[3] Robert H. Smith School of Business & Institute for Systems Research, University of Maryland
meenal@vt.edu, sanmay@wustl.edu, iryzhov@rhsmith.umd.edu

Abstract

A seller with unlimited inventory of a digital good interacts with potential buyers with i.i.d. valuations. The seller can adaptively quote prices to each buyer to maximize long-term profits, but does not know the valuation distribution exactly. Under a linear demand model, we consider two information settings: partially censored, where agents who buy reveal their true valuations after the purchase is completed, and completely censored, where agents never reveal their valuations. In the partially censored case, we prove that myopic pricing with a Pareto prior is Bayes optimal and has finite regret. In both settings, we evaluate the myopic strategy against more sophisticated look-aheads using three valuation distributions generated from real data on auctions of physical goods, keyword auctions, and user ratings, where the linear demand assumption is clearly violated. For some datasets, complete censoring actually helps, because the restricted data acts as a "regularizer" on the posterior, preventing it from being affected too much by outliers.

1 Introduction

Consider a seller with an unlimited supply of identical goods (e.g., digital goods, which have zero or very low marginal costs of production/storage) who interacts with a stream of buyers with i.i.d. valuations. The seller does not know the precise distribution from which valuations are drawn, but can learn it indirectly through observing the outcomes of interactions with buyers. Lack of information about the valuation introduces an exploration vs. exploitation trade off: the seller wants to maximize long-term revenue, but it may be necessary to compromise on some profit in order to learn the demand curve.

Dynamic pricing is a fundamental problem class within revenue management [Talluri and Van Ryzin, 2005], and the notion of *learning* an unknown demand curve has attracted considerable attention within this community recently. There are numerous variations of this problem: Harrison *et al.* (2012) consider a simple hypothesis testing problem with only two possible demand curves; Farias and Van Roy (2010) allow the demand to be a function of an unknown positive scalar, which is learned using Bayesian updating; Cope (2006) assumes a fixed number of allowed prices with a Dirichlet prior on the response probabilities. Most variations assume sequentially arriving, independent buyers [Blum *et al.*, 2004; Kleinberg and Leighton, 2003]. Extensions include non-stationary demand [Besbes and Zeevi, 2011] and finite inventory [den Boer and Zwart, 2015].

The optimal policy in all of these problems can be characterized using Bellman's equation, but exact solutions are typically intractable. In many cases (see, e.g., [Blum *et al.*, 2004; Broder and Rusmevichientong, 2012; Keskin and Zeevi, 2014]), a simple myopic (greedy) pricing policy is sufficient to achieve sublinear regret over time, possibly requiring minor modifications to avoid "confounding prices" that provide no new information [Harrison *et al.*, 2012]. In some cases where special structure is available, such a policy may even achieve finite (bounded) regret [Mersereau *et al.*, 2009]. Other proposed approaches reduce the pricing problem to an instance of the well-known multi-armed bandit problem [Rothschild, 1974; Kleinberg and Leighton, 2003; Leloup and Deveaux, 2001]; the set of allowable prices is discretized, with each price representing an arm of the bandit, allowing the decision-maker to draw on the rich literature on bandit algorithms with desirable regret bounds [Kleinberg and Leighton, 2003; Kleinberg, 2005; Auer *et al.*, 2002]. Because these techniques may not always work well in practice [Chhabra and Das, 2011; Conitzer and Garera, 2006; Vermorel and Mohri, 2005], one may also consider alternate approaches such as knowledge gradient policies [Ryzhov *et al.*, 2010; Powell and Ryzhov, 2012], which are simple to implement and often produce competitive empirical performance. However, both theoretical and empirical results in this area are highly dependent on the modeling assumptions made in defining the problem; for instance, Carvalho and Puterman (2005) find that a one-step look-ahead strategy achieves significantly better finite-time performance than a myopic strategy in an empirical study.

We focus on a different issue in dynamic pricing, namely the problem of *censored information*. The decision-maker typically does not observe valuations directly, and only receives partial information. Censored information presents challenges for both statistical inference (it is difficult to perform Bayesian updating under censoring) and optimization (traditional algorithms and results may no longer apply). Harrison *et al.* (2012) circumvent this issue by using a binary prior; however, if the set of possible valuation distributions is continuous (e.g., normal or uniform with unknown parameters), a censored observation will only lead to a conjugate belief model in certain special cases. Das and Magdon-Ismail (2008) and Chhabra and Das (2011) use moment-matching approximations to collapse the posterior into the same form as the prior, whereas Qu *et al.* (2012) use a density projection technique (minimizing KL divergence between the true posterior and the approximate distribution), but such methods are necessarily approximate.

Contributions First, we consider partially censored information, where customers reveal their true valuations only after buying the product; they do not do so if they do not buy. Intuitively, once the customer has purchased a digital good, she will presumably never need to purchase it again, and thus would be willing to reveal her true valuation to the seller. The partially censored case admits a conjugate Bayesian belief; with a Pareto prior, the posterior remains Pareto for both buys and no-buys. When the true valuation distribution is uniform, corresponding to linear demand, we show that myopic pricing is optimal and that the cumulative regret incurred by this strategy is finite.

In the case of fully censored information, we derive two statistical approximations (moment-matching and Kullback-Leibler divergence minimization) to collapse the posterior to a Pareto distribution. We experimentally compare the myopic policy against a knowledge gradient (one-step lookahead) policy with respect to several baselines on real datasets (consisting of valuations from three different real-world data sources) where the valuation distribution differs significantly from the assumed uniform distribution. We find that the myopic policy now underperforms under fully censored information; when the valuation distributions are heavy-tailed, full censoring can actually help by acting as a regularizer that prevents overreaction of the (misspecified) prior to outliers. These results provide insight into both when and why myopic algorithms can be successful in dynamic pricing problems, as well as when it is better to turn to algorithms with more lookahead.

2 The Model

We assume that the seller interacts with one buyer at a time, and wishes to set prices sequentially so as to maximize infinite horizon discounted revenue $\pi = \sum_{t=0}^{\infty} \delta^t \pi_t$, where δ is a discount factor and π_t is the profit from the transaction that occurs at time t. The buyers' valuations are i.i.d draws from a known distribution function $f_v(x)$ with unknown parameters. At each time t, the seller quotes a price q_t. An arriving buyer with valuation v_t sees the quoted price, and chooses not to buy if $v_t < q_t$ and chooses to buy otherwise. If she chooses to buy, she reveals her valuation v_t to the seller in the partially censored setting, but not in the fully censored setting.

We assume that the underlying valuation distribution is uniform on $[0, Z]$, but Z is unknown. The seller faces an exploration-exploitation dilemma due to the uncertainty in the value of Z. This is equivalent to a linear demand assumption in the single-unit posted price setting, since the probability of a purchase is linear in the price; linear demand models are standard in the literature [McLennan, 1984; Harrison *et al.*, 2012]. The Pareto distribution is a natural choice for representing our prior beliefs about Z [DeGroot, 1969], since it is conjugate in the case where the actual observations from the distribution are revealed (although not necessarily in the censored cases we examine).

There exists a single optimal price if the valuation distribution $f_v(x)$ is known to the seller. In our model, this is equivalent to knowing the value of Z *a priori*. First, note that $\Pr(\text{Buy}|q, Z) = 1 - \frac{q}{Z}$ if $q < Z$, and 0 otherwise. Then the single optimal price can be calculated as $\pi_{\text{opt}} = \max_q (q \Pr(\text{Buy}|q, Z)) \Rightarrow q_{\text{opt}} = \frac{Z}{2}$.

Since the underlying valuation distribution is uniform, we model the seller as maintaining a Pareto prior on Z, the unknown parameter of the uniform distribution. The Pareto is the conjugate prior for the uniform in this case. The seller's beliefs are then fully represented by the two parameters (a and b) of the Pareto distribution $f_Z(x; a, b) = \frac{ab^a}{x^{a+1}} \quad x > b$. Note, that the expected value of Z is not finite if $a < 1$. In order for the seller to always have finite mean belief distribution we assume that $a > 1$ at all times.

3 Pricing with partially censored information

Here we deal with a setting in which the buyer discloses her true valuation if and only if she decides to buy.

3.1 The Seller's Bayesian Updates

The seller quotes a price q to an arriving buyer, and the buyer either decides to buy and disclose her valuation to the seller, or leave. Depending on the information, the seller gets from the buyer's decision, she updates her belief on Z as follows for the two cases:

The buyer buys: When the buyer chooses to buy, she also discloses her valuation, which comes from a uniform distribution. With some algebra, it is straightforward to show that the posterior remains Pareto:

$$f_Z(x|\text{Buy}) = \begin{cases} f_Z(x; a+1, \max(v, b)) & x > \max(v, b) \\ 0 \quad \text{otherwise} \end{cases}$$

The buyer does not buy: As long as the current price is less than the parameter b of the Pareto distribution, the posterior is also Pareto. We argue later that prices above b should be excluded from consideration; thus, in the partially censored setting, we will always have a Pareto posterior.

$$\Pr(\neg\text{Buy}|q, Z = x) = \begin{cases} \frac{q}{x} & q < x \\ 1 & \text{otherwise} \end{cases}$$

$$\Pr(\neg\text{Buy}|q) = \int_0^{\infty} \Pr(\neg\text{Buy}|q, Z = x) f_Z(x)\, dx$$

$$f_Z(x|\neg\text{Buy}) = \frac{f_Z(x) \Pr(\neg\text{Buy}|x)}{\Pr(\neg\text{Buy})} = \begin{cases} \frac{q f_Z(x)}{x \Pr(\neg\text{Buy}|q)} & x \geq q \\ \frac{f_Z(x)}{\Pr(\neg\text{Buy}|q)} & x < q \end{cases}$$

Therefore, we see that if $b > q$ (the price) then the posterior is $f_Z(x; a+1, b)$ (i.e., Pareto with parameters $a+1$ and b).

3.2 Strategies

Now we turn to the problem of choosing how to set the price q_t at any point in time t. Two strategies are discussed: (1) the Bayesian myopic pricing strategy and; (2) the knowledge-gradient (i.e., one-period Bayesian look-ahead) strategy.

Bayesian myopic pricing strategy

The simplest strategy is to price the item greedily, in order to maximize the expected profit from the next interaction with a buyer. At time t: $\mathbb{E}(\pi_{t_{\text{myopic}}}) = \max_{q_t}(\mathbb{E}(\pi_t)) = \max_{q_t}(q_t \Pr(\text{Buy}|q_t))$. This myopic profit is maximized when $\max(b, q_t) = b$ and $q_t = \frac{b(a+1)}{2a}$ and it is:

$$\mathbb{E}(\pi_{t_{\text{myopic}}}) = \frac{b(a+1)}{4a} \tag{1}$$

For details, see the supplementary information.

Knowledge-Gradient

The knowledge-gradient (KG) strategy is a one-period Bayesian look-ahead strategy. The seller assumes that, starting from the next round, her beliefs will be fixed forever, and she will choose the myopic price every time after that. Mathematically: $\pi_{KG} = \max_{q_t}\left(\pi_t + \frac{\delta}{1-\delta}\left(\pi_{t+1_{\text{Myopic}}}\right)\right)$. Then:

$$\begin{aligned}\mathbb{E}(\pi_{KG}) = \max_{q_t} \Big(\mathbb{E}(\pi_t) + \frac{\delta}{1-\delta}\big(&\Pr(\neg\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1|\neg\text{Buy}})) \\ &+ \Pr(\text{Buy}|q_t)\max(\mathbb{E}(\pi_{t+1|\text{Buy}}))\big)\Big)\end{aligned} \tag{2}$$

Surprisingly, the KG optimal price (from Equation 2) is equal to the myopic price.

Theorem 1. *The KG optimal price for the uniform $[0, Z]$ valuation distribution with left-censored observations is equal to the myopic price when the seller maintains a Pareto prior on Z.*

The proof can be found in the supplementary information; however, we give a brief sketch. To calculate the price set by the KG policy, we first calculate the expected myopic profit for the $(t+1)^{\text{st}}$ round. As seen in Section 3.1, the posterior distribution when the buyer does not buy depends on the value of $\max(b, q_t)$, so it is necessary to evaluate two cases, depending on which of these two quantities is larger.

3.3 Optimality of the myopic strategy

To argue the optimality of the Bayesian myopic strategy, we first restrict the set of allowable strategies in the following manner. We consider only those strategies that, for any t, choose prices q_t below the current value of the Pareto parameter b. From Section 3.2, we know that the myopic strategy satisfies this condition.

$$q_{t_{\text{Myopic}}} = \frac{b(a+1)}{2a} < b \quad \text{for } a > 1$$

Now consider a model where the seller gets perfect information after each interaction with a buyer. That is, the seller observes the exact valuation of the buyer regardless of whether the buyer purchases the product. The updates in this complete-information setting for both buying and not buying are Pareto:

$$f_Z(x; a, b|v) = f_Z(x; a+1, \max(b, v))$$

After k buyers have visited, let $m(k)$ represent the maximum of the valuations of these k buyers. If the seller starts with initial parameters $(a = a_0, b = b_0)$ for a Pareto prior on the value of Z ($v \sim U[0, Z]$) at time $t = 0$, the posterior distribution after k buyers have visited is given by:

$$f_Z(x; a_k, b_k|k) = f_Z(x; a_0 + k, \max(b_0, m(k)))$$

Note that the parameter b_0 also provides the information that the seller initially assumes $Z > b_0$ because $f_Z(x; a_0, b_0) = 0$ for $x < b_0$ for the Pareto distribution, so it is advisable to choose a small value for b_0.

Now, return to our setting, where the buyer discloses her valuation if she chooses to buy, but not otherwise. Since we are only considering strategies that price below the current value of the Pareto parameter b, the conjugacy of our belief model is preserved even with censored information. Let $m_b(k)$ represent the maximum of all the valuations observed among those who chose to buy after a sequence of buyers has arrived. After k buyers have visited the seller, the posterior distribution is given by:

$$f_Z(x|k) = f_Z(x; a_0 + k, \max(b_0, m_b(k)))$$

Observe that the a parameter of the Pareto distribution is always equal (incremented by 1) in both the complete-information and partial-information models. Let $b_c(k)$ and $b_p(k)$ represent the b parameter for the Pareto distribution after k buyers have arrived for complete and partial information respectively. We now show that $b_p(k)$ is actually equal to $b_c(k)$ for any allowable strategy, as long as the two sellers start with the same prior.

Theorem 2. *At any time $k > 0$, the parameters of the posterior distribution in our settings (partially-censored information about buyers' valuation) for a Pareto prior are equal to the parameters of the posterior distribution with complete information using any pricing strategy that prices the item less than the b parameter of the Pareto distribution about the buyers' valuation if the initial value of the parameters is chosen to be equal (i.e., $b_p(k) = b_c(k)$ if the same b_0 is chosen for both complete information and partial information model).*

The proof proceeds by induction on the time index k; in each such time index, it is again necessary to separately consider two cases, depending on whether or not the next buyer buys. The details can be found in the supplementary information.

Corollary 3. *Among all strategies in the allowable set, the myopic pricing strategy is optimal.*

Since every strategy in the allowable set produces the same sequence of posterior distributions under either complete information or partial censoring, it follows that the same expected revenue is achieved in both cases. However, the myopic strategy is optimal in the case of complete information, since there is no benefit from exploration. It follows that it continues to be optimal under partial censoring.

It remains to argue that the myopic strategy continues to be optimal when we expand the set of allowable strategies to allow arbitrary pricing decisions (include those above the Pareto parameter b). We offer the following intuition in support of this idea:

1. Non-myopic decisions can be optimal if they allow us to collect new information that compensates for lost revenue over time. However, in our setting, higher prices increase the likelihood of a lost sale, and thus the risk that we will receive less information rather than more.
2. Reducing our ability to collect information should also reduce the revenue we can generate. Thus, the maximum achievable revenue under partial information should not be greater than the maximum revenue under complete information. However, the revenues generated by the myopic strategy are the same in both settings.

We conclude that the myopic strategy is optimal for the uniform-Pareto pricing problem with censored information. This result relates to earlier work by Harrison *et al.* (2012) and Mersereau *et al.* (2009) on greedy and semi-greedy policies in Bayesian dynamic pricing.

3.4 Regret Bounds

For a multi-armed bandit with independent arms the regret grows as $\Omega(\log T)$ [Lai and Robbins, 1985]. Mersereau *et al.* (2009) demonstrate the existence of finite regret in a special case with correlated arms where the expected reward of each arm is a linear function of an unknown scalar with known prior. The existence of finite regret in general means that the algorithm learns very fast. We show that the myopic Bayesian algorithm in our model also has finite regret, implying that it learns Z quickly.

Theorem 4. *The regret for the myopic Bayesian policy for the partially-censored information setting (i.e, the buyers disclose their valuation in case of a buy) when the true valuations are drawn from the uniform distribution in* $[0, Z]$*, where Z is unknown and the seller maintains a Pareto prior, is finite.*

Essentially, the proof first derives the bound for the complete-information case, and then observes that the expected revenue of the myopic policy is the same under both complete information and partial censoring, meaning that the same bound holds in both cases. Details can be found in the supplementary information.

4 Pricing with fully censored information

Now, we turn to the case of completely censored information. The additional complicating factor for the Pareto prior arises when an arriving agent chooses to buy, in which case the posterior

$$f_Z(x|\text{Buy}) = \begin{cases} \frac{(1-\frac{q}{x})f_Z(x)}{\Pr(\text{Buy}|q)} & x > \max(q,b) \\ 0 \quad \text{otherwise} \end{cases} \tag{3}$$

is no longer Pareto. If the arriving agent does not buy, the updates are the same as in the partially censored case. The posterior is Pareto $(f_Z(x; a+1, b))$ if $b > q$, otherwise it is not.

$$f_Z(x|\neg\text{Buy}) = \frac{f_Z(x)\Pr(\neg\text{Buy}|x)}{\Pr(\neg\text{Buy})} = \begin{cases} \frac{q f_Z(x)}{x \Pr(\neg\text{Buy}|q)} & x \geq q \\ \frac{f_Z(x)}{\Pr(\neg\text{Buy}|q)} & x < q \end{cases} \tag{4}$$

One way to tackle non-conjugacy is to constrain the posterior to have the same form as the prior. We consider two possibilities, moment matching on the first two moments (following [Chhabra and Das, 2011; Das and Magdon-Ismail, 2008]) which can in many cases be proved to be consistent [Chen and Ryzhov, 2016], and density projection, where the KL divergence of the approximate distribution from the true distribution is minimized [Qu *et al.*, 2012], an approach also often used for model evaluation [Burnham and Anderson, 2002].

4.1 Moment matching approximation

Let $\mathbb{E}_{\text{tp}}(Z)$ and $\mathbb{E}_{\text{tp}}(Z^2)$ represent the first two moments about the mean of the true posterior distribution. We find (a', b') of the Pareto distribution $f_p(x; a', b')$ such that its first two moments are the same as those of the posterior distribution. The derivation is algebraic, and can be found in the supplementary information.

4.2 Approximation using density projection

Another possibility is to choose the Pareto distribution that minimizes the KL divergence (or relative entropy) from the true distribution. Formally, given two densities p and q, $D(p||q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx = -H(p) + H(p,q)$ where $H(p)$ is the entropy of p and $H(p,q)$ is the cross entropy of p and q. In our case, p, the true distribution, is the exact posterior distribution calculated using Bayes rule in Equations 3 and 4 above, and q is the Pareto distribution which we want to use as a proxy to the true distribution in order to characterize the belief state of the seller. In this technique, we want to find parameters (a', b') to the Pareto distribution such that $D(p||q)$ is minimized. Here, minimizing KL divergence is equivalent to minimizing cross entropy($H(p,q)$) because the entropy ($H(p)$) of the true posterior is constant. The details of the computation are deferred to the supplementary information.

4.3 Pricing strategies

Myopic The myopic price is only dependent on the current state parameters; hence, it is the same for the completely and the partially censored case. From Equation 1, if the state parameters are a and b then the myopic price is $\frac{b(a+1)}{2a}$.

Knowledge Gradient (KG) optimal In order to compute the KG optimal price at any time t, we first have to compute the myopic profit at time t + 1. For doing so, we first calculate the approximate posterior distribution using either of the two methods discussed above. Then the myopic profit can be calculated using Equation 1. The KG price can then be calculated by maximizing the expected one-step-look-ahead profit using Equation 2.

5 Experimental Analysis

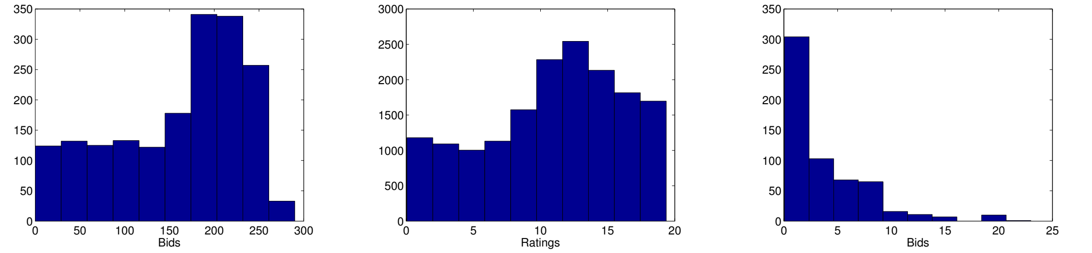
5.1 Data

In order to evaluate the algorithms when the uniformity assumption is violated, we conduct experiments using three different datasets. Space restrictions preclude a full explanation of the datasets and preprocessing here, but are available in the supplementary information. We use three datasets that can be considered to give valuation information. These are (1) **eBay auctions**, which contains bidding information from eBay for several auctions of Palm Pilot M515 PDAs, (2) **Jester**, which contains ratings for a set of 100 jokes by 29,483 users on a scale from -10.00 to 10.00 (continuous rating) from Jester, an online joke recommender system [Goldberg *et al.*, 2001], and (3) **Yahoo! advertiser bids**, which contains data on advertisers' bids on the top 1000 keywords for sponsored search auctions from June 15, 2002 to June 14, 2003. For datasets (1) and (3), we take each user's single highest bid and assume that to be their valuation. As can be seen from the examples in Figure 1, the resulting distributions are significantly different from uniform.

5.2 Baseline Algorithms

Partially censored observations For the partially censored case we consider a modified Gittins index strategy as a baseline. We first discretize the prices — each price is equivalent to an arm of a multi-armed bandit. The seller maintains a Beta prior $B(\alpha, \beta)$ on the probability of success of each arm. Using this, the Gittins indices (or the dynamic allocation indices) are calculated for all the arms and the arm with the highest index is selected [Leloup and Deveaux, 2001; Chhabra and Das, 2011]. On playing the arm, the parameter α is incremented if the buyer decides to buy and β is incremented if the buyer decides not to buy.

This Gittins strategy is optimal for a multi-armed bandit when all the arms are independent. While the arms for the pricing problem are not independent, this approach has shown good results in past work [Leloup and Deveaux, 2001; Chhabra and Das, 2011]. We experiment with some variations of the Gittins strategy to incorporate dependence.



(a) eBay auction dataset

(b) Normalized rating of a joke from the Jester dataset

(c) Sponsored search keyword bids from Yahoo!

Figure 1: Example histograms showing the value distribution for eBay data, one joke from Jester, and one search keyword.

In order to discretize the arms from the dataset of unique bidders, we find the minimum and the maximum bids and discretize the price uniformly in that range. We fix the number of arms at 20. We report on two variants of the Gittins index strategy: (a) *Partially censored Gittins*, where, when the seller observes the true valuation of the good (in case of a buy), we update the beliefs on all the arms based on the true valuation seen. In case of a no buy, the seller increments the β parameter of all the arms with price greater than or equal to the current arm (since higher prices would also have failed). (b) *Complete information Gittins*, which we use as an upper bound of the performance of a Gittins-index style strategy – in this case, we give the seller additional information by also assuming the prospective buyer reports her true valuation in the case of a “no buy.”

Fully censored observations For fully censored observations, we use (a) a similar modification of the Gittins-index based strategy, originally suggested by Chhabra and Das (2011), which uses “smart” priors. (b) UCB1, which is a multi-armed bandit algorithm with logarithmic regret in time; (c) The LLVD algorithm of [Chhabra and Das, 2011], which also assumes linear demand, but uses a Beta prior and is significantly more complex in its pricing strategy. For the bandit-based algorithms we select 20 prices corresponding to 20 arms of the bandit between the minimum and the maximum values of the the specific dataset uniformly. We use similar priors for the LLVD and Pareto-based strategies ($\alpha = \beta = 1.5$ for the Beta; $a = 2.1$ and $b = 2.3$ for the Pareto). We constrain $a \geq 1.01$ for the Pareto distribution, so the seller's belief distribution always has a finite mean.

5.3 Experimental Setup and Results

We show average “learning curves” for three representative datasets and also investigate performance in the large by averaging revenues across all the different Jester and Yahoo! keywords data.

Learning curves We used the data from eBay auctions, one Yahoo! keyword, and one joke from Jester. For eBay and Yahoo!, we average our results over multiple random permutations of the bids. The Jester dataset is larger, so we instead repeatedly sampled 500 unique buyers from the valuation data. We ran 1000 experiments for each dataset, allowing our seller strategies to interact with a stream of potential buyers with valuations in the permuted order. The discount factor, δ, was set to 0.95.

Figures 2 and 3 show the results for the partially censored and fully censored cases respectively. In each figure, the X axis represents time (equivalent to number of buyers who have visited so far) and the Y axis represents the ratio of the profit earned until that time and total profit achievable by a single fixed price strategy with prior knowledge of the true underlying valuation distribution.

With partially censored information, the myopic strategy (which is equivalent to KG), performs better than not only the partial-information Gittins strategy but also sometimes the complete information Gittins strategy for the eBay and Jester data (2 (a) & (b)). However, in Figure 2 (c), we see that the Pareto-prior based strategy does not perform well, converging to a highly suboptimal price; this is because the Yahoo! bids are fat-tailed in a way that interacts badly with the Pareto prior. The true optimal price is $5 but the seller sometimes observes a higher valuation during a buy; the uniformity assumption pushes the seller's prior to a bad belief, and the seller is unable to re-adjust appropriately.

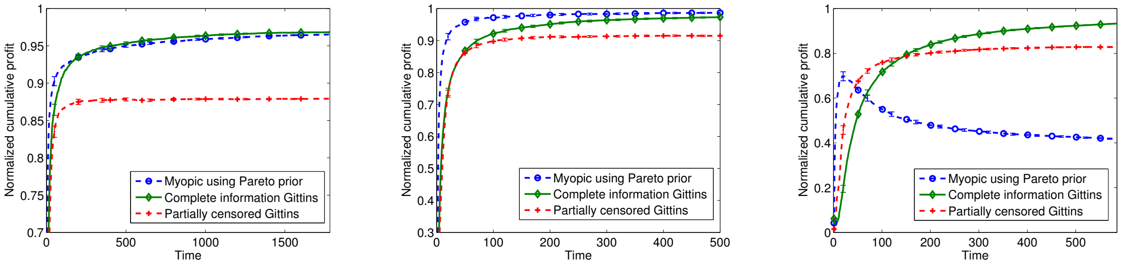
Interestingly, we do not observe this kind of bad behavior for the myopic or KG strategies using the moment-matching approximation in Figure 3. The censoring of data is actually playing a regularization role here, preventing the seller's belief from being pushed too far in the wrong direction from observing outliers.

Large Scale Analysis We want to test the performance of the algorithms “in the large”, so we take all 100 jokes from Jester and the 100 keywords with the highest numbers of unique bidders from Yahoo! (Winsorizing the data at the 1st and 99th percentiles). For each dataset created in this manner, we run multiple iterations, selecting 500 unique users for each iteration for each joke, and randomly permuting the bidders for each Yahoo! keyword. We compute the average discounted and undiscounted profits (normalized by the profit of the clairvoyant fixed-price strategy), as well as counting how many times each algorithm “won” a dataset by being the best performer on it. The discounted profit is the measure that the algorithm is trying to optimize, while the undiscounted profit gives a better idea of learning performance in the long run.

Strategy	Jester dataset		Yahoo! dataset	
	$\pi_{\text{discounted}}$, #wins	$\pi_{\text{undiscounted}}$, #wins	$\pi_{\text{discounted}}$, #wins	$\pi_{\text{undiscounted}}$, #wins
PC Gittins	0.7577±0.0020, 0	0.9162±0.0014, 0	0.5945±0.0053, 9	0.7487±0.0042, 22
CI Gittins	0.7677±0.0024, 0	0.9734±0.0003, 39	0.5437±0.0079, 8	0.7963±0.0064, 59
Myopic	0.9072±0.0024, 100	0.9735±0.0009, 61	0.6500±0.0065, 83	0.5819±0.0128, 19

Table 1: Average discounted and undiscounted profits achieved by different strategies on the ratings of 100 jokes from Jester and the bidding data for 100 keywords from Yahoo! sponsored search auctions in the partially censored setting. PC Gittins is the partially censored strategy and CI Gittins is the Gittins index strategy with access to all valuations. These results are averaged over 1000 iterations.

In the partially censored setting (Table 1), the myopic strategy performed best on both datasets in terms of discounted revenue. However, in terms of undiscounted revenue, the Gittins-index strategies performed better on the Yahoo! datasets, probably because of

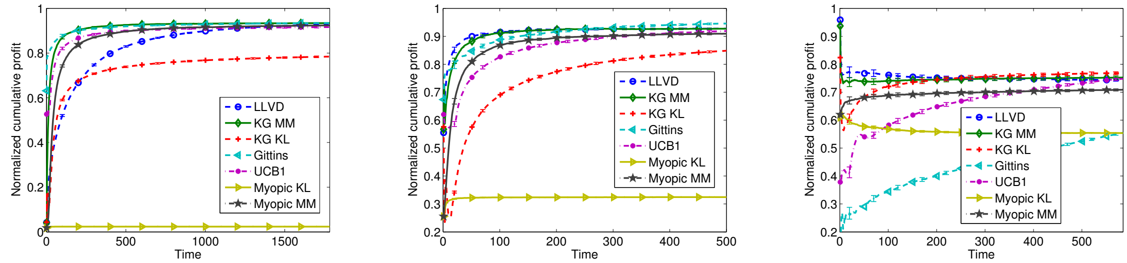


(a) eBay auction data

(b) Joke rating dataset

(c) Yahoo! sponsored search dataset

Figure 2: Learning curves in the partially censored case. Each graph shows the normalized cumulative profit averaged over 1000 iterations and 95% confidence intervals. In (a) and (b) the Pareto prior based myopic strategy for the partially censored case is either very close to or outperforms even the *completely-informed* Gittins-index-based strategy. The distribution for (c) is the one that misleads the myopic strategy the most. At $t = 0$: $a = 3$ and $b = 0.1$.



(a) eBay auction dataset

(b) Jester rating dataset

(c) Yahoo! keyword bidding data

Figure 3: Learning curves in the completely censored case. Each graph shows the normalized cumulative profit averaged over 1000 iterations and 95% confidence intervals. LLVD and a KG-based algorithm using moment-matching approximation perform consistently well in these cases even when the underlying uniform distribution assumption on the valuation data is violated.

the problem with the prior on fat-tailed data mentioned above. Surprisingly, the Gittins-index strategy that only uses partial information has higher average normalized discounted profit compared to the one with access to complete valuation information on the Yahoo! datasets. While the complete information variant learns better in the long-term (see the undiscounted profits), the partial information variant is clearly performing better early on.

Strategy	Jester dataset		Yahoo! dataset	
	$\pi_{\text{discounted}}$, #wins	$\pi_{\text{undiscounted}}$, #wins	$\pi_{\text{discounted}}$, #wins	$\pi_{\text{undiscounted}}$, #wins
LLVD	0.8646±0.0037, 74	0.9254±0.0064, 37	0.6656±0.0245, 19	0.6626±0.0270, 14
KG-MM	0.8322±0.0027, 0	0.9232±0.0051, 0	0.6680±0.0219, 22	0.6830±0.0230, 15
KG-KL	0.4769±0.0113, 0	0.8369±0.0119, 0	0.5914±0.0412, 21	0.6994±0.0356, 31
Myopic-MM	0.7195±0.0034, 0	0.9072±0.0064, 0	0.6094±0.0321, 6	0.6313±0.0301, 9
Myopic-KL	0.3332±0.0101, 0	0.3399±0.0100, 0	0.5443±0.0501, 9	0.5311±0.0409, 6
Gittins	0.8099±0.0149, 26	0.9365±0.0038, 58	0.4778±0.0355, 14	0.5465±0.0298, 4
UCB1	0.6792±0.0038, 0	0.9200±0.0031, 5	0.5560±0.0192, 9	0.6345±0.0227, 21

Table 2: Average discounted and undiscounted profits achieved by different strategies on the ratings of 100 jokes from Jester and the bidding data for 100 keywords from Yahoo! sponsored search auctions in the completely censored setting. These results are averaged over 100 iterations.

The fully censored setting (Table 2) yields several interesting observations: (1) Myopic strategies are outperformed by those that look ahead and try to balance exploration and exploitation. Thus, it is clear that there is a benefit to not being myopic in practice. (2) LLVD and KG-MM perform better in terms of discounted profit on Yahoo! in the completely censored setting than the myopic algorithm (which is also KG) in the partially censored setting. This is because the absence of valuation information prevents the prior from being driven far afield by outliers or observations from the heavy tails of the keyword valuation distribution. The completely censored information updates are more robust to violations of the modeling assumptions. (3) Moment matching outperforms the density projection approach (similar to results reported by Zhang and Song (2017)), which can suffer some spectacular failures. This shows that the form of the Pareto distribution is in itself not a problem, but the interaction of the prior and the approximate update can be.

6 Conclusions

Recent work in dynamic pricing with learning has found that myopic strategies are surprisingly effective, but has focused on regret bounds and asymptotic optimality. We show that, for the important class of uniform valuation distributions (which model linear demand) and the appropriate conjugate prior (Pareto), the myopic strategy is, in fact, Bayes optimal even when the seller receives partially censored information. We extend the model to a practical Knowledge Gradient (KG) algorithm for the fully censored setting, and then evaluate the algorithms on realistic datasets, where the demand model is violated. Our results help in achieving a deeper understanding of when myopic strategies can be optimal or close to it, and when they fail (for example, with heavy-tailed distributions and partially censored information). An interesting insight is that fully censored information can sometimes be beneficial, by acting as a regularizer that helps make the algorithm robust to model misspecification.

7 Acknowledgement

The authors are grateful for support from the US National Science Foundation through awards IIS-1303350, 1414452 and 1527037 (Das), and EAGER award CMMI-1745198 (Ryzhov).

References

[Auer *et al.*, 2002] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.

[Besbes and Zeevi, 2011] O. Besbes and A. Zeevi. On the minimax complexity of pricing in a changing environment. *Operations Research*, 59(1):66–79, 2011.

[Blum *et al.*, 2004] A. Blum, V. Kumar, A. Rudra, and F. Wu. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.

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

[Burnham and Anderson, 2002] Kenneth P Burnham and David R Anderson. *Model selection and multi-model inference: A practical information-theoretic approach*. Springer, 2002.

[Carvalho and Puterman, 2005] A.X. Carvalho and M.L. Puterman. Learning and pricing in an Internet environment with binomial demands. *Journal of Revenue & Pricing Management*, 3(4):320–336, 2005.

[Chen and Ryzhov, 2016] Ye Chen and Ilya O Ryzhov. Approximate Bayesian inference as a form of stochastic approximation: A new consistency theory with applications. In *Winter Simulation Conference (WSC), 2016*, pages 534–544. IEEE, 2016.

[Chhabra and Das, 2011] M. Chhabra and S. Das. Learning the demand curve in posted-price digital goods auctions. In *Proceedings of AAMAS*, pages 63–70, 2011.

[Conitzer and Garera, 2006] V. Conitzer and N. Garera. Learning algorithms for online principal-agent problems (and selling goods online). In *Proceedings of ICML*, pages 209–216, 2006.

[Cope, 2006] E. Cope. Bayesian strategies for dynamic pricing in e-commerce. *Naval Research Logistics (NRL)*, 54(3):265–281, 2006.

[Das and Magdon-Ismail, 2008] Sanmay Das and Malik Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. In *Advances in NIPS*, pages 361–368, 2008.

[DeGroot, 1969] M. H. DeGroot. *Optimal Statistical Decisions*. Wiley, 1969.

[den Boer and Zwart, 2015] Arnoud V den Boer and Bert Zwart. Dynamic pricing and learning with finite inventories. *Operations Research*, 63(4):965–978, 2015.

[Farias and Van Roy, 2010] V.F. Farias and B. Van Roy. Dynamic pricing with a prior on market response. *Operations Research*, 58(1):16–29, 2010.

[Goldberg *et al.*, 2001] Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

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

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

[Kleinberg and Leighton, 2003] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for on-line posted-price auctions. In *Proceedings of FOCS*, pages 594–605, 2003.

[Kleinberg, 2005] R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in NIPS*, volume 18, pages 697–704, Cambridge, MA, 2005. MIT Press.

[Lai and Robbins, 1985] T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[Leloup and Deveaux, 2001] B. Leloup and L. Deveaux. Dynamic pricing on the internet: Theory and simulations. *Electronic Commerce Research*, 1(3):265–276, 2001.

[McLennan, 1984] A. McLennan. Price dispersion and incomplete learning in the long run. *Journal of Economic Dynamics and Control*, 7(3):331–347, 1984.

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

[Powell and Ryzhov, 2012] W.B. Powell and I.O. Ryzhov. *Optimal Learning*. Wiley, 2012.

[Qu *et al.*, 2012] Huashuai Qu, Ilya O Ryzhov, and Michael C Fu. Ranking and selection with unknown correlation structures. In *Winter Simulation Conference*, pages 1–12, 2012.

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

[Ryzhov *et al.*, 2010] I.O. Ryzhov, P.I. Frazier, and W.B. Powell. On the robustness of a one-period look-ahead policy in multi-armed bandit problems. *Procedia Computer Science*, 1(1):1635–1644, 2010.

[Talluri and Van Ryzin, 2005] K. T. Talluri and G. J. Van Ryzin. *The Theory and Practice of Revenue Management*. Springer, 2005.

[Vermorel and Mohri, 2005] J. Vermorel and M. Mohri. Multi-armed bandit algorithms and empirical evaluation. In *Proceedings of ECML*, pages 437–446. Springer, 2005.

[Zhang and Song, 2017] Qiong Zhang and Yongjia Song. Moment-matching-based conjugacy approximation for Bayesian ranking and selection. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 27(4):26, 2017.