Dynamic pricing with Bayesian updates from online reviews *

José Correa[1], Mathieu Mari[2], and Andrew Xia[3]

[1]Universidad de Chile, Chile. correa@uchile.cl.
[2]LIRMM, Université de Montpellier, CNRS, Montpellier, France.
mathieu.mari@lirmm.fr.
[3]Flagship, USA. axia@alum.mit.edu.

April 2024

Abstract

When launching new products, firms face uncertainty about market reception. Online reviews provide valuable information not only to consumers but also to firms, allowing firms to adjust the product characteristics, including its selling price. In this paper, we consider a pricing model with online reviews in which the quality of the product is uncertain, and both the seller and the buyers Bayesianly update their beliefs to make purchasing & pricing decisions. We model the seller's pricing problem as a basic bandits' problem and show a close connection with the celebrated Catalan numbers, allowing us to efficiently compute the overall future discounted reward of the seller. With this tool, we analyze and compare the optimal static and dynamic pricing strategies in terms of the probability of effectively learning the quality of the product.

1 Introduction

As a key part of modern online platforms, online decision-making plays a crucial role in a variety of settings, particularly related to the Internet. Two landmark examples that have been widely studied are *dynamic pricing* and *online reviews*. Online review systems constitute powerful platforms for users to get informed about the product and for the firm to understand how a given market is receiving the product. The study of these systems has been vast for the last two decades [6, 10], and more recently, modeling simple like/dislike reviews as bandits problems have become standard [1, 2, 3, 13, 16, 18]. Dynamic pricing, on the other hand, is an active area of research in economics, computer science, and operations research [12, 14], and has become a common practice in several industries such as transportation and retail.

There has been a growing interest in combining the two areas as a way to design more effective pricing mechanisms that gather information from current reviews to update prices and make the product more attractive [5, 11, 17]. In particular, [5] considers social learning with non-Bayesian agents in a market with like & dislike reviews, and the resulting pricing decision of a monopolist. [17] considers a setting when the volume of sales is large and

*An extended abstract of a preliminary version of this paper has been presented in the NeurIPS Workshop on ML for Economic Policy in 2020.

optimizes revenue via fluid dynamics and ODEs. The general idea here is to use online reviews to update the market's belief about the quality of the product, thus influencing its pricing. However, the complexity of many models limits their practicality as the Bayesian updating of beliefs becomes intractable.

In this paper, we continue on this line of research. We consider a straightforward model that precisely determines the optimal pricing strategies from information elicited by online reviews. The main message of this paper is to show that online reviews not only influence how successful a product is but also help find more effective dynamic pricing strategies. These dynamic pricing strategies lead to more efficient allocations. Dynamic pricing with online reviews gives a foundation for the common practice of temporarily pricing a product below its production cost, leading to short-term revenue losses. These losses come with a potential boost in future purchases, even at a higher price, and ultimately may lead to more revenue.

2 Preliminaries

2.1 Our model.

We first give a general description of our model and then define its features precisely. We defer a discussion of the most closely related models [4, 5, 11, 17] to Section 3.

Consider a seller marketing a new product. Neither the seller nor the buyers are informed about the quality of the product but only receive a public signal, representing the prior probability of the quality of the product. Based on this estimation and the product price, buyers, that arrive in an online fashion, estimate their expected utility and decide to buy the product or not. If no purchase occurs, the buying process terminates. Otherwise, after buying, the buyer experiences the product and may like it or not; in both cases, he submits an online review with either a like or dislike. These reviews allow following users to update their priors on the quality of the product.

Product. As it is common in the literature [1, 11] we assume that the product may be either *good* or *bad*. A *good* product will be liked by a user with probability p (so that roughly a fraction p of the market is satisfied with the product), while a *bad* product will be liked by a user with probability $q < p$. We also assume that the product has a fixed production cost c per unit. To avoid trivial cases, we assume in the paper that $q < c < p$. Finally, the product has a price π that may evolve along the process. The *prior* probability on the quality of the product, denoted by x, is the probability that the product is good. All p, q, and x are common knowledge.

Buyers. The market is composed of an infinite stream of users arriving at times $t = 0, 1, 2, \ldots$, which are offered the product at a certain, possibly time-dependent, price π. Upon receiving the offer, a user (which we assume risk-neutral) evaluates his expected utility for buying the product. Initially, since the first buyer's prior is x, his expected utility can be evaluated as $xp + (1 - x)q$. If this quantity exceeds the current price π, then he decides to buy the product.

Priors. After experiencing the product, the buyer submits an online review in the like/dislike format. If the buyer likes (👍) the product then, given a current prior x,

we update the prior to $L(x)$ as follows:

$$L(x) := \mathbb{P}_x(\text{good} \mid \text{👍}) = \frac{\mathbb{P}_x(\text{👍} \mid \text{good})\mathbb{P}_x(\text{good})}{\mathbb{P}_x(\text{👍})} = \frac{x \cdot p}{x \cdot p + (1-x) \cdot q}.$$

Similarly, given a dislike (👎), we update the prior as follows:

$$D(x) := \mathbb{P}_x(\text{good} \mid \text{👎}) = \frac{\mathbb{P}_x(\text{👎} \mid \text{good})\mathbb{P}_x(\text{good})}{\mathbb{P}_x(\text{👎})} = \frac{x \cdot (1-p)}{x \cdot (1-p) + (1-x) \cdot (1-q)}.$$

In particular, the prior increases after a like, and decreases after a dislike: $D(x) < x < L(x)$. An interesting feature of our model is that the updated prior after a sequence of likes and dislikes only depends on the number of reviews and not the sequence of reviews. This is in contrast with most models of online reviews, and it allows both the seller and the users to update their beliefs based solely on these figures.

Lemma 1. *Given a prior x, the updated prior after a sequence of ℓ likes and d dislikes does not depend on the order and this value is $x_{\ell,d} = \frac{xp^\ell(1-p)^d}{xp^\ell(1-p)^d + (1-x)q^\ell(1-q)^d}$. Furthermore, the probability of each such sequence only depends on ℓ and d.*

Proof. Let us start by calculating $D(L(x))$, the value of the updated prior after seeing a like and then a dislike:

$$\begin{aligned} D(L(x)) &= \frac{(1-p)L(x)}{(1-p)L(x) + (1-q)(1-L(x))} = \frac{(1-p)\frac{p \cdot x}{p \cdot x + q \cdot (1-x)}}{(1-p)\frac{p \cdot x}{p \cdot x + q \cdot (1-x)} + (1-q)(1 - \frac{p \cdot x}{p \cdot x + q \cdot (1-x)})} \\ &= \frac{p(1-p) \cdot x}{p(1-p) \cdot x + q(1-q) \cdot (1-x)} \end{aligned}$$

Clearly, this expression when swapping p (resp. q) with $1-p$ (resp. $1-q$) is unchanged, thus $L(D(x)) = D(L(x))$.

The proof of the formula for $x_{\ell,d}$, with ℓ likes and d dislikes, uses induction and works similarly.

We now show that the probability of observing a given sequence of likes and dislikes is independent of the order. To see this, we calculate

$$\begin{aligned} \mathbb{P}_{L(x)}(\text{👎}) \cdot \mathbb{P}_x(\text{👍}) &= (L(x)(1-p) + (1-L(x))(1-q)) \cdot (xp + (1-x)q) \\ &= \left(\frac{xp}{xp + (1-x)q}(1-p) + (1 - \frac{xp}{xp + (1-x)q})(1-q)\right) \cdot (xp + (1-x)q) \\ &= xp(1-p) + (1-x)q(1-q). \end{aligned}$$

And we easily get the same expression for $\mathbb{P}_{D(x)}(\text{👍}) \cdot \mathbb{P}_x(\text{👎})$. The result follows by induction. □

Seller's problem. A basic problem faced by the seller is to find an optimal pricing strategy. One alternative for the seller is to adopt a *static* price: the price π offered to each consumer is fixed at the beginning and can not be changed. In this situation, users will buy the product as long as the current prior x satisfies $xp + (1-x)q \geq \pi$ and yields positive expected value. Whenever $xp + (1-x)q < \pi$, the process will stop forever. In other words, the sales process will continue until the prior reaches $x_{\min} = \frac{\pi - q}{p - q}$. The *local reward* for the seller for each sale is $R(x) = \pi - c$. Given the prior x, the seller, who discounts the future

at rate δ, can express her *expected revenue* or *expected global reward* recursively as $V(x) = 0$ if $x < x_{\min}$, and otherwise

$$V(x) = R(x) + \delta \cdot \mathbb{P}_x(\text{👍}) \cdot V(L(x)) + \delta \cdot \mathbb{P}_x(\text{👎}) \cdot V(D(x)). \tag{1}$$

Finally, the seller will optimize this function over the values of $\pi \in (c, xp + (1-x)q]$.[1]

On the other hand, the seller may opt for a *dynamic* pricing approach. In this setting the price may be adjusted according to the current reviews the product has received. To maximize revenue, the seller will just make the user indifferent to purchase whenever she decides to continue selling the product, offering the product at time t with prior x at exactly $\pi_t = xp + (1-x)q$. Thus, with dynamic prices, the local reward for the seller after each sale depends on the prior with value $R(x) = xp + (1-x)q - c$.[2] The decision of when to stop, however, becomes more involved. Although at times $xp + (1-x)q < c$ may hold, and therefore in the next sale the seller will incur in a loss, it may still be worth to continue selling. The reason for this relies on the information gain provided by one more sale and the impact this information has on future purchases. In this scenario, given a prior x, to express her expected revenue, the seller can either decide to stop selling, and her reward is thus $V(x) = 0$; or can decide to continue selling, and then her revenue can be expressed as in equation (1). To maximize her profit, she will pick the best of these two options, and thus her *expected global reward* satisfies the following recursive equation:

$$V(x) = \max\Big(0, R(x) + \delta \cdot \mathbb{P}_x(\text{👍}) \cdot V(L(x)) + \delta \cdot \mathbb{P}_x(\text{👎}) \cdot V(D(x))\Big), \tag{2}$$

where $R(x) = xp + (1-x)q - c$.

Therefore, with dynamic prices, the seller stops selling when the current prior x satisfies $x < x^*$, where x^* is defined as the largest that $V(x^*) = 0$, (where V is the solution of (2)). This value determines the stopping time of the seller under dynamic pricing.

Thus, the expected global reward in both pricing scenarios can be expressed with the same recursion:

$$V(x) = \begin{cases} 0, & \text{if } x < x_{\text{stop}} \\ R(x) + \delta \cdot \mathbb{P}_x(\text{👍}) \cdot V(L(x)) + \delta \cdot \mathbb{P}_x(\text{👎}) \cdot V(D(x)), & \text{otherwise.} \end{cases} \tag{3}$$

Here, for static pricing we have $x_{\text{stop}} = x_{\min} = \frac{\pi - q}{p - q}$ and $R(x) = \pi - c$, while for dynamic pricing we have $x_{\text{stop}} = x^*$ and $R(x) = xp + (1-x)q - c$.

2.2 Our Results

In this paper we study dynamic and static pricing with online reviews in detail. We first show that we can formulate the underlying dynamic process as a simple multi-armed bandit problem. Perhaps surprisingly, this problem appears to be unexplored in the literature. The core of the paper is devoted to the computation of the stopping prior x^* in the dynamic price setting and the computation of the global expected reward in both settings. For this, we propose two approaches.

[1] Clearly π has to be at least c for the revenue to be positive. Also, if $\pi > xp + (1-x)q$, then no user will ever buy, and then the revenue is zero.

[2] Note that, slightly abusing notation, we always use $R(x)$ for the local reward, although its value depends on whether we are considering static or dynamic pricing.

We first present a fast dynamic programming approach in Section 4 that computes an approximation of x^* and an approximation of the value of the global expected reward in both pricing scenarios. The backside of this fast heuristic is that we are unable to provide a good guarantee on the quality of the solutions produced.

To resolve this issue, in Section 5, we take a combinatorial approach using the classic Gittins index [8], and we explicitly determine the optimal underlying stopping time and the global excepted rewards. Specifically, we obtain a closed-form formula for these quantities. To this end, we uncover a connection with the classic Catalan numbers. With this, we are able to compute an arbitrarily good approximation for x^* and the global expected reward in both pricing scenarios.

In Section 6, we study, for each pricing strategy, the probability of achieving full efficiency. Note that this fully efficient situation occurs when the product is good and is sold forever[3], i.e., the product is good, and the market learns this fact. To this end, we exploit that the stochastic process governing the prior updates is a martingale, and we can thus use the optional stopping theorem.

We finally discuss in Section 7 an extension of the model in which the quality of the product is not restricted to be *good* or *bad*, but it can take arbitrary values in a set $Q \subseteq [0, 1]$. For this more general model we note that the history independence property still holds and that, in essence, our dynamic programming approach can still be used to compute the seller's revenue. We additionally observe that when the set of possible product qualities Q is a continuous set and the initial prior quality distribution is sufficiently smooth, the dynamic program is very effective since already a short sequence of reviews gives a very good estimate about the true quality of the product.

2.3 The Bandits Connection

We can model our dynamic pricing problem using a bandit framework. To the best of our knowledge, an optimal strategy for that bandit problem was previously unknown. Imagine there is a single slot machine (bandit), which could either be a *good* or *bad* machine. This machine costs c to play and will yield a return of 1 with some probability and 0 with some other probability.[4] If the machine is good, it has a fixed, known probability of p of returning 1, while if it is bad it has known probability of q of returning 1. We have a prior x that the machine is good, however we do not know for sure if the machine is good or not. Thus, given a prior x, the expected earning of a single pull is $xp + (1 - x)q - c$. Finally, we discount the future at rate δ, so that the value of earnings in time t is discounted by a factor of δ^t. As we play, we update our prior using Bayes rule and the problem we consider is that of determining the prior, x^*, under which we should stop playing.

The correspondence between the dynamic pricing problem and the bandits problem is straightforward. The only apparent difference in the problems comes from the available information. In bandit problems, we typically assume that we have the whole history of pulls, whereas in the dynamic pricing problem, we only want to assume that the users get to see the number of likes and dislikes the product has received so far. However, as observed in Lemma 1 this is not an issue since the updated prior after a number of reviews only depends on the number of likes and dislikes and not on the sequence itself.

[3]This is the best possible situation for the whole market, considering both the seller and the buyers.

[4]Again, to avoid trivial cases we assume that $0 \leq q < c < p \leq 1$.

3 Related models

Our model adds to the literature on pricing with online reviews. Most of these existing models try to make simplifying assumptions, but even with these, they end up being extremely difficult to solve and analyze. On the contrary, the model we consider is relatively simple (though still realistic) and can be solved exactly. In what follows we present a more detailed comparison between our model and the closest related work in the literature.

First, let us discuss the model of Crapis et al. [5]. As in our model, they consider an infinite stream of buyers purchasing a product. The quality p of the product is also initially unknown and can take values in an interval (so, in this sense, it is similar to our extended model). As opposed to our model (and actually to most models in social learning) they consider heterogeneous buyers. The utility of buyer i is given by $u_i = \alpha_i q - p$, where p is the price charged by the seller, and the α_i are i.i.d. random variables drawn from a known distribution F. They also consider that buyers arrive according to an independent Poisson process, but this does not significantly affect the results. Initially, all buyers have some common prior on the quality of the product, q_0. Consumers report likes and dislikes depending on whether their utility was nonnegative, taking into account the true quality of the product (which was discovered upon buying it). The information available to the buyer upon making his purchasing decision includes reviews made by all of his predecessors and knowledge about the order in which they acted. The seller's problem is to find the price maximizing her discounted expected revenue. We remark that the information structure here is quite involved, so, in order to tackle the seller's problem, the authors make some simplifying assumptions and resort to mean-field approximations.

In subsequent work, Ifrach et al. [11] further refine the latter model. In particular, they reduce the possible *qualities* that the product can take to two possible values: *high* and *low*. Accordingly they modify the form of the utility of buyers to take an additive form. Namely, a buyer's utility equals to the quality of the product, minus the price paid, plus the buyer's type which are represented by i.i.d. random variables. Again, a like/dislike represents whether the buyer's utility was positive/negative. With these assumptions, the information structure gets somewhat simplified, although it is still quite complex. In this paper, as in ours, the authors additionally assume that the product has a cost c and observe that this cost plays an important role in the optimal dynamic pricing policy, and on whether consumers ultimately learn the true quality of the product. Closely related to the model of Ifrach et al., is that of Acemoglu et al., [1]. Similar to our work they consider static and dynamic pricing strategies for the seller and give conditions under which asymptotic learning occurs, and at which speed. A key distinction in Acemoglu et al.'s model refers to whether buyers have access to summary statistics or to the full history. We note that in our model setup (which is different) this distinction is unnecessary as we prove that both settings coincide.

Very recently, Shin et al. [17] take a slightly different approach to the problem. They consider a finite horizon model in which at each point in time a new buyer shows up. In their model buyers are homogeneous and the utility of buyer i is given by $u_i = q_i - p_i$, where q_i is the experienced quality of the product and p_i the price paid. This paper takes a different approach to modeling the quality of the product. This is assumed to be a Gaussian random variable of mean μ and standard deviation 1. However, only the seller known the mean while the buyers' use online reviews to learn this parameter. Another different feature of this paper is that online reviews may take numerical values beyond like/dislike (say star rating). As the authors note, the resulting dynamic pricing problem is extremely hard to

solve analytically so they end up looking at asymptotically optimal pricing policies.

Another related paper is the work of Chawla et al. [4]. This paper, however, follows a different language that makes the comparison slightly more difficult. The authors consider a buyer who repeatedly interacts with a seller. The buyer does not know his valuation, and every time he purchases the product, he updates his valuation V_t. This value can be thought of as the prior on the quality in our model. The evolution of V_t is assumed to follow a martingale, which naturally holds in our Bayesian updating. However, Chawla et al. make some assumptions about the variance and the step size of the random process (that then affect the main results). Another difference is that this paper does not consider the cost of the product, which plays an important role in our and other dynamic pricing models. The main result of Chawla et al. is that, under suitable conditions, a simple pricing strategy in which the product is given away for free up until some point, and then a fixed price is used, recovers a good fraction of the optimal revenue.

4 A Dynamic Programming Approach

In this section, we propose a fast dynamic programming approach to compute an approximation of the threshold prior x^* and the global expected reward in the dynamic pricing scenario. Then, we explain how to adapt this algorithm to compute an approximation of the global expected reward in the static price scenario.

Let us denote the set of possible updated priors by $\mathcal{P}(x) := \{x_{\ell,d} \mid \ell, d \geq 0\}$. We say that a set $S \subseteq [0,1]$ is *discrete* if all elements of S have a neighborhood that contains no other elements of S. The next result characterizes when the set of possible updated priors is discrete. We note that whenever it is not, then $\mathcal{P}(x)$ is dense in $[0,1]$.

Proposition 2. *The space of possible updated priors $\mathcal{P}(x)$ is discrete if and only if*

$$\frac{\log(\frac{p}{q})}{\log(\frac{1-q}{1-p})} \in \mathbb{Q}.$$

Furthermore, whenever $\mathcal{P}(x)$ is not a discrete set, it is dense in $[0,1]$.

Proof. First, assume that

$$\gamma := \frac{\log(\frac{p}{q})}{\log(\frac{1-q}{1-p})} = \frac{a}{b},$$

where a and b are two integers such that $gcd(a,b) = 1$. This is equivalent to $\left(\frac{1-q}{1-p}\right)^a = \left(\frac{p}{q}\right)^b$. We show that $\mathcal{P}(x)$ is discrete, by arranging its elements as an infinite increasing sequence $(x_i)_{i\in\mathbb{Z}}$ such that $x_0 = x$ and for all $i \in \mathbb{Z}$ we have $L(x_i) = x_{i+a}$ and $D(x_i) = x_{i-b}$.

We first show that the updated prior after a sequence of b likes and a dislikes is unchanged, i.e. $x_{b,a} = x_{0,0}$. Indeed, by Lemma 1, we have

$$x_{b,a} = \frac{xp^b(1-p)^a}{xp^b(1-p)^a + (1-x)q^b(1-q)^a} = \frac{x}{x + (1-x)(\frac{q}{p})^b(\frac{1-q}{1-p})^a} = x.$$

Then, for any $i \in \mathbb{Z}$, we can set $x_i := x_{\ell,d}$ where ℓ, d is any pair of integers such that $\ell \cdot a - d \cdot b = i$. The sequence $(x_i)_{i\in\mathbb{Z}}$ is strictly increasing because $a\ell - bd > a\ell' - bd'$ if and only if $(\frac{q}{p})^\ell(\frac{1-q}{1-p})^d > (\frac{q}{p})^{\ell'}(\frac{1-q}{1-p})^{d'}$, i.e., if and only if $x_{\ell,d} > x_{\ell',d'}$.

Now, we show that $\mathcal{P}(x)$ is dense in $[0,1]$ when γ is irrational. Since $x_{\ell,d}$ can be re-written as

$$x_{\ell,d} = \frac{x}{x + (1-x)\exp(\log(p/q) \cdot (d\gamma^{-1} - \ell))}$$

and $y \mapsto \dfrac{x}{x + (1-x)\exp(\log(p/q) \cdot y)}$ is a continuous one-to-one function from $\mathbb{R}$ to $(0,1)$ of bounded derivative, it is enough to show that the set $\{d\gamma^{-1} - \ell \mid \ell, d \geq 0\}$ is dense in $\mathbb{R}$. To show that, it is in fact sufficient to show that the set $H = \{h(d), d \geq 0\}$ is dense in $[0,1]$, where $h(d) := d\gamma^{-1} - \lfloor d\gamma^{-1} \rfloor$.

The first step to prove that H is dense in $[0,1]$ is to observe that it is infinite. Otherwise, there would exist two distinct integers d and d' such that $h(d) = h(d')$, which would contradict our assumption that γ is irrational.

Next, let $c \in [0,1]$ and $\epsilon > 0$. We show that that there exists $h \in H$ such that $|h - c| < \epsilon$. Since H is infinite and compact, there exist two elements of H that are at a distance less than ϵ from each other, i.e., there are two integers d and d' with $d > d'$ such that $0 < |h(d) - h(d')| < \epsilon$. First, suppose that $h(d) > h(d')$. Then, $h(d - d') = h(d) - h(d') < \epsilon$ and the element

$$h\left(\left\lfloor \frac{c}{h(d-d')} \right\rfloor (d-d')\right) = \left\lfloor \frac{c}{h(d-d')} \right\rfloor h(d-d')$$

is in H and is at distance less than ϵ from c, what we wanted to find. In the other case we have $h(d) < h(d')$ and then $1 - \epsilon < h(d - d') < 1$. In that case, the element

$$h\left(\left\lfloor \frac{1-c}{1-h(d-d')} \right\rfloor (d-d')\right) = \left\lfloor \frac{1-c}{1-h(d-d')} \right\rfloor h(d-d')$$

is in H and is at distance ϵ from c. This completes the proof. □

Corollary 3. *Let a and b be two integers such that $gcd(a,b) = 1$ and assume that*

$$\frac{\log(\frac{1-q}{1-p})}{\log(\frac{p}{q})} = \frac{a}{b}.$$

Then there exists an infinite increasing sequence $(x_i)_{i \in \mathbb{Z}}$ such that for all $i \in \mathbb{Z}$ we have $L(x_i) = x_{i+a}$ and $D(x_i) = x_{i-b}$.

In this setting, we can re-write eq. (3) as $V(x_i) = 0$ if $x < x_{\text{stop}}$ and otherwise:

$$V(x_i) = R(x_i) + \delta \mathbb{P}_{x_i}(\text{👍})V(x_{i+a}) + \delta \mathbb{P}_{x_i}(\text{👎})V(x_{i-b})$$

The algorithm. To apply a dynamic programming approach, let us fix an $\epsilon > 0$ that corresponds to the threshold between the precision of the solution returned and the running time: the smaller ϵ, the greater the precision and the running time. We note that this approach is not mathematically rigorous since its validity requires smoothness conditions on the function $V(\cdot)$ that do not follow from eq. (3). In the next section we derive a formal approach while now we continue with this approach that is computationally tractable.

First, assuming that V is differentiable at $x = 1$ we get from eq. (3) that $V(1) = \frac{p-c}{1-\delta}$ and $V'(1) = \frac{p-q}{1-\delta}$, as the following proposition shows.

Proposition 4. *Assume that the solution V of eq. (2) is differentiable at $x = 1$, then we have $V(1) = \frac{p-c}{1-\delta}$ and $V'(1) = \frac{p-q}{1-\delta}$.*

Proof. If $x = 1$, then the product must be good; therefore the expected global reward is

$$V(1) = \sum_{t \geq 0} (p - c)\delta^t = \frac{p - c}{1 - \delta}.$$

Now fix a small $\epsilon > 0$. Assume that the function $V(x)$ admits a derivative in $x = 1$, we can write $V(1 - \epsilon) = V(1) - \epsilon V'(1) + o(\epsilon)$. With eq. (2) we have for $x = 1 - \epsilon$ close to 1:

$$\begin{aligned}
V(1-\epsilon) &= R(1-\epsilon) + \delta \cdot \mathbb{P}_x(👍) \cdot V(L(1-\epsilon))\delta \cdot \mathbb{P}_x(👎) \cdot V(D(1-\epsilon)) \\
&= R(1-\epsilon) + \delta(p - \epsilon(p-q))(V(1 - \frac{q}{p}\epsilon + o(\epsilon))) \\
&\quad + \delta((1-p) + \epsilon(p-q))(V(1 - \frac{1-q}{1-p}\epsilon + o(\epsilon))) \\
&= R(1-\epsilon) + \delta(p - \epsilon(p-q))(V(1) + \frac{q}{p}\epsilon V'(1) + o(\epsilon)) \\
&\quad + \delta((1-p) + \epsilon(p-q))(V(1) + \frac{1-q}{1-p}\epsilon V'(1) + o(\epsilon)) \\
&= R(1-\epsilon) + \delta V(1) - \epsilon\delta V'(1) + o(\epsilon) \\
&= p - c + \delta V(1) + \epsilon\left(-(p-q) + \delta V'(1)\right) + o(\epsilon) \\
&= V(1) - \epsilon V'(1) + o(\epsilon).
\end{aligned}$$

Thus, $V'(1) = \frac{p-q}{1-\delta}$. □

With Proposition 4, for all indices i such that $x_i > 1 - \epsilon$, we can make the estimation that $V(x_i) \approx V(1) - (1 - x_i)V'(1) = \frac{p-c}{1-\delta} - (1 - x_i)\frac{p-q}{1-\delta}$. Let i_{start} be the greatest index i such that $x_{i_{start}} < 1 - \epsilon$. We have the following estimate: $i_{start} = O_x(a \cdot \log_{p/q}(1/\epsilon))$.

Then, for all $i \leq i_{start}$, we recursively compute $V(x_i)$, in decreasing order, with:

$$V(x_i) := \frac{V(x_{i+b}) - R(x_{i+b}) - \delta \cdot \mathbb{P}_{x_{i+b}}(👍)V(x_{i+b+a})}{\delta \cdot \mathbb{P}_{x_{i+b}}(👎)}$$

until we have $V(x_i) \leq 0$. We call this index i^* and then set $V(x_i) := 0$ for all $i \leq i^*$. The prior x_{i^*} gives an estimation of x^*. This method is fast but it is difficult to prove an upper bound on the precision of the results obtained. Figure 1 presents an example that showcases the properties we have described.

Static price scenario. We can easily adapt this algorithm to compute an approximation of the global expected reward in the static price scenario. Here we have $R(x) = \pi - c$ which implies $V(1) = \frac{\pi - c}{1-\delta}$ and $V'(1) = 0$. Then, we compute values $V(x_i)$ similarly, for all $i \leq i_{start}$, until $x_i \leq x_{\min}$.

In the next section, we give a combinatorial method to compute x^* and the global expected reward, that enables us to provide a strong guarantee on the solution.

5 A Combinatorial Approach

In this section we give a good estimation of the value $V(x)$ in both pricing scenarios. To compute $V(x)$, we make use of the concepts of *Catalan's triangle* and *Catalan's trapezoid*. Conceptually, *Catalan's triangle* is a number triangle whose entries $C(\ell, d)$ correspond to the number of strings such that there are ℓ "likes" and d "dislikes" such that no initial segment

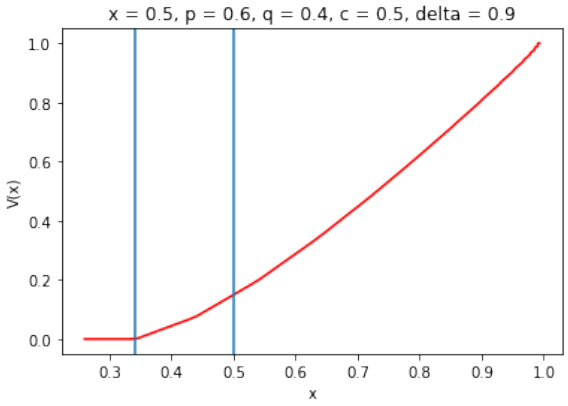

Figure 1: An example of computing $V(x)$, under dynamic prices, through our dynamic programming approach. Note that $x^* = 0.33$, which is lower than $x = 0.5 = \frac{c-q}{p-q}$, causes the local reward to be 0.

of the string has more dislikes than likes. The well-known *Catalan numbers* correspond to $C(n,n) = \frac{1}{n+1}\binom{2n}{n}$.

The so-called *Catalan's trapezoid* is an extension of Catalan's triangle, in which $C_m(\ell,d)$ counts the number of strings with ℓ likes and d dislikes such that every initial segment has at least m more likes than dislikes[5]. In particular the Catalan's triangle corresponds to the special case where $m = 0$. We have the following closed form for the Catalan's trapezoid [15]: $C_m(\ell,d) = \binom{\ell+d}{d}$ if $0 \leq d \leq m$, $C_m(\ell,d) = \binom{\ell+d}{d} - \binom{\ell+d}{d-m-1}$ if $m < d \leq \ell + m$ and $C_m(\ell,d) = 0$ otherwise.

Definition 5 (Catalan's quadrilateral)**.** *Given integers ℓ, d and parameters a, b and m we denote $C_m^{a,b}(\ell,d)$ the number of strings consisting of w L-s and d D-s such that in every initial segment of the string that consists of ℓ' L-s and d' D-s, the value $a \cdot \ell' - b \cdot d'$ is always at least $-m$.*

Let us say that these numbers form *Catalan's quadrilateral* since the Catalan's trapezoid corresponds to $a = b = 1$. Figure 2 provides a geometrical interpretation of these numbers.

We have the following induction to compute these numbers: $C_m^{a,b}(\ell,d) = 0$ whenever $a \cdot \ell - b \cdot d < -m$. Otherwise $C_m^{a,b}(\ell,d) = 1$ when $\ell = 0$ or $d = 0$. And generally: $C_m^{a,b}(\ell,d) = C_m^{a,b}(\ell-1,d) + C_m^{a,b}(\ell,d-1)$.

Thus, computing $C_m^{a,b}(\ell,d)$ can be done in time $O(\ell \cdot d)$ with a simple dynamic program implementing the above inductive formulation. No non-recurrence-based formula exists for Catalan's quadrilaterals, though computation of partial values in the quadrilateral are derived in [7].[6] The purpose of defining these particular numbers lies in the following lemma that will enable us to provide an expression of the global expected reward.

Given an initial prior x, we let X_t be the (random) updated prior after t reviews. In particular, let $X_0 = x$. Recall that $x_{\ell,d}$ denotes the updated prior after a sequence of ℓ likes and d dislikes from an initial prior x, and $R(x)$ is the local reward[7] when the prior is x. The

[5]We use here a slightly different definition than in [15]. We have $C'_m(\ell,d) = C_{m-1}(\ell,d)$ where $C'_m(\ell,d)$ denote the original Catalan Trapezoid numbers.

[6]See also [9] for some recent developments.

[7]Recall that $R(x) = \pi - c$ in the static price setting and $R(x) = xp + (1-x)q - c$ in the dynamic price setting

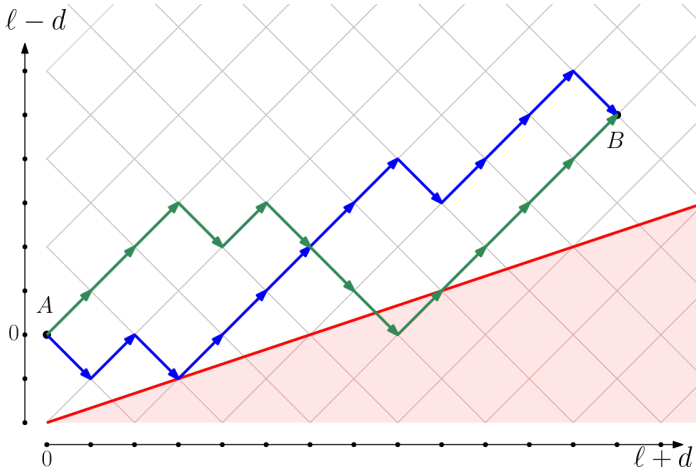

Figure 2: The number of paths from A to B along the grid, that do not enter the light red area (like the blue path but not like the green one) is the Catalan's quadrilateral number $C_3^{1,2}(9,4) = 570$ (for comparison, the unconditional number of paths from A to B is $\binom{9+4}{4} = 715$). The slope of the boundary depends on parameters a, b. The classic Catalan's trapezoid arises when the boundary is horizontal.

global expected prior from a prior x is given by

$$V(x) = \sum_{\ell,d \geq 0, t=\ell+d} \delta^t \cdot \mathbb{P}_x(X_t = x_{\ell,d}) \cdot R(x_{\ell,d}).$$

We now use the Catalan's quadrilateral to provide an expression of $\mathbb{P}_x(X_t = x_{\ell,d})$. Let x_{stop} denote the value of the prior when the selling process stops: in the dynamic pricing scenario, we have $x_{\text{stop}} = x^*$ and in the fixed price scenario, we have $x_{\text{stop}} = x_{min} = \frac{\pi - q}{p - q}$.

Lemma 6. *Let ℓ, d two integers, and $t = \ell + d$. Given any prior x, we have that*

$$\mathbb{P}_x(X_t = x_{\ell,d}) = C_m^{a,b}(\ell, d) \cdot p_x(\ell, d),$$

where

$$p_x(\ell, d) := xp^\ell(1-p)^d + (1-x)q^\ell(1-q)^d$$

is the probability of having a given ordered sequence of ℓ likes and d dislikes; $a = \log(p/q)$; $b = \log(\frac{1-q}{1-p})$; and $m = \log_{\frac{1-q}{1-p}}\left(\frac{x(1-x_{stop})}{x_{stop}(1-x)}\right)$.

Proof. By Lemma 1, the updated prior after a sequence of ℓ likes and d dislikes only depends on ℓ and d so as the probability of such each sequence. Therefore, $\mathbb{P}_x(X_t = x_{\ell,d})$ is the product of the probability of one sequence and the number of such sequences.

We first show that the probability of having a given ordered sequence of ℓ likes and d dislikes is

$$p_x(\ell, d) = xp^\ell(1-p)^d + (1-x)q^\ell(1-q)^d.$$

We proceed by induction on the length $\ell + d$ of the sequence $\ell + d$. The base case of the induction follows from Lemma 1. Then, using the induction hypothesis, we obtain

$$\begin{aligned} p_x(\ell+1, d) &= p_{L(x)}(\ell, d) \cdot \mathbb{P}_x(\text{👍}) = (L(x)p^\ell(1-p)^d + (1-L(x))q^\ell(1-q)^d)(xp + (1-x)q) \\ &= \Big(\frac{xp}{xp + (1-x)q}p^\ell(1-p)^d + (1 - \frac{xp}{xp + (1-x)q})q^\ell(1-q)^d\Big)(xp + (1-x)q) \\ &= xp^{\ell+1}(1-p)^d + (1-x)q^{\ell+1}(1-q)^d. \end{aligned}$$

The calculation for $p_x(\ell, d+1) = xp^\ell(1-p)^{d+1} + (1-x)q^\ell(1-q)^{d+1}$ works similarly. Thus we have established the induction.

Now for any integers ℓ, d, it is easy to see that $x_{\ell,d} < x_{\text{stop}}$ if and only if $a \cdot \ell - b \cdot d < -m$ where $a = \log(p/q)$, $b = \log(\frac{1-q}{1-p})$ and $m = \log_{\frac{1-q}{1-p}}\left(\frac{x(1-x_{\text{stop}})}{x_{\text{stop}}(1-x)}\right)$. Thus, the number of sequences of ℓ likes and d dislikes such that the prior at any time is at least x_{stop} is equal to the Catalan's quadrilateral number $C_m^{a,b}(\ell, d)$. □

In the dynamic pricing scenario, to compute a good approximation of $V(x)$, we first need to compute a good approximation of the threshold x^*.

Computing x^*. By definition, x^* is the prior for which stopping or continuing to play gives the same global expected reward. Assuming that the initial prior is $x = x^*$, we have $m = 0$ and we obtain after simplification the following equation :

$$0 = V(x^*) = \sum_{\ell,d\geq 0} \delta^{\ell+d} C_0^{a,b}(\ell, d) \cdot (x^* p^\ell (p-c)(1-p)^d + (1-x^*)q^\ell(q-c)(1-q)^d),$$

where $a = \log(p/q)$ and $b = \log(\frac{1-q}{1-p})$. If we set $\Phi(p) := \sum_{\ell,d\geq 0} \delta^{\ell+d} \cdot C_0^{a,b}(\ell, d) \cdot (p-c)p^\ell(1-p)^d$, the above equation becomes $0 = x^*\Phi(p) + (1-x^*)\Phi(q)$. Thus we can express x^* as

$$x^* = \frac{\Phi(q)}{\Phi(q) - \Phi(p)}.$$

To get a precise estimate of the value $\Phi(p)$, we only need to focus on sequences of likes and dislikes that do not exceed a certain length. More precisely, fix any $\epsilon > 0$. Since $\Phi(p)$ is defined as a series of positive terms, we know that there exists an integer t_ϵ such that

$$\sum_{\ell,d\geq 0, \ell+d\geq t_\epsilon} \delta^{\ell+d} \cdot C_0^{a,b}(\ell, d) \cdot (p-c)p^\ell(1-p)^d \leq \epsilon$$

and since this series is upper bounded by a convergent geometric series, we have the following estimate $t_\epsilon = O(\log 1/\epsilon)$. Thus, we can compute an ϵ-estimate $\widehat{x^*}$ of x^*, i.e. $|\widehat{x^*} - x^*| < \epsilon$, in time $O(\log(1/\epsilon)^2)$. When the ratio a/b is a rational number, the set of possible updated priors from x is discrete, so that choosing an ϵ sufficiently small enables to compute an exact value for x^*.

Computing $V(x)$. Once we have a precise estimation of x^*, we can proceed similarly to compute an arbitrarily close estimation of $V(x)$ for any prior x. For any $\epsilon > 0$, there exists $t_\epsilon = O(\log 1/\epsilon)$, such that

$$\sum_{\ell,d\geq 0, \ell+d\geq t_\epsilon} \delta^{\ell+d} \cdot \mathbb{P}_x(X_t = x_{\ell,d}) \cdot R(x_{\ell,d}) \leq \epsilon.$$

We can then use x_{stop} and the values of the Catalan trapezoid to compute the sum

$$\hat{V}(x) := \sum_{\ell,d\geq 0, t=\ell+d\leq t_\epsilon} \delta^t \cdot \mathbb{P}_x(X_t = x_{\ell,d}) \cdot R(x_{\ell,d})$$

and we have $|V(x) - \hat{V}(x)| \leq \epsilon$.

In the *symmetric* case, when the values of p and q are such that $q = 1-p$, we can even get a closed expression for x^* and $V(x)$. Indeed, we have $a = b = 1$ and we can use the closed formula of the coefficients of the Catalan Trapezoid.

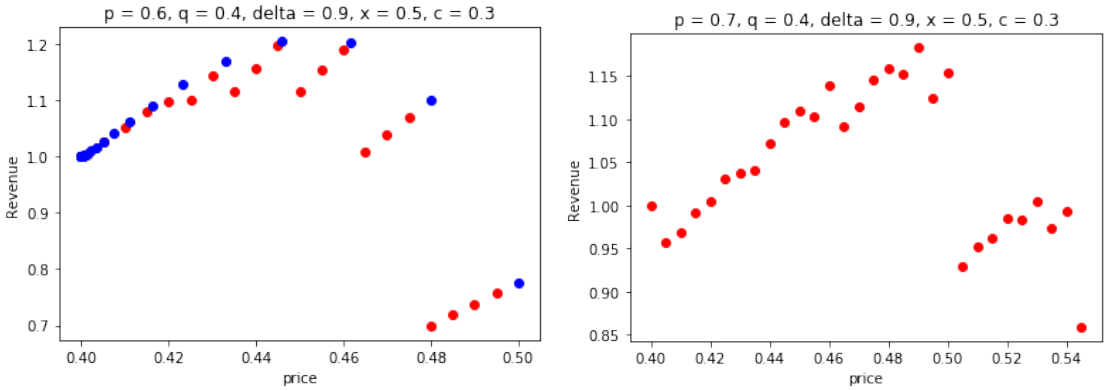

Figure 3: The value of global expected reward depending on the fixed price π. In the symmetric setting (left), the blue points represent the revenue on the efficient frontier, which is the maximum possible price per discrete x_{stop}. The red points are not optimal because such prices yield the same number of possible net dislikes before the buyers stop buying. In the general setting (right), we see that revenue as a function of price is not as well-defined.

Computing the optimal static price. For each fixed price π, we can thus compute the global expected reward for the seller. In order to maximize her revenue, the seller can optimize this function over the values of $\pi \in (c, xp + (1-x)q]$. Clearly π has to be at least c for the revenue to be positive. Also if $\pi > xp + (1-x)q$ then no user will ever buy and then the revenue is zero.

In the symmetric setting, note that for a price π, buyers can initially tolerate up to a net of m_π dislikes, where m_π is a function decreasing in π before no longer buying. Thus, for a fixed integer m, we can maximize the global expected reward by setting the maximum price π such that $m_\pi = m$. This efficient frontier appears to be concave, as shown in Figure 3, so finding the optimal price is simply a binary search procedure along the efficient frontier. In the asymmetric setting, however, m_π is a function of both the number of likes and dislikes, due to the behavior of Catalan's quadrilateral. Figure 3, shows that computing the optimal price in the asymmetric case requires searching over a larger set of prices.

6 Success and Failure of Learning

In this section we investigate the probability that the market learns the true value of the product, i.e. the probability of stopping in finite time when the product is bad and the probability of selling forever when the product is good. If dynamic pricing is used, learning occurs with larger probability. Our main conclusion is to express this additional gain as a function of the primitives of the model, proving a simple quantification of the potential gains of dynamic pricing over static pricing.

A first observation is that in both models the market will always discover if the product is bad. Recall that we assume that $q \leq c$.

Lemma 7. *Assuming that the product is bad, we stop selling in finite time almost surely.*

Proof. Let $D_t = a\ell_t - bd_t$ denote the random variable that corresponds to the weighted difference between the number of likes and dislikes after t reviews, where $a = \log(p/q)$ and

$b = \log((1-q)/(1-p))$. We define the stopping time τ as the first time t when $D_t < -m$ where m depends on the original prior x and the threshold prior x_{stop}.

Given that the product is bad, we have: $\mathbb{E}(D_t - D_{t-1}) = a \cdot \mathbb{P}(\text{👍}|\text{bad}) - b \cdot \mathbb{P}(\text{👎}|\text{bad}) = aq - b(1-q) =: \mu < 0$ for any $0 < q < p$. Then, $\mathbb{E}(D_t) = \mu \cdot t \to_t -\infty$.

We deduce, for t sufficiently large, using Bienaymé-Tchebychev inequality, that

$$\mathbb{P}(\tau \geq t) \leq \mathbb{P}\left(|D_t - \mathbb{E}(D_t)| \geq -\mu \cdot t - m\right) \leq O(1/t)$$

where $x_{\text{stop}} = x^*$ for the dynamic pricing model and $x_{\text{stop}} = x_{\min}$ for the single price model.[8] □

On the other hand, when the product is good, learning may fail to occur. To quantify this efficiency loss, recall that x_{stop} corresponds to the threshold prior from which the users stop buying. In the dynamic pricing model we have $x_{\text{stop}} = x^*$ and in the static price model, we have $x_{\text{stop}} = x_{\min} = \frac{\pi - q}{p - q}$. The main result of this section, Lemma 8, establishes that when the product is good the probability of learning it is at least $\frac{x - x_{\text{stop}}}{x(1 - x_{\text{stop}})}$.

Given an initial prior x, for all time $t \geq 0$ we define the (random) variable X_t that is the updated prior after t reviews. $(X_t)_{t \geq 0}$ is a martingale with $X_0 = x$. Now, let τ denote the (random) time at which the selling process stops.

Lemma 8. *Given an initial prior x, we have the following estimation on the probability that the process continues forever: $\frac{x - x_{stop}}{1 - x_{stop}} < \mathbb{P}_x(X_\tau = \infty) \leq \frac{x - D(x_{stop})}{1 - D(x_{stop})}$. Additionally, if we assume that the product is good, then the probability of learning that the product is good is $\mathbb{P}_x(X_\tau = \infty \mid \textit{good}) > \frac{x - x_{stop}}{x(1 - x_{stop})}$.*

Proof. Let us fix $\epsilon > 0$. The random time τ_ϵ at which X_t reaches x_{stop} or $1 - \epsilon$ is a stopping time. Since τ_ϵ has finite expectation, by the optional stopping theorem, the expected value of X_{τ_ϵ} is equal to the initial prior, i.e., $\mathbb{E}(X_{\tau_\epsilon}) = x$. Then we get

$$x = \mathbb{P}_x(X_{\tau_\epsilon} < x_{\text{stop}}) \mathbb{E}(X_{\tau_\epsilon} | X_{\tau_\epsilon} < x_{\text{stop}}) + \mathbb{P}_x(X_{\tau_\epsilon} > 1 - \epsilon) \mathbb{E}(X_{\tau_\epsilon} | X_{\tau_\epsilon} > 1 - \epsilon).$$

We know that $D(x_{\text{stop}}) \leq \mathbb{E}(X_{\tau_\epsilon} | X_{\tau_\epsilon} < x_{\text{stop}}) < x_{\text{stop}}$ and $1 - \epsilon < \mathbb{E}(X_{\tau_\epsilon} | X_{\tau_\epsilon} > 1 - \epsilon) \leq L(1 - \epsilon)$.

Thus, when ϵ goes to zero, we obtain $\mathbb{P}_x(\tau = \infty) = \frac{x - \mathbb{E}(X_\tau | X_\tau < x_{\text{stop}})}{1 - \mathbb{E}(X_\tau | X_\tau < x_{\text{stop}})}$. The estimation then follows from the fact that $D(x_{\text{stop}}) \leq \mathbb{E}(X_\tau \mid X_\tau < x_{\text{stop}}) < x_{\text{stop}}$.

To prove the second part of the statement we simply use:

$$\mathbb{P}_x(\tau = \infty) = \mathbb{P}_x(\tau = \infty \mid \text{good}) \cdot \mathbb{P}_x(\text{good}) + \mathbb{P}_x(\tau = \infty \mid \text{bad}) \cdot \mathbb{P}_x(\text{bad}) = \mathbb{P}_x(\tau = \infty \mid \text{good}) \cdot x$$

where $\mathbb{P}_x(\tau = \infty \mid \text{bad}) = 0$ holds by Lemma 7. □

Notice that since $x^* < x_{\min}$, we will learn that the product is bad in the static pricing scenario earlier than in the dynamic pricing model. Conversely, we learn that the product is good with higher probability in the dynamic pricing model.

In the symmetric case, i.e. when $q = 1 - p$, we can have a better estimation. Indeed, if $\tau = t$ then necessarily, $X_{t-1} = x_{\text{stop}}$ and we observe a dislike at time $t - 1$. Thus, $\mathbb{E}(X_\tau | X_\tau < \text{stop}) = D(x_{\text{stop}})$ so that $\mathbb{P}_x(X_\tau = \infty) = \frac{x - D(x_{\text{stop}})}{1 - D(x_{\text{stop}})}$. See also Lemma 11 in the Appendix.

[8] We now give an exact expression of this probability in the symmetric case, in Lemma 11.

With these results we can bound the ratio of not learning under the considered pricing strategies. This happens exactly when the product is good but the market does not discover it and stops buying in finite time. Let FN_{static} and FN_{dynamic} be the probabilities of stopping when the product is good in the static and in the dynamic prices scenarios, respectively. In the case when $p = 1 - q$ we have $\frac{FN_{\text{static}}}{FN_{\text{dynamic}}} = \frac{1/D(x^*)-1}{1/D(x_{\min})-1}$; see Lemma 12 in the Appendix.

7 Extension

We now consider the problem when the product has a quality $q \in Q \subseteq [0, 1]$. Again, a product has quality $q \in Q$ if it is liked by a q fraction of people, or equivalently, if the probability that a given person likes the product is q. In this model, the prior on the quality of the product becomes a random variable $X \in Q$. Given the current prior X, we update the prior to $L(X)$ after a like (👍) and to $D(X)$ after a dislike (👎) as follows:

$$\mathbb{P}(L(X) = q) := \mathbb{P}_X(X = q \mid 👍) = \frac{\mathbb{P}_X(👍 \mid X = q)\mathbb{P}_X(X = q)}{\mathbb{P}_X(👍)} = \frac{q}{\mathbb{E}(X)}\mathbb{P}(X = q);$$

$$\mathbb{P}(D(X) = q) := \mathbb{P}_X(X = q \mid 👎) = \frac{\mathbb{P}_X(👎 \mid X = q)\mathbb{P}_X(X = q)}{\mathbb{P}_X(👎)} = \frac{1 - q}{1 - \mathbb{E}(X)}\mathbb{P}(X = q).$$

These equations correspond to the case in which Q is a discrete set and should be replaced by the corresponding probability density functions in case Q is a continuous set. Again, the prior after a sequence of likes and dislikes is independent, of the order in which the likes and dislikes happened. Therefore we have a simple expression of the prior after a given sequences of reviews.

Lemma 9. *Given a prior X, the distribution of the updated prior $X_{\ell,d}$ after a sequence of ℓ likes and d dislikes is given by:*

$$\mathbb{P}(X_{\ell,d} \leq q) = \frac{\mathbb{E}(X^\ell(1 - X)^d | X \leq q)\mathbb{P}(X \leq q)}{\mathbb{E}(X^\ell(1 - X)^d)},$$

for all $q \in Q$.

Proof. By iterating the calculations above for the distribution of $L(X)$ and $D(X)$ one can obtain that $\mathbb{P}(X_{\ell,d} = q) = \frac{q^\ell(1-q)^d}{\mathbb{E}(X^\ell(1-X)^d)}\mathbb{P}(X = q)$, for any $q \in Q$. Of course, if Q is a continuous set, we need to replace the probabilities by the corresponding densities. To conclude the lemma, we simply need to integrate this equation for possible quality values in Q that are below q. □

Initially, suppose that the quality prior is X_0. At any time, both the seller and the buyers have access to the current prior X. So the buyer will buy the product if its utility $\mathbb{E}(X) - \pi$ is non-negative, where π is the current price of the product. In the dynamic price setting, in order to maximize her revenue, the seller must set the product at price $\pi := \mathbb{E}(X)$. Of course, at any point in time, the seller may decide to stop selling the product, and she will do this if the future expected reward is negative. With this, we can again write a dynamic program to estimate the seller's optimal revenue under a dynamic pricing strategy. First

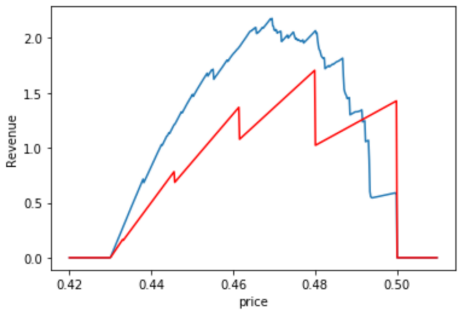

Figure 4: Seller's revenue depending on the static price. The red plot corresponds to the good/bad (binary) model with $p = 0.6, q = 0.4, c = 0.43, \delta = 0.99$, and the prior is such that the product is good/bad with probability 0.5. The blue plot corresponds to the extended model with $Q = [0.4, 0.6]$ and uniform prior distribution.

note that given a prior X, the value of the total expected reward (for the seller) satisfies the following equation:

$$V(X) = \max\Big(0, \mathbb{E}(X) - c + \delta \cdot \big(\mathbb{P}_X(\text{👍})V(L(X)) + \mathbb{P}_X(\text{👎})V(D(X))\big)\Big), \tag{4}$$

where c is the production cost per unit. Define the estimator $\widehat{V}(X) := \frac{\mathbb{E}(X)-c}{1-\delta}$ that does not take into account the reviews. We now describe a dynamic programming approach to compute an approximation $\widetilde{V}(X)$ of $V(X)$. For this, we consider a constant M, to be specified later. We set $\widetilde{V}(X_{\ell,d}) = \widehat{V}(X_{\ell,d})$, for all ℓ, d such that $\ell + d = M$. (As we note later, for M relatively large and a well behaved initial prior distribution X_0, the random variable $X_{\ell,d}$ is highly concentrated around its mean, which is roughly $\ell/(\ell+d)$, and thus $\widehat{V}(X_{\ell,d})$ is very close to $V(X_{\ell,d})$). For $i = M-1, \ldots, 0$ and all ℓ, d such that $\ell + d = i$, we compute $\widetilde{V}(X_{\ell,d})$ using (4), where we have replaced V by $\widetilde{V}$. This can be done in time $O(M^2)$.

Proposition 10. $|\widetilde{V}(X) - V(X)| \leq \frac{1-c}{1-\delta}\delta^M$.

Proof. We prove by induction that $|\widetilde{V}(X_{\ell,d}) - V(X_{\ell,d})| \leq \frac{1-c}{1-\delta}\delta^{M-i}$, where $i = \ell + d$. We trivially have $|\widetilde{V}(X_{\ell,d}) - V(X_{\ell,d})| \leq \frac{1-c}{1-\delta}$ when $\ell + d = M$. Now, when $i = \ell + d < M$, we have

$$\begin{aligned}
&|\widetilde{V}(X_{\ell,d}) - V(X_{\ell,d})| \\
&\leq \delta \cdot \Big(\mathbb{P}_{X_{\ell,d}}(\text{👍})|\widetilde{V}(X_{\ell+1,d}) - V(X_{\ell+1,d})| + \mathbb{P}_{X_{\ell,d}}(\text{👎})|\widetilde{V}(X_{\ell,d+1}) - V(X_{\ell,d+1})|\Big) \\
&\leq \delta\Big(\mathbb{P}_{X_{\ell,d}}(\text{👍})\frac{1-c}{1-\delta}\delta^{M-(i+1)} + \mathbb{P}_{X_{\ell,d}}(\text{👎})\frac{1-c}{1-\delta}\delta^{M-(i+1)}\Big) = \frac{1-c}{1-\delta}\delta^{M-i}.
\end{aligned}$$

□

The previous result provides a general bound on the quality of our dynamic programming approach. When the initial prior is discrete, the problem is combinatorial and one can use similar ideas from Section 5 to obtain improved guarantees. On the other hand, when the set of possible values is continuous, we can obtain significantly better bounds with additional properties on the initial prior. In particular, when the initial prior admits a continuous and strictly increasing distribution with bounded density, we have that for $\ell + d$ large, $X_{\ell,d}$ is highly concentrated around $\ell/(\ell+d)$. For instance, its variance is $Var(X_{\ell,d}) = O(\frac{1}{\ell+d})$.

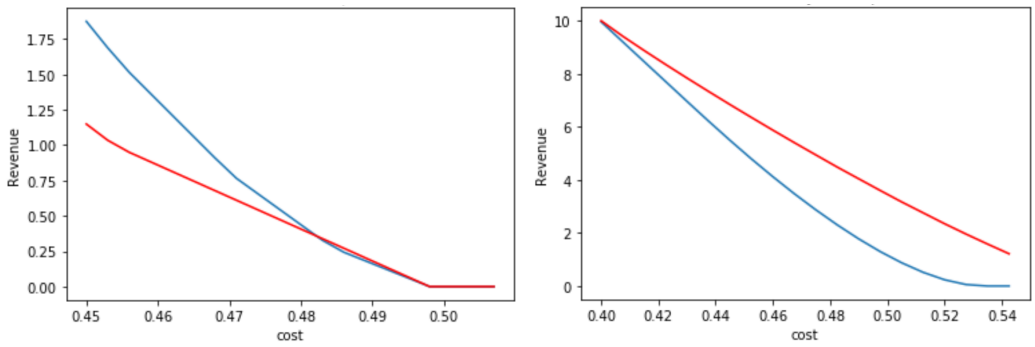

Figure 5: Seller's revenue as a function of the cost in the same instances as in Figure 4. The red function corresponds to the good/bad model, and the blue function corresponds to the extended model. On the left, we plot revenues resulting from the optimal static pricing (the price is optimized for each possible cost), and on the right, we plot revenues resulting from dynamic pricing.

Additionally, if $\mathbb{E}(X)$ is bounded away from c, using Bienaymé-Tchebychev inequality (or simply the law of large numbers), one can show that the probability that the updated prior X' obtained after M reviews is such that $\mathbb{E}(X') < c$ is $O(1/M^2)$. With this, the bound of Proposition 10 improves to $O(\frac{\delta^M}{M(1-\delta)})$.

In terms of static pricing, we can easily adapt the approach to this more general model. Again, the seller wants to fix a price π, and the buyer buys if $\mathbb{E}(X) \geq \pi$, where X is the public prior distribution. Based on this, we can devise a simple dynamic program similar to the algorithm of Proposition 10, to obtain a good estimation of the total revenue for the seller. Again, we can prove that if the product is bad, i.e., the true value of the product is smaller than the cost, then the process will stop almost surely. In Figure 4 we plot the revenue as a function of the static price in both models, setting parameters to make them comparable (equal expected quality and equal range of product quality).

Finally, we evaluate the expected revenue of both the static and dynamic pricing policies in both models in the comparable setting just described. The situation is depicted in Figure 5. A surprising result is that while for the optimal static price, the seller's revenue in the extended model is mostly higher than that in the binary model, the situation changes dramatically for dynamic pricing. The intuition is that in dynamic pricing, if $c > 0.4$, then the expected quality minus c is higher in the good/bad model. However, when considering static prices this effect is less relevant than the faster learning process that occurs in the extended model. Indeed, in that case, if c is relatively small, then the extended model gets higher revenue since the product is more interesting, and the buyers quickly learn that it is worth buying. However, as the cost approaches 0.5, the product becomes less interesting and then the faster learning implies that the seller's revenue decreases faster in the extended model.

References

[1] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Learning from reviews: The selection effect and the speed of learning. *Econometrica*, 90(6):2857–2899, 2022.

[2] O. Besbes and M. Scarsini. On information distortions in online ratings. *Operations Research*, 66(3):597–610, 2018.

[3] T. Bonald and A. Proutiere. Two-target algorithms for infinite-armed bandits with bernoulli rewards. In *proceedings of NIPS*, 2013.

[4] S. Chawla, N. Devanur, A. Karlin, and B. Sivan. Simple pricing schemes for consumers with evolving values. In *proceedings of SODA*, 2016.

[5] D. Crapis, B. Ifrach, C. Maglaras, and M. Scarsini. Monopoly pricing in the presence of social learning. *Management Science*, 63(11):3586–3608, 2017.

[6] K. Dave, S. Lawrence, and D.M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *proceedings of WWW*, 2003.

[7] Y. Fukukawa. Counting generalized dyck paths. `https://arxiv.org/abs/1304.5595`, 2013.

[8] J. Gittins and D. Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

[9] Benjamin Hackl, Clemens Heuberger, and Helmut Prodinger. Counting ascents in generalized dyck paths. In *proceedings of AofA*, 2018.

[10] M. Hu and B. Liu. Mining and summarizing customer reviews. In *proceedings of KDD*, 2004.

[11] B. Ifrach, C. Maglaras, M. Scarsini, and A. Zseleva. Bayesian social learning from consumer reviews. *Operations Research*, 67(5):1209–1221, 2019.

[12] Kanishka Misra, Eric M. Schwartz, and Jacob D. Abernethy. Dynamic online pricing with incomplete information using multiarmed bandit experiments. *Marketing Science*, 38:226–252, 2019.

[13] F. Monachou and I. Ashlagi. Discrimination in online markets: Effects of social bias on learning from reviews and policy design. In *proceedings of NeurIPS*, 2019.

[14] Yiangos Papanastasiou and Nicos Savva. Dynamic pricing in the presence of social learning and strategic consumers. *Management Science*, 63(4):919–939, April 2017.

[15] S. Reuveni. Catalan's trapezoids. *Probability in the Engineering and Informational Sciences*, 28(3):353–361, 2014.

[16] V. Shah, R. Johari, and J. Blanchet. Semi-parametric dynamic contextual pricing. In *proceedings of NeurIPS*, 2019.

[17] D. Shin, S. Vacarri, and A. Zeevi. Dynamic pricing with online reviews. *Management Science*, 69(2):824–845, 2023.

[18] Stefano Vaccari, Costis Maglaras, and Marco Scarsini. Social learning from online reviews with product choice. *SSRN Electronic Journal*, 01 2018.

Appendix

Lemma 11. *Assume that $q = 1 - p$. Then, when the product is good, the probability of selling forever is equal to $1 - \left(\frac{1-\sqrt{1-4p(1-p)}}{2p}\right)^m$ where m is the smallest integer such that after m dislikes the prior goes below x_{stop}.*

Proof. (Lemma 11) Let p_m denote this probability. We have $p_m = 0$ if $m \leq 0$; $p_m = p \cdot p_{m+1} + (1-p) \cdot p_{m-1}$ otherwise, and $\lim_m p_m = 1$. The roots of the polynomial $pX^2 - X + (1-p)$ are 1 and $\frac{1-\sqrt{1-4p(1-p)}}{2p} < 1$. Thus, we deduce easily the expected expression. □

Lemma 12. *Let FN_{static} and $FN_{dynamic}$ be the probabilities of stopping when the product is good, respectively in the static and in the dynamic prices scenarios. We have $FN_{static} \geq FN_{dynamic} > 0$ and in the case when $p = 1 - q$:*

$$\frac{FN_{static}}{FN_{dynamic}} = \frac{D(x_{\min})}{D(x^*)} \cdot \frac{1 - D(x^*)}{1 - D(x_{\min})} = \frac{1/D(x^*) - 1}{1/D(x_{\min}) - 1}.$$

Proof. (Lemma 12) By Lemma 8, in the symmetric case, the probability of having a false negative is $\frac{D(x_{\text{stop}})}{1-D(x_{\text{stop}})} \cdot \frac{1-x}{x}$ and the formula follows. □