Dynamic Pricing and Learning with Bayesian Persuasion

Shipra Agrawal
Columbia University, sa3305@columbia.edu

Yiding Feng
The University of Chicago, yidingfeng@uchicago.edu

Wei Tang
Columbia University, wt2359@columbia.edu

We consider a novel dynamic pricing and learning setting where in addition to setting prices of products in sequential rounds, the seller also ex-ante commits to 'advertising schemes'. That is, in the beginning of each round the seller can decide what kind of signal they will provide to the buyer about the product's quality upon realization. Using the popular Bayesian persuasion framework to model the effect of these signals on the buyers' valuation and purchase responses, we formulate the problem of finding an optimal design of the advertising scheme along with a pricing scheme that maximizes the seller's expected revenue. Without any apriori knowledge of the buyers' demand function, our goal is to design an online algorithm that can use past purchase responses to adaptively learn the optimal pricing and advertising strategy. We study the regret of the algorithm when compared to the optimal clairvoyant price and advertising scheme.

Our main result is a computationally efficient online algorithm that achieves an $O(T^{2/3}(m\log T)^{1/3})$ regret bound when the valuation function is linear in the product quality. Here m is the cardinality of the discrete product quality domain and T is the time horizon. This result requires some natural monotonicity and Lipschitz assumptions on the valuation function, but no Lipschitz or smoothness assumption on the buyers' demand function. For constant m, our result matches the regret lower bound for dynamic pricing within logarithmic factors, which is a special case of our problem. We also obtain several improved results for the widely considered special case of additive valuations, including an $\tilde{O}\left(T^{2/3}\right)$ regret bound independent of m when $m \leq T^{1/3}$.

Key words: dynamic pricing, advertising, online learning, regret minimization

1. Introduction

Dynamic pricing is a key strategy in revenue management that allows sellers to anticipate and influence demand in order to maximize revenue and/or utility. When the customer valuation and demand response for a product is apriori unknown, price variation can also be used to observe and learn the demand function in order to adaptively optimize price and revenue over time. This learning and optimization problem has been a focus of much recent literature that uses exploration-

exploitation and multi-armed bandit techniques with dynamic pricing algorithms (e.g., see Kleinberg and Leighton (2003), Besbes and Zeevi (2009), Keskin and Zeevi (2014), Babaioff et al. (2015)).

In practice, there is another important tool available to sellers in the form of *advertising*, using which the sellers can inform and shape customers' valuations of a product. It has been theoretically Nelson (1970, 1974) and empirically Sahni and Nair (2020), Manchanda et al. (2006) shown that advertisements can serve as a credible signal of the quality or characteristics of the advertised product. Sellers can use advertising to provide partial information about a product in order to better position the product in the market and potentially increase customers' chances of purchasing the product. For example, as a common strategy to drive subscriptions, online newspapers may use a "teaser" that selectively includes previews of some news articles that are likely to entice readers to subscribe for access to the full story; in the online used-car market, the dealer can advertise the used car by emphasizing different aspects of the car, such as fuel efficiency/mileage/unique features, or selectively disclose history-report information from reputable third parties, catering to specific buyer interests; a film distributor may advertise the movie by selectively showing footage from the film.

However, advertising must be carefully designed to achieve the desired gains. At one glance up-selling or inflating the product quality by selectively disclosing only favorable information might appear as a profitable advertising strategy. But such strategies carry the disadvantage of not being very effective in modifying customer beliefs as customers may not trust that the provided information accurately reflects the product's true quality. Also, the design of the advertising strategy needs to interact with the design of the pricing strategy and account for the demand function. For example, to sell highly-priced products or under heavy competition/low demand, the customer may need to be convinced of a good match through more information and thorough insights about the product characteristics. On the other hand, in markets with high demand or for very low-priced goods, the seller may get away with revealing very little information. An extreme example of this phenomenon is the concept of *mystery deal boxes* sold by some retailers like Amazon/Woot, where the customers are not even made aware of the exact contents of the low-cost box that they are purchasing.[1]

In this paper, we use a *Bayesian persuasion framework* (Kamenica and Gentzkow 2011) to model the effect of an advertising strategy on customers' beliefs about product quality and consequently their purchase decisions. Our novel formulation combines the Bayesian persuasion model with dynamic pricing and learning in order to quantify the tradeoffs between the design of the pricing and advertising strategies and their combined impact on the revenue outcomes. Without any apriori knowledge of the demand function, our goal is to design an online algorithm that can use past

[1] For example, when selling the opaque products, the precise product features or characteristics are hidden from the customers (Elmachtoub and Hamilton 2021).

customer responses to adaptively learn a joint pricing and advertising strategy that maximizes the seller's revenue.

Bayesian persuasion is a popular framework for information design with several different settings considered in the literature (Kamenica and Gentzkow 2011, Dughmi 2017, Kamenica 2019, Bergemann and Morris 2019). We consider a Bayesian persuasion model where the sender (seller) ex-ante commits to an information policy (advertising strategy) that prescribes the distribution of signals the sender will provide to the receiver (buyer) on observing the true state of the world (product quality). The receiver, on observing the sender's signal, uses Bayes' rule to form a posterior on the state of the world. The receiver's action (purchase decision) then optimizes their expected utility under this posterior.

1.1. Problem Formulation

Specifically, our *dynamic pricing and advertising problem* is formulated as follows. There are T sequential and discrete rounds. In each round, a fresh product is offered by the seller, with a public prior distribution λ on the product quality $\omega \in \Omega \subseteq [0,1]$. At the beginning of each round t, before observing the realized quality of the t^{th} product, the seller commits to a price $p_t \in [0,U]$ and an advertising strategy ϕ_t, where $\phi_t(\sigma|\omega)$ prescribes the distribution of signal $\sigma \in \Sigma$ where Σ is an arbitrary signaling space given the realized product quality $\omega \in \Omega$.

A buyer arrives in each round t with *private* type θ_t generated i.i.d. from a distribution with CDF $F(\cdot)$ and support[2] $\Theta = [0,1]$. The CDF $F(\cdot)$ (or equivalently the demand function $D(\cdot) \triangleq 1 - F(\cdot)$) is fixed but unknown to the seller. For a buyer of type $\theta \in \Theta$, the valuation of a product with product quality ω is given by function $v(\theta, \omega)$.

The t^{th} product quality $\omega_t \sim \lambda$ is then realized and observed by the seller. The buyer cannot observe the realized product quality, but only a signal $\sigma_t \sim \phi_t(\cdot|\omega_t)$ provided by the seller. The buyer uses this signal along with the prior λ to formulate a Bayesian posterior distribution on the product quality $\mu_t(\omega|\sigma_t) \propto \phi_t(\sigma_t|\omega) \cdot \lambda(\omega)$. The buyer then purchases the product if and only if the expected valuation under this posterior $\mathbb{E}_{\omega \sim \mu_t(\cdot|\sigma_t)}[v(\theta_t, \omega)]$ is greater than or equal to the price p_t.[3] We denote the buyer decision at time t by $a_t \in \{0,1\}$ with $a_t = 1$ denoting purchase.

A summary of the game timeline is as follows: at $t \in [T]$,

[2] Our results can be generalized to the setting where Θ is any compact interval $[\underline{\theta}, \bar{\theta}] \subseteq \mathbb{R}^+$, or unbounded (see Remark 1).

[3] We consider the setting with Bayesian rational receivers who use only the prior distribution and signal in the current round, and not the signals in the past rounds, to make their decisions. This is motivated by the fact that at each round, the buyer is facing a fresh product, whose quality is drawn independently across time. Thus, previous signals do not carry information about the current product. Fresh products with i.i.d. qualities are common in many real-world applications such as second-hand markets, mystery boxes sold by Amazon/Woot, etc. This similar problem structure has also been studied in the online/sequential Bayesian persuasion literature with repeated interactions of sender and receivers, for example, see Zu et al. (2021), Castiglioni et al. (2021, 2020), Feng et al. (2022), Wu et al. (2022).

1. the seller commits to a price $p_t \in [0, U]$ and an advertising strategy ϕ_t;
2. a buyer t with private type $\theta_t \sim F$ arrives;
3. a product with quality $\omega_t \sim \lambda$ is realized; the seller sends signal $\sigma_t \sim \phi_t(\cdot|\omega_t)$ to the buyer;
4. the buyer formulates Bayesian posterior $\mu_t(\omega|\sigma_t)$, and the buyer purchases the product (denoted by $a_t = 1$) to generate revenue p_t if and only if her expected value exceeds the price, i.e., $\mathbb{E}_{\omega\sim\mu_t(\cdot|\sigma_t)}[v(\theta_t, \omega)] \geq p_t$.

Our goal is to design an online learning algorithm that sequentially chooses the price and advertising strategy p_t, ϕ_t in each round t based on the buyers' responses in the previous rounds, in order to optimize total expected revenue over a time horizon T without apriori knowledge of the distribution F. Let $\mathsf{Rev}(p_t, \phi_t) \triangleq \mathbb{E}[p_t \cdot a_t; p_t, \phi_t]$ denote the expected revenue at time t under the price p_t and advertising strategy ϕ_t. Here the expectation is over realizations of customer type $\theta_t \sim F$, product quality $\omega_t \sim \lambda$ and advertising signal $\sigma_t \sim \phi_t(\cdot|\omega_t)$. Note that since product quality and types are i.i.d. across time, for any given $p_t = p, \phi_t = \phi$, the expected per-round revenue $\mathsf{Rev}(p, \phi)$ does not depend on time, and a static price and advertising strategy maximizes total expected revenue over the time horizon T. Therefore, we can measure the performance of an algorithm in terms of *regret* that compares the total expected revenue of the algorithm to that of the best static pricing and advertising strategy. In particular, We define the following regret measure:

$$\text{Regret}(T) \triangleq T \max_{p,\phi} \mathsf{Rev}(p, \phi) - \sum_{t=1}^{T} \mathsf{Rev}(p_t, \phi_t) \ .$$

1.2. Our Contributions

In this work, we present a computationally efficient online pricing and advertising algorithm that achieves an $O\left(T^{2/3}(m \log T)^{1/3}\right)$ bound on the regret in time T, where $m = |\Omega|$ is the cardinality of the (discrete) product quality space. Importantly, we achieve this result without any assumptions like Lipschitz or smoothness on the demand function $D(\cdot) = 1 - F(\cdot)$. However, our results require certain assumptions on the valuation function. Following the literature, we make the common assumption that the function $v(\theta, \omega)$ is linear in the product quality ω. Furthermore, we assume the following monotonicity and Lipschitz properties on the valuation function.

Assumption 1. *Buyer's valuation function $v(\cdot, \cdot)$ satisfies:*

1a *Fix any buyer type θ, function $v(\theta, \omega)$ is non-decreasing w.r.t. quality ω.*

1b *Fix any quality ω, function $v(\theta, \omega)$ is increasing and 1-Lipschitz[4] w.r.t. type θ.*

Such assumptions are in fact common in literature and natural in many economic situations where the valuation of a product increases with the product quality and buyer's type (paying ability/need).

[4] 1 Lipschitz constant here is for exposition simplicity, arbitrary Lipschitz constant can be treated similarly.

Existing literature on Bayesian persuasion/dynamic pricing often makes even stronger assumptions about the receiver's utility function. For example, both the additive functions $v(\theta,\omega)=\theta+\omega$ (cf. Ifrach et al. 2019, Kolotilin et al. 2017, Crapis et al. 2017) and the multiplicative functions $v(\theta,\omega)=\theta\omega+\theta$ (cf. Candogan and Strack 2021, Liu et al. 2021), are linear in ω and satisfy Assumption 1. Our main result is then summarized as follows:

THEOREM 1. *For any type CDF F, given a valuation function $v(\theta,\omega)$ that is linear in product quality ω and satisfies Assumption 1, Algorithm 1 with parameter $\varepsilon=\Theta((m\log T/T)^{1/3})$ has an expected regret of $O\left(T^{2/3}(m\log T)^{1/3}\right)$. Here, m is the cardinality of the discrete quality space Ω.*

Furthermore, we obtain several improved results for the widely considered special case of additive valuations, i.e., for $v(\theta,\omega)=\theta+\omega$. See Section 5 for the formal statements and analysis.

1. **(Theorem 2)** Consider discrete sets Ω that are 'equally-spaced', e.g., $\Omega=\{0,1\}$ or $\Omega=[m]$. Given such a product quality space Ω and additive valuation function, we show that the regret of Algorithm 1 is bounded by $O(T^{2/3}(\log T)^{1/3})$ when $m\le(T/\log T)^{1/3}$, and by $O(\sqrt{mT\log T})$ for larger m.
2. **(Theorem 3)** For any arbitrary (discrete or continuous) product quality space Ω, given additive valuation functions, we have a slightly modified algorithm (see Algorithm 3 in Section C.2) with an expected regret of $O(T^{3/4}(\log T)^{1/4})$ independent of m.

We might compare our results to the best regret bounds available for the well-studied dynamic pricing and learning problem with unlimited supply Kleinberg and Leighton (2003), Babaioff et al. (2015), which is a special case of our problem if the product quality is deterministic, i.e., $m=1$, and the advertising scheme reveals no information and thus has no impact on the buyer's purchase decision. For the dynamic pricing problem a lower bound of $\Omega(T^{2/3})$ on regret is known (Kleinberg and Leighton 2003). Therefore the dependence on T in our results cannot be improved. In fact, our result matches this lower bound in the case of binary or constant size quality space, which are common settings in information design literature (Kamenica and Gentzkow 2011, Bonatti et al. 2022, Au and Kawai 2020, Feng et al. 2022, Bergemann et al. 2022a).

High-level descriptions of the proposed algorithm and challenges. Our problem can be viewed as a very high dimensional combinatorial multi-armed bandit problem, where each arm is a pair of a price and a feasible advertising strategy: the set of feasible advertising strategies being the set of all possible conditional distributions $\{\phi(\cdot|\omega),\omega\in\Omega\}$ over signal space Σ. As a first step towards obtaining a more tractable setting, we present an equivalent reformulation of the problem which uses the observation that advertising affects the buyer's decision only via the posterior distribution over quality. By the linearity of valuation function $v(\cdot,\cdot)$ over product quality ω, seller's choice of

advertising strategy in every round can be further simplified to selecting a distribution over posterior means that is subject to a feasibility constraint.

From here, the seller's decision space now becomes two-dimensional (a price and a distribution of posterior means). Viewing seller's expected revenue as an unknown (nonlinear) function over this two-dimensional decision space, one may consider applying algorithms in contextual bandits or Lipschitz bandits to get sublinear regrets, e.g., $\tilde{O}(T^{3/4}\mathsf{poly}(m))$ regret for two-dimensional decision space if there exists a Lipschitz property of reward function relative to seller's decision space. However, it is unclear whether one can establish such Lipschitz property given that we do not assume Lipschitzness or smoothness on demand function and we have complex constraints on the feasibility of advertising space. Instead, in our algorithm we use a 'model-based approach': we use buyers' purchase responses to explore the demand function over the (discretized) type space and jointly learn the optimal advertising and pricing. To explore the demand function over the one-dimensional (discretized) type space, we propose a novel discretization scheme such that it enables near-optimal price and advertising strategy even without Lipschitzness or smoothness assumption and with the complex feasibility constraint on the advertising space. These treatments lead us to the optimal $\tilde{O}(T^{2/3}m^{1/3})$ regret.

2. Related work

Our work is related to several streams of research. Below we briefly review the some of these connections.

In our setting, the seller can utilize her information advantage to design an advertisement to signal the product quality to the buyer. There is a long line of research in the literature, from both empirical and theoretical perspective, dedicated to study how to use advertisement as a signal to steer buyers' evaluations of advertised goods (Nelson 1970, 1974, Kihlstrom and Riordan 1984, Milgrom and Roberts 1986, Judd and Riordan 1994, Sahni and Nair 2020, Kawai et al. 2022). In our problem, we follow the literature in information design, a.k.a., Bayesian persuasion (Kamenica and Gentzkow 2011) (also see the recent surveys by Dughmi 2017, Kamenica 2019, Bergemann and Morris 2019), where the seller can commit to an information policy that can strategically disclose product information to the buyer so that to influence buyer's belief about the product quality. Similar formulation for advertising has also appeared in Bro Miltersen and Sheffet (2012), Emek et al. (2014), Arieli and Babichenko (2019), Hwang et al. (2019), Bergemann et al. (2022c,b). Our work differs from the these works in several ways. First, the seller's offline problem in our setting is a joint pricing and advertising problem. Second, we focus on an online setting where the seller has no apriori knowledge of the demand function and has to use past buyers' purchase responses to adaptively learn optimal pricing and advertising strategy.

Our problem shares similarity to the problem on sale of information in economics and computer science literature (Babaioff et al. 2012, Bergemann and Bonatti 2015, Bergemann et al. 2018, Chen et al. 2020, Cai and Velegkas 2021, Liu et al. 2021, Bergemann et al. 2022a, Zheng and Chen 2021, Li 2022, Chen and Zhang 2020). In particular, similar to the problem on sale of information, the seller, in our setting, also commits to design an information structure to reveal information about the realized state to the decision maker (i.e., buyer); and the buyer, in our setting, then makes the payment based on the declared information structure, not for specific realizations of the seller's informative signals (Bergemann et al. 2022a, Cai and Velegkas 2021, Liu et al. 2021). However, different from these works, the seller in our setting is selling a product with some inherent value and not just information. The valuation of the product can be shaped by providing information. This gives new interesting tradeoffs between information and revenue in our problem that are absent in the settings where only information is being sold. Moreover, in most literature of selling information, the buyers' type distribution is usually assumed to be known. We consider a more practical data-driven setting where the underlying buyers' type distribution (demand function) is apriori unknown to the seller and needs to be learnt from observations.

Facing unknown buyer's preference (i.e., buyer's private type), the seller's dynamic advertising problem also relates to the growing line of work in information design on relaxing one fundamental assumption in the canonical Bayesian persuasion model – the sender perfectly knows receiver's preference. The present paper joins the recent increased interests on using online learning approach to study the regret minimization when the sender repeatedly interacts with receivers (Castiglioni et al. 2020, 2021, Zu et al. 2021, Feng et al. 2022) without knowing receivers' preferences. Moreover, our work also conceptually relates to research on Bayesian exploration in multi-armed bandit (Kremer et al. 2014, Mansour et al. 2022, 2015, Immorlica et al. 2020) which also studies an online setting where one player can utilize her information advantage to persuade another player to take the desired action. Our work departs from this line of work in terms of both the setting and the application domain. Particularly, the above works typically consider an online setting on how to learn optimal signaling scheme whereas in our setting the optimal policy is a joint pricing and signaling (advertising) scheme.

When there is no uncertainty in the product quality, the seller's problem in our setting reduces to a standard dynamic pricing and learning problem with unknown non-parametric demand function (Kleinberg and Leighton 2003, Besbes and Zeevi 2009, Keskin and Zeevi 2014, Babaioff et al. 2015). However, given any non-trivial product quality space and prior distributions, in our problem, in addition to a price, the seller needs to choose a non-trivial advertising strategy in order to maximize revenue. This makes the seller's decision space high dimensional and (as we discuss in the next section) introduces significant complexities and difficulties so that the typical techniques (like uniform or

adaptive discretization) used in pricing and continuous/combinatorial multi-armed bandit literature cannot be directly applied.

3. Algorithm Design

In this section, we present our main algorithm for the dynamic pricing and advertising problem. In subsection 3.1, we present an equivalent reformulation for tractable advertising strategies, then in subsection 3.2, we discuss many important challenges even after this simplification, and finally, in subsection 3.3 we present our algorithm.

3.1. An equivalent reformulation for tractable advertising strategies

Recall that in every round, the buyer t sees the offered price p_t and advertising strategy ϕ_t that specifies the distributions over signals $\phi_t(\cdot|\omega) \in \Delta^{\Sigma}$ that the seller will send for each possible value ω of the realized product quality. After the product quality ω_t is realized, the buyer t sees a signal $\sigma_t \sim \phi_t(\cdot|\omega_t)$ from the seller's declared advertising strategy. The buyer uses this signal along with the prior λ to form a Bayesian posterior $\mu_t(\cdot|\sigma_t) \in \Delta^{\Omega}$ on the product quality. The Bayesian rational buyer then takes the action $a_t \in \{0, 1\}$, based on expected utility maximization. In particular, we have

$$a_t = \begin{cases} 1 \text{ if } \mathbb{E}_{\omega \sim \mu_t(\cdot|\sigma_t)}[v(\theta_t, \omega)] \geq p_t \\ 0 \quad \text{otherwise} \end{cases}$$

From the decision formula above, it is clear that the choice of advertising strategy affects the buyer t's decision only through the realized posterior $\mu_t(\cdot|\sigma_t)$. Consequently, the seller's choice of advertising strategy in time t can be reduced to selecting a distribution over posteriors μ_t. Seller's choice can in fact be further simplified in the case where the buyer's valuation function is linear in the product quality ω since in that case we have

$$\mathbb{E}_{\omega \sim \mu_t(\cdot|\sigma_t)}[v(\theta_t, \omega)] = v(\theta_t, \mathbb{E}_{\omega \sim \mu_t(\cdot|\sigma_t)}[\omega]) = v(\theta_t, q_t)$$

where q_t is the realized posterior mean $q_t \triangleq \mathbb{E}_{\omega \sim \mu_t(\cdot|\sigma_t)}[\omega]$. Here $q_t \in [0, 1]$ since $\omega \in \Omega \subseteq [0, 1]$. Therefore the buyer purchases $(a_t = 1)$ if and only if $v(\theta_t, q_t) \geq p_t$. As a result, we can reduce the seller's advertising in round t to the choice of distribution (pdf) $\rho_t(\cdot) \in \Delta^{[0,1]}$ over posterior means.

However, the choice of ρ_t must be restricted to only feasible distributions of posterior means, that is, all possible distributions over posterior means that can be induced by any advertising scheme given the prior λ. It is well known that the distribution over posterior means ρ is feasible if and only if it is the mean-preserving contraction of the prior (Blackwell and Girshick 1979, Aumann et al. 1995). This condition can be equivalently written in terms of a Bayes-consistency condition (Kamenica and Gentzkow 2011) on the conditional means $\rho(\cdot|\omega), \omega \in \Omega$. For simplicity of exposition, we consider

discrete quality space $\Omega = \{\bar{\omega}_1, \ldots, \bar{\omega}_m\} \subseteq [0,1]$, where $0 = \bar{\omega}_1 < \bar{\omega}_2 < \ldots < \bar{\omega}_m = 1$ and $m = |\Omega|$ is the cardinality of the quality space. Then, a distribution ρ over posterior means is feasible if one can construct a set of conditional distributions $(\rho_i)_{i\in[m]}$ satisfying the following Bayes-consistency condition, and vice versa (Kamenica and Gentzkow 2011):

$$\frac{\sum_{i\in[m]} \lambda_i \rho_i(q)\bar{\omega}_i}{\sum_{i\in[m]} \lambda_i \rho_i(q)} = q, \quad \forall q \in \mathsf{supp}(\rho)\ . \tag{BC}$$

Throughout this paper, we use the collection of distributions $(\rho_i)_{i\in[m]}$ satisfying (BC) condition as a convenient way to construct feasible distributions over posterior means: $\rho(q) = \sum_i \rho_i(q)\lambda_i$.

With the above observations, we can without loss of generality assume that seller's advertising strategy is to directly choose a distribution ρ_t over the posterior means that satisfies (BC), without considering the design of the underlying signaling scheme $\{\phi(\sigma|\omega), \Sigma\}$.

We summarize the new equivalent game timeline as follows: at $t \in [T]$,

1. the seller commits to a price $p_t \in [0, U]$ and an advertising strategy $\rho_t = (\rho_{i,t})_{i\in[m]}$ satisfying (BC);
2. a buyer t with private type $\theta_t \sim F$ arrives;
3. a product with quality $\omega_t \sim \lambda$ is realized; a posterior mean $q_t \sim \rho_t$ is realized;
4. the buyer observes the posterior mean q_t; the buyer purchases the product ($a_t = 1$) to generate revenue p_t if only if $v(\theta_t, q_t) \geq p_t$.

Note that the seller knows the form of the buyer's valuation function v. Moreover, the seller observes the realized product quality ω_t, the realized posterior mean q_t, and the buyer's purchase decision a_t, but does not know type CDF F (i.e., the demand function D) and the realized buyer type θ_t.

Revenue and regret. Given the new formulation, we can also rewrite the revenue and regret in terms of the choices of $\rho_t, t = 1, \ldots, T$. We define the following function $\kappa(p, q)$, which we refer to as the *critical type* for a given price p and posterior mean q.

Definition 1 (Critical type). For any $p \in [0, U]$ and $q \in [0, 1]$, define function $\kappa(\cdot, \cdot)$ as $\kappa(p, q) \triangleq \min\{\theta \in \Theta : v(\theta, q) \geq p\}$.

Now under Assumption **1b**, due to the monotonicity of the valuation function in buyer's type, we have that given any p, q, θ, $\mathbf{1}[v(\theta, q) \geq p] = \mathbf{1}[\theta \geq \kappa(p, q)]$. Therefore, the buyer t will purchase the product if and only if $\theta_t \geq \kappa(p_t, q_t)$.

Therefore, given the price, advertising $p_t = p, \rho_t = \rho$ and prior distribution λ, the expected revenue in any round t is given by [5]

[5] Here, we slightly abuse the notation to redefine Rev as a function of price and ρ, instead of price and ϕ defined earlier.

$$\mathsf{Rev}(p,\rho) = \mathbb{E}_{\omega\sim\lambda,\theta\sim F,q\sim\rho}[p\cdot\mathbf{1}[\theta\geq\kappa(p,q)]] = p\sum_i \lambda_i \int_0^1 \rho_i(q) D(\kappa(p,q))dq \tag{1}$$

Let the seller's online policy offer price p_t and advertising ρ_t at time t, where p_t, ρ_t can depend on the history of observations/events up to time t. Then expected regret defined in Section 1 can be equivalently written as

$$\text{Regret}[T] = T\mathsf{Rev}(p^*,\rho^*) - \sum_{t=1}^{T}\mathbb{E}[\mathsf{Rev}(p_t,\rho_t)] \ .$$

Here, the expectation is taken with respect to any randomness in the algorithm's choice of p_t, ρ_t; and $p^*, \rho^* = (\rho_i^*)_{i\in[m]}$ are defined as the best price and advertising strategy for a given F (and $\kappa(\cdot,\cdot)$ which is determined by F). Given the expression for $\mathsf{Rev}(p,\rho)$ derived above, these can be characterized by the following optimization program

$$\begin{aligned} p^*, \rho^* \triangleq \arg\max_{p,\rho} \quad & \mathsf{Rev}(p,\rho) \\ \text{s.t.} \quad & \frac{\sum_{i\in[m]}\lambda_i\rho_i(q)\bar{\omega}_i}{\sum_{i\in[m]}\lambda_i\rho_i(q)} = q, \quad q\in[0,1]; \quad \rho_i\in\Delta^{[0,1]}, \quad i\in[m] \ . \end{aligned} \tag{$\mathsf{P}_{\mathsf{OPT}}$}$$

where the first constraint is due to (BC).

3.2. Algorithm design: challenges and ideas

Challenge: high-dimensional continuous decision space. In the last section, we obtained a considerable simplification of the problem by reducing the seller's advertising strategy in every round t to a *distribution* $\rho_t \in \Delta^{[0,1]}$ over posterior means satisfying the (BC) condition. However, the decision space (a.k.a space of arms) still remains high dimensional and therefore a naive application of (uniform or adaptive) discretization-based bandit techniques, e.g. from Lipschitz bandit literature (Kleinberg et al. 2008, Slivkins 2011), would not achieve the desired results.

Algorithm design idea: exploring over one-dimesnional type space. Our algorithm uses a 'model-based approach' instead, where we use buyer purchase responses to develop (upper confidence bound) estimates of the demand model $D(\theta) = 1 - F(\theta)$ on the points of a *discretized type space* $\mathcal{S} \subseteq \Theta$. We then use these upper confidence bounds to construct a piecewise-constant demand function that is an upper confidence bound (UCB) for the demand function D. Then, in each round we solve for the optimal price and advertising strategy by solving an optimization problem similar to $\mathsf{P}_{\mathsf{OPT}}$, but with the UCB demand function.

Challenge: efficient discretization of type space. The challenge then is to design a discretization scheme for the type space such that we have a) efficient learning, i.e., the discretized space can be efficiently explored to accurately estimate the demand function on those points, and b) Lipschitz

property, i.e., the optimal revenue with the UCB estimate of demand function is close to the true optimal revenue as long as the estimation error on the discretized space is small.

There are two main difficulties in achieving this:

1. Lack of any smoothness/Lipschitz assumption on CDF F.
2. Sensitivity of the Bayes-consistency condition (BC).

To see these difficulties, recall that given a price p_t and realized posterior mean q_t, the t^{th} buyer's purchase decision is given by $a_t = \mathbf{1}[\theta_t \geq \kappa(p_t, q_t)]$; thus the seller can obtain demand function estimate at point $\kappa(p_t, q_t)$. Without any smoothness or Lipschitz property of demand function, estimates of demand function cannot be extrapolated accurately to other points. This means that in our revenue optimization problem (estimated version of $\mathsf{P_{OPT}}$), we need to solve to find a price and advertising strategy that we can only use estimates of demand function on a discretized type space, say $\mathcal{S} \subseteq \Theta$. However, if we restrict to a discretized type space $\mathcal{S}$, then the support of posterior mean distributions (a.k.a advertising strategy) must be restricted to the points q such that the corresponding critical types $\kappa(p, q)$ are in the set $\mathcal{S}$.

Unfortunately, the (BC) condition makes the set of feasible advertising strategies very sensitive to their support. In particular, if we use uniform-grid based discretization (which is commonly used in previous dynamic pricing literature such as Kleinberg and Leighton (2003), Babaioff et al. (2015)), it is easy to construct examples of prior distribution and valuation function such that there are no or very few feasible advertising strategies with the corresponding restricted support.

EXAMPLE 1. Consider additive valuation function, i.e., $v(\theta, \omega) = \theta + \omega$, and thus $\kappa(p, q) = p - q$. Consider quality space $\Omega = \{\bar{\omega}_i\}_{i \in [3]}$. Given a small ε, consider a uniform-grid based discretization for the type space $\mathcal{S}$, i.e., $\mathcal{S} = \{0, \varepsilon, 2\varepsilon, \ldots, 1\}$. If we also use a price p that is from uniform-grid based discretized price space, i.e., $p = k\varepsilon$ for some $k \in N^+$, then to ensure $\kappa(p, q) \in \mathcal{S}$, the support of advertising strategy (i.e., the distribution of the posterior means) must also be restricted within the set $\mathcal{S}$. However, if the prior distribution λ has negligible probabilities on qualities $\bar{\omega}_1, \bar{\omega}_3$, and quality $\bar{\omega}_2 \notin \mathcal{S}$, then we cannot construct any posterior distribution with the mean in the set $\mathcal{S}$. Therefore, there does not exist any feasible advertising strategy.

Note that this difficulty cannot be fixed by simply modifying the discretized type space to $\mathcal{S} \cup \Omega$, because even then we would need to construct an advertising such that $\kappa(p, q) = p - q \in \mathcal{S} \cup \Omega$ in the grid for all p. That would need that the support of the strategy (i.e., posterior means q) is restricted to be in $\{k\epsilon - \bar{\omega}_i\}_{k \in \mathbb{N}^+, i \in [3]}$; such posterior means again may not be achieved here.

Algorithm design idea: novel discretization scheme. Our algorithm employs a carefully-designed *quality-and-price-dependent discretization* scheme. The above example shows that we cannot uniformly discretize price and type using ε-grids, as we may not have any feasible advertising strategy

under such discretization. And furthermore, it also shows that this difficulty cannot be fixed by simply adding the m points in quality space to the discretized type space $\mathcal{S}$. Instead, in our discretization scheme, we first uniformly discretize the price space to an ε-grid $\mathcal{P}$. Then to construct a discretized type space $\mathcal{S}$, in addition to the points on an ε-grid over $[0,1]$, we also include points $\{\kappa(p,\omega)\}$ for every price $p \in \mathcal{P}$ and quality $\omega \in \Omega$. This gives us a discretized type space $\mathcal{S}$ of size mU/ε. We claim that our construction ensures that there exist near-optimal price and advertising strategy with support in the discretized type space $\mathcal{S}$. The proof of this claim requires a careful rounding argument, which forms one of the main technical ingredients for our regret analysis in Section 4.

3.3. Details of the proposed algorithm

Our dynamic pricing and advertising algorithm jointly discretizes the price space and type space using the following *quality-and-price-dependent discretiztion* scheme: given parameter ε, we define the following set:

$$\begin{aligned} \mathcal{P} &\triangleq \{\varepsilon, 2\varepsilon, \ldots, U\} \\ \mathcal{S} &\triangleq \{0, \varepsilon, \ldots, 1-\varepsilon, 1\} \cup \{(\kappa(p,\omega) \wedge 1) \vee 0\}_{p\in\mathcal{P},\omega\in\Omega} \ . \end{aligned} \tag{2}$$

At time t, we restrict the price and advertising strategies (p_t, ρ_t) to the set of $(p, \rho = (\rho_i)_{i\in[m]})$ such that $p \in \mathcal{P}$, and given price p, each conditional distribution ρ_i has restricted support $\mathcal{Q}_p$ defined as

$$\mathcal{Q}_p \triangleq \{q : \kappa(p,q) \in \mathcal{S}\} \ .$$

Given the price and advertising p_t, ρ_t, let the realized posterior mean at time t be $q_t \sim \rho_t$, and let the corresponding critical type be $x_t \triangleq \kappa(p_t, q_t) \in [0,1]$. Then, note that the above restrictions on price and advertising strategies guarantee that $x_t \in \mathcal{S}$.

Next, to compute the offered price and advertising strategy in round t, we optimize an upper confidence bound on the revenue function that we develop using upper confidence bounds $D^{\mathsf{UCB}}(x), x \in \mathcal{S}$ of the demand function computed as follows. For every type $x \in \mathcal{S}$, let $\mathcal{N}_t(x)$ denote the set of time rounds before time t that the induced critical type is exactly x, and let $N_t(x)$ be the number of such time rounds. That is,

$$\mathcal{N}_t(x) \triangleq \{\tau < t : \kappa(p_\tau, q_\tau) = x\}, N_t(x) \triangleq |\mathcal{N}_t(x)|, x \in \mathcal{S} \ .$$

Recall that buyer's purchase decision follows $a_\tau = \mathbf{1}[\theta_\tau \geq \kappa(p_\tau, q_\tau)]$. We estimate the demand function at x as:

$$\bar{D}_t(x) \triangleq \frac{\sum_{\tau\in\mathcal{N}_t(x)} a_\tau}{N_t(x)} \ .$$

We can now define the following UCB index:

$$D_t^{\mathsf{UCB}}(x) = \min_{x'\in\mathcal{S}:x'\leq x} \bar{D}_t(x') + \sqrt{\frac{16\log T}{N_t(x')}} + \frac{\sqrt{(1+N_t(x'))\ln(1+N_t(x'))}}{N_t(x')} \wedge 1, \quad x \in \mathcal{S} \ . \tag{3}$$

Then, for any pair of discretized price $p \in \mathcal{P}$ and advertising strategy with discretized support for that price $\rho = (\rho_i \in \Delta^{\mathcal{Q}_p}, i \in [m]\}$, we define the following seller's revenue estimates:

$$\mathsf{Rev}_t^{\mathsf{UCB}}(p,\rho) \triangleq p \sum_{i\in[m]} \lambda_i \int_0^1 \rho_i(q) D_t^{\mathsf{UCB}}(\kappa(p,q))dq \ .$$

Above is well-defined since by definition $\kappa(p,q) \in \mathcal{S}$ for each such $(p,q) \in \mathcal{P} \times \mathcal{Q}_p$. Finally, we let p_t, ρ_t be the optimal solution to the following optimization problem:

$$\begin{aligned}(p_t,\rho_t) = \arg\max_{p,\rho} \ & \mathsf{Rev}_t^{\mathsf{UCB}}(p,\rho) \\ \text{s.t.} \quad & p \in \mathcal{P}; \quad \frac{\sum_{i\in[m]} \lambda_i \rho_i(q)\bar{\omega}_i}{\sum_{i\in[m]} \lambda_i \rho_i(q)} = q, \quad q \in \mathcal{Q}_p; \quad \rho_i \in \Delta^{\mathcal{Q}_p}, \quad i \in [m] \ .\end{aligned} \tag{$\mathsf{P}_t^{\mathsf{UCB}}$}$$

Algorithm 1: Algorithm for Dynamic Pricing and Advertising with Demand Learning.

Input: Discretization parameter ε.

For the first $|\mathcal{S}|$ rounds, for each $x \in \mathcal{S}$, offer a price p with any no information advertising s.t. $\kappa(p, \mathbb{E}_{\omega\sim\lambda}[\omega]) = x$. `// No information advertising provides completely uninformative signal - the distribution` $\phi(\cdot|\omega)$ `of signals does not depend on the realized quality` ω

for *each round* $t = |\mathcal{S}|+1, |\mathcal{S}|+2, \ldots, T$ **do**

  For all $x \in \mathcal{S}$, compute $D_t^{\mathsf{UCB}}(x)$ as defined in (3).

  Offer the price p_t and an advertising ρ_t computed as an optimal solution to program $\mathsf{P}_t^{\mathsf{UCB}}$.

  `/*` p_t, ρ_t `satisfies` $\kappa(p_t, q) \in \mathcal{S}$ `for every` $q \in \mathsf{supp}(\rho_t)$. `*/`

  Observe realized posterior mean $q_t \sim \rho_t$ and buyer's purchase decision $a_t \in \{0,1\}$.

  Update $\left\{\mathcal{N}_{t+1}(x), N_{t+1}(x), \bar{D}_{t+1}(x)\right\}_{x\in\mathcal{S}}$.

We summarize our algorithm as Algorithm 1.

The main computational bottleneck of Algorithm 1 is to solve the high-dimensional program $\mathsf{P}_t^{\mathsf{UCB}}$ at each time $t \geq |\mathcal{S}|+1$. As we illustrate in Proposition 1, there exists an efficient method to optimally solve this program. The proof of this result utilizes the monotoncity of the function D_t^{UCB}.

PROPOSITION 1 **(Adopted from Arieli et al. (2020), Candogan (2019))**. *Let ε be the discretization parameter for the set $\mathcal{P}$ defined in* (2)*. There exists a polynomial time (in $|\mathcal{S}|U/\varepsilon$) algorithm that can solve the program $\mathsf{P}_t^{\mathsf{UCB}}$.*

The proof of the above result utilizes the monotoncity of the function D_t^{UCB}, and is deferred to Section A.

We conclude this section with the following remark on the extension to unbounded type support.

REMARK 1. Our algorithm and analysis can be extended to the case with unbounded type support (e.g., $\Theta = [0, \infty)$). In particular, since the price is bounded by $[0, U]$, and the quality is bounded within $[0, 1]$, by the monotoncity of the valuation function, we know the critical types $\kappa(p, q)$ induced by any possible $p \in [0, U]$ and $q \in [0, 1]$ is bounded within $[\kappa(0, 1), \kappa(U, 0)]$. Thus, an instance with unbounded type support is equivalent to an instance with bounded type support $[\kappa(0, 1), \kappa(U, 0)]$.

4. Regret Analysis: Proof Overview of Theorem 1

In this section, we present our main regret bound for Algorithm 1, as stated in Theorem 1. Specifically, we show that for any type CDF F, given a valuation function $v(\theta, \omega)$ that is linear in product quality ω and satisfies Assumption 1, our algorithm (Algorithm 1 with parameter $\varepsilon = \Theta((m \log T/T)^{1/3})$), has an expected regret of $O\left(T^{2/3}(m \log T)^{1/3}\right)$. Importantly, for this result we do not assume any smoothness or Lipschitz properties of distribution F.

For this result, we consider arbitrary but discrete quality space Ω of cardinality m. Later in Section 5, we show improved regret bounds for the case of additive valuations and equally-spaced quality space (see Theorem 2), and also extend to arbitrary large and continuous quality spaces (see Theorem 3).

4.1. Proof Outline

Recall that in every round t, Algorithm 1 sets the price p_t and advertising strategy ρ_t as an optimal solution of program $\mathsf{P}_t^{\mathsf{UCB}}$ that approximates the benchmark $\mathsf{P}_{\mathsf{OPT}}$ in two ways. Firstly, it restricts the price and support of advertising strategy to be in a discretized space $\mathcal{P} \times \{\mathcal{Q}_p, p \in \mathcal{P}\}$. Secondly, it approximates the true demand function with an upper bound D_t^{UCB}. Our proof consists of two main steps that bound the errors due to each of the above approximations. Due to the space limit, all missing proofs in this section are deferred to Section B.

- **Step 1: bounding the discretization error using a rounding argument (see Section 4.2).** To separate the discretization error from the error due to demand function estimation, we consider an intermediate optimization problem $\widetilde{\mathsf{P}}$ (in Section 4.2) obtained on replacing the UCB demand function D_t^{UCB} with the true demand function D (while keeping the discretized space for p, ρ). Let $\widetilde{p}^*, \widetilde{\rho}^*$ be an optimal solution of program $\widetilde{\mathsf{P}}$. We show that the revenue $\mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*)$ is sufficiently close (within 2ε) to the optimal revenue $\mathsf{Rev}(p^*, \rho^*)$. This bound is obtained using a careful *rounding* argument: we show that the optimal price p^* and the optimal advertising ρ^* can be rounded to a new price $p^\dagger$ and a new advertising $\rho^\dagger$ that satisfy
  (i) feasibility (**Lemma 3**): $p^\dagger \in \mathcal{P}, \mathsf{supp}(\rho^\dagger) \subseteq \mathsf{supp}(\mathcal{Q}_{p^\dagger})$; and
  (ii) revenue guarantee (**Lemma 4**): $\mathsf{Rev}(p^\dagger, \rho^\dagger) \geq \mathsf{Rev}(p^*, \rho^*) - 2\epsilon$.

- **Step 2: bounding estimation error and establishing optimism (see Section 4.3).** Next, we show that the UCB estimates of the demand function $D_t^{\mathsf{UCB}}(x), x \in \mathcal{S}$ converge to the true demand function D with high probability, along with concentration bounds on the gap between the true and estimated function (**Lemma 5**). This allows us to show that

  (i) Revenue optimism (**Lemma 6**): we show that the algorithm's revenue estimates are (almost) optimistic, i.e.,

$$\mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t) \geq \mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \geq \mathsf{Rev}(p^*, \rho^*) - 2\varepsilon$$

  (ii) Revenue approximation (**Lemma 7**): we show that the optimistic revenue estimates are close to the true revenue in round t, with the gap between the two being inversely proportional to the number of observations, in particular,

$$\mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t) - \mathsf{Rev}(p_t, \rho_t) \leq 5p_t \mathbb{E}_{q\sim\rho_t}\left[\sqrt{\log T / N_t(\kappa(p_t, q))}\right]$$

Finally, in Section 4.4 we put it all together to bound the regret as stated in Theorem 1. This essentially involves using the above observations to show that regret over each round can be roughly bounded as $2\epsilon T + 5p_t \mathbb{E}_{q\sim\rho_t}\left[\sqrt{\log T / N_t(\kappa(p_t, q))}\right]$. Then, using the constraint that $\sum_{x\in\mathcal{S}} N_T(x) \leq T$, we show that in the worst case, total regret over time T is bounded by $O(T\epsilon + \sqrt{|\mathcal{S}| T \log T})$. The theorem statement is then obtained by substituting $|\mathcal{S}| = O(m/\epsilon)$ and optimizing the parameter ε.

4.2. A rounding procedure to bound the discretization error

In this subsection, we bound the loss in revenue due to discretization. Specifically, let $\widetilde{p}^*, \widetilde{\rho}^*$ be the solution of the following program:

$$\begin{aligned} \max_{p\in\mathcal{P}} \max_{\rho} \quad & p \sum_{i\in[m]} \lambda_i \int_0^1 \rho_i(q) D(\kappa(p,q)) dq \\ \text{s.t.} \quad & \frac{\sum_{i\in[m]} \lambda_i \rho_i(q) \bar{\omega}_i}{\sum_{i\in[m]} \lambda_i \rho_i(q)} = q, && q \in \mathcal{Q}_p \\ & \rho_i \in \Delta^{\mathcal{Q}_p}, && i \in [m] \ . \end{aligned} \tag{$\widetilde{\mathsf{P}}$}$$

The main result of this subsection is then summarized as follows:

Proposition 2. *For any type CDF F, we have* $\mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \leq 2\varepsilon$.

We prove the above result by showing that we can use a rounding procedure (see Procedure 2) to round the optimal price p^* and the optimal advertising ρ^* to a new price $p^\dagger$ and a new advertising $\rho^\dagger$ that satisfy: (i) $p^\dagger \in \mathcal{P}, \mathsf{supp}(\rho^\dagger) \subseteq \mathcal{Q}_{p^\dagger}$; and (ii) the revenue loss $\mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(p^\dagger, \rho^\dagger) \leq 2\varepsilon$. It is worth noting that Procedure 2, which we believe is of independent of interest, works for any buyer valuation function that is linear in quality and satisfies Assumption 1. And it only uses the knowledge

of critical-type function $\kappa(\cdot,\cdot)$ and prior distribution λ. In particular, Procedure 2 does not depend on any knowledge or estimates about the unknown demand function. Indeed, Proposition 2 still holds if we replace the demand function D in the revenue formulation (1) with any monotone non-increasing function. A graphic illustration of Procedure 2 is provided in Figure 1.

In this subsection, we provide details of our rounding procedure and its graphical illustration.

Procedure 2: $\mathsf{Rounding}(p,\rho)$: A critical-type guided procedure to round the strategy p,ρ

Input: ε, a price p such that $p\geq 2\varepsilon$, and an advertising ρ such that p,ρ satisfy Lemma 1 and Lemma 2.

Output: A price $p^\dagger\in\mathcal{P}$, an advertising $\rho^\dagger$ satisfy $\mathsf{supp}(\rho^\dagger)\subseteq\mathcal{Q}_{p^\dagger}$

Initialization: Let the set $\mathcal{Q}\leftarrow\emptyset$. // The set $\mathcal{Q}$ contains the support of the advertising $\rho^\dagger$.

Define price $p^\dagger\leftarrow\max\{p'\in\mathcal{P}: p-2\varepsilon\leq p'\leq p-\varepsilon\}$.

for *each posterior mean* $q\in\mathsf{supp}(\rho)$ **do**

  if $q\in\mathcal{Q}_{p^\dagger}$ **then** // Namely, for this case $\kappa(p^\dagger,q)\in\mathcal{S}$

    $\mathcal{Q}\leftarrow\mathcal{Q}\cup\{q\}$, and let $\rho^\dagger(q)=\rho(q)$, and let $\{i'\in[m]:\rho^\dagger_{i'}(q)>0\}=\{i'\in[m]:\rho_{i'}(q)>0\}$.

  else

    Suppose $\{i'\in[m]:\rho_{i'}(q)>0\}=\{i,j\}$ where $i<j$.

    Let $x\triangleq\kappa(p,q)$, and let $x^\dagger\triangleq\kappa(p^\dagger,q)\in((z-1)\varepsilon,z\varepsilon)$ for some $z\in\mathbb{N}^+$.

    Let q_L,q_R satisfy $\kappa(p^\dagger,q_L)=z\varepsilon$, $\kappa(p^\dagger,q_R)=(z-1)\varepsilon$.

    Let $q_L^\dagger\triangleq q_L\vee\bar{\omega}_i$, and let $q_R^\dagger\triangleq q_R\wedge\bar{\omega}_j$.

    $\mathcal{Q}\leftarrow\mathcal{Q}\cup\{q_L^\dagger,q_R^\dagger\}$.

    /* The conditional probabilities below are constructed to satisfy (BC). */

    Let $\rho_i^\dagger(q_L^\dagger)=\frac{\bar{\omega}_j-q_L^\dagger}{\bar{\omega}_j-\bar{\omega}_i}\frac{1}{\lambda_i}\frac{\rho(q)(q_R^\dagger-q)}{q_R^\dagger-q_L^\dagger}$ and $\rho_i^\dagger(q_R^\dagger)=\rho_i(q)-\rho_i^\dagger(q_L^\dagger)$; $\rho_j^\dagger(q_L^\dagger)=\frac{q_L^\dagger-\bar{\omega}_i}{\bar{\omega}_j-\bar{\omega}_i}\frac{1}{\lambda_j}\frac{\rho(q)(q_R^\dagger-q)}{q_R^\dagger-q_L^\dagger}$ and $\rho_j^\dagger(q_R^\dagger)=\rho_j(q)-\rho_j^\dagger(q_L^\dagger)$.

Details and guarantees of Procedure 2. Procedure 2 takes as input an input advertising strategy ρ that satisfies $|\{i'\in[m]:\rho_{i'}(q)>0\}|\leq 2$ for any posterior mean $q\in\mathsf{supp}(\rho)$. This structural requirement says that in advertising ρ, the realized signal either fully reveals the product quality, or randomizes buyer's uncertainty within two product qualities. Indeed, we can show that there exists an optimal advertising strategy for program $\mathsf{P_{OPT}}$ that satisfies this structural requirement:

LEMMA 1 **(see, e.g., Feng et al. (2022))**. *There exists an optimal advertising strategy ρ^* satisfying that $|\{i\in[m]:\rho_i^*(q)>0\}|\leq 2$ for every $q\in\mathsf{supp}(\rho^*)$.*

Intuitively, the above result is an implication of the fact that the extreme points of the distributions with fixed expectations are binary-supported distributions. Meanwhile, we can also deduce the following property for the optimal price p^* and optimal advertising ρ^*:

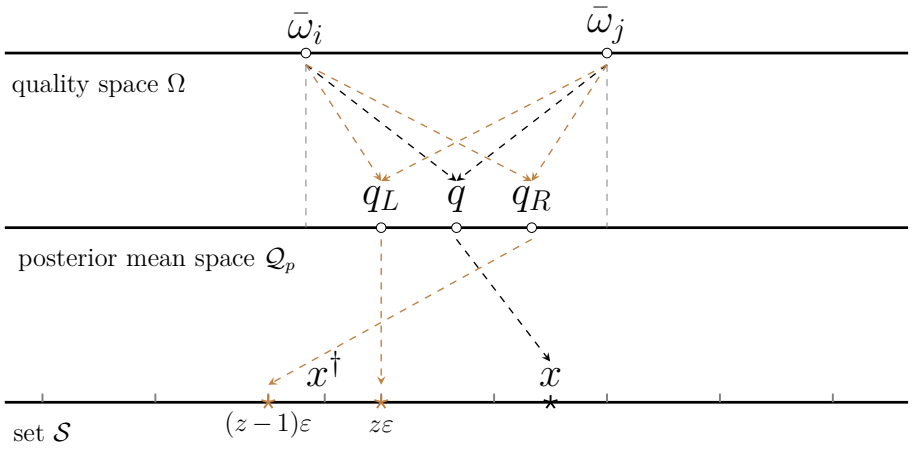

Figure 1 Graphical illustration for Procedure 2. Given the input price and advertising (p, ρ), fix a posterior mean $q \in \mathsf{supp}(\rho)$ where $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{i, j\}$ (drawn in black dashed line). According to the procedure, we first identify $x = \kappa(p, q)$, and $x^\dagger = \kappa(p^\dagger, q) \in ((z-1)\varepsilon, z\varepsilon)$ where the constructed price $p^\dagger$ is defined as in the procedure. We then find two posterior means q_L, q_R (here $q_L \geq \bar{\omega}_i, q_R \leq \bar{\omega}_j$) such that $\kappa(p^\dagger, q_L) = z\varepsilon$ and $\kappa(p^\dagger, q_R) = (z-1)\varepsilon$ (drawn in brown dashed line), and $\kappa(p^\dagger, q_R) < \kappa(p^\dagger, q_L) < \kappa(p, q)$.

LEMMA 2. *There exist an optimal price p^* and optimal advertising ρ^* such that for any posterior mean $q \in \mathsf{supp}(\rho^*)$ and $q \notin \Omega$, we have that $p^* \leq \max_{\theta \in \Theta} v(\theta, q)$.*

The above property follows from the observation that if there exists a posterior mean $q \in \mathsf{supp}(\rho^*)$ and $q \notin \Omega$, then from Lemma 1, it must be the case $\{i' \in [m] : \rho^*_{i'}(q) > 0\} = \{i, j\}$ for some $i < j$ such that $\bar{\omega}_i < q < \bar{\omega}_j$. Now, if $p^* > \max_{\theta \in \Theta} v(\theta, q)$, then for all types of buyers, the valuation at this posterior mean is below the given price p^* so that this posterior mean does not contribute to the revenue; therefore one can decompose the probability over this posterior mean $\rho^*(q)$ to probabilities over $\bar{\omega}_i, \bar{\omega}_j$ without losing any revenue and thus obtain a p^*, ρ^* with the desired property.

Proof of Lemma 2. Let us fix the optimal price p^* and optimal advertising ρ^*. Suppose there exists a posterior mean $q \in \mathsf{supp}(\rho^*)$ and $q \notin \Omega$, then from Lemma 1, it must be the case $\{i' \in [m] : \rho^*_{i'}(q) > 0\} = \{i, j\}$ for some $i < j$ that $\bar{\omega}_i < q < \bar{\omega}_j$. Suppose $p^* > \max_{\theta \in \Theta} \kappa(\theta, q)$, then it is easy to see that the revenue contributed from this posterior mean $p^* \sum_i \lambda_i \rho^*_i(q) D(\kappa(p^*, q)) = 0$. Thus, decoupling this posterior mean q to the states $\bar{\omega}_i$ and $\bar{\omega}_j$ will not lose any revenue. □

With the above Lemma 1 and Lemma 2, we now formally present two guarantees on the price and advertising strategy obtained from Procedure 2.

LEMMA 3 **(Feasibility guarantee)**. *Given an input price and advertising strategy p, ρ satisfying the properties stated in Lemma 1 and Lemma 2, the output price $p^\dagger$ and the advertising strategy $\rho^\dagger$ from Procedure 2 satisfies: $p^\dagger \in \mathcal{P}$, $\rho^\dagger$ is a feasible advertising and satisfies $\kappa(p^\dagger, q) \in \mathcal{S}$ for every $q \in \mathsf{supp}(\rho^\dagger)$.*

LEMMA 4 **(Revenue guarantee).** *Fix a price $p \geq 2\varepsilon$ and a feasible advertising strategy ρ, let $p^\dagger, \rho^\dagger = \mathsf{Rounding}(p, \rho)$ be the output from Procedure 2, then we have $\mathsf{Rev}(p, \rho) - \mathsf{Rev}(p^\dagger, \rho^\dagger) \leq 2\varepsilon$.*

The proofs of the above two lemmas are provided in Section B.1. With these two guarantees, we can now prove Proposition 2:

Proof of Proposition 2. Let $p^\dagger, \rho^\dagger = \mathsf{Rounding}(p^*, \rho^*)$, then we have $\mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(p^\dagger, \rho^\dagger) \leq 2\varepsilon$ where the first inequality follows from the feasibility guarantee of price $p^\dagger$ and advertising $\rho^\dagger$ in Lemma 3 and the definition of $\widetilde{p}^*, \widetilde{\rho}^*$, and the second inequality follows from revenue guarantee in Lemma 4. □

4.3. Estimation error and optimism

We begin our estimation error analysis by showing that $D_t^{\mathsf{UCB}}(x)$ provides an upper confidence bound on the true demand function $D(x)$ for all $x \in \mathcal{S}$, and deriving a bound on how large it can be compared to $D(x)$.

LEMMA 5. *For every $t \geq |\mathcal{S}| + 1$, the following holds with probability at least $1 - 1/T^2$:*

$$D_t^{\mathsf{UCB}}(x) \geq D(x), \qquad \forall x \in \mathcal{S}; \tag{4}$$

$$D_t^{\mathsf{UCB}}(x) - D(x) \leq 2\sqrt{\frac{16 \log T}{N_t(x)}} + \frac{2\sqrt{(1 + N_t(x)) \ln(1 + N_t(x))}}{N_t(x)}, \quad \forall x \in \mathcal{S}. \tag{5}$$

To prove the inequalities for the points $x \in \mathcal{S}$, we first show that the empirical estimates $\bar{D}_t(x), \forall x \in \mathcal{S}$ concentrate around the true demand value $D(x)$ as $N_t(x)$ increases. We prove this concentration bound by using a uniform bound given by scalar-valued version of self-normalized martingale tail inequality (Abbasi-Yadkori et al. 2012). The proof of the above lemma is provided in Section B.2.

We now analyze how close the seller's optimistic revenue estimates using the upper confidence bound D_t^{UCB} is to the true revenue. In particular, we have the following result.

LEMMA 6. *For every time $t \geq |\mathcal{S}| + 1$, with probability at least $1 - 2/T^2$, we have*

$$\mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}_t^{\mathsf{UCB}}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t)$$

The above results follow from the bounds in Lemma 5 where we established that $D_t^{\mathsf{UCB}}(x) \geq D(x)$ with high probability. The proof of the above Lemma 6 is provided in Section B.3.

Next we show that we can also upper bound $\mathsf{Rev}_t^{\mathsf{UCB}}(p_t, p_t) - \mathsf{Rev}(p_t, \rho_t)$ by applying the results in Lemma 5 again.

LEMMA 7. *For every time $t \geq |\mathcal{S}| + 1$, with probability at least $1 - 2/T^2$, we have*

$$\mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t) - \mathsf{Rev}(p_t, \rho_t) \leq 5p_t \sum_{q \in \mathsf{supp}(\rho_t)} \rho_t(q) \sqrt{\frac{\log T}{N_t(\kappa(p_t, q))}}$$

The proof of above Lemma 7 is provided in Section B.4. Intuitively, the difference between the estimated seller's revenue $\mathsf{Rev}_t^{\mathsf{UCB}}(p_t, p_t)$, and the true expected revenue $\mathsf{Rev}(p_t, \rho_t)$, can be bounded by a weighted sum (weighted by probabilities $\rho_t(q), q \in \mathsf{supp}(\rho_t)$) of errors in demand estimates on the points $\kappa(p_t, q)$ for $q \in \mathsf{supp}(\rho_t)$: $|D_t^{\mathsf{UCB}}(\kappa(p_t, q)) - D(\kappa(p_t, q))|$.

4.4. Putting it all together

We can now combine the above lemmas to prove Theorem 1.

Proof of Theorem 1. We have that with probability at least $1 - O(1/T)$,

$$\begin{aligned}
\mathrm{Regret}[T] &\le |\mathcal{S}| + \mathbb{E}\left[\sum_{t=|\mathcal{S}|+1}^{T} \mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(p_t, \rho_t)\right] \\
&\overset{(a)}{\le} |\mathcal{S}| + 2\varepsilon T + \mathbb{E}\left[\sum_{t=|\mathcal{S}|+1}^{T} \mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t) - \mathsf{Rev}(p_t, \rho_t)\right] \\
&\overset{(b)}{\le} |\mathcal{S}| + 2\varepsilon T + 5U\mathbb{E}\left[\sum_{t=|\mathcal{S}|+1}^{T} \sum_{q \in \mathsf{supp}(\rho_t)} \rho_t(q) \sqrt{\frac{\log T}{N_t(\kappa(p_t, q))}}\right] \\
&\overset{(c)}{=} |\mathcal{S}| + 2\varepsilon T + 5U\mathbb{E}\left[\sum_{x \in \mathcal{S}} \sum_{t=|\mathcal{S}|+1}^{T} \beta_t(x) \sqrt{\frac{\log T}{N_t(x)}}\right] , \qquad (6)
\end{aligned}$$

where the first inequality follows from the definition of regret. Inequality (a), follows from Lemma 6 and Proposition 2, and inequality (b) follows from Lemma 7 along with upper bound U on prices p_t in all rounds. For inequality (c), we use that by construction $\kappa(p_t, q) \in \mathcal{S}$ for every posterior mean $q \in \mathsf{supp}(\rho_t)$. and define distribution $\beta_t \in \Delta^{\mathcal{S}}$ over the set $\mathcal{S}$ as

$$\beta_t(x) = \sum_{q \in \mathsf{supp}(\rho_t): \kappa(p_t, q) = x} \rho_t(q), \quad x \in \mathcal{S} .$$

Define Bernoulli random variable $X_t(x) = \mathbf{1}[x_t = x]$, where $x_t = \kappa(p_t, q_t)$. Then, from the definition of $\beta_t(x)$ observe that $\mathbb{P}[X_t(x) = 1 | N_t(x)] = \beta_t(x)$. Also, by definition

$$N_{t+1}(x) = 1 + \sum_{\ell=|\mathcal{S}|+1}^{t} X_\ell(x) \le 2N_t(x) .$$

We use these observations below to obtain a bound on the third term in the RHS of (6):

$$\begin{aligned}
\mathbb{E}\left[\sum_{x \in \mathcal{S}} \sum_{t=|\mathcal{S}|+1}^{T} \beta_t(x) \sqrt{\frac{\log T}{N_t(x)}}\right] &= \mathbb{E}\left[\sum_{x \in \mathcal{S}} \sum_{t=|\mathcal{S}|+1}^{T} \mathbb{E}\left[X_t(x) \sqrt{\frac{\log T}{N_t(x)}} \,\middle|\, N_t(x)\right]\right] \\
&\le \mathbb{E}\left[\sum_{x \in \mathcal{S}} \sum_{t=|\mathcal{S}|+1}^{T} X_t(x) \sqrt{\frac{2\log T}{N_{t+1}(x)}}\right] \\
&= \mathbb{E}\left[\sum_{x \in \mathcal{S}} \sum_{n=2}^{N_{T+1}(x)} \sqrt{\frac{2\log T}{n}}\right]
\end{aligned}$$

$$\leq \mathbb{E}\left[\sum_{x\in\mathcal{S}} \sqrt{8N_{T+1}(x)\log(T)}\right]$$
$$\leq 2\sqrt{2|\mathcal{S}|T\log(T)}\ ,$$

Substituting this bound in (6), we obtain that with probability $1-O(1/T)$,

$$\text{Regret}[T] \leq |\mathcal{S}| + 2\varepsilon T + 10U\sqrt{2|\mathcal{S}|T\log(T)}\ .$$

Now, by construction, the set $\mathcal{S}$ has the cardinality of $O(mU/\varepsilon)$. Optimizing $\varepsilon = \Theta((m\log T/T)^{1/3})$ in the above regret bound, we have $\text{Regret}[T] \leq O\left(T^{2/3}(m\log T)^{1/3}\right)$. □

5. Improved Regret bounds For Additive Valuations

In this section, we discuss improved regret bounds for Algorithm 1 in the case when valuation function is additive, i.e., $v(\theta,\omega) = \theta+\omega$.

First we consider an additional assumption that the product quality domain Ω is an 'equally-spaced set', which include many natural discrete ordered sets like $\Omega = \{0,1\}$ or $\Omega = [m]$ that are commonly used in the Bayesian persuasion literature.

Definition 2 (Equally-spaced sets). We say that a discrete ordered set $\Omega = \{\bar{\omega}_1,\ldots,\bar{\omega}_m\}$ is equally-spaced if for all $i\in[m-1]$, $\bar{\omega}_{i+1}-\bar{\omega}_i = c$ for some constant c.

With this definition, we prove the following improved regret bound.

Theorem 2. *Given an additive valuation function, $v(\theta,\omega)=\theta+\omega$, and equally-spaced product quality domain, Ω Algorithm 1 with parameter $\varepsilon = \Theta((\log T/T)^{1/3}\wedge 1/m)$ has an expected regret of $O\left(T^{2/3}(\log T)^{1/3}+\sqrt{mT\log T}\right)$.*

Note that a corollary of the above theorem is that the regret is bounded by $O(T^{2/3}(\log T)^{1/3})$ when $m\leq (T/\log T)^{1/3}$ and by $O(\sqrt{mT\log T})$ for larger m. The high-level idea behind the above result is as follows. In the previous section (see Section 4.4) we show that the expected regret of Algorithm 1 is bounded by $O(T\varepsilon+\sqrt{|\mathcal{S}|T\log T})$. To prove Theorem 2 we show that in case of additive valuation and the equally-spaced qualities, there exists a discretization parameter $\varepsilon = \Theta((\log T/T)^{1/3}\wedge 1/m)$ such that $\{\kappa(p,\omega)\}_{p\in\mathcal{P},\omega\in\Omega}\subset\{0,\varepsilon,2\varepsilon,\ldots,1\}$. Thus, the constructed set $\mathcal{S}$ satisfies $|\mathcal{S}| = O(m+1/\varepsilon)$. Substituting the value of ε then gives the result in Theorem 2. A formal proof of Theorem 2 is provided in Section C.1.

Furthermore, for additive valuation functions, we can also handle arbitrary large or continuous quality spaces to obtain an $\tilde{O}(T^{3/4})$ regret independent of size of quality space m.

Theorem 3. *Given an additive valuation function $v(\theta,\omega)=\theta+\omega$, and arbitrary (discrete or continuous) product quality space Ω, there exists an algorithm (Algorithm 3 in Section C.2) that has expected regret of $O(T^{3/4}(\log T)^{1/4})$.*

The proposed Algorithm 3 that achieves the above result is essentially a combination of a pre-processing step and Algorithm 1. In this pre-processing step, we pool the product qualities that are "close enough". This gives us a new problem instance with a smaller discrete product quality space so that we can apply Algorithm 1. With additive valuation function, we show that this reduction does not incur too much loss in revenue. A formal description of the algorithm and proof of Theorem 3 is provided in Section C.2.

References

Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pages 1–9. PMLR, 2012.

Itai Arieli and Yakov Babichenko. Private bayesian persuasion. *Journal of Economic Theory*, 182:185–217, 2019.

Itai Arieli, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita. Optimal persuasion via bi-pooling. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 641–641, 2020.

Pak Hung Au and Keiichi Kawai. Competitive information disclosure by multiple senders. *Games and Economic Behavior*, 119:56–78, 2020.

Robert J Aumann, Michael Maschler, and Richard E Stearns. *Repeated games with incomplete information*. MIT press, 1995.

Moshe Babaioff, Robert Kleinberg, and Renato Paes Leme. Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 92–109, 2012.

Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg, and Aleksandrs Slivkins. Dynamic pricing with limited supply. 3(1), 2015.

Dirk Bergemann and Alessandro Bonatti. Selling cookies. *American Economic Journal: Microeconomics*, 7 (3):259–94, 2015.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American economic review*, 108(1):1–48, 2018.

Dirk Bergemann, Yang Cai, Grigoris Velegkas, and Mingfei Zhao. Is selling complete information (approximately) optimal? *arXiv preprint arXiv:2202.09013*, 2022a.

Dirk Bergemann, Tibor Heumann, and Stephen Morris. Screening with persuasion. *arXiv preprint arXiv:2212.03360*, 2022b.

Dirk Bergemann, Tibor Heumann, Stephen Morris, Constantine Sorokin, and Eyal Winter. Optimal information disclosure in auctions. *American Economic Review: Insights*, 2022c.

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

David A Blackwell and Meyer A Girshick. *Theory of games and statistical decisions*. Courier Corporation, 1979.

Alessandro Bonatti, Munther Dahleh, Thibaut Horel, and Amir Nouripour. Selling information in competitive environments. *arXiv preprint arXiv:2202.08780*, 2022.

Peter Bro Miltersen and Or Sheffet. Send mixed signals: earn more, work less. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 234–247, 2012.

Yang Cai and Grigoris Velegkas. How to sell information optimally: An algorithmic study. In *Proceedings of the12th Innovations in Theoretical Computer Science Conference*, volume 185, 2021.

Ozan Candogan. Persuasion in networks: Public signals and k-cores. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 133–134, 2019.

Ozan Candogan and Philipp Strack. Optimal disclosure of information to a privately informed receiver. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 263–263, 2021.

Matteo Castiglioni, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Online bayesian persuasion. *Advances in Neural Information Processing Systems*, 33, 2020.

Matteo Castiglioni, Alberto Marchesi, Andrea Celli, and Nicola Gatti. Multi-receiver online bayesian persuasion. In *International Conference on Machine Learning*, pages 1314–1323. PMLR, 2021.

Yanlin Chen and Jun Zhang. Signalling by bayesian persuasion and pricing strategy. *The Economic Journal*, 130(628):976–1007, 2020.

Yiling Chen, Haifeng Xu, and Shuran Zheng. Selling information through consulting. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2412–2431. SIAM, 2020.

Davide Crapis, Bar Ifrach, Costis Maglaras, and Marco Scarsini. Monopoly pricing in the presence of social learning. *Management Science*, 63(11):3586–3608, 2017.

Shaddin Dughmi. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2): 2–24, 2017.

Adam N Elmachtoub and Michael L Hamilton. The power of opaque products in pricing. *Management Science*, 67(8):4686–4702, 2021.

Yuval Emek, Michal Feldman, Iftah Gamzu, Renato PaesLeme, and Moshe Tennenholtz. Signaling schemes for revenue maximization. *ACM Transactions on Economics and Computation (TEAC)*, 2(2):1–19, 2014.

Yiding Feng, Wei Tang, and Haifeng Xu. Online bayesian recommendation with no regret. EC '22, page 818–819, New York, NY, USA, 2022. Association for Computing Machinery.

Ilwoo Hwang, Kyungmin Kim, and Raphael Boleslavsky. Competitive advertising and pricing. 2019.

Bar Ifrach, Costis Maglaras, Marco Scarsini, and Anna Zseleva. Bayesian social learning from consumer reviews. *Operations Research*, 67(5):1209–1221, 2019.

Nicole Immorlica, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu. Incentivizing exploration with selective data disclosure. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 647–648, 2020.

Kenneth L Judd and Michael H Riordan. Price and quality in a new product monopoly. *The Review of Economic Studies*, 61(4):773–789, 1994.

Emir Kamenica. Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Kei Kawai, Ken Onishi, and Kosuke Uetake. Signaling in online credit markets. *Journal of Political Economy*, 130(6):000–000, 2022.

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.

Richard E Kihlstrom and Michael H Riordan. Advertising as a signal. *Journal of Political Economy*, 92(3): 427–450, 1984.

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

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. Persuasion of a privately informed receiver. *Econometrica*, 85(6):1949–1964, 2017.

Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

Yingkai Li. Selling data to an agent with endogenous information. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 664–665, 2022.

Shuze Liu, Weiran Shen, and Haifeng Xu. Optimal pricing of information. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 693–693, 2021.

Puneet Manchanda, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108, 2006.

Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 565–582, 2015.

Yishay Mansour, Alex Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. *Operations Research*, 70(2):1105–1127, 2022.

Paul Milgrom and John Roberts. Price and advertising signals of product quality. *Journal of political economy*, 94(4):796–821, 1986.

Phillip Nelson. Information and consumer behavior. *Journal of political economy*, 78(2):311–329, 1970.

Phillip Nelson. Advertising as information. *Journal of political economy*, 82(4):729–754, 1974.

Navdeep S Sahni and Harikesh S Nair. Does advertising serve as a signal? evidence from a field experiment in mobile search. *The Review of Economic Studies*, 87(3):1529–1564, 2020.

Aleksandrs Slivkins. Contextual bandits with similarity information. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 679–702. JMLR Workshop and Conference Proceedings, 2011.

Jibang Wu, Zixuan Zhang, Zhe Feng, Zhaoran Wang, Zhuoran Yang, Michael I. Jordan, and Haifeng Xu. Sequential information design: Markov persuasion process and its efficient reinforcement learning. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, EC '22, page 471–472, New York, NY, USA, 2022. Association for Computing Machinery.

Shuran Zheng and Yiling Chen. Optimal advertising for information products. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 888–906, 2021.

You Zu, Krishnamurthy Iyer, and Haifeng Xu. Learning to persuade on the fly: Robustness against ignorance. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 927–928, 2021.

Appendix

A. Missing Proof of Section 3.3

PROPOSITION 1 **(Adopted from Arieli et al. (2020), Candogan (2019))**. *Let ε be the discretization parameter for the set $\mathcal{P}$ defined in* (2)*. There exists a polynomial time (in $|\mathcal{S}|U/\varepsilon$) algorithm that can solve the program $\mathsf{P}_t^{\mathsf{UCB}}$.*

Proof. Since function D_t^{UCB} is monotone with discontinuities at the points in the set $\mathcal{S}$, when we fix a price $p \in \mathcal{P}$, the function $D_t^{\mathsf{UCB}}(\kappa(p,\cdot))$ is also monotone with discontinuities at the points in $\mathcal{Q}_p = \{q : \kappa(p,q) = x\}_{x\in\mathcal{S}}$. Given a prior λ, optimizing a monotone function with discontinuities over all feasible advertising strategies induced from the prior λ subject to the constraint where the support of advertising strategies must be in the set $\mathcal{Q}_p$ has been studied in Arieli et al. (2020), Candogan (2019). It has been shown that there exists a polynomial (w.r.t. the number of discontinuities) algorithm based on convex programming that can find an optimal advertising strategy (see Proposition 2 in Arieli et al. 2020). Thus, an exhaustively search over the discretized price space $\mathcal{P}$ can lead to an optimal solution to the program $\mathsf{P}_t^{\mathsf{UCB}}$. □

B. Detailed Discussions and Proofs of Section 4

B.1. Proofs of Lemma 3 and Lemma 4

In this subsection, we provide proofs of Lemma 3 and Lemma 4. At a high-level the argument is as follows. Given an input price p, Procedure 2 outputs the closest price $p^\dagger \in \mathcal{P}$ that satisfies $p - 2\varepsilon \le p^\dagger \le p - \varepsilon$. Given an input advertising ρ, for every posterior mean $q \in \mathsf{supp}(\rho)$ and $q \notin \mathcal{Q}_{p^\dagger}$, with Lemma 1, there must exist two qualities $\bar{\omega}_i, \bar{\omega}_j$ where $i < j$ such that $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{i, j\}$. For such posterior mean q, Procedure 2 first identifies the critical type $x = \kappa(p,q)$ and $x^\dagger = \kappa(p^\dagger, q)$ where $x^\dagger$ lies within a grid $((z-1)\varepsilon, z\varepsilon)$ for some $z \in \mathbb{N}^+$. Then, Procedure 2 utilizes the critical-type function $\kappa(p^\dagger,\cdot)$ for the constructed price $p^\dagger$ to find two posterior means q_L, q_R such that they satisfy: $\kappa(p^\dagger, q_L) = z\varepsilon$ and $\kappa(p^\dagger, q_R) = (z-1)\varepsilon$. [6] To construct a feasible advertising strategy, we then round q_L up to be $\bar{\omega}_i$ when $q_L < \bar{\omega}_i$ happens and round q_R down to be $\bar{\omega}_j$ when $q_R > \bar{\omega}_j$ happens. By Assumption 1, and together with Lemma 2, we can also show that the constructed two posterior means q_L, q_R satisfy (a): $q_L < q < q_R$; and moreover (b): $\kappa(p^\dagger, q_L) \le \kappa(p,q)$, $\kappa(p^\dagger, q_R) \le \kappa(p,q)$. The relation (a) enables us to decompose the probability over this posterior mean $\rho(q)$ into probabilities over the two posterior means q_L, q_R while still satisfying (BC) condition. Together with the monotonicity of demand function D, the relation (b) can guarantee that the revenue of the output from Procedure 2 is 2ε-approximate of the revenue of the input.

[6] To see that such q_L and q_R always exist, note that because the valuation function is assumed to be monotone increasing in type θ (see Assumption **1b**), given any θ, p, q, if we have $v(\theta, q) = p$, then $\kappa(p,q) = \theta$. Therefore, q_L and q_R are the values of q satisfying $v(z\epsilon, q) = p^\dagger$ and $v((z-1)\epsilon, q) = p^\dagger$, respectively. Now, under linearity in quality, $v(\theta, q)$ is continuous in q for any given θ, which means that such solutions q_L and q_R always exist.

LEMMA 3 **(Feasibility guarantee)**. *Given an input price and advertising strategy p, ρ satisfying the properties stated in Lemma 1 and Lemma 2, the output price $p^\dagger$ and the advertising strategy $\rho^\dagger$ from Procedure 2 satisfies: $p^\dagger \in \mathcal{P}$, $\rho^\dagger$ is a feasible advertising and satisfies $\kappa(p^\dagger, q) \in \mathcal{S}$ for every $q \in \mathsf{supp}(\rho^\dagger)$.*

Proof. $p^\dagger \in \mathcal{P}$ holds trivially by construction. In below, we first show that the output $\rho^\dagger$ is indeed a feasible advertising strategy, and then prove that $\kappa(p^\dagger, q') \in \mathcal{S}$ for every $q' \in \mathsf{supp}(\rho^\dagger)$. In below analysis, let the price p and the advertising strategy ρ be the input of Procedure 2, and we will focus on an arbitrary posterior mean $q \in \mathsf{supp}(\rho)$ and analyze the corresponding construction for $\rho^\dagger$ from the posterior mean q.

$\rho^\dagger$ as a feasible advertising strategy: Clearly, a strategy $\rho^\dagger$ is a feasible advertising strategy must satisfy that $\rho^\dagger \in \Delta([0,1])$, i.e., $\rho^\dagger$ is indeed a distribution over $[0,1]$; and the associated conditional distributions $(\rho_i^\dagger)_{i\in[m]}$ must be Bayes-consistent as defined in (BC). In below analysis, by Lemma 2, we assume that $p^* \le v(1,q)$ for every $q \in \mathsf{supp}(\rho^*)$ and $q \notin \Omega$.

We first prove that the constructed advertising strategy $\rho^\dagger$ is indeed a feasible distribution. We focus on the case where $q \notin \mathcal{Q}_{p^\dagger}$. In this case, we must have $p^\dagger \in (v(0,q), v(1,q))$, otherwise it either $p^\dagger \le v(0,q)$ or $p^\dagger = v(1,q)$ which both cases falls into the scenario $q \in \mathcal{Q}_{p^\dagger}$. We first show the following claim: For any posterior mean $q \in \mathsf{supp}(\rho)$ with $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{i, j\}$, we have $q_L^\dagger \le q \le q_R^\dagger$. To see this, by definition, we have $v(x^\dagger, q) = p^\dagger, v(z\varepsilon, q_L) = p^\dagger, v((z-1)\varepsilon, q_R) = p^\dagger$, where $x^\dagger \in ((z-1)\varepsilon, z\varepsilon)$. By Assumption 1 where buyer's valuation $v(\cdot,\cdot)$ is monotone non-decreasing, we know $q_L \le q \le q_R$. Now we show that $\rho_i^\dagger(q_L^\dagger) \le \rho_i(q)$ (similar analysis can also show that $\rho_j^\dagger(q_L^\dagger) \le \rho_j(q)$). To see this, notice that from Lemma 1, we must have

$$\rho(q) = \lambda_i \rho_i(q) + \lambda_j \rho_j(q), \quad \frac{\lambda_i \rho_i(q)\bar{\omega}_i + \lambda_j \rho_j(q)\bar{\omega}_j}{\rho(q)} = q \ .$$

Thus, we must have $\rho_i(q) = \frac{\rho(q)\cdot(\bar{\omega}_j - q)}{\lambda_i(\bar{\omega}_j - \bar{\omega}_i)}$. Hence,

$$\begin{aligned}\rho_i(q) - \rho_i^\dagger(q_L^\dagger) &= \frac{\rho(q)\cdot(\bar{\omega}_j - q)}{\lambda_i(\bar{\omega}_j - \bar{\omega}_i)} - \frac{\bar{\omega}_j - q_L^\dagger}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{1}{\lambda_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger} \\ &= \frac{\rho(q)\cdot(\bar{\omega}_j - q_L^\dagger)}{\lambda_i(\bar{\omega}_j - \bar{\omega}_i)}\cdot\left(\frac{\bar{\omega}_j - q}{\bar{\omega}_j - q_L^\dagger} - \frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger}\right) \ge 0 \ ,\end{aligned}$$

where the last inequality follows from the fact that $\bar{\omega}_i \le q_L^\dagger \le q \le q_R^\dagger \le \bar{\omega}_j$. Together with the fact that $\rho_i(q) \le 1$, this shows that value $\rho_i^\dagger(q_L^\dagger) \in [0,1]$.

We now argue that in the constructed advertising strategy $\rho^\dagger$, the summation of all conditional probabilities for realizing all possible posterior mean in $\mathcal{Q}$ indeed equals to 1. Notice that from Procedure 2, for any posterior mean $q \in \mathsf{supp}(\rho)$ with $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{\bar{\omega}_i, \bar{\omega}_j\}$, the constructed

advertising strategy $\rho^\dagger$ included two posterior means $q_L^\dagger, q_R^\dagger$, and the probabilities for realizing posterior means $q_L^\dagger, q_R^\dagger$ are $\rho^\dagger(q_L^\dagger) = \lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_R^\dagger)$ (resp. $\rho^\dagger(q_R^\dagger) = \lambda_i\rho_i^\dagger(q_R^\dagger) + \lambda_j\rho_j^\dagger(q_R^\dagger)$). By construction, we have $\rho^\dagger(q_L^\dagger) + \rho^\dagger(q_R^\dagger) = \rho(q)$. Hence, from

$$\sum_{q\in\mathsf{supp}(\rho)} \rho^\dagger(q_L^\dagger) + \rho^\dagger(q_R^\dagger) = \sum_{q\in\mathsf{supp}(\rho)} \rho(q) = 1,$$

we know the constructed advertising strategy $\rho^\dagger$ is indeed a feasible distribution.

We now show that the constructed advertising strategy $\rho^\dagger$ indeed satisfies the condition (BC). In other words, we want to prove that for every $q' \in \mathsf{supp}(\rho^\dagger)$, we have

$$\frac{\sum_{i\in[m]} \lambda_i\rho_i^\dagger(q)\bar{\omega}_i}{\sum_{i\in[m]} \lambda_i\rho_i^\dagger(q)} = q'$$

Notice that when $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{i\}$, the condition (BC) holds trivially. When $\{i' \in [m] : \rho_{i'}(q) > 0\} = \{i, j\}$, Procedure 2 adds two posterior means $q_L^\dagger, q_R^\dagger$ to the support of $\rho^\dagger$. For the posterior mean $q_L^\dagger$:

$$\begin{aligned}\frac{\lambda_i\rho_i^\dagger(q_L^\dagger)\bar{\omega}_i + \lambda_j\rho_j^\dagger(q_L^\dagger)\bar{\omega}_j}{\lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_L^\dagger)} &= \frac{\frac{\bar{\omega}_j - q_L^\dagger}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger}\cdot\bar{\omega}_i + \frac{q_L^\dagger - \bar{\omega}_i}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger}\cdot\bar{\omega}_j}{\frac{\bar{\omega}_j - q_L^\dagger}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger} + \frac{q_L^\dagger - \bar{\omega}_i}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger}} \\ &= \frac{\bar{\omega}_j - q_L^\dagger}{\bar{\omega}_j - \bar{\omega}_i}\cdot\bar{\omega}_i + \frac{q_L^\dagger - \bar{\omega}_i}{\bar{\omega}_j - \bar{\omega}_i}\cdot\bar{\omega}_j = q_L^\dagger\end{aligned}$$

On the other hand, we observe

$$\begin{aligned}\rho^\dagger(q_L^\dagger)q_L^\dagger + \rho^\dagger(q_R^\dagger)q_R^\dagger &= (\lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_L^\dagger))q_L^\dagger + (\lambda_i(\rho_i(q) - \rho_i^\dagger(q_L^\dagger)) + \lambda_j(\rho_j(q) - \rho_j^\dagger(q_L^\dagger)))q_R^\dagger \\ &= (\lambda_i\rho_i(q) + \lambda_j\rho_j(q))q_R^\dagger - (\lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_L^\dagger))(q_R^\dagger - q_L^\dagger) \\ &= \rho(q)q_R^\dagger - \left(\frac{\bar{\omega}_j - q_L^\dagger}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger} + \frac{q_L^\dagger - \bar{\omega}_i}{\bar{\omega}_j - \bar{\omega}_i}\cdot\frac{\rho(q)\cdot(q_R^\dagger - q)}{q_R^\dagger - q_L^\dagger}\right)(q_R^\dagger - q_L^\dagger) \\ &= \rho(q)q_R^\dagger + \rho(q)(q_R^\dagger - q) = \rho(q)q\ .\end{aligned}$$

Thus, for the posterior mean $q_R^\dagger$, we have

$$\begin{aligned}\frac{\lambda_i\rho_i^\dagger(q_R^\dagger)\bar{\omega}_i + \lambda_j\rho_j^\dagger(q_R^\dagger)\bar{\omega}_j}{\lambda_i\rho_i^\dagger(q_R^\dagger) + \lambda_j\rho_j^\dagger(q_R^\dagger)} &= \frac{\lambda_i(\rho_i(q) - \rho_i^\dagger(q_L^\dagger))\bar{\omega}_i + \lambda_j(\rho_j(q) - \rho_j^\dagger(q_L^\dagger))\bar{\omega}_j}{\lambda_i(\rho_i(q) - \rho_i^\dagger(q_L^\dagger)) + \lambda_j(\rho_j(q) - \rho_j^\dagger(q_L^\dagger))} \\ &= \frac{\lambda_i\rho_i(q)\bar{\omega}_i + \lambda_j\rho_j(q)\bar{\omega}_j - q_L^\dagger\cdot(\lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_L^\dagger))}{\rho(q) - (\lambda_i\rho_i^\dagger(q_L^\dagger) + \lambda_j\rho_j^\dagger(q_L^\dagger))} \\ &= \frac{\rho(q)q - \rho^\dagger(q_L^\dagger)q_L^\dagger}{\rho(q) - \rho^\dagger(q_L^\dagger)} = \frac{\rho^\dagger(q_R^\dagger)q_R^\dagger}{\rho^\dagger(q_R^\dagger)} = q_R^\dagger\ .\end{aligned}$$

We now have shown that the constructed advertising strategy $\rho^\dagger$ indeed satisfies condition (BC).

$\kappa(p^\dagger, q') \in \mathcal{S}$ for every $q' \in \mathsf{supp}(\rho^\dagger)$: Fix a posterior mean $q \in \mathsf{supp}(\rho)$, we focus on the case $q \notin \mathcal{Q}_{p^\dagger}$ (the other case is trivial by construction), we know that in Procedure 2, the corresponding posterior

means $q_L^\dagger, q_R^\dagger \in \mathsf{supp}(\rho^\dagger)$. And either $q_L^\dagger = q_L$ or $q_L^\dagger = \bar{\omega}_i$, either $q_R^\dagger = q_R$ or $q_R^\dagger = \bar{\omega}_j$. When $q_L^\dagger = q_L$, we have $\kappa(p^\dagger, q_L^\dagger) = \kappa(p^\dagger, q_L) = z\varepsilon \in \mathcal{S}$. When $q_L^\dagger = \bar{\omega}_i$, we have $\kappa(p^\dagger, q_L^\dagger) = \kappa(p^\dagger, \bar{\omega}_i) \in \mathcal{S}$ as $p^\dagger \in \mathcal{P}$. Similar analysis also shows that $\kappa(p^\dagger, q_R^\dagger) \in \mathcal{S}$. The proof completes.

□

LEMMA 4 **(Revenue guarantee).** *Fix a price $p \geq 2\varepsilon$ and a feasible advertising strategy ρ, let $p^\dagger, \rho^\dagger = \mathsf{Rounding}(p, \rho)$ be the output from Procedure 2, then we have $\mathsf{Rev}(p, \rho) - \mathsf{Rev}(p^\dagger, \rho^\dagger) \leq 2\varepsilon$.*

Proof. We provide the proof when the input to Procedure 2 is p^*, ρ^*. The proof only utilizes the monotoncity of the function D. In below analysis, let $p^\dagger, \rho^\dagger$ be the advertising strategy output from Procedure 2 with the input p^*, ρ^*. We now fix a posterior mean $q \in \mathsf{supp}(\rho^*)$ and consider the following two cases:

Case 1: $q \in \mathcal{Q}_{p^\dagger}$**.** In this case, we have $\kappa(p^\dagger, q) \leq \kappa(p^*, q)$.

Case 2: $q \notin \mathcal{Q}_{p^\dagger}$**.** Let $\{i' \in [m] : \rho^*_{i'}(q) > 0\} = \{i, j\}$. Let $q_L^\dagger$, $q_R^\dagger$ be the corresponding counterpart in the new advertising strategy $\rho^\dagger$, we now show the following claim:

CLAIM 1. $\kappa(p^\dagger, q_R^\dagger) \leq \kappa(p^\dagger, q_L^\dagger) \leq \kappa(p^*, q)$.

To see the above claim, recall that in construction, we have $v(x, q) \leq p^*$, $v(x^\dagger, q) = p^\dagger$, and by construction, we have $p^\dagger \leq p^* - \varepsilon$. Thus, by Assumption **1b**, we have $\varepsilon \leq p^* - p^\dagger \leq v(x, q) - v(x^\dagger, q) \leq x - x^\dagger$, which implies that $z\varepsilon \leq x^\dagger + \varepsilon \leq x$. Fix any price p, from Assumption **1a**, we know that the function $\kappa(p, \cdot)$ is monotone non-increasing. Recall that in previous analysis, we have shown $q_L \leq q_L^\dagger \leq q \leq q_R^\dagger \leq q_R$. We thus have $\kappa(p^\dagger, q_R^\dagger) \leq \kappa(p^\dagger, q_L^\dagger) \leq \kappa(p^\dagger, q_L) = z\varepsilon \leq x = \kappa(p^*, q)$.

With the above observations, we have

$$
\begin{aligned}
& \mathsf{Rev}(p^*, \rho^*) - \mathsf{Rev}(p^\dagger, \rho^\dagger) \\
= \; & p^* \sum_{i \in [m]} \lambda_i \int_0^1 \rho_i^*(q) D(\kappa(p^*, q)) dq - p^\dagger \sum_{i \in [m]} \lambda_i \int_0^1 \rho_i^\dagger(q) D(\kappa(p^\dagger, q)) dq \\
= \; & p^* \int_0^1 \rho^*(q) D(\kappa(p^*, q)) dq - p^\dagger \int_0^1 \rho^\dagger(q) D(\kappa(p^\dagger, q)) dq \\
\overset{(a)}{\leq} \; & p^* \int_0^1 \rho^*(q) D(\kappa(p^*, q)) dq - (p^* - 2\varepsilon) \int_0^1 \rho^\dagger(q) D(\kappa(p^\dagger, q)) dq \\
\overset{(b)}{=} \; & p^* \sum_{q \in \mathsf{supp}(\rho^*)} \rho^*(q) D(\kappa(p^*, q)) - p^* \sum_{q \in \mathsf{supp}(\rho^*)} \left(\rho^\dagger(q_L^\dagger) D(p^\dagger, q_L^\dagger) + \rho^\dagger(q_R^\dagger) D(p^\dagger, q_R^\dagger) \right) + 2\varepsilon \\
\overset{(c)}{=} \; & p^* \sum_{q \in \mathsf{supp}(\rho^*)} \left(\rho^*(q) D\left(\kappa(p^*, q)\right) - \left(\rho^\dagger(q_L^\dagger) D\left(\kappa(p^\dagger, q_L^\dagger)\right) + \rho^\dagger(q_R^\dagger) D\left(\kappa(p^\dagger, q_R^\dagger)\right) \right) \right) + 2\varepsilon \\
\overset{(d)}{\leq} \; & \sum_{q \in \mathsf{supp}(\rho^*)} p^* \left(\rho^*(q) D(\kappa(p^*, q)) - \left(\rho^\dagger(q_L^\dagger) D(\kappa(p^*, q)) + \rho^\dagger(q_R^\dagger) D(\kappa(p^*, q)) \right) \right) + 2\varepsilon \\
\overset{(e)}{=} \; & \sum_{q \in \mathsf{supp}(\rho^*)} p^* \left(\rho^*(q) D(\kappa(p^*, q)) - \rho^*(q) D(\kappa(p^*, q)) \right) + 2\varepsilon = 2\varepsilon \; ,
\end{aligned}
$$

where inequality (a) holds since $p^\dagger \geq p^* - 2\varepsilon$; in equality (b), we, for simplicity, focus on **else** case, the analysis for other scenarios is the same; equality (c) holds by the construction of the strategy $\rho^\dagger$, inequality (d) holds from Claim 1; inequality (e) holds since by construction, we have for any $q \in \mathsf{supp}(\rho^\dagger)$, we have $\rho^\dagger(q_L^\dagger) + \rho^\dagger(q_R^\dagger) = \rho^*(q)$. □

B.2. Proof of Lemma 5

We use the following self-normalized martingale tail inequality to prove the high-probability bounds. In particular, we use the following results obtained in Abbasi-Yadkori et al. (2012):

LEMMA 8 **(Uniform bound for self-normalized bound for martingales, see Abbasi-Yadkori et al. 2012)** *Let $\{\mathcal{F}_t\}_{t=1}^\infty$ be a filtration. Let $\{Z_t\}_{t=1}^\infty$ be a sequence of real-valued variables such that Z_t is $\mathcal{F}_t$-measurable. Let $\{\eta_t\}_{t=1}^\infty$ be a sequence of real-valued random variables such that η_t is $\mathcal{F}_{t+1}$-measurable and is conditionally R-sub-Gaussian. Let $V > 0$ be deterministic. For any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$:*

$$\left|\sum_{s=1}^t \eta_s Z_s\right| \leq R\sqrt{2\left(V + \sum_{s=1}^t Z_s^2\right)\ln\left(\frac{\sqrt{V + \sum_{s=1}^t Z_s^2}}{\delta\sqrt{V}}\right)} \tag{7}$$

LEMMA 5. *For every $t \geq |\mathcal{S}| + 1$, the following holds with probability at least $1 - 1/T^2$:*

$$D_t^{\mathsf{UCB}}(x) \geq D(x), \qquad \forall x \in \mathcal{S}; \tag{4}$$

$$D_t^{\mathsf{UCB}}(x) - D(x) \leq 2\sqrt{\frac{16\log T}{N_t(x)}} + \frac{2\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)}, \quad \forall x \in \mathcal{S}. \tag{5}$$

Proof. To prove the results, we first show the following concentration inequality for the empirical demand estimates of the points in the set $\mathcal{S}$: the following holds with probability at least $1 - 1/T^2$,

$$|\bar{D}_t(x) - D(x)| \leq \sqrt{\frac{16\log T}{N_t(x)}} + \frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)}, \quad \forall x \in \mathcal{S}\ . \tag{8}$$

To prove the above inequality, we fix an arbitrary $x \in \mathcal{S}$. We define the random variable $Z_t(x) = \mathbf{1}[\kappa(p_t, q_t) = x]$, We also define random variable $\eta_t(x) = a_t(x) - D(x)$ if $Z_t(x) = 1$ at time step t. Then by definition, we know that the sequence $\{\sum_{s=1}^t \eta_t(x)Z_t(x)\}$ is a martingale adapted to $\{\mathcal{F}_{t+1}\}_{t=0}^\infty$. Moreover, the sequence of the variable $\{Z_t(x)\}_{t=1}^\infty$ is $\mathcal{F}_t$-measurable, and the variable $\eta_t(x)$ is 1-sub-Gaussian. Now take $V = 1$ and substitute for $\eta_t(x) = a_t(x) - D(x)$, apply the uniform bound obtained in Lemma 8, we have for any $t \geq |\mathcal{S}| + 1$, the following holds with probability at least $1 - \delta$,

$$\left|\sum_{s=1}^t (a_s(x) - D(x))Z_s(x)\right| \leq \sqrt{2\left(1 + \sum_{s=1}^t Z_s(x)^2\right)\ln\left(\frac{\sqrt{1 + \sum_{s=1}^t Z_s(x)^2}}{\delta}\right)}$$

Observe that in the above inequality, the term $\left|\sum_{s=1}^{t}(a_s(x)-D(x))Z_s(x)\right|$ is exactly $|\sum_{s\in\mathcal{N}_t(x)}a_s(x)-N_t(x)D(x)|$, and the term $\sum_{s=1}^{t}Z_s(x)^2$ exactly equals to $N_t(x)$. Dividing both sides with $N_t(x)$, substituting for $\sum_{s=1}^{t}Z_s(x)^2=N_t(x)$, we obtain

$$\begin{aligned}\left|\bar{D}_t(x)-D(x)\right| &\le \frac{1}{N_t(x)}\sqrt{2\left(1+N_t(x)\right)\ln\left(\frac{\sqrt{1+N_t(x)}}{\delta}\right)}\\ &=\sqrt{\frac{2\left(1+N_t(x)\right)\ln\frac{1}{\delta}+\left(1+N_t(x)\right)\ln\left(1+N_t(x)\right)}{N_t(x)^2}}\\ &\le\sqrt{\frac{4\ln\frac{1}{\delta}}{N_t(x)}}+\frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)}\end{aligned}$$

where in last inequality we use the fact that $1+N_t(x)\le 2N_t(x)$, and $\sqrt{u+v}\le\sqrt{u}+\sqrt{v}$ for any $u,v\ge 0$. Setting $\delta=T^{-5}$, we know that the above inequality holds with probability at least $1-1/T^5$. Taking the union bound over all choices of t and over all choices of $x\in\mathcal{S}$, we obtain that the first statement holds with probability at least $1-1/T^2$ as long as $|\mathcal{S}|\le T$, which is the case for us.

For the inequality (4), for notation simplicity, let $\mathsf{CR}_t(x)\triangleq\sqrt{\frac{16\log T}{N_t(x)}}+\frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)}$ be the high-probability error, and we also write $\mathcal{S}=\{x^{(1)},\ldots,x^{(|\mathcal{S}|)}\}$ where $x^{(i)}<x^{(j)}$ for any $i<j$. Now fix an arbitrary $x^{(i)}\in\mathcal{S}$, fix a time round $t\ge|\mathcal{S}|+1$. Denote the random variable $i^\dagger=\arg\min_{i':i'\le i}\bar{D}_t(x^{(i')})+\mathsf{CR}_t(x^{(i')})\wedge 1$.

$$\begin{aligned}\mathbb{P}\Big[D_t^{\mathsf{UCB}}(x^{(i)})\ge D(x^{(i)})\Big] &= 1-\sum_{j=1}^{i}\mathbb{P}\big[i^\dagger=j\big]\,\mathbb{P}\Big[D_t^{\mathsf{UCB}}(x^{(i)})<D(x^{(i)})\mid i^\dagger=j\Big]\\ &=1-\sum_{j=1}^{i}\mathbb{P}\big[i^\dagger=j\big]\,\mathbb{P}\Big[\bar{D}_t(x^{(j)})+\mathsf{CR}_t(x^{(j)})<D(x^{(i)})\mid i^\dagger=j\Big]\\ &\overset{(a)}{\ge}1-\sum_{j=1}^{i}\mathbb{P}\big[i^\dagger=j\big]\,\mathbb{P}\Big[\bar{D}_t(x^{(j)})+\mathsf{CR}_t(x^{(j)})<D(x^{(j)})\mid i^\dagger=j\Big]\\ &=1-\sum_{j=1}^{i}\mathbb{P}\Big[\bar{D}_t(x^{(j)})+\mathsf{CR}_t(x^{(j)})<D(x^{(j)}),i^\dagger=j\Big]\\ &\ge 1-\sum_{j=1}^{i}\mathbb{P}\Big[\bar{D}_t(x^{(j)})+\mathsf{CR}_t(x^{(j)})<D(x^{(j)})\Big]\\ &\overset{(b)}{\ge}1-|\mathcal{S}|\delta\ge 1-T^{-4}\end{aligned}$$

where inequality (a) holds since $D(x^{(j)})\ge D(x^{(i)})$ for any $j\le i$, and inequality (b) holds follows from earlier analysis where for a fixed t and fixed $x\in\mathcal{S}$, we have $\mathbb{P}\Big[\bar{D}_t(x)+\mathsf{CR}_t(x)<D(x)\Big]\le\delta$. Taking the union bound over all choices of t and over all choices of $x\in\mathcal{S}$ finishes the proof.

For the inequality (5), from triangle inequality, we have

$$\left|D_t^{\mathsf{UCB}}(x)-D(x)\right|\le\left|D_t^{\mathsf{UCB}}(x)-\bar{D}_t(x)\right|+\left|\bar{D}_t(x)-D(x)\right|$$

$$\leq \sqrt{\frac{16\log T}{N_t(x)}} + \frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} + \left|\bar{D}_t(x) - D(x)\right|$$
$$\overset{(a)}{\leq} 2\sqrt{\frac{16\log T}{N_t(x)}} + \frac{2\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} ,$$

where the inequality (a) holds with probability at least $1 - 1/T^2$ according to the first statement we just proved. □

B.3. Proof of Lemma 6

LEMMA 6. *For every time $t \geq |\mathcal{S}|+1$, with probability at least $1-2/T^2$, we have*

$$\mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}_t^{\mathsf{UCB}}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}_t^{\mathsf{UCB}}(p_t, \rho_t)$$

Proof. We begin our analysis by defining the following event. For all $t = |\mathcal{S}|+1, \ldots, T$, define events E_t

$$\mathsf{E}_t \triangleq \bigcup_{x\in\mathcal{S}} \left\{ D_t^{\mathsf{UCB}}(x) < D(x) \text{ or } D_t^{\mathsf{UCB}}(x) > D(x) + \sqrt{\frac{16\log T}{N_t(x)}} + \frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} \right\}$$

From union bound, it follows that

$$\begin{aligned}
\mathbb{P}[\mathsf{E}_t] &\leq \sum_{x\in\mathcal{S}} \mathbb{P}\left[D_t^{\mathsf{UCB}}(x) < D(x) \text{ or } D_t^{\mathsf{UCB}}(x) > D(x) + \sqrt{\frac{16\log T}{N_t(x)}} + \frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} \right] \\
&\leq \sum_{x\in\mathcal{S}} \mathbb{P}\left[D_t^{\mathsf{UCB}}(x) < D(x) \right] + \\
&\quad \sum_{x\in\mathcal{S}} \mathbb{P}\left[D_t^{\mathsf{UCB}}(x) > D(x) + \sqrt{\frac{16\log T}{N_t(x)}} + \frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} \right] \\
&\overset{(a)}{\leq} \frac{2}{T^2}
\end{aligned}$$

where the inequality (a) follows from inequalities (4) and (5) in Lemma 5. Recall that whenever $\mathbf{1}[\mathsf{E}_t^c] = 1$, we have

$$\begin{aligned}
&\mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) - \mathsf{Rev}_t^{\mathsf{UCB}}(\widetilde{p}^*, \widetilde{\rho}^*) \\
&= \widetilde{p}^* \sum_{i\in[m]} \lambda_i \int_0^1 \widetilde{\rho}_i^*(q) D(\kappa(\widetilde{p}^*, q)) dq - \widetilde{p}^* \sum_{i\in[m]} \lambda_i \int_0^1 \widetilde{\rho}_i^*(q) D_t^{\mathsf{UCB}}(\kappa(\widetilde{p}^*, q)) dq \\
&= \widetilde{p}^* \sum_{i\in[m]} \lambda_i \int_0^1 \widetilde{\rho}_i^*(q) \left(D(\kappa(\widetilde{p}^*, q)) - D_t^{\mathsf{UCB}}(\kappa(\widetilde{p}^*, q)) \right) dq \leq 0
\end{aligned}$$

Thus, whenever $\mathbf{1}[\mathsf{E}_t^c] = 1$, we have

$$\mathsf{Rev}(\widetilde{p}^*, \widetilde{\rho}^*) \leq \mathsf{Rev}_t^{\mathsf{UCB}}(\widetilde{p}^*, \widetilde{\rho}^*) \overset{(a)}{\leq} \mathsf{Rev}(p_t, \rho_t)$$

where inequality (a) follows from our algorithm design. □

B.4. Proof of Lemma 7

LEMMA 7. *For every time $t \geq |\mathcal{S}|+1$, with probability at least $1-2/T^2$, we have*

$$\mathsf{Rev}_t^{\mathsf{UCB}}(p_t,\rho_t) - \mathsf{Rev}(p_t,\rho_t) \leq 5p_t \sum_{q\in\mathsf{supp}(\rho_t)} \rho_t(q)\sqrt{\frac{\log T}{N_t(\kappa(p_t,q))}}$$

Proof. Follow from the definition of the event E_t^c, when $\mathbf{1}[\mathsf{E}_t^c]=1$, we have

$$\begin{aligned}
&\mathsf{Rev}_t^{\mathsf{UCB}}(p_t,\rho_t) - \mathsf{Rev}(p_t,\rho_t)\\
\leq\ & p_t \sum_{i\in[m]} \lambda_i \int_0^1 \rho_{i,t}(q)\left(D_t^{\mathsf{UCB}}(\kappa(p_t,q)) - D(\kappa(p_t,q))\right) dq\\
\overset{(a)}{\leq}\ & p_t \sum_{i\in[m]} \lambda_i \int_0^1 \rho_{i,t}(q)\left(\sqrt{\frac{16\log T}{N_t(\kappa(p_t,q))}} + \frac{\sqrt{(1+N_t(\kappa(p_t,q)))\ln(1+N_t(\kappa(p_t,q)))}}{N_t(\kappa(p_t,q))}\right) dq\\
\overset{(b)}{\leq}\ & 5p_t \sum_{i\in[m]} \lambda_i \int_0^1 \rho_{i,t}(q)\sqrt{\frac{\log T}{N_t(\kappa(p_t,q))}} dq\\
\overset{(c)}{=}\ & 5p_t \sum_{q\in\mathsf{supp}(\rho_t)} \rho_t(q)\sqrt{\frac{\log T}{N_t(\kappa(p_t,q))}}\ ,
\end{aligned}$$

where inequality (a) follows from the definition of event E_t^c, inequality (b) follows from the fact that $N_t(x)\leq T, \forall t$ and thus, $\frac{\sqrt{(1+N_t(x))\ln(1+N_t(x))}}{N_t(x)} \leq \sqrt{\frac{\log T}{N_t(x)}}$, and in last equality (c), we have $\rho_t(q) = \sum_{i\in[m]} \lambda_i \rho_{i,t}(q)$. □

C. Missing Proofs of Section 5

C.1. Proof of Theorem 2

THEOREM 2. *Given an additive valuation function, $v(\theta,\omega)=\theta+\omega$, and equally-spaced product quality domain, Ω Algorithm 1 with parameter $\varepsilon = \Theta((\log T/T)^{1/3} \wedge 1/m)$ has an expected regret of $O\left(T^{2/3}(\log T)^{1/3} + \sqrt{mT\log T}\right)$.*

Proof. For additive valuation, we know $\kappa(p,q) = ((p-q)\wedge 1)\vee 0$. Since $q\in[0,1]$, we know that $\bar{v}=2, \underline{v}=0$.

We first prove the regret $O(T^{2/3}(\log T)^{1/3})$ when $m \leq (T/\log T)^{1/3}+1$. Define the following discretization parameter that will be used to define the discretized price space $\mathcal{P}$ and the discretized type space $\mathcal{S}$ in (2).

$$\varepsilon = \max\left\{\varepsilon' \geq 0 : \frac{1/m}{\varepsilon'} \in \mathbb{N}^+ \wedge \varepsilon' \leq \left(\frac{\log T}{T}\right)^{1/3}\right\} \tag{9}$$

We now argue that the above $\varepsilon = \Theta((\log T/T)^{1/3})$. To see this, let the integers $k_1, k_2 \in \mathbb{N}^+$ satisfy

$$\left\lfloor \frac{1/(m-1)}{(\frac{\log T}{T})^{1/3}} \right\rfloor = k_1, \quad \left\lfloor \frac{1/(m-1)}{\frac{1}{2}(\frac{\log T}{T})^{1/3}} \right\rfloor = k_2\ .$$

By assumption, we have $\frac{1}{m-1} \geq (\log T/T)^{1/3}$, implying $k_1 \geq 1$, and $k_2 \geq 2$. Thus, there must exist an $\varepsilon' \in [(\frac{1}{2}(\log T/T)^{1/3}, (\log T/T)^{1/3}]$ such that $\frac{1/(m-1)}{\varepsilon'} \in [k_1 : k_2]$, which implies that the above defined $\varepsilon = \Theta((\log T/T)^{1/3})$. Suppose $K_\varepsilon \in \mathbb{N}^+$ such that $K_\varepsilon \varepsilon = 1/(m-1)$. By definition of uniformly-spaced qualities, we know that $\bar{\omega}_i = \frac{i-1}{(m-1)}, \forall i \geq 2$. For a discretized price space $\mathcal{P} = \{\varepsilon, 2\varepsilon, \ldots, 2-\varepsilon, 2\}$, we know that for any price $p = k_p\varepsilon \in \mathcal{P}$ for some integer $k_p \in \mathbb{N}^+$, we have $\kappa(k_p\varepsilon, \bar{\omega}_i) = k_p\varepsilon - \bar{\omega}_i = k_p\varepsilon - (i-1)K_\varepsilon\varepsilon \in \{0, \varepsilon, \ldots, 1\}$. Thus, for the set $\mathcal{S}$ defined in (2) we have $|\mathcal{S}| = O(1/\varepsilon)$. With ε defined in (9), Algorithm 1 has the desired regret upper bound.

We now prove the regret $O(\sqrt{mT\log T})$ when number of qualities $m > (T/\log T)^{1/3} + 1$. For this case, we can simple feed the Algorithm 1 with discretization parameter $\varepsilon = 1/(m-1)$. Then, according to the proof of Theorem 1, the regret of Algorithm 1 can be bounded as $O(T/m + \sqrt{Tm\log T}) = O(\sqrt{Tm\log T})$ as desired. □

C.2. Missing Algorithm and Proof of Theorem 3

The detailed algorithm description when the number of qualities is large is provided in Algorithm 3.

Algorithm 3: Algorithm for arbitrary size m of product quality space.

Input: Discretization parameter ε and pooling precision parameter $\widehat{\varepsilon}$.

Input: Instance $\mathcal{I}$ with quality space Ω and prior λ.

Construct instance $\mathcal{I}^\dagger$ as follows: Let the quality space $\Omega^\dagger = \{\bar{\omega}_i^\dagger\}_{i\in[\lceil 1/\widehat{\varepsilon}\rceil+1]}$ where $\bar{\omega}_1^\dagger = 0, \lambda_1^\dagger = \lambda_1$; and $\bar{\omega}_{i+1}^\dagger = \mathbb{E}_{\omega\sim\lambda}[\omega \mid \omega \in ((i-1)\widehat{\varepsilon}, i\widehat{\varepsilon}]]$, and let the prior $\lambda^\dagger = (\lambda_i^\dagger)_{i\in[\lceil 1/\widehat{\varepsilon}\rceil+1]}$ where $\lambda_{i+1}^\dagger = \mathbb{P}_{\omega\sim\lambda}[\omega \in ((i-1)\widehat{\varepsilon}, i\widehat{\varepsilon}]]$ for all $1 \leq i \leq \lceil 1/\widehat{\varepsilon}\rceil$.

Run Algorithm 1 on instance $\mathcal{I}^\dagger$ with discretization parameter ε.

In below, we provide a regret bound that is independent of the size of quality space and it holds for valuation function beyond the additive one as long as it satisfies the following assumption:

Assumption 2. *Function $\kappa(p,\cdot)$ satisfies that for any price $p \in [0,U]$, for any q_1, q_2 where $q_1 \leq q_2$, $\kappa(p,q_1) - \kappa(p,q_2) \leq q_2 - q_1$.* [7]

Notice that additive valuation $v(\theta,\omega) = \theta + \omega$, which has $\kappa(p,q) = p - q$, satisfies the above assumption.

Proposition 3. *With Assumption 1 and Assumption 2, Algorithm 3 with $\widehat{\varepsilon} = \varepsilon = (\log T/T)^{1/4}$ has an expected regret of $O(T^{3/4}(\log T)^{1/4})$ independent of the size m of quality space.*

[7] We can also relax the assumption to be $\kappa(p,q_1) - \kappa(p,q_1) \leq L(q_2 - q_1)$ where an arbitrary constant $L \in \mathbb{R}^+$ can be treated similarly.

Given the above Proposition 3, Theorem 3 simply follows as additive valuation function satisfies Assumption 2.

Proof of Proposition 3. We fix a small $\widehat{\varepsilon} \in (0,1)$. Let $\mathcal{I}$ be an instance with quality space Ω and prior $\lambda \in \Delta^{\Omega}$. For exposition simplicity, let us assume that for each $i \in [\lceil 1/\widehat{\varepsilon} \rceil]$, there exists at least one quality $\omega \in \Omega$ such that $\omega \in ((i-1)\widehat{\varepsilon}, i\widehat{\varepsilon}]$. We now construct a new instance $\mathcal{I}^{\dagger}$ with quality space $\Omega^{\dagger} = (\bar{\omega}_i^{\dagger})_{i \in [\lceil 1/\widehat{\varepsilon} \rceil + 1]}$ and prior $\lambda^{\dagger} = (\lambda_i^{\dagger})_{i \in [\lceil 1/\widehat{\varepsilon} \rceil + 1]}$ as follows:

- for $i = 1$: $\bar{\omega}_i^{\dagger} = 0, \lambda_i^{\dagger} = \lambda_1$;
- for $2 \le i \le \lceil 1/\widehat{\varepsilon} \rceil + 1$: $\bar{\omega}_i^{\dagger} = \mathbb{E}_{\omega \sim \lambda}[\omega \,|\, \omega \in ((i-2)\widehat{\varepsilon}, (i-1)\widehat{\varepsilon}]], \lambda_i^{\dagger} = \mathbb{P}_{\omega \sim \lambda}[\omega \in ((i-2)\widehat{\varepsilon}, (i-1)\widehat{\varepsilon}]]$.

Essentially, the instance $\mathcal{I}^{\dagger}$ is constructed by pooling all product qualities that are "close enough" with each other (i.e., qualities in a grid $((i-1)\widehat{\varepsilon}, i\widehat{\varepsilon}]$). By construction, we know that $|\Omega^{\dagger}| = O(1/\widehat{\varepsilon})$. Given a price p and an advertising ρ, let $\mathsf{Rev}_{\mathcal{I}}(p,\rho)$ be the seller's revenue for problem instance $\mathcal{I}$. In below, we have the following revenue guarantee between these two problem instances $\mathcal{I}, \mathcal{I}^{\dagger}$.

LEMMA 9. *Let p^*, ρ^* be the optimal price and optimal advertising for instance $\mathcal{I}$, with Assumption 2, there exists a price $p^{\dagger}$ and advertising $\rho^{\dagger}$ for instance $\mathcal{I}^{\dagger}$ such that $\mathsf{Rev}_{\mathcal{I}}(p^*,\rho^*) \le \mathsf{Rev}_{\mathcal{I}^{\dagger}}(p^{\dagger},\rho^{\dagger}) + \widehat{\varepsilon}$.*

The proof of the above Lemma 9 utilizes Assumption 2 and is provided subsequently. With Lemma 9, by feeding Algorithm 1 with new instance $\mathcal{I}^{\dagger}$, the total expected regret for instance $\mathcal{I}$ can be bounded as follows

$$\mathrm{Regret}_{\mathcal{I}}[T] \le O\left(T\widehat{\varepsilon} + T\varepsilon + \sqrt{|\mathcal{S}|T\log T}\right) = O\left(T\widehat{\varepsilon} + T\varepsilon + \sqrt{\frac{1}{\varepsilon\widehat{\varepsilon}} T\log T}\right) \le O\left(T^{3/4}(\log T)^{1/4}\right)$$

where the term $T\widehat{\varepsilon}$ is from Lemma 9 and due to reducing the instance $\mathcal{I}$ to the new instance $\mathcal{I}^{\dagger}$, the term $T\varepsilon + \sqrt{|\mathcal{S}|T\log T}$ is the incurred regret of Algorithm 1 for the new instance $\mathcal{I}^{\dagger}$ where the number of discretized types $|\mathcal{S}|$ for the new instance $\mathcal{I}^{\dagger}$ equals $\frac{1}{\varepsilon\widehat{\varepsilon}}$, and in the last inequality, we choose $\widehat{\varepsilon} = \varepsilon = (\log T/T)^{1/4}$. □

In below, we provide the proof for Lemma 9.

Proof of Lemma 9. Let us fix the problem instance $\mathcal{I}$ with quality space $\Omega, |\Omega| = m$ and prior distribution λ. Let $\mathcal{I}^{\dagger}$ be the constructed instance (see Line 3 in Algorithm 3). In the proof, we construct a price $p^{\dagger}$ and an advertising strategy $\rho^{\dagger}$ for instance $\mathcal{I}^{\dagger}$ based on p^*, ρ^*. Consider a price $p^{\dagger} = p^* - \widehat{\varepsilon}$. In below, we show that how to construct advertising strategy $\rho^{\dagger}$ from the advertising strategy ρ^*. In particular, for each posterior mean $q \in \mathsf{supp}(\rho^*)$, we construct a corresponding posterior mean $q^{\dagger} \in \mathsf{supp}(\rho^{\dagger})$, and furthermore, with Assumption 1 and Assumption 2, we also show that we always have $\kappa(p^*, q) \ge \kappa(p^{\dagger}, q^{\dagger})$. Recall that from Lemma 1, the advertising strategy ρ^* satisfies $\{i \in [m] : \rho_i^*(q) > 0\} \le 2$ for all $q \in \mathsf{supp}(\rho^*)$. Our construction based on threes cases of $\{i \in [m] : \rho_i^*(q) > 0\}$.

- **Case 1** – if $\{i \in [m] : \rho_i^*(q) > 0\} = \{i'\}$, in this case, suppose $\bar{\omega}_{i'} \in ((j-1)\widehat{\varepsilon}, j\widehat{\varepsilon}]$ for some $j \in [\lceil 1/\widehat{\varepsilon} \rceil]$, then consider

$$\rho_{j+1}^\dagger(q^\dagger) = \frac{\lambda_{i'}\rho_{i'}^*(q)}{\lambda_{j+1}^\dagger}; \quad \text{where } q^\dagger = \bar{\omega}_{j+1}^\dagger.$$

From the above construction, we know that $\kappa(p^*, q) = \kappa(p^*, \bar{\omega}_{i'})$, and

$$\kappa(p^*, \bar{\omega}_{i'}) \overset{(a)}{\geq} \kappa(p^\dagger, \bar{\omega}_{i'}) + \widehat{\varepsilon} \overset{(b)}{\geq} \kappa(p^\dagger, \bar{\omega}_{j+1}^\dagger) = \kappa(p^\dagger, q^\dagger)$$

where inequality (a) holds since $\widehat{\varepsilon} = p^* - p^\dagger \leq \kappa(p^*, \bar{\omega}_{i'}) - \kappa(p^\dagger, \bar{\omega}_{i'})$ due to Assumption **1b**, and inequality (b) holds since $|\kappa(p^\dagger, \bar{\omega}_{j+1}^\dagger) - \kappa(p^\dagger, \bar{\omega}_{i'})| \leq |\bar{\omega}_{j+1}^\dagger - \bar{\omega}_{i'}| \leq \widehat{\varepsilon}$ due to Assumption 2.

- **Case 2** – if $\{i \in [m] : \rho_i^*(q) > 0\} = \{i', i''\}$ where $i' < i''$, in this case, suppose both $\bar{\omega}_{i'}, \bar{\omega}_{i''} \in ((j-1)\widehat{\varepsilon}, j\widehat{\varepsilon}]$ for some $j \in [\lceil 1/\widehat{\varepsilon} \rceil]$, then consider

$$\rho_{j+1}^\dagger(q^\dagger) = \frac{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)}{\lambda_{j+1}^\dagger}; \quad \text{where } q^\dagger = \bar{\omega}_{j+1}^\dagger.$$

From the above construction, we know that

$$\kappa(p^*, q) \overset{(a)}{\geq} \kappa(p^\dagger, q) + \widehat{\varepsilon} \overset{(b)}{\geq} \kappa(p^\dagger, \bar{\omega}_{j+1}^\dagger) = \kappa(p^\dagger, q^\dagger)$$

where inequality (a) holds since $\widehat{\varepsilon} = p^* - p^\dagger \leq \kappa(p^*, q) - \kappa(p^\dagger, q)$ due to Assumption **1b**, and inequality (b) holds since $|\kappa(p^\dagger, \bar{\omega}_{j+1}^\dagger) - \kappa(p^\dagger, q)| \leq |\bar{\omega}_{j+1}^\dagger - q| \leq \widehat{\varepsilon}$ due to Assumption 2 and the fact that $q = \frac{\lambda_{i'}\rho_{i'}^*(q)\bar{\omega}_{i'} + \lambda_{i''}\rho_{i''}^*(q)\bar{\omega}_{i''}}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} \in ((j-1)\widehat{\varepsilon}, j\widehat{\varepsilon}]$.

- **Case 3** – if $\{i \in [m] : \rho_i^*(q) > 0\} = \{i', i''\}$ where $i' < i''$, in this case, suppose $\bar{\omega}_{i'} \in ((j'-1)\widehat{\varepsilon}, j'\widehat{\varepsilon}]$ and $\bar{\omega}_{i''} \in ((j''-1)\widehat{\varepsilon}, j''\widehat{\varepsilon}]$ for some $j', j'' \in [\lceil 1/\widehat{\varepsilon} \rceil]$ where $j' < j''$, then consider

$$\rho_{j'+1}^\dagger(q^\dagger) = \frac{\lambda_{i'}\rho_{i'}^*(q)}{\lambda_{j'+1}^\dagger}, \ \rho_{j''+1}^\dagger(q^\dagger) = \frac{\lambda_{i''}\rho_{i''}^*(q)}{\lambda_{j''+1}^\dagger};$$
$$\text{where } q^\dagger = \frac{\lambda_{j'+1}^\dagger\rho_{j'+1}^\dagger(q^\dagger)\bar{\omega}_{j'+1}^\dagger + \lambda_{j''+1}^\dagger\rho_{j''+1}^\dagger(q^\dagger)\bar{\omega}_{j''+1}^\dagger}{\lambda_{j'+1}^\dagger\rho_{j'+1}^\dagger(q^\dagger) + \lambda_{j''+1}^\dagger\rho_{j''+1}^\dagger(q^\dagger)}$$

From the above construction, we know that

$$\kappa(p^*, q) \overset{(a)}{\geq} \kappa(p^\dagger, q) + \widehat{\varepsilon} \overset{(b)}{\geq} \kappa(p^\dagger, q^\dagger)$$

where inequality (a) holds since $\widehat{\varepsilon} = p^* - p^\dagger \leq \kappa(p^*, q) - \kappa(p^\dagger, q)$ due to Assumption **1b**, and inequality (b) holds due to Assumption 2 and the following fact:

$$\begin{aligned}
|q - q^\dagger| &= \left| \frac{\lambda_{i'}\rho_{i'}^*(q)\bar{\omega}_{i'} + \lambda_{i''}\rho_{i''}^*(q)\bar{\omega}_{i''}}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} - \frac{\lambda_{j'+1}^\dagger\rho_{j'+1}^\dagger(q^\dagger)\bar{\omega}_{j'+1}^\dagger + \lambda_{j''+1}^\dagger\rho_{j''+1}^\dagger(q^\dagger)\bar{\omega}_{j''+1}^\dagger}{\lambda_{j'+1}^\dagger\rho_{j'+1}^\dagger(q^\dagger) + \lambda_{j''+1}^\dagger\rho_{j''+1}^\dagger(q^\dagger)} \right| \\
&= \left| \frac{\lambda_{i'}\rho_{i'}^*(q)\bar{\omega}_{i'} + \lambda_{i''}\rho_{i''}^*(q)\bar{\omega}_{i''}}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} - \frac{\lambda_{i'}\rho_{i'}^*(q)\bar{\omega}_{j'+1}^\dagger + \lambda_{i''}\rho_{i''}^*(q)\bar{\omega}_{j''+1}^\dagger}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} \right| \\
&\leq \frac{\lambda_{i'}\rho_{i'}^*(q)|\bar{\omega}_{j'+1}^\dagger - \bar{\omega}_{i'}| + \lambda_{i''}\rho_{i''}^*(q)|\bar{\omega}_{j''+1}^\dagger - \bar{\omega}_{i''}|}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} \\
&\leq \frac{\lambda_{i'}\rho_{i'}^*(q)\widehat{\varepsilon} + \lambda_{i''}\rho_{i''}^*(q)\widehat{\varepsilon}}{\lambda_{i'}\rho_{i'}^*(q) + \lambda_{i''}\rho_{i''}^*(q)} = \widehat{\varepsilon}
\end{aligned}$$

We also note that by construction, for any posterior mean $q \in \mathsf{supp}(\rho^*)$, the corresponding constructed posterior mean $q^\dagger \in \mathsf{supp}(\rho^\dagger)$ satisfies that

$$\rho^\dagger(q^\dagger) = \sum_{i \in [\lceil 1/\widehat{\varepsilon} \rceil + 1]} \rho_i^\dagger(q^\dagger)\lambda_i^\dagger = \rho^*(q) \tag{10}$$

Armed with the above observation $\kappa(p^*, q) \geq \kappa(p^\dagger, q^\dagger)$, we are now ready to show $\mathsf{Rev}_{\mathcal{I}}(p^*, \rho^*) \leq \mathsf{Rev}_{\mathcal{I}^\dagger}(p^\dagger, \rho^\dagger) + \widehat{\varepsilon}$:

$$\begin{aligned}
\mathsf{Rev}_{\mathcal{I}}(p^*, \rho^*) - \mathsf{Rev}_{\mathcal{I}^\dagger}(p^\dagger, \rho^\dagger) &= p^* \int_q \rho^*(q) D(\kappa(p^*, q))dq - p^\dagger \int_{q^\dagger} \rho^\dagger(q^\dagger) D(\kappa(p^\dagger, q^\dagger))dq^\dagger \\
&\overset{(a)}{\leq} p^* \int_q \rho^*(q) D(\kappa(p^*, q))dq - p^* \int_{q^\dagger} \rho^\dagger(q^\dagger) D(\kappa(p^\dagger, q^\dagger))dq^\dagger + \widehat{\varepsilon} \\
&= p^* \left(\int_q \rho^*(q) D(\kappa(p^*, q))dq - \int_{q^\dagger} \rho^\dagger(q^\dagger) D(\kappa(p^\dagger, q^\dagger))dq^\dagger \right) + \widehat{\varepsilon} \\
&\overset{(b)}{\leq} \widehat{\varepsilon}
\end{aligned}$$

where inequality (a) holds since we have $p^\dagger = p^* - \widehat{\varepsilon}$, and inequality (b) holds by the observation $\kappa(p^*, q) \geq \kappa(p^\dagger, q^\dagger)$ and (10). □