PASTA: A Unified Framework for Offline Assortment Learning*

Juncheng Dong[§1] Weibin Mo[§2] Zhengling Qi[3] Cong Shi[4]
Ethan X. Fang[1] Vahid Tarokh[1]

[§]Equal Contribution
[1]Duke University
[2]Purdue University
[3]George Washington University
[4]University of Miami

Abstract

We study a broad class of assortment optimization problems in an offline and data-driven setting. In such problems, a firm lacks prior knowledge of the underlying choice model, and aims to determine an optimal assortment based on historical customer choice data. The combinatorial nature of assortment optimization often results in insufficient data coverage, posing a significant challenge in designing provably effective solutions. To address this, we introduce a novel Pessimistic Assortment Optimization (PASTA) framework that leverages the principle of pessimism to achieve optimal expected revenue under general choice models. Notably, PASTA requires only that the offline data distribution contains an optimal assortment, rather than providing the full coverage of all feasible assortments. Theoretically, we establish the first finite-sample regret bounds for offline assortment optimization across several widely used choice models, including the multinomial logit and nested logit models. Additionally, we derive a minimax regret lower bound, proving that PASTA is minimax optimal in terms of sample and model complexity. Numerical experiments further demonstrate that our method outperforms existing baseline approaches.

1 Introduction

In business operations, a key challenge is selecting a product assortment that maximizes the seller's expected revenue, a problem commonly known as assortment optimization. Solving this problem requires understanding customer choice behavior for a given assortment, often modeled using established frameworks such as the widely studied multinomial logit (MNL) model [35]. This work tackles the offline assortment optimization problem by leveraging historical observational data. We propose a novel algorithm that efficiently identifies the optimal assortment under a broad class of choice models, without requiring prior knowledge of the model parameters.

Offline assortment optimization is of particular interest in the era of Big Data, as many companies have access to extensive customer data, and it is in their best interest to leverage

*This is an extended version of arXiv:2302.03821.

this existing asset to inform better decisions. In contrast, online learning for assortment optimization can be costly, time-consuming, and sometimes even infeasible due to various constraints [16]. However, existing works on data-driven assortment optimization mainly study the dynamic/online setting [5], where decision-makers adaptively learn the underlying choice model and optimize assortments in a trial-and-error fashion. Online algorithms often build upon the principle of optimism, prioritizing exploration of assortments with the potential to generate large revenues [11]. Such an exploration strategy ensures sufficient coverage of assortments, preventing the suboptimal ones from being mistakenly selected due to the estimation error of choice models. In comparison, the offline data distribution can hardly cover all potential assortments. For example, some assortments with obvious suboptimal performance are unlikely to be offered to customers and, therefore, are never observed in the offline data. Then, the revenues of these assortments are often poorly estimated, leading to likely suboptimal decisions. One cannot address this challenge through active assortment exploration because it is infeasible in the offline setting. This challenge, unique to the offline setting, goes by the name of *insufficient data coverage*. Due to this challenge, offline assortment optimization requires a different treatment compared with the optimism-based online algorithms, whose success hinges on sufficient assortment exploration.

To this end, we adopt the principle of pessimism, an approach that finds great success in offline learning problems [27]. Building on this principle, we propose a novel *Pessimistic ASsorTment leArning* (PASTA for short) framework for offline assortment optimization under *general* choice models. To identify an optimal assortment under insufficient data coverage, PASTA first constructs an uncertainty set for the underlying choice model based on the likelihood ratio test [37]. It then evaluates each assortment based on its worst-case revenue across all models within the uncertainty set and selects the assortment that maximizes the worst-case revenue by solving a max-min problem. To shed some light, we note that insufficient data coverage can potentially result in large estimation errors of the underlying choice model, where some suboptimal assortments' corresponding revenues can be over-estimated. PASTA addresses this challenge by leveraging robust optimization to prevent the selection of suboptimal assortments. In particular, when the uncertainty set contains the true choice model, the worst-case optimization objective serves as a valid lower bound for the true revenue function. If a suboptimal assortment is under-explored, the potential estimation error can cause significant underestimation of its revenue. Conversely, when an optimal assortment is sufficiently explored, the risk of underestimation is negligible. As a result, PASTA guarantees to rule out under-explored suboptimal assortments and recover an optimal assortment, without requiring sufficient exploration of all feasible assortments but only an optimal one. We note that assuming sufficient exploration of an optimal assortment is reasonable, as many historical assortments in the offline data are from sales experts with the goal of maximizing revenue.

From a technical perspective, the success of PASTA depends on the precise control of the size of its uncertainty set, governed by a scalar parameter $\alpha > 0$. Note that we construct the uncertainty set from a likelihood ratio test, where α serves as the test's critical value, guiding the selection of choice models likely to include the true underlying one. By selecting α appropriately, we show that PASTA is effective across a broad class of choice models. In particular, we demonstrate that when α is sufficiently large relative to the choice model of interest (e.g., the MNL model), it provides an upper bound on PASTA's regret, which we aim

to minimize. We characterize α following empirical process theory for a general model class [47], encompassing the celebrated MNL [36], latent class logit [26], and nested logit [52] models as special cases. Specifically, we derive an efficient choice of α that scales with D/n, where n denotes the sample size of the offline data, and D is the effective dimension of the choice model. Under this selection, PASTA achieves a regret upper bound of the order $\sqrt{\alpha} = \sqrt{D/n}$. We further show that this rate is minimax optimal, as it matches the lower bound under the MNL model.

We highlight that our theoretical guarantees of PASTA hold even in the challenging setting of insufficient data coverage. Rather than assuming the offline data includes all feasible assortments, PASTA only requires that an optimal assortment appears in the data with positive probability. For a model-agnostic framework with general regret guarantees, this is a necessary condition to identify an optimal assortment. The positivity assumption is indeed realistic, as logged assortments often include expert-selected, high-revenue sets. We note that this requirement can be weakened under additional model structure, but that would narrow PASTA's generality.

Major Contributions. Our contribution is threefold: (1) We introduce PASTA, a unified framework for offline assortment optimization under a broad class of choice models. To illustrate its generality, we apply PASTA to several widely used choice models, including the MNL, latent class logit, and nested logit (NL) models. For these models, we establish finite-sample regret bounds in the offline setting, which, to the best of our knowledge, are novel to the literature. (2) Under the MNL model, we derive the first minimax regret lower bound for assortment optimization in the offline setting, which justifies the minimax optimality of the proposed framework. (3) We thoroughly validate our framework with simulations that allow exact computation of the optimal assortment and regret.

Paper Organization. We first briefly discuss related works in Section 2. In Section 3, we formalize the offline assortment optimization problem and discuss its unique challenges. In Section 4, we elaborate on our PASTA framework and establish its general regret bounds. In Section 5, we instantiate PASTA on various choice models and establish regret guarantees for PASTA under these particular models. After presenting simulation studies in Section 6, we conclude the paper.

2 Related Work

Assortment Optimization. The assortment optimization problem is a fundamental decision-making challenge in business operations, owing to its broad applicability [31]. Consequently, extensive works aim to address this problem under various choice models, including the MNL model without constraints [45, 49], MNL model with unimodular constraints [12], NL model [13], ranking-based preference model [19], and consider-then-choose model [2]. These works assume known choice models and focus on the combinatorial optimization challenges associated with assortment planning [e.g., 15, 20, 22, 33, 50], whereas we consider the settings with unknown model parameters that require estimation based on the offline data.

Data-driven Assortment Optimization. On the other hand, in the current era of Big Data, we observe an increasing volume of works regarding data-driven assortment optimization [5].

The existing literature mainly focuses on dynamic assortment optimization (DAO) [8] where we repeatedly update the assortment based on the observed choices of customers, with the goal that the offered assortment converges to an optimal one. A wealth of research addresses this problem under a range of application scenarios. In particular, [41] study DAO under MNL model with capacity constraints; [11] propose an optimal algorithm for DAO under MNL model without constrains, achieving cumulative regrets independent of the total number of items that matches the minimax regret lower bound in terms of the horizon; [9] consider DAO under NL model; for more complex scenarios, [28] and [10] study DAO under MNL with high-dimensional context information and changing context information; [7] investigate the tiered DAO problems. In comparison with the aforementioned works, our work is the first to focus on the offline setting where the seller selects an assortment relying on pre-collected historical data, and does not actively explore new assortments. While we use "offline" to describe the source of data for data-driven decisions, some existing works in assortment optimization use "offline" to refer to either physical locations, i.e., brick-and-mortar stores [18] or decision-making problems without data [48]. These works are orthogonal to ours.

Pessimism. Our proposed PASTA framework builds upon the principle of pessimism, which has been successfully applied in offline reinforcement learning (RL) to find optimal policies using historical datasets. On the empirical side, it improves the performance of both the model-based and value-based approaches in the offline setting [e.g., 29, 32, 57]. The effectiveness of pessimism has also been analyzed and verified theoretically in the setting of RL [23, 27]. Notably, while the principle of pessimism has proven effective for offline learning, its exact implementations for different problems require careful instance-specific designs. To this end, the main contribution of our work is to take a pessimistic approach to offline assortment optimization problems and demonstrate its empirical and theoretical values. Moreover, our work differs from the above works by focusing on a decision-making problem with actions of combinatorial structures.

Distributionally Robust Optimization (DRO) and Robust Assortment Optimization (RAO). As detailed in Section 4, PASTA tackles the insufficient data coverage in offline learning by solving max–min programs over data-driven uncertainty sets. While this closely relates to work in DRO and RAO [3, 4, 14, 17, 42, 44, 51], our aim is fundamentally different: instead of robustness, PASTA seeks to identify the optimal assortment with regret guarantees using large-scale historical data, whereas, to the best of our knowledge, such optimality cannot be guaranteed by existing DRO and RAO approaches. Although DRO approaches also adopt max-min formulations and require the uncertainty set to contain the true choice model, PASTA further requires the uncertainty set to satisfy a finite-sample concentration condition (see Assumption 1). RAO likewise uses max–min programs, but its main focus is to efficiently solve specific robust optimization problems and find an assortment that achieves robustness. In comparison, our contribution is statistical: we use pessimism to address limited coverage and prove that the worst-case solution recovers the optimal assortment and achieves efficient finite-sample regret.

3 Offline Assortment Optimization under Insufficient Data Coverage

In Section 3.1, we introduce and formally define the offline assortment optimization problem. In Section 3.2, we explore the unique challenges rising from insufficient data coverage and illustrate them using a motivating example.

3.1 Offline Assortment Optimization

Let $[N] = \{1, 2, .., N\}$ denote the set of N distinct items. Denote the collection of assortments under consideration by $\mathbb{S} \subseteq 2^{[N]} \setminus \{\emptyset\}$. In the assortment optimization problem, when presented with an assortment $S \in \mathbb{S}$ offered by the seller, a customer makes a purchase $A \in S \cup \{0\}$ with $A = 0$ if no purchase is made. The seller subsequently receives a revenue R. The ultimate goal of assortment optimization is to find an optimal assortment $s^\star \in \mathbb{S}$ to maximize the expected revenue. Specifically, let $p_0(a|s)$ denote an unknown choice model which represents the probability of the customer purchasing $a \in s \cup \{0\}$ when s is offered, with $p_0(a|s) = 0$ for all $a \notin s$. The assortment optimization problem is to find an optimal assortment $s^\star$ that

$$s^\star \in \arg\max_{s \in \mathbb{S}} \left\{ \mathcal{V}_0(s) = \sum_{a \in s} p_0(a|s) r(s, a) \right\}, \tag{1}$$

where $r(s, a) = \mathbb{E}[R|S = s, A = a]$ represents the expected revenue when the customer is presented with the assortment $s \in \mathbb{S}$ and purchases the item $a \in s$. Here, $\mathcal{V}_0(s)$ in (1) represents the true expected revenue of assortment s. For optimization tractability, it is often assumed that the revenue $r(s, a) = r(a)$ only depends on the purchased item but not on the offered assortment [12, 22, 45]. We assume that the optimization problem in the form of (1) can be tractably solved. Our goal is to develop an offline learning method that further enjoys performance guarantees under the challenge of insufficient data coverage. We also assume that $r(s, a)$ is a known function to the seller, which incorporates the special case that the revenue R is a deterministic and known function of (s, a).

In the offline data-driven context, we choose assortments based on an offline dataset $\mathcal{D} = \{S_i, A_i\}_{i=1}^n$, consisting of n pairs of historically offered assortments and corresponding customer choices. We assume that they are independent and identically distributed (i.i.d.) samples of the random tuple (S, A), following the probability mass function $(s, a) \mapsto \pi_S(s) p_0(a|s)$ where $\pi_S(s)$ represents the probability of observing s in the offline data. When presented with the offline dataset, a common approach is to first estimate the true choice model p_0 with an estimator $\widehat{p}_n$, such as the maximum likelihood estimator (MLE), and then to solve (1) by replacing p_0 with $\widehat{p}_n$. However, this *as-if* optimization approach may fail due to the insufficient data coverage problem, which is the most critical challenge for offline learning [27, 43]. Specifically, insufficient data coverage implies that certain feasible assortments are entirely absent from the offline dataset, meaning $\pi_S(s) = 0$ for some $s \in \mathbb{S}$. In the next subsection, we present a simple example to demonstrate how insufficient data coverage poses a unique challenge for offline assortment optimization.

3.2 An Example of Insufficient Data Coverage

In the following example, we show that if an assortment never appears in the data due to insufficient data coverage, we cannot recover its expected revenue, and the optimal assortment may therefore remain ambiguous.

Consider a simple assortment optimization problem with two items, denoted as 1 and 2, and all possible assortments are $\{1\}, \{2\}, \{1,2\}$. Suppose that a customer makes the purchasing choice according to the following model, which is a special case of the latent class logit model. In particular, for $j \in \{1,2\}$, we assume

$$\begin{aligned} p_0(j|\{j\}) &= \frac{1}{2}\left\{\frac{v_j}{1+v_j} + \frac{\widetilde{v}_j}{1+\widetilde{v}_j}\right\}, \\ p_0(0|\{j\}) &= \frac{1}{2}\left\{\frac{1}{1+v_j} + \frac{1}{1+\widetilde{v}_j}\right\}, \\ p_0(j|\{1,2\}) &= \frac{1}{2}\left\{\frac{v_j}{1+v_1+v_2} + \frac{\widetilde{v}_j}{1+\widetilde{v}_1+\widetilde{v}_2}\right\}, \\ p_0(0|\{1,2\}) &= \frac{1}{2}\left\{\frac{1}{1+v_1+v_2} + \frac{1}{1+\widetilde{v}_1+\widetilde{v}_2}\right\}, \end{aligned} \tag{2}$$

where the choice of 0 means that the customer does not make a purchase, and the item-specific parameters $v_1, v_2, \widetilde{v}_1, \widetilde{v}_2 > 0$ are unknown to the seller. Then given the item-specific revenues $r(1), r(2) > 0$, the seller aims to choose from assortments $\{1\}, \{2\}, \{1,2\}$ to maximize the expected revenue by utilizing the offline data.

Suppose that in the historical dataset, the seller can only observe the samples of singleton assortments $\{1\}$ and $\{2\}$, i.e., $\pi_S(\{1\}), \pi_S(\{2\}) > 0$, while their union assortment $\{1,2\}$ is never observed, i.e., $\pi_S(\{1,2\}) = 0$. To better understand the challenge, suppose that we can precisely recover the customer choice probability for the two singleton assortments $\{1\}$ and $\{2\}$ and have $p_0(1|\{1\}) = p_0(0|\{1\}) = p_0(2|\{2\}) = p_0(0|\{2\}) = 1/2$, i.e., any singleton assortment yields a 50% purchase rate. These equalities imply that the unknown parameters $v_1, v_2, \widetilde{v}_1, \widetilde{v}_2$ satisfy

$$p(j|\{j\}) = \frac{1}{2}\left\{\frac{v_j}{1+v_j} + \frac{\widetilde{v}_j}{1+\widetilde{v}_j}\right\} = \frac{1}{2} \quad \Leftrightarrow \quad \widetilde{v}_j = \frac{1}{v_j}; \quad j = 1,2.$$

In other words, all the parameters in the parameter space $\{(v_1, v_2, \widetilde{v}_1, \widetilde{v}_2) = (v_1, v_2, v_1^{-1}, v_2^{-1}) : v_1, v_2 > 0\}$ satisfy the data equally well. For example, both of the parameters $(1,1,1,1)$ and $(10, 1/10, 1/10, 10)$ are feasible parameters to yield a purchase rate of 50% for singleton assortments. But these different parameters give different customer choice probabilities for $\{1,2\}$, leading to different expected revenues. As a consequence of this ambiguity, it is challenging for the seller to determine which assortment is optimal.

For example, consider a scenario where item 1 has revenue $r(1) = \frac{1}{5}$, and item 2 has revenue $r(2) = 1$. In this case, the revenues of the singleton assortments are given by $\mathcal{V}_0(\{1\}) = \frac{1}{10}$ and $\mathcal{V}_0(\{2\}) = \frac{1}{2}$. To determine the optimal assortment, the seller needs to compare assortment $\{1,2\}$'s expected revenue $\mathcal{V}_0(\{1,2\})$ with $\mathcal{V}_0(\{1\})$ and $\mathcal{V}_0(\{2\})$. However, the seller cannot determine the optimal assortment: if the true parameter is $(1,1,1,1)$, then $\mathcal{V}_0(\{1,2\}) = \frac{2}{5} <$

$\mathcal{V}_0(\{2\})$, implying that $\{2\}$ is the optimal assortment; conversely, if the true parameter is $(10, \frac{1}{10}, \frac{1}{10}, 10)$, then $\mathcal{V}_0(\{1,2\}) = \frac{303}{555} > \mathcal{V}_0(\{2\})$, making $\{1,2\}$ the optimal assortment.

This motivating example illustrates that, when certain assortments are missing from the offline data due to insufficient coverage, their revenues cannot be recovered, and thus it is challenging to determine the optimal assortment. This challenge also makes it difficult to establish performance guarantees for the as-if optimization approach. For instance, in the example above, even with an infinite amount of offline data, the MLE is only guaranteed to converge to some point within the parameter set $\{(v_1, v_2, v_1^{-1}, v_2^{-1}) : v_1, v_2 > 0\}$ [40], based on which the seller cannot guarantee finding an optimal assortment. It remains unclear how to perform assortment optimization in this case.

4 Pessimistic Assortment Learning

Despite the aforementioned challenges of insufficient coverage, our insight is that finding an optimal assortment may not necessarily require $\pi_S(s) > 0$ everywhere but only at one optimal assortment $s^\star$. In particular, when computing for an optimal assortment $s^\star$ following (1), it is unnecessary to estimate $p_0(a|s)$ well for all $s \neq s^\star$, as long as we can rule out those suboptimal assortments. Building on this insight, we propose a novel PASTA framework, where we only require the coverage at optimum, i.e., $\pi_S(s^\star) > 0$. This is a much weaker and more realistic assumption than that of uniform coverage required by the as-if approach, and much more likely to hold in practice.

4.1 PASTA Framework

Let $\mathcal{P}$ be a set of conditional mass functions $p(a|s)$ for $A|S$, representing a family of models that includes the true choice model $p_0(a|s)$. For example, $\mathcal{P}$ can be a family of MNL models with different parameters. Following the principle of pessimism, PASTA first constructs an uncertainty set for p_0, which serves as a set estimate of p_0. Specifically, given an offline dataset $\mathcal{D} = \{S_i, A_i\}_{i=1}^n$ where n is the sample size, we let the likelihood-based loss function $\widehat{L}_n(p)$ (smaller the better) be

$$\widehat{L}_n(p) := -\frac{1}{n}\sum_{i=1}^{n} \log p(A_i|S_i); \quad p \in \mathcal{P}.$$

Let $\widehat{p}_n \in \arg\min_{p\in\mathcal{P}} \widehat{L}_n(p)$ be the maximum likelihood estimator (MLE). For a set estimate of the true choice model p_0, we consider the set of α_n-minimizers of the loss function $\widehat{L}_n(\cdot)$ defined as

$$\Omega_n(\alpha_n) := \left\{ p \in \mathcal{P} : \widehat{L}_n(p) - \widehat{L}_n(\widehat{p}_n) \leq \alpha_n \right\}, \tag{3}$$

where $\alpha_n > 0$ is the critical value to control the size of the uncertainty set $\Omega_n(\alpha_n)$. Note that $\Omega_n(\alpha_n)$ is essentially a confidence region induced by the likelihood ratio test. Without identification assumptions, Manski and Tamer [34] point out that the consistency of MLE $\widehat{p}_n$ is ambiguous, while the set estimate $\Omega_n(\alpha_n)$ is consistent with a proper choice of α_n that $\mathbb{P}\{\Omega_n(p_0) \in \alpha_n\} \to 1$. We provide the details of how to determine α_n in Section 4.3 later. In the sequel, we drop α_n in $\Omega_n(\alpha_n)$ when there is no confusion.

After constructing uncertainty set Ω_n, PASTA solves the robust optimization problem

$$\widehat{s}_{\mathfrak{p},n} \in \arg\max_{s\in\mathbb{S}} \min_{p\in\Omega_n} \left\{ \mathcal{V}(s;p) := \sum_{a\in s} p(a|s) r(s,a) \right\}. \tag{4}$$

Here, for any $p \in \mathcal{P}$, we let $\mathcal{V}(s;p)$ be the value function of assortment s under choice model p. For a given assortment $s \in \mathbb{S}$, the inner layer of minimization computes the worst-case revenue among all possible choice models within the uncertainty set. In particular, when the uncertainty set contains the true model, i.e., $p_0 \in \Omega_n(\alpha_n)$, the optimal value $\min_{p\in\Omega_n} \mathcal{V}(s;p)$ is a proper lower bound of the true value $\mathcal{V}_0(s)$ that $\mathcal{V}_0(s) = \mathcal{V}(s;p_0) \geq \min_{p\in\Omega_n} \mathcal{V}(s;p)$. In this way, the max-min value attained by (4) is also a valid lower bound of the optimal revenue $\max_{s\in\mathbb{S}} \mathcal{V}_0(s)$ attained by the true optimal assortment. To shed more intuition of PASTA, we compare it with the commonly used as-if approach that

$$\widehat{s}_{\mathrm{ML},n} \in \arg\max_{s\in\mathbb{S}} \left\{ \mathcal{V}(s;\widehat{p}_n) \right\}, \tag{5}$$

where we replace p_0 with an estimate $\widehat{p}_n$ as if $\widehat{p}_n$ is the true model. Note that a less observed or never observed assortment s in the data often incurs some larger estimation error for $\widehat{p}_n(\cdot|s)$. Consequently, this potentially leads to some overestimation of its true expected revenue $\mathcal{V}(s;p_0)$. That is, the estimated revenue $\mathcal{V}(s;\widehat{p}_n)$ can be much larger than its true value $\mathcal{V}(s;p_0)$. Such estimation errors increase the risk of the as-if approach mistakenly selecting a suboptimal assortment. In contrast, PASTA quantifies the estimation error via the uncertainty set Ω_n, and robustifies assortment optimization against the estimation error. Specifically, when the estimated revenue $\mathcal{V}(s;\widehat{p}_n)$ for an assortment s is highly uncertain, the worst-case revenue $\min_{p\in\Omega_n} \mathcal{V}(s;p)$ substantially underestimates $\mathcal{V}_0(s)$. Consequently, the outer layer of (4) selects an alternative assortment with a relatively higher worst-case revenue. In this way, PASTA prevents the suboptimal assortments from being mistakenly selected due to their highly uncertain estimated revenues. To be successful, PASTA only needs that the underestimation of the revenue at an optimal assortment is mild. This explains why the PASTA algorithm only requires some coverage at the optimum, a much weaker and thus more realistic assumption than the uniform coverage assumption.

In what follows, we establish the theoretical guarantees for our proposed framework.

4.2 General Regret Guarantees

We show that the proposed PASTA framework (4) achieves a general regret guarantee under the weak assumption of coverage at optimum, where we only require that $\pi_S(s^\star) > 0$ for an optimal assortment $s^\star$ that solves (1). We defer the detailed proofs in this subsection to the AppendixC.

Notations For any vector x, let $x^\intercal$ and $\|x\|_2$ respectively denote the transpose and ℓ_2-norm of x. For any finite set A, let $|A|$ denote the cardinal number of A. For any two sequences $\{\varpi(n)\}_{n\geq 1}$ and $\{\gamma(n)\}_{n\geq 1}$, we write $\varpi(n) \gtrsim \gamma(n)$ (respectively $\varpi(N) \lesssim \gamma(n)$) whenever there exist constants $c_1 > 0$ (resp. $c_2 > 0$) such that $\varpi(n) \geq c_1\gamma(n)$ (respectively $\varpi(n) \leq c_2\gamma(n)$). Moreover, we write $\varpi(n) \eqsim \gamma(n)$ whenever $\varpi(n) \gtrsim \gamma(n)$ and $\varpi(n) \lesssim \gamma(n)$.

Our analysis involves the likelihood ratio-based uncertainty quantification, and we adopt the Hellinger distance [6, 53] for such quantification. In particular, given an assortment s and two choice models $p_1(a|s), p_2(a|s)$, let

$$h^2(p_1, p_2, s) = \frac{1}{2} \sum_{a \in s \cup \{0\}} \left(\sqrt{p_1(a|s)} - \sqrt{p_2(a|s)} \right)^2$$

be their squared Hellinger distance. Meanwhile, note that the offline assortment S is stochastic. We further consider the *generalized squared Hellinger distance* between two choice models p_1 and p_2 by taking the expectation over S with respect to the data distribution π_S that

$$H^2(p_1, p_2) = \mathbb{E}_S[h^2(p_1, p_2, S)]. \tag{6}$$

To evaluate offline assortment optimization algorithms, we adopt the following regret as the performance metric. In particular, for any estimated optimal assortment $\widehat{s}$, we let its regret be

$$\mathcal{R}(\widehat{s}) = \mathcal{V}_0(s^\star) - \mathcal{V}_0(\widehat{s}),$$

which represents the gap between the expected revenues of an optimal assortment and the estimated one $\widehat{s}$. Note that by definition (4), $\mathcal{V}(s; p_0) = \mathcal{V}_0(s)$ for all $s \in \mathbb{S}$.

We then establish an upper bound for $\mathcal{R}(\widehat{s}_{\mathfrak{p},n})$, where $\widehat{s}_{\mathfrak{p},n}$ is the solution from PASTA in (4). We begin with the following lemma that sheds light on our proof techniques via pessimistic optimization.

Lemma 4.1. If $\pi_S(s^\star) > 0$ for an optimal assortment $s^\star$, and the uncertainty set (3) includes the true choice model, i.e., $p_0 \in \Omega_n$, then

$$\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \le \max_{p \in \Omega_n} \left\{ \mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p) \right\} \le \mathcal{C} \max_{p \in \Omega_n} \sqrt{H^2(p, p_0)},$$

where $\mathcal{C} = r_{s^\star} \sqrt{1/\pi_S(s^\star)}$ is a constant independent of the sample size n, and $r_{s^\star} = \max_{j \in s^\star} r(s^\star, j)$ is the maximal revenue among the items in $s^\star$.

Proof. Proof of Lemma 4.1 To prove the first inequality, we have

$$\begin{aligned} \mathcal{V}(s^\star; p_0) - \mathcal{V}(\widehat{s}_{\mathfrak{p},n}; p_0) &\le \mathcal{V}(s^\star; p_0) - \min_{p \in \Omega_n} \mathcal{V}(\widehat{s}_{\mathfrak{p},n}; p) \\ &\le \mathcal{V}(s^\star; p_0) - \min_{p \in \Omega_n} \mathcal{V}(s^\star; p) = \max_{p \in \Omega_n} \left\{ \mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p) \right\}, \end{aligned}$$

where the first inequality above holds since $p_0 \in \Omega_n$, and the second inequality holds because $\widehat{s}_{\mathfrak{p},n}$ solves (4). In the Appendix C.1, by direct applications of Lemmas E.3 and E.4, we prove that for any $p \in \mathcal{P}$,

$$|\mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p)| \le \mathcal{C} \sqrt{H^2(p, p_0)},$$

which concludes the proof. □

We compare Lemma 4.1 with the typical theoretical justifications for the as-if approaches [30]. Specifically, in the regret analysis for the as-if optimization, the regret is typically upper bounded by a uniform estimation error $\sup_{s \in \mathbb{S}} \left| \mathcal{V}(s; \widehat{p}_n) - \mathcal{V}(s; p_0) \right|$ for values across all $s \in \mathbb{S}$ [38].

This is to control the error caused by the greedy policy, i.e., $\left|\mathcal{V}(\widehat{s}_{\mathrm{ML},n}, \widehat{p}_n) - \mathcal{V}(\widehat{s}_{\mathrm{ML},n}, p_0)\right|$, where $\widehat{s}_{\mathrm{ML},n}$ is the solution of the as-if optimization in (5). Such an upper bound typically entails that we estimate the value function $\mathcal{V}(s;p)$ uniformly well for all $s \in \mathbb{S}$. It further explains why an estimate-then-optimize approach requires the strong assumption that $\pi_S(s) > 0$ for all $s \in \mathbb{S}$. In contrast, Lemma 4.1 suggests that we only need to control the value function's estimation error at $s^\star$, i.e., $\max_{p\in\Omega_n} \left\{\mathcal{V}(s^\star;p_0) - \mathcal{V}(s^\star;p)\right\}$, as long as the uncertainty set Ω_n contains the true model p_0. This explains why PASTA only requires the coverage $\pi_S(s^\star) > 0$ at the optimal assortment $s^\star$.

To leverage Lemma 4.1 for establishing theoretical guarantees of PASTA, we further need *(i)* the validity of the uncertainty set Ω_n that $p_0 \in \Omega_n$ with high probability, and *(ii)* an upper bound of the set estimation error $\max_{p\in\Omega_n} H(p, p_0)$, measured by the maximum generalized Hellinger distance in (6) between the uncertain models within Ω_n and the true model p_0. We summarize the assumptions and conditions in Assumption 1, which serve as sufficient conditions for establishing the finite-sample regret guarantees of PASTA. Without loss of generality, we assume that the revenue function is non-negative, i.e., $r(a) \geq 0$ for all $a \in [N]$, and $r(0) = 0$ that no purchase incurs zero revenue.

Assumption 1.
(I) The true choice model is correctly specified by the family $\mathcal{P}$, i.e., $p_0 \in \mathcal{P}$.
(II) [Coverage at Optimum] For the offline data generating process, there exists $s^\star$ as an optimal assortment solving (1), such that $\pi_S(s^\star) > 0$.
(III) [Uncertainty Set Validity] For any $0 < \delta < 1$, with probability at least $1-\delta$, there exists a constant $\alpha_n(\delta)$ depending on δ such that $p_0 \in \Omega_n(\alpha_n(\delta))$, and $\sup_{p\in\Omega_n(\alpha_n(\delta))} H^2(p, p_0) \lesssim \alpha_n(\delta)$.

In Assumption 1 (I), we focus on a correctly specified and fixed model family $\mathcal{P}$, although our arguments can be extended to a family growing in n that approximates p_0 via sieve methods [53]. In Assumption 1 (II), we highlight that PASTA only requires the coverage at optimum, a much weaker condition compared to the uniform coverage that $\inf_{s\in\mathbb{S}} \pi_S(s) > 0$. As PASTA is a unified framework for general choice models, and our goal is to establish a generic regret guarantee without relying on the specific choice model assumptions, we believe that this is the minimal requirement on the offline dataset to identify an optimal assortment.

In Assumption 1 (III), we present the uncertainty quantification validity required by PASTA. In Section 5, we show that Assumption 1 (III) holds under different choice model families. In particular, our likelihood-based uncertainty set $\Omega_n(\alpha_n)$ depends on a critical value α_n, which needs to be selected in an *efficient* manner. On one hand, α_n cannot be too small, such that $\Omega_n(\alpha_n)$ contains p_0 with high probability. In fact, α_n needs to be at least as large as the difference $\widehat{L}_n(p_0) - \widehat{L}_n(\widehat{p}_n)$, which is associated with the convergence rate of the likelihood ratio test statistic. On the other hand, α_n cannot be too large, such that the entire set $\Omega_n(\alpha_n)$ remains close to the truth p_0. In Section 4.3, based on the empirical process theory, we derive an efficient α_n for any choice model family $\mathcal{P}$.

Notably, the distance H defined in (6) only depends on the distance between $p_1(\cdot|s)$ and $p_2(\cdot|s)$ at the assortment s with positive coverage $\pi_S(s) > 0$. As a result, the last part of Assumption 1 (III) only needs that for any $p \in \Omega_n(\alpha_n)$, $p(\cdot|s)$ is close to $p_0(\cdot|s)$ for s with $\pi_S(s) > 0$. This eliminates the need to identify or accurately estimate $p_0(\cdot|s)$ for those assortments s with no

coverage, thus effectively addressing the issue of insufficient data coverage.

We then establish the generic regret bound for PASTA under Assumption 1.

Theorem 4.1. Let $r_{s^\star}$ be defined as in Lemma 4.1. Under Assumption 1, for any $0 < \delta < 1$, with probability at least $1 - \delta$, we have $\mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p) \lesssim r_{s^\star} \sqrt{\alpha_n(\delta)/\pi_S(s^\star)}$ uniformly for all $p \in \Omega_n(\alpha_n(\delta))$. Following Lemma 4.1, it further implies that, with probability at least $1 - \delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ satisfies

$$\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \lesssim r_{s^\star} \sqrt{\frac{\alpha_n(\delta)}{\pi_S(s^\star)}}.$$

We point out that the two constants in the regret bound, $\pi_S(s^\star)$ and $r_{s^\star}$, only depend on the optimal $s^\star$. In particular, the regret decreases as $\pi_S(s^\star)$ increases, that is, when $s^\star$ is observed more frequently. In addition, we note that the choice of critical value α_n is crucial for obtaining a valid and sharp regret bound. Specifically, the smallest α_n such that Assumption 1 (III) is fulfilled leads to the tightest regret bound in Theorem 4.1. In the next subsection, we derive an efficient choice of the critical value α_n, which minimizes the regret while satisfying Assumption 1 (III).

4.3 Validity of Uncertainty Set

We present an efficient approach to determining the critical value such that Assumption 1 (III) holds for general choice model family $\mathcal{P}$. Building upon the empirical process theory [47], we first characterize the complexity of $\mathcal{P}$ by calculating its bracketing number. Next, we derive a balance equation to determine an efficient choice of the critical value α_n for the uncertainty set, which is also the learning rate of conditional density estimation of $\mathcal{P}$ via MLE [6]. Finally, we show that Assumption 1 (III) holds in Theorem 4.2. Notably, our justifications only rely on the complexity of $\mathcal{P}$ instead of the particular form of the choice model, making our analysis applicable to a wide class of choice models.

We first introduce the definition of bracketing number, which is the tool we use to characterize the complexity of $\mathcal{P}$. Note that $\mathcal{P}$ is essentially a class of measurable functions from domain $\mathbb{S} \times [N]$ to $[0, 1]$. In particular, $p \in \mathcal{P}$ is a mapping from $(s, a) \in \mathbb{S} \times [N]$ to $p(a|s) \in [0, 1]$.

Definition 4.1 (Bracketing Number)**.** A class of measurable functions $\mathcal{B} \subseteq \{\mathbb{S} \times [N] \to [0, 1]\}$ is an H-bracket-cover of $\mathcal{P}$ at scale $\varepsilon > 0$, if for any $p \in \mathcal{P}$, there exists $b_1, b_2 \in \mathcal{B}$, such that $H(b_1, b_2) \leq \varepsilon$, and for any $(s, a) \in \mathbb{S} \times [N]$, $b_1(a|s) \leq p(a|s) \leq b_2(a|s)$. The H-bracketing number at scale $\varepsilon > 0$ is the cardinality of the smallest $\mathcal{B}$ satisfying the above, and is denoted as $\mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H)$.

In the literature, $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H)$ is often referred to as the *bracketing entropy* [47]. Its dependency on ε as $\varepsilon \to 0^+$ typically characterizes the minimax learning rate (lower bound) on the function class $\mathcal{P}$ [6, 55], and also determines the convergence rate (upper bound) of MLE [53]. One typical example is the parametric entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H) \lesssim D \log(1/\varepsilon)$, which is achievable by a parametric class $\mathcal{P}$ with D-dimensional unrestricted parameters, or by a Vapnik–Chervonenkis (VC) function class with VC-dimension D. Another example is the nonparametric entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H) \lesssim \varepsilon^{-\eta}$, which is achievable by an infinite-dimensional

class of smooth functions on a compact domain differentiable up to a certain order, where $\eta > 0$ is a constant related to the smoothness of functions.

For the assortment optimization problem of our interest, the domain of $\mathcal{P}$ is $\mathbb{S} \times [N]$, which is finite. This implies that we can parameterize $\mathcal{P}$ as a finite-dimensional family with dimension up to $N \cdot 2^N$. Thus, we consider parametric entropy in all of our applications in the next section. Meanwhile, to obtain sharp bounds, we devote most of our discussion to characterize the effective dimension D that is potentially much smaller than $N \cdot 2^N$. We note that our theoretical results are extendable to an infinite-dimensional customer choice model family $\mathcal{P}$. Such results are of interest if we consider a nonparametric contextual assortment optimization problem, where the customer choice depends on the presented assortment as well as some context-specific covariates. We leave this extension as a future work.

Using the bracketing entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H)$ to measure the complexity of $\mathcal{P}$, we now elaborate on our approach for an efficient choice of $\alpha_n > 0$ for any model family $\mathcal{P}$. Specifically, we determine the critical value α_n based on the following *balance equation*:

$$\int_{C_1\alpha_n}^{C_2\sqrt{\alpha_n}} \sqrt{\log \mathcal{N}_{[]}\left(u/C_3, \mathcal{P}, H\right)} \mathrm{d}u \leq C_4 \sqrt{n} \alpha_n, \tag{7}$$

where $C_1, C_2, C_3, C_4 \in (0, +\infty)$ are some universal constants. In Theorem 4.2 below, we prove that any α_n satisfying (7) leads to an uncertainty set $\Omega_n(\alpha_n)$ satisfying Assumption 1 (III) with high probability. Moreover, Lemma 4.2 shows that the solutions to (7) guarantee a monotonic property that allows us to find the most efficient choice of α_n.

Lemma 4.2. Suppose that $\alpha_n > 0$ satisfies (7). Then every $\alpha \geq \alpha_n$ also satisfies (7).

By Lemma 4.2, it is natural to identify a smallest $\alpha_n > 0$ satisfying (7), such that the same relationship holds for every $\alpha \geq \alpha_n$. This approach results in the smallest critical value for the uncertainty set and thus leads to the tightest regret upper bound for PASTA, as implied by Theorem 4.1. To present a concrete example, if we have a parametric entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H) \lesssim D \log(1/\varepsilon)$, then a critical value solving (7) is $\alpha_n \eqsim (D/n) \log(n/D)$, which is also the typical parametric estimation rate.

Note that (7) also characterizes the convergence rate of the MLE for both unconditional [53] and conditional [6] density estimation. In particular, Bilodeau et al. [6, Theorem 5] established that

$$H^2(\widehat{p}_n, p_0) \lesssim \alpha_n \tag{8}$$

for MLE $\widehat{p}_n$. However, this is not yet sufficient for any performance guarantee for the as-if optimization, since $\widehat{p}_n(\cdot|s)$ and $p_0(\cdot|s)$ can exhibit arbitrary relationships at s with no coverage, i.e., when $\pi_S(s) = 0$. Consequently, although the MLE $\widehat{p}_n$ converges to p_0 in the metric H, the challenge of insufficient data coverage, as demonstrated by the motivating example in Section 3.2, remains unsolved for the as-if optimization approach.

In comparison, we establish theoretical guarantees of PASTA under the challenge of insufficient data coverage, only requiring the conditions in Assumption 1. It is important to note that the existing results for the convergence rates of MLE [6] are not sufficient for the statements required by Assumption 1 (III). Specifically, we note that by the definition of $\Omega_n(\alpha_n)$, MLE $\widehat{p}_n$ belongs to $\Omega_n(\alpha_n)$. Then the requirement $\max_{p \in \Omega_n(\alpha_n)} H^2(p, p_0) \lesssim \alpha_n$ in Assumption 1 (III)

is a stronger condition than (8). Furthermore, the condition $p_0 \in \Omega_n(\alpha_n)$ in Assumption 1 (III) is equivalent to $\widehat{L}_n(p_0) - \widehat{L}_n(\widehat{p}_n) \lesssim \alpha_n$, which does not follow from (8) immediately. To this end, we establish these conditions below under (7). For simplicity, we present below a simplified version of the theorem, while the full statement can be found in Theorem A.1 in Appendix A.

Theorem 4.2. Consider the set estimate $\Omega_n(\alpha)$ in (3). Suppose that the critical value $\alpha_n > 0$ satisfies (7). Then for every $\alpha \geq \alpha_n$, with probability at least $1 - C_5 e^{-C_6 n\alpha}$, we have

$$p_0 \in \Omega_n(\alpha), \quad \text{and} \quad \sup_{p \in \Omega_n(\alpha)} H^2(p, p_0) \leq C_7 \alpha.$$

Here, $\{C_j\}_{j=1}^7$ are universal constants characterized in Theorem A.1.

We remark that our critical value determination is tight in terms of satisfying Assumption 1 (III) for various families $\mathcal{P}$. In particular, Bilodeau et al. [6, Theorem 2] has established that for $\mathcal{P}$ with a parametric entropy $\log \mathcal{N}_{[]}(\epsilon, \mathcal{P}, H) \asymp D \log(1/\epsilon)$, any estimate $\widetilde{p}_n$ of p_0 satisfies $H^2(\widetilde{p}_n, p_0) \gtrsim D/(n \log n)$ in the worst-case problem instance. This further implies that for any set estimate $\widetilde{\Omega}_n$, we have $\sup_{p \in \widetilde{\Omega}_n} H^2(p, p_0) \gtrsim D/(n \log n)$ in the worst-case. Here, the minimax rate $D/(n \log n)$ matches (up to logarithmic factors) with the upper bound $\sup_{p \in \Omega_n(\alpha)} H^2(p, p_0) \lesssim (D/n) \log(n/D)$ obtained in Theorem 4.2 from (7), demonstrating that our critical value determination through (7) is tight. A similar minimax rate optimality also holds for a nonparametric entropy $\log \mathcal{N}_{[]}(\epsilon, \mathcal{P}, H) \asymp \epsilon^{-\eta}$ with $0 < \eta \leq 2$, corresponding to a sufficiently smooth family $\mathcal{P}$.

Based on Theorem 4.2, we conclude the regret guarantee for PASTA based on the critical value α_n from (7) as in Corollary 4.1 below. To obtain a high-probability statement as in Assumption 1 (III), for every fixed $\delta > 0$, we further need $\alpha \geq \frac{1}{C_6 n} \log \frac{C_5}{\delta}$ in Theorem 4.2. Therefore, we inflate the critical value from (7) by such an amount, which is mild since α_n satisfying (7) is at least in the order of n^{-1} even in the parametric entropy case.

Corollary 4.1. Suppose that the assumptions of Theorem 4.2 hold, and let $\alpha_n^\circ > 0$ satisfy (7). Fix $0 < \delta < 1$, and let $\alpha_n(\delta) = \alpha_n^\circ + \frac{1}{C_6 n} \log \frac{C_5}{\delta}$. Under Assumption 1 (I) $p_0 \in \mathcal{P}$ and (II) $\pi_S(s^\star) > 0$, with probability at least $1 - \delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ satisfies

$$\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \lesssim r_{s^\star} \sqrt{\frac{\alpha_n(\delta)}{\pi_S(s^\star)}}.$$

Corollary 4.1 concludes the performance guarantee of PASTA with three required ingredients: the correct specification of the choice model family $\mathcal{P}$, the coverage at optimum, as well as the critical value determination via (7). For the applications in specific choice models, it remains to determine the critical value α_n in (7), which involves the proper characterization of the complexity of $\mathcal{P}$ via the bracketing entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H)$. To this end, we instantiate on several of the most widely used choice model families $\mathcal{P}$, and establish the corresponding regret guarantees in the next section.

5 Applications

We apply Corollary 4.1 to several widely used choice models and establish the regret of PASTA under these models. We consider three of the most commonly used choice models – the MNL [36], latent logit [26], and nested logit [52] models. Throughout this section, we assume that Assumption 1 (I) $p_0 \in \mathcal{P}$ and (II) $\pi_S(s^\star) > 0$ hold. For each of the choice model family $\mathcal{P}$ to consider, we first characterize its complexity via the bracketing entropy $\log \mathcal{N}_{[]}(\varepsilon, \mathcal{P}, H)$, then we determine the critical value from (7), and finally apply Corollary 4.1 to establish PASTA's regret guarantee. All the proofs are provided in the Appendix D.

For a given family $\mathcal{P}$, we let $\theta \in \Theta$ denote the underlying parameter, and Θ is a pre-specified parameter space. The form of a customer choice model $p(a|s;\theta)$ is known up to the unknown parameter θ. Hence, the family of choice models is $\mathcal{P} = \{p(a|s;\theta) : \theta \in \Theta\}$. For each model discussed in this section, there exist $\mathcal{P}$-dependent constants $D_{\mathcal{P}}, C_{\mathcal{P}} < +\infty$, where $D_{\mathcal{P}}$ is the effective dimension of the family $\mathcal{P}$, and $C_{\mathcal{P}}$ is based on the boundedness of the problem. Let n be the sample size of the offline dataset. For fixed $0 < \delta < 1$, with probability at least $1-\delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ takes the form

$$\begin{aligned} &\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \lesssim r_{s^\star} \sqrt{\frac{\alpha_n(\delta)}{\pi_S(s^\star)}}; \\ &\alpha_n(\delta) \asymp \frac{D_{\mathcal{P}}}{n} \log \frac{C_{\mathcal{P}} n}{D_{\mathcal{P}}} + \frac{1}{n} \log \frac{1}{\delta}. \end{aligned} \tag{9}$$

Next, we investigate specific choice models.

5.1 Multinomial Logit (MNL) Model

The MNL model is arguably the most widely used discrete choice model in assortment optimization literature [21]. Given the item-specific features $\{x_j\}_{j\in[N]}$ where $x_j \in \mathbb{R}^d$, an MNL model with parameter $\theta \in \Theta \subset \mathbb{R}^d$ assumes that customer's preference for the j-th item is proportional to $\exp(x_j^\intercal \theta)$. Specifically, the customer choice probability under MNL with parameter θ is

$$p(a|s;\theta) = \begin{cases} \frac{\exp(x_a^\intercal \theta)}{1+\sum_{j\in s}\exp(x_j^\intercal \theta)} & a \in s; \\ \frac{1}{1+\sum_{j\in s}\exp(x_j^\intercal \theta)} & a = 0. \end{cases} \tag{10}$$

We impose the following mild boundedness assumptions. In particular, we assume that the item-specific feature vectors $\{x_j\}_{j\in[N]} \subseteq \mathbb{R}^d$ are deterministic and uniformly bounded, and the parameter space Θ is also bounded.

Assumption 2. In the MNL model (10), there exist constants $C_x, C_\Theta < +\infty$, such that $\max_{j\in[N]} \|x_j\|_2 \le C_x$ and $\sup_{\theta\in\Theta} \|\theta\|_2 \le C_\Theta$.

We are ready to establish the regret of PASTA under the MNL model.

Theorem 5.1. Consider the MNL model (10) satisifying Assumption 2. Let $D_{\mathcal{P}} = d$ and $C_{\mathcal{P}} = C_x$. Under Assumptions 1 (I) and (II), for $n \gtrsim D_{\mathcal{P}} \log(n/D_{\mathcal{P}})$ and every $\delta \in (0,1)$, with probability at least $1-\delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ satisfies (9).

5.2 Latent Class Logit (LCL) Model

The LCL model is closely related to the MNL model, as it assumes K distinct MNL models (10) as components, and the LCL model is the mixture of them [26]. Specifically, let $\Delta(K) = \{(\lambda_1, \cdots, \lambda_K) : \lambda_1, \cdots, \lambda_K \geq 0, \lambda_1 + \cdots + \lambda_K = 1\}$ represent the probability simplex in $\mathbb{R}^K$. An LCL model is parametrized by $\theta = (\theta_1, \cdots, \theta_K; \lambda_1, \cdots, \lambda_K)$, where $\theta_k \in \widetilde{\Theta} \subset \mathbb{R}^d$ is the parameter in the k-th MNL model for $k \in [K]$, and $(\lambda_1, \ldots, \lambda_K) \in \Delta(K)$ consists of the mixture probabilities. The parameter space can be written as $\Theta = \widetilde{\Theta}^K \times \Delta(K) \subset \mathbb{R}^{Kd+K}$. With item features $x_j \in \mathbb{R}^d$ for $j \in [N]$, the custsomer choice probability under LCL with parameter θ is

$$p(a|s;\theta) = \begin{cases} \sum_{k=1}^K \frac{\lambda_k \exp(x_a^\intercal \theta_k)}{1+\sum_{j\in s} \exp(x_j^\intercal \theta_k)} & a \in s; \\ \sum_{k=1}^K \frac{\lambda_k}{1+\sum_{j\in s} \exp(x_j^\intercal \theta_k)} & a = 0. \end{cases} \tag{11}$$

Theorem 5.2. Consider the LCL model (11) with the component MNL models satisfying Assumption 2. Let $D_{\mathcal{P}} = Kd + K - 1$ and $C_{\mathcal{P}} = KN \exp(C_\Theta C_x)$. Under Assumptions 1 (I) and (II), for $n \gtrsim D_{\mathcal{P}} \log(n/D_{\mathcal{P}})$ and every $\delta \in (0,1)$, with probability at least $1-\delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ satisfies (9).

5.3 Nested Logit (NL) Model

Despite its widespread use, the MNL model assumes the Independence of Irrelevant Alternatives (IIA) [39], which can limit its applicability in more complex assortment problems. To this end, the NL model [52] assumes a partition of all available items into K exclusive baskets $\{B_k\}_{k=1}^K$ such that $B_k \subset [N]$, $B_i \bigcap B_j = \emptyset$ for $i \neq j$, and $\bigcup_{k\in[K]} B_k = [N]$ [24]. The NL model is parametrized by $\theta = (\widetilde{\theta}; \lambda_1, \ldots, \lambda_K)$ where $\widetilde{\theta} \in \widetilde{\Theta} \subset \mathbb{R}^d$ and $0 < \lambda_1, \cdots, \lambda_k \leq 1$. Thus, the parameter space is $\Theta = \widetilde{\Theta} \times (0,1]^K \subset \mathbb{R}^{d+K}$. The seller offers an assortment $s = \{s_k\}_{k=1}^K$ consisting of the sub-assortments $s_k \subset B_k$ selected from the K baskets. For $k \in [K]$, let $v_k = \sum_{i\in s_k} \exp(x_i^\intercal \widetilde{\theta}/\lambda_k)$ be the basket-level attraction parameter. With item features $x_j \in \mathbb{R}^d$ for $j \in [N]$, the custsomer choice probability under LCL with parameter θ is

$$p(a|s;\theta) = \begin{cases} \frac{v_j^{\lambda_j - 1}}{1+\sum_{k=1}^K v_k^{\lambda_k}} \exp\left(\frac{x_a^\intercal \widetilde{\theta}}{\lambda_j}\right) & a \in s_j,\ j = 1, \cdots, K; \\ \frac{1}{1+\sum_{k=1}^K v_k^{\lambda_k}} & a = 0. \end{cases} \tag{12}$$

In order to establish the regret guarantee, the required boundedness assumption becomes the following.

Assumption 3. In the NL model (12), there exist constants $C_x, C_\Theta, C_\lambda < +\infty$, such that $\max_{j\in[N]} \|x_j\|_2 \leq C_x$, $\sup_{\theta\in\widetilde{\Theta}} \|\widetilde{\theta}\|_2 \leq C_\Theta$, and $1/\lambda_k \leq C_\lambda$ for $k \in [K]$.

Note that $\lambda_1, \cdots, \lambda_K$ are used to parametrize the within-basket *dissimilarity* among items [52]. The assumption on their boundedness away from zero implies that the within-baskets items should be strictly distinguishable from each other.

Theorem 5.3. Consider the NL model (12) satisfying Assumption 3. Let $D_{\mathcal{P}} = d + K$ and $C_{\mathcal{P}} = C_x C_\Theta C_\lambda$. Under Assumptions 1 (I) and (II), for $n \gtrsim D_{\mathcal{P}} \log(n/D_{\mathcal{P}})$ and every $\delta \in (0,1)$, with probability at least $1-\delta$, the regret of PASTA with the uncertainty set $\Omega_n(\alpha_n(\delta))$ satisfies (9).

5.4 Minimax Optimality

To understand the tightness of the aforementioned regret guarantees, we establish the first minimax regret lower bound for the offline uncapacitated assortment optimization problem. Specifically, we consider the scenarios where the choice model is an MNL model (10) with the true parameter $\theta_0 \in \mathbb{R}^d$ and item features $\{x_j\}_{j\in[N]} \subset \mathbb{R}^d$. Here, N is the total number of items.

Theorem 5.4. Assume the offline data sample size $n \geq \min(d, N)$. There exists a universal constant c such that, for any measurable function $\widehat{s}$ of the offline dataset $\mathcal{D}_n$, there exists a worst-case offline assortment optimization problem instance with bounded item features $\{x_j\}_{j\in[N]}$, an MNL parameter θ_0, an item-only reward function $r : [N] \to [0, +\infty)$, and an offline assortment data distribution π_S with $\pi_S(s^\star) > 0$, such that the expected regret of $\widehat{s}$ is lower bounded by $c \cdot \sqrt{\min(d, N)/n}$.

Notably, the minimax regret in Theorem 5.4 matches the regret guarantee of PASTA in Theorem 5.1 in terms of the sample size n up to logarithmic factors, demonstrating that PASTA is provably efficient in terms of sample complexity for offline assortment optimization. Moreover, in the common scenarios where $d \leq N$, PASTA is also provably efficient in terms of the choice model complexity.

6 Empirical Studies

In this section, we conduct experiments on simulation datasets to empirically demonstrate the effectiveness of PASTA. We consider offline assortment optimization under two of the most popular choice models: the MNL and NL models. To facilitate empirical studies, in Section 6.1, we first propose an iterative algorithm for heuristically solving the max-min problem given by PASTA in (4). We next discuss the data generating process and baselines in Section 6.2, and subsequently present all the empirical results in Section 6.3.

6.1 PASTA Algorithm

Our algorithm PASTA is applicable for scenarios where the choice model is a parametric model $p(a|s;\theta)$ with parameters $\theta \in \Theta \subset \mathbb{R}^d$ and $p(a|s;\theta)$ is differentiable with respect to θ for all s and $a \in s \cup \{0\}$. These scenarios include many practical settings, such as the MNL and NL models. Notably, even when $\mathcal{P}$ is nonparametric, it is standard to approximate $\mathcal{P}$ with a parametric family so that the optimization is tractable. Hence, an algorithm for parametric models is sufficient and without loss of generality. Moreover, we follow the convention of assortment optimization literature to assume that the revenue of a selected item remains the same across distinct assortments, i.e., there exists r_j for $j \in [N]$ such that for any $s \in \mathbb{S}$, $r(s, j) = r_j$ if $j \in s$ and $r(s, j) = 0$ if $j \notin s$.

Specifically, let $\widehat{L}_n(\theta) = -(1/n)\sum_{i=1}^n \log p(a_i|s_i;\theta)$ and $\widehat{\theta}_n \in \arg\min_{\theta\in\Theta} \widehat{L}_n(\theta)$ be the MLE. We aim to solve the following optimization problem

$$\widehat{s}_{\mathfrak{p},n} \in \arg\max_{s\in\mathbb{S}} \min_{\theta\in\Omega_n(\alpha_n)} \mathcal{V}(s;\theta),$$

where $\Omega_n(\alpha_n) := \left\{\theta \in \Theta : \widehat{L}_n(\theta) - \widehat{L}_n(\widehat{\theta}_n) \leq \alpha_n\right\}$ and $\mathcal{V}(s;\theta) = \mathcal{V}(s; p(\cdot|s;\theta)) = \sum_{j\in s} r_j p(j|s;\theta)$.

The proposed algorithm PASTA is executed for a maximum of T iterations. For initialization, set $\theta^{(0)} = \widehat{\theta}_n$ and randomly choose $s^{(0)} \in \mathbb{S}$. At the t-th iteration, given $s^{(t-1)}$ and $\theta^{(t-1)}$ from the previous iteration, we consecutively execute the following two steps:

$\mathcal{A}_1$. Compute the optimal assortment $s^{(t)}$ given $\theta^{(t-1)}$: $s^{(t)} \in \arg\max_{s\in\mathbb{S}} \mathcal{V}(s, \theta^{(t-1)})$;

$\mathcal{A}_2$. Compute the optimal $\theta^{(t)}$ using $s^{(t)}$: $\theta^{(t)} \in \arg\min_{\theta\in\Omega_n} \mathcal{V}(s^{(t)}, \theta)$.

In particular, the step $\mathcal{A}_1$ corresponds to finding an optimal assortment for a known choice model with parameter $\theta^{(t-1)}$. As discussed in Section 2, there exists a streamline of works that address this problem efficiently. In our experiments, we use two existing linear programming(LP)-based algorithms that respectively solve step $\mathcal{A}_1$ under the MNL and NL models [12, 13]. Due to the space constraint and their tangential connection to our contribution, we summarize them in Appendix B.

We next propose a procedure for the step $\mathcal{A}_2$. For a given optimized assortment $s^{(t)}$, we aim to search for the worst-case model parameter $\theta^{(t)}$ from the uncertainty set Ω_n that minimizes the expected revenue. In particular, we employ a gradient descent with line search (GDLS) method to compute $\theta^{(t)}$ by solving the following problem.

$$\theta^{(t)} \in \arg\min_{\theta\in\Omega_n} \mathcal{V}(s^{(t)}; \theta). \tag{13}$$

Given a feasible initial parameter $\theta^{(t,0)} \in \Omega_n$, we run at most M gradient descent steps. Note that feasible $\theta^{(t,0)}$ is easy to obtain as $\widehat{\theta}_n$ and $\theta^{(t-1)}$ are guaranteed to be feasible by definition. Suppose β_m is the step size for gradient descent in the m-th step. At each step $m \in \{1, 2, \cdots, M\}$, we perform a line search to maintain the feasibility. In particular, given $\theta^{(t,m-1)} \in \Omega_n$, we first evaluate the gradient as $\xi_m = \nabla_\theta \mathcal{V}(s^{(t)}; \theta^{(t,m-1)})$. Then we initiate β_m with a pre-specified step size $\beta_m = \widetilde{\beta}$, and check whether $\theta^{(t,m)} = \theta^{(t,m-1)} - \beta_m \xi_m$ is feasible, i.e. $\theta^{(t,m-1)} \in \Omega_n$. If not, we set $\beta_m \leftarrow c\beta_m$ for some pre-specified $c \in (0, 1)$, and recompute $\theta^{(t,m)} = \theta^{(t,m-1)} - \beta_m \xi_m$. Such a search is repeated until $\theta^{(t,m)}$ is feasible. After M steps, we set $\theta^{(t)} = \theta^{(t,M)}$ and the step $\mathcal{A}_2$ finishes. We provide the pseudocode in Algorithm 1 for the overall process. In all of our numerical studies, we set $M = 3$, $\widetilde{\beta} = 0.01$ and $c = \frac{1}{5}$, which perform well.

6.2 Data Generation Process and Baselines

We compare PASTA to assortment optimization without pessimism, which in the sequel goes by the name baseline method. Our method and the baseline method are evaluated on synthetic data for which the optimal assortment $s^\star$ and true choice model parameter θ_0 are known so that we can compute the regret. We study two popular choice models, the MNL and NL models. We describe the data generation process and the baseline method in details below.

The MNL Model For the MNL model, we consider the assortment optimization scenarios described by N, K, d, n and p, where N is the total number of available products; K is the cardinality constraint of the assortments, i.e., $\mathbb{S} = \{s : |s| \le K\}$; d is the dimension of choice model parameter θ_0 and item features $\{x_j\}_{j=1}^N$; n is the sample size of the offline dataset; p is the probability of observing the optimal assortment $s^\star$. Similar to [10], we first generate the true parameter θ_0 as a uniformly random unit d-dim vector. Recall that r_j and

Algorithm 1 Gradient Descent with Line Search (GDLS)

Input: assortment $s^{(t)}$; uncertainty set hyperparameter α_n; initial parameter $\theta^{(t,0)}$; initial step size $\widetilde{\beta}$; step shrinkage constant c; number of descent steps M
Output: the optimized parameter $\theta^{(t)}$
$\widehat{L}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log p(A_i|S_i;\theta)$
$\widehat{\theta}_{\mathrm{ML},n} \leftarrow \arg\min_{\theta\in\Theta} \widehat{L}_n(\theta)$; $\widehat{L}_{\mathrm{ML}} \leftarrow \widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n})$
for $m = 1$ **to** M **do**
    $\xi_m = \nabla_\theta \mathcal{V}(s^{(t)};\theta^{(t,m-1)})$ `/*compute the gradient*/`
    $\beta_m \leftarrow \widetilde{\beta}$
    $\theta^{(t,m)} \leftarrow \theta^{(t,m-1)} - \beta_m \xi_m$
    while $n \cdot \left(\widehat{L}_n(\theta^{(t,m)}) - \widehat{L}_{\mathrm{ML}}\right) > \alpha_n$ **do** `/* check model feasibility */`
        $\beta_m \leftarrow c\beta_m$ `/* decrease the step size */`
        $\theta^{(t,m)} \leftarrow \theta^{(t,m-1)} - \beta_m \xi_m$
    end while
end for
$\theta^{(t)} \leftarrow \theta^{(t,M)}$

x_j denote the reward and feature of product j, respectively. For $j \in \{1,\ldots,N\}$, we generate the r_j uniformly from the range $[5,8]$ and x_j as uniformly random unit d-dim vector such that $\exp(x_j^\intercal \theta_0) \le \exp(-0.6)$ to avoid degenerate cases, where the optimal assortments include too few items. Given such information, the true optimal assortment $s^\star$ can be computed. Then, we generate an offline dataset $\mathcal{D} = \{(S_i, A_i)\}_{i=1}^n$ with n samples. For $i \in \{1,\ldots,n\}$, we generate S_i following the distribution π_S such that $\pi_S(s^\star) = p$ and $\pi_S(s) = (1-p)/(|\mathbb{S}|-1)$ for $s \neq s^\star$, where $0 < p < 1$ is the probability of observing the optimal assortment $s^\star$. After the assortment S_i is sampled, the customer choice A_i is sampled according to the probability computed by the MNL model as in (10) with the true parameter θ_0.

The NL Model For the NL model, we assume without loss of generality that items are partitioned into K nests, each containing an equal number of κ items [13]. Thus, the total number of items is $K \cdot \kappa$, and we characterize the assortment optimization problems by K, κ, d, n and p, where K is the total number of nests and κ is the number of products in each nest; n is the sample size of the offline dataset; p is the probability for sampling the optimal assortment; d is the dimension of item features. As presented in Section 5, the NL model has parameters $\theta_0 = (\widetilde{\theta}_0, \lambda_0)$ with $\widetilde{\theta}_0 \in \mathbb{R}^d$ and $\lambda_0 \in (0,1)^M$ We consider unconstrained assortment optimization for the NL model, i.e., $\mathbb{S} = 2^{[N]} \setminus \{\emptyset\}$. The data generating process is similar to that of the MNL model case. We generate the true parameters $\widetilde{\theta}_0$ as a uniformly random unit d-dim vector and λ_0 uniformly within $(0,1)^M$. For notation simpilicity, let r_{kj} and x_{kj} respectively denote the reward and feature of the j-th item in the k-th nest. For $k \in \{1,\ldots,K\}$ and $j \in \{1,\ldots,\kappa\}$, we generate r_{kj} uniformly from the range $[5,8]$ and x_{kj} as uniformly random unit d-dim vector such that $\exp(x_{kj}^\intercal \theta_0) \le \exp(-0.6)$. The generating process of the offline dataset $\mathcal{D}$ is identical to that of the MNL model case.

Baseline Methods For both the MNL and NL models, we use assortment optimization without pessimism, i.e., the as-if approach defined in (5), as the baseline method. In our

experiments, we use the gradient descent method to find the MLE $\widehat{\theta}_{\mathrm{ML},n}$ that minimizes the empirical negative log-likelihood function $\widehat{L}_n$. Then given $\widehat{\theta}_{\mathrm{ML},n}$, the baseline method solves the assortment optimization problem in (5).

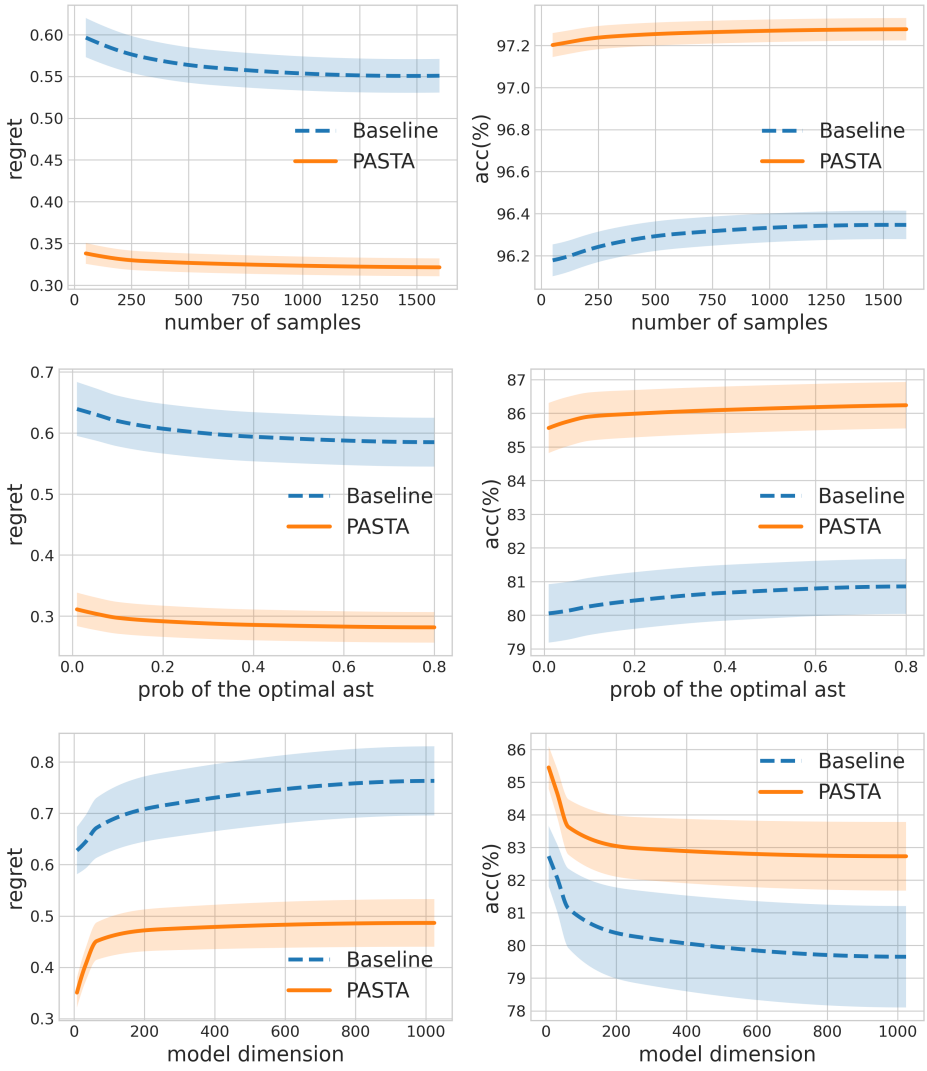

Figure 1: Performance comparison between PASTA and the baseline method on the MNL model. The left panel shows average regret (lower is better), while the right panel presents assortment accuracy (higher is better).Our proposed method PASTA consistently outperforms the baseline method in both metrics. The performance gain remains consistent across varying probabilities of observing the optimal assortment (middle row) and varying model complexity, measured by the dimension of the model parameter (bottom row).

6.3 Performance Comparison

To measure the statistical variance of the results, we repeat the data generation process in Section 6.2 to randomly generate 50 offline datasets. The solutions of PASTA and the baseline method are recorded in these experiments. We measure the performance with two metrics:

(1) the *regret* of the solutions which indicates how far the performance of the solutions is to that of the optimal performance, i.e., revenue of the optimal assortment $s^\star$; (2) the assortment *accuracy* of the solutions with respect to the optimal assortment $s^\star$. The assortment accuracy of an assortment s is defined as the ratio of the number of correctly chosen products to the number of products in $s^\star$. For hyper-parameters, we set $\alpha_n = 2\widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n})$ and the maximum of iteration $T = 30$.

The MNL Model We first set $N = 40$, $K = 10$, $d = 8$, $p = 0.1$ and gradually increase the number of samples n. In Figure 1, we observe that PASTA consistently outperforms the baseline method in both regret and assortment accuracy. Other than the number of samples, another important parameter for the offline assortment optimization is the probability of observing the optimal assortment $\pi_S(s^\star)$. To this end, Figure 1 shows the performance of PASTA under varying $\pi_S(s^\star)$ from as low as 0.01 to 0.8. We observe that the gain of PASTA is consistent under a wide range of $\pi_S(s^\star)$.

The NL Model We first set $K = 3$, $\kappa = 5$, $d = 8$ and $p = 0.1$, and gradually increase the number of samples n. In Figure 2, we observe that PASTA consistently outperforms the baseline method. While the performance of the baseline method improves with increasing number of samples, PASTA maintains a regret that is less than 25% of the baseline regret. Next, we study the robustness of PASTA against the probability of observing the optimal assortment $\pi_S(s^\star)$. In Figure 2, we set $K = 3$, $\kappa = 5$, $d = 8$, $n = 200$ and increase $\pi_S(s^\star)$. As shown in Figure 2, the gain of PASTA over the baseline method is consistent. Lastly, we study how the model complexity d effects the performance. We set $K = 3$, $\kappa = 5$, $p = 0.1$, $n = 1000$ and increase d from 8 to 1024. We present the results in Figure 2. It can be observed that, while the regrets of PASTA and the baseline method both increase as d increases, the gain of PASTA maintains.

7 Conclusion

We propose PASTA, a unified framework for offline assortment optimization under general choice models. Notably, the success of PASTA requires only that the offline data distribution includes an optimal assortment. We demonstrate the effectiveness of PASTA by applying it to several popular choice models, including the MNL, LCL and NL models, and establish the first finite-sample regret guarantees for the offline assortment optimization problem under these models. Furthermore, we derive the minimax regret lower bound under MNL, justifying the minimax optimality of PASTA in terms of the sample and model complexities. During the preparation of our manuscript, we notice a concurrent work [25] that explores assortment optimization for a specialized MNL model using the observational data. Their approach is ranking-based and requires only that the offline data contains the items in the optimal assortment, due to a stronger model assumption. It remains an interesting research problem to investigate how to generalize the similar optimal item coverage for a broader class of choice models under a unified framework such as PASTA.

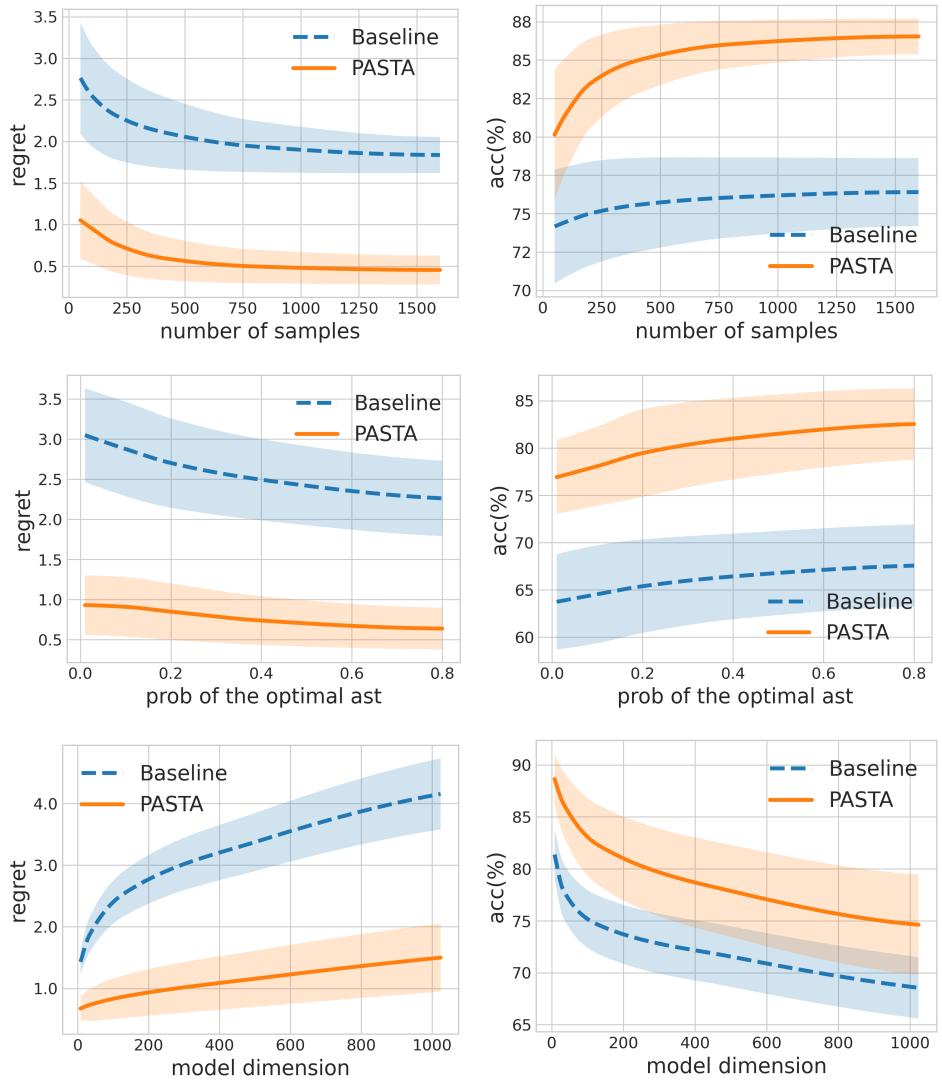

Figure 2: PASTA also consistently outperforms the baseline method when the underlying choice model is an NL model. The performance gain in terms of the regret difference increases with increasing model complexity (bottom row).

A Main Technical Theorem: Hellinger Concentration of MLE for Conditional Density

We note that the results in this section are applicable to general scenarios that involve estimating conditional probability density function with MLE. Hence, in this section, we will use Y and X to respectively denote A and S to underscore that the results are generally applicable. Let $\mathcal{P}$ be a family of conditional densities for $Y|X$. Assume a dataset D consisting of n independent and identically distributed samples of (X, Y) following the probability density function (PDF) $(x, y) \mapsto \pi(x)p_0(y|x)$ where p_0 is the true data generating conditional PDF.

The goal is to estimate p_0 with the observed data D. In particular, let the MLE $\widehat{p}_n$ of p_0 be defined as

$$\widehat{p}_n \in \arg\min_{p\in\mathcal{P}} \left\{ \widehat{L}_n(p) := -\frac{1}{n}\sum_{i=1}^{n} \log p(Y_i|X_i) \right\}. \tag{14}$$

Next we present our main theorem.

Theorem A.1. Define

$$\kappa(u) = \begin{cases} \frac{2\left(e^{|u|/2}-1-|u|/2\right)}{(1-e^{-u/2})^2}, & u \neq 0; \\ 1, & u = 0. \end{cases}$$

Let $\tau > 0$ be an arbitrary but fixed smoothing constant satisfying $\kappa(\tau) < 8(1-e^{-\tau})^2$, which is equivalent to $0.5518 < \tau < 3.1926$. Let n denote the number of samples. Suppose $\sigma_n, \alpha_n > 0$ satisfy the following relationships: $\alpha_n = \sigma_n^2/8\kappa(\tau)$, $\sigma_n^2 \geq n^{-1}$, and

$$\int_{\sigma_n^2/2^{10}}^{\sigma_n} \sqrt{\log \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}, H\right)}\mathrm{d}u \leq \frac{1}{2^{27}}\sqrt{n}\sigma_n^2. \tag{15}$$

Let $c(\tau) = 2\left\{\frac{1-e^{-\tau}}{4} - \frac{\kappa(\tau)}{32(1-e^{-\tau})}\right\} > 0$ and $c_1 = 63/(257\times 2^{14})$. Consider the set estimate

$$\Omega_n(\alpha) := \left\{ p \in \mathcal{P} : \widehat{L}_n(p) - \widehat{L}_n(\widehat{p}_n) \leq \alpha \right\}.$$

Then for every $\alpha \geq \alpha_n/4$, with probability at least $1 - c\cdot e^{-2c_1 n\alpha}$, it holds that

$$(i) p_0 \in \Omega_n(\alpha), \text{ and } (ii) \sup_{p\in\Omega_n(\alpha)} H^2(p, p_0) \leq \frac{\alpha}{c(\tau)},$$

where c is a universal constant. In particular, at $\tau = 1.7024$, $c(\tau)$ is maximized as 0.1802.

Remark 1. To connect with the statement in Theorem 4.2, we have $C_1 = 2^{-7}\kappa(\tau)$, $C_2 = 2\sqrt{2\kappa(\tau)}$, $C_3 = 2\sqrt{2}e^{\tau/2}$, $C_4 = 2^{-24}\kappa(\tau)$, $C_5 = c$, $C_6 = 2c_1$, $C_7 = 1/c(\tau)$.

Proof Sketch of Theorem A.1. We defer the detailed proof to the Appendix C and only present a sketch of proof below. All the intermediate results referred below are also deferred to the Appendix due to space constraint.

The key difficulty is that likelihood–ratio functions are not uniformly bounded, which obstructs direct empirical-process control. We therefore introduce a smoothed class $\mathcal{P}_\tau = \{(1-e^{-\tau})p + e^{-\tau}p_0 : p \in \mathcal{P}\}$ and the smoothed risks $L_\tau, \widehat{L}_{\tau,n}$; this guarantees a pointwise bound on the log-likelihood ratio (Lemma E.9) while preserving correct specification ($p_0 \in \mathcal{P}_\tau$). We localize the analysis to the Hellinger ball $\{p : H^2(p, p_0) \leq \alpha\}$ and consider the corresponding class of smoothed log-likelihood ratios $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$. By Lemma E.5, this localized class satisfies a Bernstein condition, yielding sharp concentration for $\|(\mathbb{P}_n - \mathbb{P})\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)}$ (Theorem C.2). Next, we identify that a special empirical "margin" event implies that any p with $H^2(p, p_0) \geq \alpha$ must have strictly larger empirical risk than p_0; since $\widehat{p}_n$ minimizes $\widehat{L}_n$, it must lie in the local region $H^2(\widehat{p}_n, p_0) \leq \alpha$. On this event, an empirical smoothed risk inequality together with $L_\tau(p_0) \leq L_\tau(\widehat{p}_n)$ bounds the empirical excess $\widehat{L}_n(p_0) - \widehat{L}_n(\widehat{p}_n)$ by the localized empirical-process norm. A standard peeling argument then transfers this pointwise control to the whole uncertainty set $\Omega_n(\alpha)$, showing that all empirically near-optimal p remain within Hellinger radius $\alpha/c(\tau)$ of p_0.

Under the numeric condition $\kappa(\tau) < 8(1-e^{-\tau})^2$ and for $\alpha \geq \alpha_n/4$, the concentration bounds deliver $\mathbb{P}\Big\{ p_0 \in \Omega_n(\alpha) \text{ and } \sup_{p\in\Omega_n(\alpha)} H^2(p, p_0) \leq \alpha/c(\tau) \Big\} \geq 1 - ce^{-c'n\alpha}$, with universal constants $c, c' > 0$.

□

B Linear Programming (LP) for Assortment Optimization Problems with Known Choice Models

Here we summarize two LP-based approaches for solving the assortment optimization problem under known choice models. Specifically, we solve the following optimization problem

$$\arg\max_{s\in\mathbb{S}} \mathcal{V}(s;\theta) = \sum_{j\in s} r_j p(a|s;\theta),$$

where $p(a|s;\theta)$ is a choice model parameterized by θ. Note that here θ is fixed and known. Different choice models require different optimization methods. This work considers two popular choice models – the MNL and NL models. In both scenarios, the assortment optimization problem can be solved with LP-based methods [12, 13].

B.1 MNL Model

Let an assortment s be represented by an N-dimensional binary vector $\gamma \in \{0,1\}^N$ where $\gamma_j = 1$ if and only if $j \in s$. Suppose that $s \in \mathbb{S}$ corresponds to the following feasible set for γ with M linear inequality constraints:

$$\Gamma = \left\{ \gamma \in \{0,1\}^N : \sum_{j\in N} a_{ij}\gamma_j \leq b_i \text{ for } i \in [M] \right\},$$

where the matrix of constraint coefficients $[a_{ij}]_{i\in[M],j\in[N]}$ is a totally unimodular matrix. In other words, based on the one-to-one correspondence between s and γ, we have $s \in \mathbb{S}$ if and only if $\gamma \in \Gamma$.

Next, we denote $v_i = \exp(x_i^\intercal \theta)$ as the preference score for the i-th item, with item feature x_i and revenue r_i. The customer choice probability under the MNL model (10) becomes $p(i|s;\theta) = \frac{v_i}{1+\sum_{j\in s} v_j}$. The assortment optimization can be formulated as

$$\max_{\gamma\in\Gamma} \frac{\sum_{i\in[N]} r_i v_i \gamma_i}{1 + \sum_{i\in[N]} v_i \gamma_i}, \tag{16}$$

which is equivalent to the following linear programming problem [12]:

$$\begin{aligned} \max_{w_j : j \in [N] \cup \{0\}} \quad & \sum_{j \in [N]} r_j w_j \\ \text{subject to} \quad & \sum_{j \in [N]} w_j + w_0 = 1 \\ & \sum_{j \in [N]} a_{ij} \frac{w_j}{v_j} \le b_i w_0 \quad \forall i \in [M] \\ & 0 \le \frac{w_j}{v_j} \le w_0 \quad \forall j \in [N]. \end{aligned} \tag{17}$$

In particular, we can recover the optimal solution to Problem (16), denoted as $\gamma^\star$, using the optimal solution to Problem (17), denoted by $w^\star$, via the following formula $\gamma_j^\star = \frac{w_j^\star}{v_j w_0^\star} \quad \forall j \in [N]$. Finally, an optimal assortment $s^\star(\theta)$ for a fixed choice $p(a|s;\theta)$ is obtained by the correspondence $j \in s^\star$ if and only if $\gamma_j^\star = 1$.

B.2 NL Model

Without loss of generality, we consider NL models with K baskets where each basket contains M items. Let an assortment s for the NL model be represented by an array of K binary vectors $\gamma = \{\gamma_k\}_{k=1}^K$ where $\gamma_k \in \{0,1\}^M$ represents the choice for the k-th basket. In particular, for $m \in [M]$, $\gamma_{k,m} = 1$ if we choose the m-th item in the k-th basket, with item feature $x_{k,m}$ and revenue $r_{k,m}$. For the purpose of optimization and without loss of generality, we assume that the items in each basket are sorted with decreasing revenues, i.e., $r_{k,m} \ge r_{k,m'}$ for $k \in [K]$, $m, m' \in [M]$, and $m \ge m'$. Let

$$\begin{aligned} V_k(\gamma_k) &= \sum_{m=1}^M \gamma_{k,m} \exp(\tilde{\theta}^\intercal x_{k,m}/\lambda_k); \\ R_k(\gamma_k) &= \frac{\sum_{m=1}^M \gamma_{k,m} \exp(\tilde{\theta}^\intercal x_{k,m}/\lambda_k) r_{k,m}}{V_k(\gamma_k)}. \end{aligned}$$

The assortment optimization problem for fixed $\tilde{\theta}, \lambda, x, r$ is therefore defined as

$$\max_{\gamma = \{\gamma_k\}_{k=1}^K} \sum_{k=1}^K \frac{V_k(\gamma_k)^{\lambda_k}}{1 + \sum_{k'=1}^K V_{k'}(\gamma_{k'})^{\lambda_{k'}}} R_k(\gamma_k). \tag{18}$$

We can efficiently solve (18) through the following LP problem [13]

$$\begin{aligned} \min_{\eta, y_1, \ldots, y_K} \quad & \eta \\ \text{s.t.} \quad & \eta \ge \sum_{k=1}^K y_k \\ & k \in [K],\ y_k \ge V_k(\gamma_k)^{\lambda_k} (R_k(\gamma_k) - \eta) \qquad \forall \gamma_k \in \Gamma, \end{aligned} \tag{19}$$

where $\Gamma = \{\widetilde{M}_1, \widetilde{M}_2, \ldots, \widetilde{M}_M\}$ has the subsets of baskets containing first m items, i.e., for $m \in [M]$ $\widetilde{M}_m = \{1, \ldots, 1, 0, \ldots, 0\} \in \{0,1\}^M$ with the first m elements as 1 and the last

$M - m$ elements as 0. After solving (19) with solutions $\eta^\star, y_k^\star$, the optimal assortment $\gamma_k^\star$ can be constructed as $\gamma_k^\star = \arg\max_{\gamma_k \in \Gamma} V_k(\gamma_k)^{\lambda_k}(R_k(\gamma_k) - \eta^\star)$.

C Proofs of Theoretical Results

C.1 Proof of Theorem 4.1

Proof of Theorem 4.1. By triangle inequality, for any $p \in \mathcal{P}$, we have $\mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p) \leq |\mathcal{V}(s^\star; p) - \mathcal{V}(s^\star; \widehat{p}_n)| + |\mathcal{V}(s^\star; \widehat{p}_n) - \mathcal{V}(s^\star; p_0)|$. If $p_0 \in \Omega_n$, we further have for any $p \in \Omega_n(\alpha_n)$

$$\mathcal{V}(s^\star; p_0) - \mathcal{V}(s^\star; p) \leq 2 \max_{p \in \Omega_n} |\mathcal{V}(s^\star; p) - \mathcal{V}(s^\star; \widehat{p}_n)| \,. \tag{20}$$

Lemma E.3 implies that for any $p \in \mathcal{P}$, $|\mathcal{V}(s^\star; p) - \mathcal{V}(s^\star; \widehat{p}_n)| \leq r_{s^\star} \sqrt{C_{s^\star} \mathbb{E}_S \left[\|p(\cdot|S) - \widehat{p}_n(\cdot|S)\|_1^2 \right]}$, where $\|\cdot\|_1$ is the L_1 norm, $r_{s^\star} = \max_{j \in s^\star} r_j$ is the largest possible revenue among all items and $C_{s^\star} = 1/\pi_S(s^\star)$. In Lemma E.4, we establish that $\mathbb{E}_S \left[\|p(\cdot|S) - \widehat{p}_n(\cdot|S)\|_1^2 \right] \leq 8H^2(p, \widehat{p}_n)$, where H^2 is the generalized squared Hellinger distance defined in (6). Combining the above two inequalities, we have that for any $p \in \mathcal{P}$, $\left| \mathcal{V}(s^\star; p) - \mathcal{V}(s^\star; \widehat{p}_n) \right| \leq r_{s^\star} \sqrt{8 C_{s^\star} H^2(p, \widehat{p}_n)}$. From Assumption 1 (III), we have that with probability at least $1 - \delta$ where $0 < \delta < 1$, for any $p \in \Omega_n(\alpha_n)$, $H^2(p, \widehat{p}_n) \leq \alpha_n$. With this upper bound for $H^2(p, \widehat{p}_n)$, we further have for probability at least $1 - \delta$,

$$\sup_{p \in \Omega_n} \left| \mathcal{V}(s^\star; p) - \mathcal{V}(s^\star; \widehat{p}_{\mathrm{ML},n}) \right| \lesssim r_{s^\star} \sqrt{\alpha_n(\delta)/\pi_S(s^\star)}. \tag{21}$$

Lastly, under Assumption 1 (III), with probability at least $1 - \delta$ we have $p_0 \in \Omega_n$. Then application of Lemma 4.1 along with the inequalities in (20) and (21) conclude the proof. □

C.2 Proof of Theorem 4.2

Proof of Theorem 4.2. This result is a simplified result of Theorem A.1. In particular, with $\kappa(\tau)$ and $c(\tau)$ defined as in Theorem A.1, choosing the smoothing constant $\tau = 2$ directly leads to Theorem 4.2. □

C.3 Proof of Theorem A.1

Our proof builds upon the empirical process on likelihood ratio functions. In particular, for a family of $\mathbb{R}$-valued functions $\mathcal{F}$, denote $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n(f) - \mathbb{P}(f)|$. Note that under the assumption of correct specification, i.e., $p_0 \in \mathcal{P}$, p_0 minimizes the population negative log-likelihood:

$$p_0 \in \arg\min_{p \in \mathcal{P}} \Big\{ L(p) := \mathbb{E}[-\log p(Y|X)] \Big\}. \tag{22}$$

To establish empirical process results without the assumption of uniform boundedness on the function class, our proof relies on a technique known as smoothing. To this end, for an arbitrary

but fixed smoothing constant $\tau > 0$, consider the following *smoothed family of conditional PDFs* $\mathcal{P}_\tau := \{(1-e^{-\tau})p + e^{-\tau}p_0 : p \in \mathcal{P}\}$. In particular, for $p = p_0$, we have $p_0 \in \mathcal{P}_\tau$. That is, $\mathcal{P}_\tau$ is also a correctly specified family. Moreover, the smoothed log-likelihood ratio is upper bounded as $\log \frac{p_0(Y|X)}{(1-e^{-\tau})p(Y|X)+e^{-\tau}p_0(Y|X)} \le \tau$. Define the corresponding smoothed negative log-likelihoods as

$$\begin{aligned} L_\tau(p) &:= \mathbb{E}\Big\{ -\log[(1-e^{-\tau})p(Y|X) + e^{-\tau}p_0(Y|X)]\Big\}; \\ \widehat{L}_{\tau,n}(p) &:= \mathbb{E}_n\Big\{ -\log[(1-e^{-\tau})p(Y|X) + e^{-\tau}p_0(Y|X)]\Big\}; \end{aligned} \tag{23}$$

To facilitate proof, for any $\tau > 0$ and $\alpha \ge 0$, we also define the local family of smoothed log-likelihood ratios $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$: for $p \in \mathcal{P}$, let $f_{p,\tau}(X,Y) = \log p_0(Y|X) - \log[(1-e^{-\tau})p(Y|X) + e^{-\tau}p_0(Y|X)]$, and define $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$ as

$$\mathcal{F}_\tau^{\mathrm{loc}}(\alpha) := \left\{ f_{p,\tau} : H^2(p, p_0) \le \alpha \right\}. \tag{24}$$

Notably, per Lemma E.5, it satisfies the Bernstein's condition. That is, for any $f_{p,\tau} \in \mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$ and $k \ge 2$, we have $\mathbb{E}|f_{p,\tau}|^k \le \frac{k!}{2} 2^{k-2} \times 8\kappa(\tau) H^2(p, p_0)$. In particular, $\mathbb{E}(f_{p,\tau}^2) \le 8\kappa(\tau) H^2(p, p_0)$.

Proof. Proof of Theorem A.1 Fix $\alpha \ge 0$ and consider the following event

$$\mathcal{E}_n(\alpha) := \left\{ \inf_{p \in \mathcal{P}: H^2(p,p_0) \ge \alpha} \left\{ \widehat{L}_n(p) - \widehat{L}_n(p_0) \right\} \le c(\tau)\alpha \right\}.$$

On $\mathcal{E}_n(\alpha)^{\complement}$, we have

$$\inf_{p \in \mathcal{P}: H^2(p,p_0) \ge \alpha} \left\{ \widehat{L}_n(p) - \widehat{L}_n(p_0) \right\} > c(\tau)\alpha \ge 0,$$

while $\widehat{L}_n(\widehat{p}_n) - \widehat{L}_n(p_0) \le 0$ because $\widehat{p}_n$ minimizes $\widehat{L}_n$. Therefore, we have $H^2(\widehat{p}_n, p_0) \le \alpha$, and hence $f_{\widehat{p}_n} \in \mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$. The following inequalities hold

$$\widehat{L}_n(p_0) - \widehat{L}_n(\widehat{p}_n) \le \frac{1}{1-e^{-\tau}} \left\{ \widehat{L}_{\tau,n}(p_0) - \widehat{L}_{\tau,n}(\widehat{p}_n) \right\} \tag{25}$$

$$\le \frac{1}{1-e^{-\tau}} \left\{ L_\tau(p_0) - L_\tau(\widehat{p}_n) + \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)} \right\} \le \frac{1}{1-e^{-\tau}} \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)}, \tag{26}$$

where the first inequality is due to Lemma E.9, the second inequality holds because $f_{\widehat{p}_n} \in \mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$, and the last inequality is due to $L_\tau(p_0) = \min_{p \in p} L_\tau(p) \le L_\tau(\widehat{p}_n)$. Let $\mathcal{A}_n(\alpha) := \left\{ \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)} \ge \frac{\kappa(\tau)}{32}\alpha \right\}$. On $\mathcal{A}_n(\alpha)^{\complement}$, we further have $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)} \le \frac{\kappa(\tau)}{32}\alpha$. That is, on $\mathcal{E}_n(\alpha)^{\complement} \cap \mathcal{A}_n(\alpha)^{\complement}$, we have that $\widehat{L}_n(p_0) - \widehat{L}_n(\widehat{p}_n) \le \frac{\kappa(\tau)}{32(1-e^{-\tau})}\alpha$, which by definition of Ω_n is equivalent to $p_0 \in \Omega_n\left(\frac{\kappa(\tau)}{32(1-e^{-\tau})}\alpha\right)$. In other words, $p_0 \in \Omega_n(\alpha)$ under the event $\mathcal{E}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)}\alpha\right)^{\complement} \cap \mathcal{A}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)}\alpha\right)^{\complement}$. Recall that on $\mathcal{E}_n(\alpha)^{\complement}$, we have

$$\inf_{p \in \mathcal{P}: H^2(p,p_0) \ge \alpha} \left\{ \widehat{L}_n(p) - \widehat{L}_n(p_0) \right\} > c(\tau)\alpha. \tag{27}$$

For any $p \in \Omega_n(c(\tau)\alpha)$, we also have $\widehat{L}_n(p) - \widehat{L}_n(p_0) \leq \widehat{L}_n(p) - \min_{\widetilde{p} \in p} \widehat{L}_n(\widetilde{p}) \leq c(\tau)\alpha$. Then for any $p \in \Omega_n$, $H^2(p, p_0)$ cannot exceed α_n, or otherwise violating (27). This implies that $\sup_{p \in \Omega_n(c(\tau)\alpha)} H^2(p, p_0) \leq \alpha$. Equivalently, $\sup_{p \in \Omega_n(\alpha)} H^2(p, p_0) \leq \alpha / c(\tau)$ on the event $\mathcal{E}_n\left(\alpha/c(\tau)\right)^{\complement}$. Thus, we have

$$\mathbb{P}\left\{p_0 \in \Omega_n(\alpha), \sup_{p \in \Omega_n(\alpha)} H^2(p, p_0) \leq \frac{\alpha}{c(\tau)}\right\}$$

lower bounded by

$$\mathbb{P}\left\{\mathcal{E}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)}\alpha\right)^{\complement} \cap \mathcal{A}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)}\alpha\right)^{\complement} \cap \mathcal{E}_n\left(\frac{\alpha}{c(\tau)}\right)^{\complement}\right\},$$

which by the same calculation in (28) is further lower bounded by

$$1 - \mathbb{P}\left\{\bigcup_{j=0}^{+\infty} \mathcal{A}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)} 2^j \alpha\right)\right\} - \mathbb{P}\left\{\bigcup_{j=1}^{+\infty} \mathcal{A}_n\left(\frac{2^j \alpha}{c(\tau)}\right)\right\}.$$

To continue, note that with the condition $\kappa(\tau) < 8(1-e^{-\tau})^2$, we have

$$\max\left\{\frac{\kappa(\tau)}{32(1-e^{-\tau})}, \frac{c(\tau)}{2}\right\} \leq \frac{1-e^{-\tau}}{4} \leq \frac{1}{4}.$$

With this fact, for any $\alpha \geq \alpha_n/4 \gtrsim n^{-1}$, by Theorem C.2, we have

$$\mathbb{P}\left\{\bigcup_{j=0}^{+\infty} \mathcal{A}_n\left(\frac{32(1-e^{-\tau})}{\kappa(\tau)} 2^j \alpha\right)\right\} \leq c \cdot \exp(-4c_2 n\alpha),$$

where c is a universal constant. Similarly, we have $\mathbb{P}\left\{\bigcup_{j=1}^{+\infty} \mathcal{A}_n\left(\frac{2^j\alpha}{c(\tau)}\right)\right\} \leq c \cdot \exp(-2c_2 n\alpha)$, where c is also a universal constants. Hence, $\mathbb{P}\left\{p_0 \in \Omega_n(\alpha),\ \sup_{p \in \Omega_n(\alpha)} H^2(p, p_0) \leq \frac{\alpha}{c(\tau)}\right\} \geq 1 - c \cdot \exp(-2c_2 n\alpha)$. Note that all c's represent different universal constants. Notably, Theorem C.2 requires that $\frac{32(1-e^{-\tau})}{\kappa(\tau)}\alpha \geq \alpha_n$; $\frac{2\alpha}{c(\tau)} \geq \alpha_n$, which are satisfied due to the condition that $\alpha \geq \alpha_n/4 \geq \max\left\{\kappa(\tau)/32(1-e^{-\tau}), c(\tau)/2\right\}\alpha_n$. □

Theorem C.1 (Rate Theorem)**.** Let $\tau, \alpha_n, c(\tau), c_2$ be defined as in Theorem A.1. Then for every $\alpha \geq \alpha_n/2$, consider the following event

$$\mathcal{E}_n(\alpha) := \left\{\sup_{p \in \mathcal{P}: H^2(p,p_0) \geq \alpha} \left\{\prod_{i=1}^{n} \frac{p(Y_i|X_i)}{p_0(Y_i|X_i)}\right\} \geq e^{-c(\tau)n\alpha}\right\},$$

which is equivalent to the event $\left\{\inf_{p \in \mathcal{P}: H^2(p,p_0) \geq \alpha} \left\{\widehat{L}_n(p) - \widehat{L}_n(p_0)\right\} \leq c(\tau)\alpha\right\}$, we have $\mathbb{P}\{\mathcal{E}_n(\alpha)\} \leq c \cdot e^{-c_2 n\alpha}$ where c is a universal constant.

Proof. Proof of Theorem C.1 Consider the local family of smoothed log-likelihood ratios $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$ in (24). Let $J(\alpha) := \min\{J \in \mathbb{N} : 2^J\alpha \geq 1\}$. Then, with Lemma E.9, we further have

$$\sup_{p:H^2(p,p_0)>\alpha} \left\{\widehat{L}_n(p_0) - \widehat{L}_n(p)\right\} \leq \frac{1}{1-e^{-\tau}} \sup_{p:H^2(p,p_0)>\alpha} \left\{\widehat{L}_{\tau,n}(p_0) - \widehat{L}_{\tau,n}(p)\right\}.$$

Let $\mathcal{P}_j(\alpha) = \{p \in \mathcal{P} : 2^{j-1}\alpha < H^2(p,p_0) \leq 2^j\alpha\}$, and we have $\sup_{p:H^2(p,p_0)>\alpha} \left\{\widehat{L}_{\tau,n}(p_0) - \widehat{L}_{\tau,n}(p)\right\} \leq \max_{1\leq j\leq J(\alpha)} \left\{ \sup_{p\in\mathcal{P}_j(\alpha)} \{L_\tau(p_0) - L_\tau(p)\} + \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(2^j\alpha)} \right\}$. Moreover, for each $j \in J(\alpha)$ and $p \in \mathcal{P}$ such that $2^{j-1}\alpha < H^2(p,p_0) \leq 2^j\alpha$, we have $L_\tau(p_0) - L_\tau(p) \leq -2H^2\big((1-e^{-\tau})p + e^{-\tau}p_0, p_0\big) \leq -\frac{(1-e^{-\tau})^2}{4}2^j\alpha$, where the first inequality is by Lemma E.11, the second inequality is by Lemma E.10 and that $H^2(p,p_0) > 2^{j-1}\alpha$. Hence, we have

$$\sup_{p:H^2(p,p_0)>\alpha} \left\{\widehat{L}_n(p_0) - \widehat{L}_n(p)\right\} \leq \frac{1}{1-e^{-\tau}} \max_{1\leq j\leq J(\alpha)} \left\{ -\frac{(1-e^{-\tau})^2}{4}2^j\alpha + \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(2^j\alpha)} \right\}.$$

Note that the event $\frac{1}{1-e^{-\tau}} \max_{1\leq j\leq J(\alpha)} \left\{-\frac{(1-e^{-\tau})^2}{4}2^j\alpha + \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(2^j\alpha)}\right\} \geq -c(\tau)\alpha$ implies the event $\exists j$ such that $1 \leq j \leq J(\alpha)$ and $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(2^j\alpha)} \geq (1-e^{-\tau})\left(\frac{1-e^{-\tau}}{4}2^j - c(\tau)\right)\alpha \geq \frac{\kappa(\tau)}{32}2^j\alpha$. Here, the last inequality is based on the fact that for all $j \geq 1$, we have $c(\tau) \leq 2^j\left\{\frac{1-e^{-\tau}}{4} - \frac{\kappa(\tau)}{32(1-e^{-\tau})}\right\}$ if and only if $(1-e^{-\tau})\left(\frac{1-e^{-\tau}}{4}2^j - c(\tau)\right) \geq \frac{\kappa(\tau)}{32}2^j$. For $\alpha \geq 0$, define $\mathcal{A}_n(\alpha) := \left\{\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)} \geq \frac{\kappa(\tau)}{32}\alpha\right\}$. Here, for all $\alpha \geq \alpha_n/4$ and $1 \leq j \leq J(\alpha)$, we have $2^j\alpha \geq \alpha_n/2$, and hence Theorem C.2 can be applied to $\mathcal{A}_n(2^j\alpha)$. That is, $\mathbb{P}\left\{\mathcal{A}_n(2^j\alpha)\right\} \leq c\cdot\exp(-2^jc_2n\alpha)$ where c is a universal constant and $c_2 = 63/(257\times 2^{14})$. With this we have

$$\mathbb{P}\{\mathcal{E}_n(\alpha)\} \leq \mathbb{P}\left\{\bigcup_{j=1}^{J(\alpha)} \mathcal{A}_n(2^j\alpha)\right\} \leq c\sum_{j=1}^{J(\alpha)} e^{-2^jc_2n\alpha}. \tag{28}$$

Lastly, because $\alpha \geq \alpha_n \gtrsim n^{-1}$, we have $\sum_{j=1}^{J(\alpha)} e^{-2^jc_2n\alpha} \leq \sum_{j=1}^{\infty}(e^{-c_2n\alpha})^{2^j} \leq c\cdot e^{-c_2n\alpha}$, where c is a universal constant that is different to the previous universal constants. □

Theorem C.2. Let α_n and $\kappa(\tau)$ be defined as in Theorem A.1. Consider the local family of smoothed log-likelihood ratios $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$ in (24). Then for every $\alpha \geq \alpha_n$,

$$\mathbb{P}\left(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)} \geq \frac{\kappa(\tau)}{32}\alpha\right) \leq Ce^{-c_1\kappa(\tau)n\alpha},$$

where C is a universal constant and $c_1 = 63/(257\times 2^{14})$.

Proof. Proof of Theorem C.2 This result is a direct consequence of Theorem C.3. By Lemma E.5, the smoothed local family $\mathcal{F}_\tau^{\mathrm{loc}}(\alpha)$ satisfies the Bernstein's condition in Theorem C.3 with $\sigma^2 = 8\kappa(\tau)\alpha$. Let σ_n^2 be defined as in Theorem A.1. Due to $\alpha \geq \alpha_n$, we have $\sigma^2 = 8\kappa(\tau)\alpha \geq 8\kappa(\tau)\alpha_n = \sigma_n^2 \geq 1/n$. Moreover, by (15) and Lemmas E.7, E.8, it further satisfies the bracketing integral condition for $s = 6$ and every $\alpha \geq \alpha_n$. Then Theorem C.3 can be applied to establish the desired result. □

Theorem C.3. Suppose $\mathcal{F}$ is a family of $\mathbb{R}$-valued random variables. Assume that there exists some $\sigma^2 < +\infty$, such that for any $f \in \mathcal{F}$, it satisfy the Bernstein's condition:

$$\mathbb{E}|f - \mathbb{E}[f]|^k \leq \frac{k!}{2} 2^{k-2}\sigma^2;\ \forall k \geq 2.$$

Let $M > 0$ and $\varepsilon \in (0,1)$. Suppose $n, M, \varepsilon, \sigma^2$ satisfy $M = \varepsilon\sigma^2/4$, $\sigma^2 \geq n^{-1}$,

$$\int_{\frac{\varepsilon M}{32}}^{\sigma} \sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{M\sqrt{n}\varepsilon^{3/2}}{2^{10}}, \tag{29}$$

where $\mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))$ is the minimal number of u-brackets to cover $\mathcal{F}$. Then $\mathbb{P}\left\{\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \geq M\right\} \leq C\exp\left\{-(1-\varepsilon)\frac{nM^2}{4(\sigma^2+M)}\right\}$, where C is a universal constant. In particular, choose $\varepsilon = 1/2^s$ for $s > 0$, the bracketing integral condition (29) becomes

$$\int_{\frac{\sigma^2}{2^{7+s/2}}}^{\sigma} \sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{2^{12+5s/2}}\sqrt{n}\sigma^2, \tag{30}$$

and for any σ^2 such that (30) is satisfied, $\mathbb{P}\left(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \geq \frac{\sigma^2}{2^{2+s}}\right) \leq C\exp\left(-\Delta n\sigma^2\right);\ \Delta = \frac{2^s-1}{2^{5+2s}(2^{2+s}+1)}$.

Proof. Proof of Theorem C.3 Consider a decreasing sequence of real numbers $\delta_0 > \delta_1 > \cdots > \delta_N > 0$. Because $\mathcal{F}$ has a finite bracketing number, then by definition of $\log \mathcal{N}_{[]}$, we have for every $\delta_j, j \in [N]$, there exists a finite set $\mathcal{F}_j \subset \mathcal{F}$ with $|\mathcal{F}_j| = \mathcal{N}_{[]}(\delta_j, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))$ such that for every $f \in \mathcal{F}$, there exists a pair of function $g_j^u(f), g_j^l(f) \in \mathcal{F}_j$ with $g_j^l(f) \leq f \leq g_j^u(f)$ almost surely with respect to $\mathbb{P}$, and $\mathbb{E}[(g_j^u - g_j^l)^2]^{1/2} \leq \delta_j$. Hence, given a fixed $f \in \mathcal{F}$, we can construct a set of functions $\{g_j^l, g_j^u\}_{j=0}^N$ for every $\delta_0 > \delta_1 > \cdots > \delta_N > 0$. For $k \in [N]$, let $u_k(f)$ and $l_k(f)$ be defined as $u_k(f) = \min_{j\leq k} g_j^u(f)$ and $l_k(f) = \max_{j\leq k} g_j^l(f)$. We will write u_k and l_k in the following proof if their dependency on f is clear in context. It is easy to observe that $g_0^l = l_0 \leq l_1 \leq l_2 \leq \cdots \leq l_N \leq f \leq u_N \leq \cdots \leq u_0 = g_0^u$. Hence, with $\{g_j^l, g_j^u\}_{j=0}^N$, we have built a sequence of brackets with decreasing width. Note that by definition, $u_j \leq g_j^u$ and $l_j \geq g_j^l$, we have $\mathbb{E}[(u_j - l_j)^2] \leq \mathbb{E}[(g_j^u - g_j^l)^2] \leq \delta_j^2$ for $j \in [N]$. Consider another decreasing sequence $a_1 > a_2 > \cdots > a_N$ which need not to be positive. We will use $\{a_i\}_{i=1}^N$ to construct truncation. Let Ω be the domain of $\mathcal{F}$, and let $A_0 = \{\omega \in \Omega : u_0 - l_0 \geq a_1\}$, $A_k = \{\omega \in \Omega : u_k - l_k \geq a_{k+1} \quad \text{and} \quad u_j - l_j < a_{j+1} \quad \text{for} \quad j = 0, \ldots, k-1\}, \quad k \in \{1, \ldots, N-1\}$, and $A_N = \left(\bigcup_{j<N} A_j\right)^{\complement}$. Note that $\{A_j\}_{j=0}^N$ forms a partition of Ω and thus $\sum_{j=0}^N \mathbf{1}_{A_j} = 1$. With this and some computation, we can write every $f \in \mathcal{F}$ as $f = u_0 + \sum_{k=1}^N (u_k - u_{k-1})\mathbf{1}_{\bigcup_{j\geq k} A_j} + \sum_{k=0}^N (f - u_k)\mathbf{1}_{A_k}$. Let $\nu_n(f) = \mathbb{P}_n f - Pf$, and we are interested in the quantity that for $M \in \mathbb{R}$, $\mathbb{P}^* = \mathbb{P}(\sup_{f\in\mathcal{F}} \nu_n(f) \geq M)$. Let $\varepsilon \in (0,1)$, and $\{\gamma_1, \ldots, \gamma_N\}$ be a sequence of real numbers such that $\gamma_j > 0$ for all $j \in \{1, \ldots, N\}$ and

$$\sum_{j=1}^N \gamma_j \leq \frac{\varepsilon M}{8}. \tag{31}$$

Then we have $\mathbb{P}^* = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n(f) \geq M) \leq \mathbb{P}_1 + \sum_{k=1}^{N}\mathbb{P}_{2,k} + \sum_{k=0}^{N-1}\mathbb{P}_{3,k} + \mathbb{P}_4$, where $\mathbb{P}_1 = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n(u_0) > (1-\frac{\varepsilon}{4})M)$, $\mathbb{P}_{2,k} = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n((u_k - u_{k-1})\mathbf{1}_{\bigcup_{j\geq k}A_k}) > \gamma_k)$, $\mathbb{P}_{3,k} = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n((f-u_k)\mathbf{1}_{A_k}) > \gamma_{k+1}), k \in \{0,\ldots,N-1\}$, and $\mathbb{P}_4 = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n((f-u_N)\mathbf{1}_{A_N}) > \frac{\varepsilon M}{8}+\gamma_N)$. The following proof is dedicated to upper bound these quantities. Specifically, we consider a carefully selected set $\{\delta_j\}_{j=0}^N$. Let δ_0 be the smallest δ such that $\log\mathcal{N}_{[]}(\delta,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \leq \frac{\varepsilon}{4}\psi(M,\sigma^2,\mathcal{F})$, i.e., $\delta_0 = \left(\log\mathcal{N}_{[]}\right)^{-1}\left(\frac{\varepsilon}{4}\psi(M,\sigma^2,\mathcal{F})\right)$. Subsequently, let $s = \varepsilon M/8$ and for $j \in \{0,\ldots,N-1\}$, $\delta_{j+1} = \max\left(s, \sup\{\delta \leq \delta_j/2 : \log\mathcal{N}_{[]}(\delta,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \geq 4\cdot\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P}))\}\right)$. Let $N = \min\{j : \delta_j \leq s\}$. Suppose that $\log\mathcal{N}_{[]}(\sigma,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \leq \frac{\varepsilon}{4}\psi(M,\sigma^2,\mathcal{F})$, then $\delta_0 \leq \sigma$ by the monotonicity of the bracketing number. With $\{\delta_j\}_{j=0}^N$, we proceed to define $\{\gamma_j\}$ and $\{a_j\}$. For $j \in \{1,\ldots,N\}$, let $\gamma_j = 4\frac{\delta_{j-1}}{\sqrt{n}}\left(\frac{\sum_{i\leq j}\log\mathcal{N}_{[]}(\delta_i,\mathcal{F},\mathcal{L}^2(\mathbb{P}))}{\varepsilon}\right)^{1/2}$ and $a_j = \frac{8\delta_{j-1}^2}{\gamma_j}$.

Now we are at the stage of presenting the upper bounds. To bound $\mathbb{P}_4$, note that for any $f \in \mathcal{F}$, because $(f-u_N)\mathbf{1}_{A_N} \leq 0$ almost surely, $\nu_n((f-u_N)\mathbf{1}_{A_N}) \leq \mathbb{P}(u_N - l_N) \leq \mathbb{E}[(u_N-l_N)^2]^{1/2} \leq \delta_N \leq s = \frac{\varepsilon M}{8}$. Hence, we have $\mathbb{P}_4 = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n((f-u_N)\mathbf{1}_{A_N}) > \frac{\varepsilon M}{8}+\gamma_N) = 0$. To bound $\mathbb{P}_3$, note that for any $f \in \mathcal{F}$ and $k \in \{0,\ldots,N-1\}$, $\nu_n((f-u_k)\mathbf{1}_{A_k}) \leq \mathbb{E}[(u_k-f)\mathbf{1}_{A_k}] \leq \sqrt{\mathbb{E}[(u_k-l_k)^2]\mathbb{E}[\mathbf{1}_{A_k}^2]} \leq \sqrt{\delta_k^2\mathbb{P}(A_k)}$. Moreover, due to construction of $\{A_k\}$, we have $u_k - l_k \geq a_{k+1}$ on A_k. Thus, due the chain of inequalities, $\mathbb{E}[(u_k-l_k)^2] \geq \mathbb{E}[(u_k-l_k)^2\mathbf{1}_{A_k}] \geq a_{k+1}^2\mathbb{E}[\mathbf{1}_{A_k}]$, we have $\mathbb{P}(A_k) \leq \frac{\mathbb{E}[(u_k-l_k)^2]}{a_{k+1}^2} \leq \frac{\delta_k^2}{a_{k+1}^2}$. Plugging this back, we have for all $f \in \mathcal{F}$, $\nu_n((f-u_k)\mathbf{1}_{A_k}) \leq \sqrt{\delta_k^2\mathbb{P}(A_k)} \leq \sqrt{\delta_k^2\frac{\delta_k^2}{a_{k+1}^2}} \leq \frac{\delta_k^2}{a_{k+1}}$. By construction of a_j, i.e., $a_j = 8\delta_{j-1}^2/\gamma_j$, we have $\gamma_j \geq \delta_{j-1}^2/a_j$. Hence, $\nu_n((f-u_k)\mathbf{1}_{A_k}) \leq \delta_k^2/a_{k+1} \leq \gamma_{k+1}$ and $\mathbb{P}_{3,k} = \mathbb{P}(\sup_{f\in\mathcal{F}}\nu_n((f-u_k)\mathbf{1}_{A_k}) > \gamma_{k+1}) = 0$.

To upper bound $\mathbb{P}_1$, we have $\mathbb{P}_1 \leq \mathbb{P}(\max_{u_0\in\mathcal{F}_0}\nu_n(u_0) > (1-\frac{\varepsilon}{4})M) \leq \sum_{u_0\in\mathcal{F}_0}\mathbb{P}(\nu_n(u_0) > (1-\frac{\varepsilon}{4})M)$. By assumption for any $u_0 \in \mathcal{F}_0 \subset \mathcal{F}$ and $k \geq 2$, $\mathbb{E}[|u_0 - \mathbb{E}[u_0]|^k] \leq \frac{k!}{2}2^{k-2}\sigma^2$. Recall that by Bernstein's Inequality, for any random random X with $\mathbb{E}[|X-\mathbb{E}[X]|^k] \leq \frac{k!}{2}2^{k-2}\sigma^2$ and $X_1,\ldots,X_n$ as random variables with the same distribution as X, we have $\mathbb{P}(\frac{1}{n}\sum_{i=1}^n(X_i - \mathbb{E}[X_i]) > t) \leq \exp\left(-\frac{nt^2}{2(\sigma^2+2t)}\right) \leq \exp\left(-\frac{nt^2}{4(\sigma^2+t)}\right)$. Hence, let $\psi(M,\sigma^2,\mathcal{F}) = nM^2/4(\sigma^2+M)$, we thus have $\mathbb{P}_1 \leq \mathcal{N}_{[]}(\delta_0,\mathcal{F},\mathcal{L}^2(\mathbb{P}))\exp(-\psi((1-\frac{\varepsilon}{4})M,\sigma^2,\mathcal{F}))$. Moreover, by definition of δ_0, we have $\mathcal{N}_{[]}(\delta_0,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \leq \exp(\frac{\varepsilon}{4}\psi(M,\sigma^2,\mathcal{F}))$. After straightforward calculation. we have $\mathbb{P}_1 \leq \exp(-(1-\varepsilon)\psi(M,\sigma^2,\mathcal{F}))$.

To upper bound $\mathbb{P}_2$, note that for any $k \in \{1,\ldots,N\}$, $\mathrm{Var}[(u_k-u_{k-1})\mathbf{1}_{\cup_{j\geq k}A_j}] \leq \mathbb{E}[((u_k - u_{k-1})\mathbf{1}_{\cup_{j\geq k}A_j})^2] \leq \mathbb{E}[(u_k-u_{k-1})^2] \leq \mathbb{E}[(u_{k-1}-l_{k-1})^2] \leq \delta_{k-1}^2$. By construction of $\{A_k\}$, we have $0 \geq u_k - u_{k-1} \geq l_{k-1} - u_{k-1} \geq -a_k$ on $\bigcup_{j\geq k}A_j$ for $k = 1,\ldots,N$. Then by Bernstein's Inequality for random variable with bounded support and finite variance, it holds that $\mathbb{P}(\nu_n((u_k-u_{k-1})\mathbf{1}_{\bigcup_{j\geq k}A_j}) > \gamma_k) \leq \exp(-\frac{n\gamma_k^2}{2(\delta_{k-1}^2+a_k\gamma_k/3)})$. Note that u_k is defined as $\min_{j\leq k}g_j^u$ where $g_j^u \in \mathcal{F}_j$. This implies that the possible number of u_k is upper bounded by $\prod_{j=0}^k|\mathcal{F}_j|$. Then we have $\prod_{j=0}^k|\mathcal{F}_j|\prod_{j=0}^{k-1}|\mathcal{F}_j| \leq \exp\left(2\sum_{j=1}^k\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P}))\right)$. This leads to $\mathbb{P}_{2,k} \leq \exp\left(2\sum_{j=1}^k\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P})) - \frac{n\gamma_k^2}{2(\delta_{k-1}^2+a_k\gamma_k/3)}\right)$. By definition of γ_k and

a_k, we also have $\frac{n\gamma_k^2}{2(\delta_{k-1}^2+a_k\gamma_k/3)} \geq \frac{2\sum_{j\leq k}\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P}))}{\varepsilon}$. Plugging this back in, we have $\mathbb{P}_{2,k} \leq \exp\left(-2\frac{1-\varepsilon}{\varepsilon}\sum_{j\leq k}\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P}))\right)$. Recall that by construction $\log\mathcal{N}_{[]}(\delta_0,\mathcal{F},\mathcal{L}^2(\mathbb{P})) = \varepsilon\psi(M,\sigma^2,\mathcal{F}))/4$ and $\log\mathcal{N}_{[]}(\delta_k,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \geq 4\log\mathcal{N}_{[]}(\delta_{k-1},\mathcal{F},\mathcal{L}^2(\mathbb{P}))$. Hence, $\log\mathcal{N}_{[]}(\delta_k,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \geq 4^{k-1}\varepsilon\psi(M,\sigma^2,\mathcal{F}))$ for $k=1,\ldots,N$. Then, $\sum_{j=1}^{k}\log\mathcal{N}_{[]}(\delta_j,\mathcal{F},\mathcal{L}^2(\mathbb{P})) \geq 4^{k-1}\varepsilon\psi(M,\sigma^2,\mathcal{F})$. This leads to $\mathbb{P}_{2,k} \leq \exp\left(-2(1-\varepsilon)4^{k-1}\psi(M,\sigma^2,\mathcal{F})\right)$, and $\sum_{k=1}^{N}\mathbb{P}_{2,k} \leq \sum_{k=1}^{\infty}\exp(-(1-\varepsilon)\psi(M,\sigma^2,\mathcal{F}))^{2^k}$. To continue let $\alpha = \exp(-(1-\varepsilon)\psi(M,\sigma^2,\mathcal{F})) < 1$ where $\psi(M,\sigma^2,\mathcal{F}) = \frac{nM^2}{4\sigma^2+4M}$. By the assumption that $\sigma^2 \geq 1/n$ and $M = \varepsilon\sigma/4$, we have $\psi(M,\sigma^2,\mathcal{F}) = n\varepsilon^2\sigma^2/16(4+\varepsilon) \geq \varepsilon^2/16(4+\varepsilon) > 0$. Then we have $\sum_{k=1}^{N}\mathbb{P}_{2,k} \leq \sum_{k=1}^{\infty}\alpha^{2^k} \leq c\cdot\exp(-(1-\varepsilon)\psi(M,\sigma^2,\mathcal{F}))$, where c is a universal constant. With all the results above, we finally have $\mathbb{P}^* \leq \mathbb{P}_1 + \sum_{k=1}^{N}\mathbb{P}_{2,k} + \sum_{k=0}^{N-1}\mathbb{P}_{3,k} + \mathbb{P}_4 \leq c\cdot\exp(-(1-\varepsilon)\psi(M,\sigma^2,\mathcal{F}))$, where c is a different universal constant. The last step is to verify that (31) holds. With Lemma 3.1 in [1] and $s = \varepsilon M/8$, this can be verified with straightforward calculation. □

C.4 Proof of Corollary 4.1

Proof of Corollary 4.1. By Theorem 4.2, consider any σ_n that satisfies (7). For any $\alpha \geq \sigma_n^2/96$, with probability at least $1 - C_1\exp(-2C_2n\alpha)$ where C_1 and C_2 are universal constants that $p_0 \in \Omega_n(\alpha)$, and $\sup_{p\in\Omega_n(\alpha)}H^2(p,p_0) \leq 4\alpha/25$. Then by Theorem 4.1, we have $\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \lesssim r_{s^\star}\sqrt{\alpha/\pi_S(s^\star)}$. Here we need $C_1\exp(-2C_2n\alpha) \leq \delta$, which for sufficiently small δ is equivalent to $\alpha \gtrsim \frac{1}{n}\log\frac{1}{\delta}$. Note that we also need $\alpha \geq \sigma_n^2/96$. Choosing $\alpha = \sigma_n^2 + \frac{1}{n}\log\frac{1}{\delta}$ satisfies these two constraints. Plugging this back into $\mathcal{R}(\widehat{s}_{\mathfrak{p},n}) \lesssim r_{s^\star}\sqrt{\alpha/\pi_S(s^\star)}$ concludes the proof. □

C.5 Proof of Lemma 4.2

Proof of Lemma 4.2. Define $J(t) = \int_{C_1\alpha_n}^{C_2t}\sqrt{\log\mathcal{N}_{[]}\left(u/C_3,\mathcal{P},H\right)}\mathrm{d}u$; $t \geq \alpha_n$. By the fact that $u \mapsto \sqrt{\log\mathcal{N}_{[]}\left(u/C_3,\mathcal{P},H\right)}$ is nonnegative and nonincreasing, $J'(t) \geq 0$ and is nonincreasing in t. This suggests that $J(t)$ is concave in t, and hence, $\alpha \mapsto J(\sqrt{\alpha})/\sqrt{\alpha}$ is nonincreasing. For any $\alpha \geq \alpha_n$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{\alpha}}\int_{C_1\alpha}^{C_2\sqrt{\alpha}}\sqrt{\log\mathcal{N}_{[]}\left(u/C_3,\mathcal{P},H\right)}\mathrm{d}u \leq \frac{1}{\sqrt{\alpha}}\int_{C_1\alpha_n}^{C_2\sqrt{\alpha}}\sqrt{\log\mathcal{N}_{[]}\left(u/C_3,\mathcal{P},H\right)}\mathrm{d}u \\
&= \frac{J(\sqrt{\alpha})}{\sqrt{\alpha}} \leq \frac{J(\sqrt{\alpha_n})}{\sqrt{\alpha_n}} = \frac{1}{\sqrt{\alpha_n}}\int_{C_1\alpha_n}^{C_2\sqrt{\alpha_n}}\sqrt{\log\mathcal{N}_{[]}\left(u/C_3,\mathcal{P},H\right)}\mathrm{d}u \leq C_4\sqrt{n\alpha_n} \leq C_4\sqrt{n\alpha},
\end{aligned}$$

where the second inequality is due to that $\alpha \mapsto \frac{J(\sqrt{\alpha})}{\sqrt{\alpha}}$ is nonincreasing. This suggests that α also satisfies (7). □

D Proofs of Results in Section 5

D.1 Proof of Theorem 5.1

Proof of Theorem 5.1. The MNL model has parameters $\theta \in \mathbb{R}^d$ and $x_i \in \mathbb{R}^d$ for $i \in [N]$ with the following conditional probability function

$$p(a \mid s; \theta) = \frac{\exp(\theta^\top x_a)\,\mathbf{1}_{\{a \in s\}} + \mathbf{1}_{\{a=0\}}}{1 + \sum_{j \in s} \exp(\theta^\top x_j)}.$$

We first establish that $p(a|s;\theta)$ is a Lipchitz continuous function of θ for all s and $a \in s \cup \{0\}$. To this end,

$$\begin{aligned}
\nabla_\theta p(a|s;\theta) &= \frac{\exp(\theta^\top x_a) x_a \cdot (1 + \sum_{j\in s} \exp(\theta^\top x_j)) - \exp(\theta^\top x_a)(\sum_{j\in s} \exp(\theta^\top x_j) x_j)}{\left(1 + \sum_{j\in s} \exp(\theta^\top x_j)\right)^2} \\
&= \frac{\exp(\theta^\top x_a)}{1 + \sum_{j\in s} \exp(\theta^\top x_j)} \cdot \frac{x_a + \sum_{j\in s} \exp(\theta^\top x_j) x_a - \sum_{j\in s} \exp(\theta^\top x_j) x_j}{1 + \sum_{j\in s} \exp(\theta^\top x_j)} \\
&= p(a|s;\theta) \cdot \left(\frac{x_a}{1 + \sum_{j\in s} \exp(\theta^\top x_j)} + \sum_{j \in s} p(j|s;\theta) x_a + \sum_{j \in s} p(j|s;\theta) x_j \right).
\end{aligned} \tag{32}$$

With this we have $||\nabla_\theta \sqrt{p(a|s;\theta)}||_2 \le \frac{1}{2}\sqrt{p(a|s;\theta)} \left(||x_a||_2 + \sum_{j\in s} p(j|s;\theta)||x_a||_2 + \sum_{j\in s} p(j|s;\theta)||x_j||_2 \right) \le \frac{3}{2}\sqrt{p(a|s;\theta)} C_x$. Hence, we have proved that the function $\sqrt{p(a|s;\theta)}$ is $3C_x/2$-Lipschitz continuous with respect to θ. Then for any θ_1, θ_2 and fixed s, their squared Hellinger distance can then be bounded as

$$\begin{aligned}
h^2(p(\cdot|s;\theta_1), p(\cdot|s;\theta_2)) &= \frac{1}{2} \sum_{a \in s} \left(\sqrt{p(a|s;\theta_1)} - \sqrt{p(a|s;\theta_2)} \right)^2 \\
&\le \frac{1}{2} \sum_{a \in s} \left(\frac{3C_x \sqrt{p(a|s;\theta)}}{2} ||\theta_1 - \theta_2||_2 \right)^2 \le \Delta \cdot ||\theta_1 - \theta_2||_2^2,
\end{aligned} \tag{33}$$

where $\Delta = 9C_x^2/8$. This implies that $H^2(\theta_1, \theta_2) \le \Delta ||\theta_1 - \theta_2||_2^2$. Then if $\theta \in \Theta$ where Θ is bounded, we have $\mathcal{N}_{[]}(u, \mathcal{P}, H) \le \mathcal{N}_{[]}(u/\Delta, \Theta, ||\cdot||) \simeq \left(\frac{\Delta}{u}\right)^d$. With this, the left hand side of the entropy integral in (7) can be upper bounded by

$$\int_{C_1\alpha_n}^{\sqrt{\alpha_n}} \sqrt{\log \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e}, \mathcal{P}, H\right)} \mathrm{d}u \le \sqrt{d} \int_0^{\sqrt{\alpha_n}} \sqrt{\log\left(\frac{\Delta \cdot 2\sqrt{2}e}{u}\right)} \mathrm{d}u \lesssim \sqrt{\alpha_n d \log \frac{\Delta \cdot 2\sqrt{2}e}{\sigma_n}},$$

where the last inequality follows from Lemma E.1 provided that $\sqrt{\alpha_n} \le \Delta \cdot 2\sqrt{2}$. Hence, to solve (7), it is sufficient to solve the equation $\sqrt{\alpha_n d \log \frac{\Delta \cdot 2\sqrt{2}e}{\sqrt{\alpha_n}}} \lesssim C_2 \sqrt{n} \alpha_n$, which is equivalent to $\alpha_n \gtrsim \frac{d}{n} \log \frac{C_x}{\alpha_n}$. Choosing $\alpha_n = \frac{d}{n} \log \frac{nC_x}{d}$, and under the assumption that $n \gtrsim d \log(n/d)$, the right hand side of the above inequality reads as

$$\frac{d}{n} \log \left(\frac{n}{d} \frac{C_x}{\log(n/d)} \right) \lesssim \frac{d}{n} \log \frac{nC_x}{d} = \alpha_n. \tag{34}$$

Thus, the above choice of α_n satisfies (7). Moreover, $\sqrt{\alpha_n} \lesssim C_x^2$ is satisfied provided that $n \gtrsim d\log(n/d)$. Then applying Corollary 4.1 concludes the proof. □

D.2 Proof of Theorem 5.2

Proof of Theorem 5.2. To avoid notation overload, only in this proof we let $g(s,a;\theta_k)$ for $\theta_k \in \mathbb{R}^d$ and $k \in [K]$ denote the conditional probability function of the MNL model $g(s,a;\theta_k) = \frac{\exp(\theta_k^\top x_a)}{1+\sum_{j\in s}\exp(\theta_k^\top x_j)}$, and let $p(a|s;\theta)$ be the conditional probability function of the LCL model as defined in (11) where $\theta = [\theta_1,\dots,\theta_K;\lambda_1,\dots,\lambda_K]$. Thus we have $p(a|s;\theta) = \sum_{k=1}^K \lambda_k \cdot g(a,s;\theta_k)$. First note that, under Assumption 2, $g(a,s;\theta)$ is uniformly lower bounded for all s and $a \in s \cup \{0\}$:

$$g(a,s;\theta) = \frac{\exp(\theta^\top x_a)}{1+\sum_{j\in s}\exp(\theta^\top x_j)} \geq \frac{\exp(\theta^\top x_a)}{1+|s|\exp(||\theta||_2 C_x)} \geq \frac{\exp(-C_\theta C_x)}{1+N\exp(C_\theta C_x)} := \Delta > 0 \quad (35)$$

With the same calculations in (32), we have

$$\nabla_{\theta_k} p(a|s;\theta) = \lambda_k \cdot g(a,s;\theta_k) \cdot \left(\frac{x_a}{1+\sum_{j\in s}\exp(\theta_k^\top x_j)} + \sum_{j\in s} g(j,s;\theta_i) x_a + \sum_{j\in s} g(j,s;\theta_k) x_j \right);$$

$$\nabla_{\lambda_k} p(a|s;\theta) = g(a,s;\theta_k).$$

Then we have

$$\begin{aligned}
&||\nabla_\theta \sqrt{p(a|s;\theta)}||_2 = \frac{1}{2}\sqrt{p(a|s;\theta)}||(\nabla_\theta p(a|s;\theta))/p(a|s;\theta)||_2 \\
&\leq \frac{1}{2}\sqrt{p(a|s;\theta)}\left(\sum_{i=1}^K ||\nabla_{\theta_i} p(a|s;\theta)/p(a|s;\theta)||_2 + ||\nabla_\lambda p(a|s;\theta)/p(a|s;\theta)||_2\right) \\
&\leq \frac{1}{2}\sqrt{p(a|s;\theta)}\left(3KC_x + \frac{\sum_{i=1}^K g(a,s;\theta_i)}{\sum_{i=1}^K \lambda_i \cdot g(a,s;\theta_i)}\right).
\end{aligned}$$

To proceed, with (35), we have $\frac{\sum_{i=1}^K g(a,s;\theta_i)}{\sum_{i=1}^K \lambda_i \cdot g(a,s;\theta_i)} \leq \frac{\sum_{i=1}^K g(a,s;\theta_i)}{\Delta \sum_{i=1}^K \lambda_i} \leq K/\Delta$. Thus, the density function of LCL model is $(3KC_x + K/\Delta)$-Lipschitz continuous. With the same computation in (33), let $C_{LCL} = \frac{K^2(3C_x+1/\Delta)^2}{2}$, it holds that $\mathcal{N}_{[]}(u,\mathcal{P},H) \simeq (u/C_{LCL})^{K(d+1)-1}$. Here we have $K(d+1)-1 = Kd+K-1$ as the effective dimension as there are only $K-1$ free parameters for $\lambda \in \Delta^{K-1}$ in the $K-1$ simplex. With this, the left hand side of the entropy integral in (7) can be upper bounded by $\int_{C_1\alpha_n}^{\sqrt{\alpha_n}} \sqrt{\log \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e},\mathcal{P},H\right)}\mathrm{d}u \lesssim \sqrt{K(d+1)-1}\int_0^{\sqrt{\alpha_n}} \sqrt{\log\left(\frac{C_{LCL}\cdot 2\sqrt{2}e}{u}\right)}\mathrm{d}u \lesssim \sqrt{\alpha_n(Kd+K-1)\log\frac{C_{LCL}\cdot 2\sqrt{2}e}{\sigma_n}}$. Hence, to solve (7), it is sufficient to solve the equation

$$\sqrt{\alpha_n(Kd+K-1)\log\frac{C_{LCL}\cdot 2\sqrt{2}e}{\sigma_n}} \lesssim \sqrt{n}\alpha_n,$$

which is equivalent to $\alpha_n \gtrsim \frac{(Kd+K-1)}{n} \log \frac{C_{LCL}}{\alpha_n}$. Using the same technique in (34), we have $\alpha_n = \frac{(Kd+K-1)}{n} \log \frac{C_{\mathcal{P}} n}{(Kd+K-1)}$ satisfies the above inequality. Then applying Corollary 4.1 concludes the proof. □

D.3 Proof of Theorem 5.3

Proof of Theorem 5.3. Let $p(a \in s_j|s;\theta)$ be the conditional choice probability function as defined in (12) where $\theta = [\tilde{\theta}, \lambda]$. Note that $p(a \in s_j|s;\theta)$ can be written as: for $a \in s_j$,

$$p(a \in s_j|s;\theta) = g_A(a, s_j; \tilde{\theta}, \lambda) \cdot g_B(s_j, s; \tilde{\theta}, \lambda),$$

where $g_A(a, s_j; \tilde{\theta}, \lambda) = \frac{\exp(\tilde{\theta}^\top x_a/\lambda_j)}{\sum_{i\in S_j} \exp(\tilde{\theta}^\top x_i/\lambda_j)}$ and $g_B(s_j, s; \tilde{\theta}, \lambda) = \frac{\left(\sum_{i\in S_j} \exp(\tilde{\theta}^\top x_i/\lambda_j)\right)^{\lambda_j}}{1+\sum_{l=1}^K \left(\sum_{i\in S_l} \exp(\tilde{\theta}^\top x_i/\lambda_l)\right)^{\lambda_l}}$.

For $j \in [K]$, let $V_j = \sum_{i\in S_j} \exp(\tilde{\theta}^\top x_i/\lambda_j)$ and $\kappa_{i,j} = \exp(\tilde{\theta}^\top x_i/\lambda_j)$. With straightforward calculation, it can be shown that

$$\nabla_{\tilde{\theta}} g_A(a, s_j; \tilde{\theta}, \lambda) = g_A(a, s_j; \tilde{\theta}, \lambda) \frac{x_a - \sum_{i\in S_j} \kappa_{i,j} x_i}{\lambda_j v_j};$$

$$\nabla_{\tilde{\theta}} g_B(s_j, s; \tilde{\theta}, \lambda) = g_B(s_j, s; \tilde{\theta}, \lambda) \frac{(\sum_{k\in[K]} v_k^{\lambda_k}) \frac{\sum_{i\in S_j} \kappa_{i,j} x_i}{v_j} - \sum_{k\in[K]} v_k^{\lambda_k} \frac{\sum_{i\in S_k} \kappa_{i,k} x_i}{v_k}}{\sum_{k\in[K]} v_k^{\lambda_k}}.$$

Moreover, with the above inequalities and the given assumption that $1/\lambda_j \le C_\lambda$ for $j \in [K]$, we have $||\nabla_{\tilde{\theta}} g_A(a, s_j; \tilde{\theta}, \lambda)|| \le g_A(a, s_j; \tilde{\theta}, \lambda) \frac{2C_x}{\lambda_j}$ and $||\nabla_{\tilde{\theta}} g_B(s_j, s; \tilde{\theta}, \lambda)|| \le 2 g_B(s_j, s; \tilde{\theta}, \lambda) C_x$. Hence, through triangular inequalities of norms, we have

$$\begin{aligned} &||\nabla_{\tilde{\theta}} p(a \in s_j|s;\theta)|| = ||\nabla_{\tilde{\theta}} g_A(a, s_j; \tilde{\theta}, \lambda) g_B(j; \tilde{\theta}, \lambda) + g_A(a, s_j; \tilde{\theta}, \lambda) \nabla_{\tilde{\theta}} g_B(s_j, s; \tilde{\theta}, \lambda)||; \\ &\le 2 g_A(a, s_j; \tilde{\theta}, \lambda) g_B(s_j, s; \tilde{\theta}, \lambda) \max(2C_x/\lambda_j, \lambda) \le 2 p(a \in s_j|s;\theta) C_x \max(2C_\lambda, 1). \end{aligned}$$

Regarding λ, it is useful to use the fact that $\nabla_{\lambda_i} \kappa_{a,j} = -\tilde{\theta}^\top x_a \cdot \exp(\tilde{\theta}^\top x_a/\lambda_j)/\lambda_j^2$ for $i = j, a \in S_j$ and $\nabla_{\lambda_i} \kappa_{a,j} = 0$ otherwise. With these, we have

$$\nabla_{\lambda_i} g_A(a, s_j; \tilde{\theta}, \lambda) = g_A(a, s_j; \tilde{\theta}, \lambda) \frac{\sum_{t\in S_j} \kappa_{t,j} \tilde{\theta}^\top x_t - \tilde{\theta}^\top x_a}{\lambda_j^2 v_j} \mathbf{1}_{\{i=j\}};$$

$$\nabla_{\lambda_i} g_B(s_j, s; \tilde{\theta}, \lambda) = g_B(s_j, s; \tilde{\theta}, \lambda) \frac{-\nabla_{\lambda_i} v_i^{\lambda_i}}{\sum_{k\in[K]} v_k^{\lambda_k}} \mathbf{1}_{\{i\neq j\}} + \frac{\sum_{k\in[K]\setminus\{j\}} v_k^{\lambda_k}}{\sum_{k\in[K]} v_k^{\lambda_k}} \frac{-\nabla_{\lambda_j} v_j^{\lambda_j}}{\sum_{k\in[K]} v_k^{\lambda_k}} \mathbf{1}_{\{i=j\}}.$$

To continue, with straightforward calculation, for $j \in [K]$, $\nabla_{\lambda_j} v_j^{\lambda_j} = v_j^{\lambda_j} \left(\frac{\sum_{i\in S_j} -\tilde{\theta}^\top x_i \kappa_{i,j}}{\lambda_j v_j} + \ln v_j \right)$. Then we have $\sum_{i\in[K]} ||\nabla_{\lambda_i} g_A(a, s_j; \tilde{\theta}, \lambda)|| \le g_A(a, s_j; \tilde{\theta}, \lambda) \frac{2C_x C_\theta}{\lambda_j^2}$ where $||\nabla_{\lambda_i} g_B(s_j, s; \tilde{\theta}, \lambda)|| \le g_B(s_j, s; \tilde{\theta}, \lambda) g_B(s_i, s; \tilde{\theta}, \lambda) \frac{C_x C_\theta}{\lambda_j}$ if $i \neq j$ and $||\nabla_{\lambda_i} g_B(s_j, s; \tilde{\theta}, \lambda)|| \le g_B(s_j, s; \tilde{\theta}, \lambda) \frac{C_x C_\theta}{\lambda_j}$ if $i = j$. Hence, $||\nabla_{\lambda_i} g_B(s_j, s; \tilde{\theta}, \lambda)|| \le g_B(s_j, s; \tilde{\theta}, \lambda) C_x C_\theta C_\lambda$ for all $i, j \in [K]$. With these, we have

$$\begin{aligned} &||\nabla_\lambda p(a \in s_j|s;\theta)||_2 = ||\nabla_\lambda g_A(a, s_j; \tilde{\theta}, \lambda) g_B(s_j, s; \tilde{\theta}, \lambda) + g_A(a, s_j; \tilde{\theta}, \lambda) \nabla_\lambda g_B(s_j, s; \tilde{\theta}, \lambda)|| \\ &\le 2 p(a \in s_j|s;\theta) C_x C_\theta \max(2C_\lambda^2, C_\lambda). \end{aligned}$$

The above results lead to

$$||\nabla_{\tilde{\theta},\lambda}\sqrt{p(a \in s_j|s;\theta)}|| \leq \frac{\sqrt{p(a \in s_j|s;\theta)}}{2}\frac{(||\nabla_\theta\sqrt{p(a \in s_j|s;\theta)}|| + ||\nabla_\lambda\sqrt{p(a \in s_j|s;\theta)}||)}{p(a \in s_j|s;\theta)}$$
$$\leq 2\sqrt{p(a \in s_j|s;\theta)}C_x(1 + C_\theta C_\lambda)\max(2C_\lambda, 1).$$

Let $C_{NL} = (2C_x(1+C_\theta C_\lambda)\max(2C_\lambda,1))^2/2$ and we have for any $s = \{s_j\}_{j=1}^K$, and any pairs of $\theta_1 = (\tilde{\theta}_1, \lambda_1), \theta_2 = (\tilde{\theta}_2, \lambda_2)$ that $h^2(p(\cdot|s;\theta_1), p(\cdot|s;\theta_2)) = \frac{1}{2}\sum_{a\in s_j, s_j\in s}\left(\sqrt{p(a \in s_j|s;\theta_1)} - \sqrt{p(a \in s_j|s;\theta_2)}\right)^2 \leq C_{NL}||\theta_1-\theta_2||_2^2$, where $\theta_1, \theta_2 \in \mathbb{R}^{K+d}$. This implies that $H^2(\theta_1,\theta_2) \leq C_{NL}||\theta_1-\theta_2||_2^2$. Then with the same steps in the proofs for the MNL and LCL models, we can choose $\alpha_n = \frac{K+d}{n}\log\frac{C_{\mathcal{P}}n}{K+1}$ so that it satisfies the entropy condition (7). Applying Corollary 4.1 concludes the proof. □

D.4 Proof of Theorem 5.4

Proof of Theorem 5.4. We present the proof for the case where $d \leq N$. The proof for the case where $d > N$ is identical. Let the four-element tuple $[\mathbf{x} = \{x_j\}_{j\in[N]}, r, \theta_0, \pi_S]$ define an instance of the offline assortment optimization problem under the MNL model where θ_0 is the true model parameter. Moreover, let $I = [\mathbf{x}, r]$ and $P = [\theta_0, \pi_S]$. Let $\mathcal{I} \ni I$ and $\mathcal{P} \ni P$ be the set of possible offline assortment optimization problems under consideration. The goal is to lower bound $\mathcal{R} = \inf_{\widehat{s}}\sup_{I\in\mathcal{I},P\in\mathcal{P}}\mathbb{E}\left[|\mathcal{V}(s^\star) - \mathcal{V}(\widehat{s})|\right]$, where $s^\star = s^\star(I,P)$ is the optimal assortment for instance $[I,P]$, and $\mathcal{V}(s) = \mathcal{V}(s;I,P) = \sum_{j\in s} r(s,j)p(j|s;\theta_0)$ is the true expected revenue function of the instance $[I,P]$. To establish a lower bound for $\mathcal{R}$, we first consider d instances of I. For the case where $d > N$, we consider N instances. All the other proof steps are identical. Specifically, let the i-th instances I_i be defined as follows: (i) $r(s,i) = r_i = 1$ and $r(s,j) = r_j = 0$ for $j \neq i$; (ii) For $i \in \{1,\ldots,d\}$, we set $x_i \in \mathbb{R}^d$ to have all zero elements except that the i-th element is set to $1/\varepsilon$ where $\varepsilon \in \mathbb{R}$ is to be chosen later , e.g., $x_1 = [1/\varepsilon, 0, \ldots, 0]$, $x_2 = [0, 1/\varepsilon, 0, \ldots, 0]$, etc. For $i \in \{d+1,\ldots,N\}$, x_i can be any vector in $\mathbb{R}^d$ as long as x_i is distinct to x_j for $j < i$.

Clearly, the optimal assortment for I_i is always the singleton set $\{i\}$, regardless of the value of θ_0. We also have $\mathcal{R} \geq \mathcal{R}_1 = \inf_{\widehat{s}}\sup_{P\in\mathcal{P}}\mathbb{E}\left[\sum_{i=1}^d \frac{1}{d}|\mathcal{V}(s^\star;I_i,P) - \mathcal{V}(\widehat{s};I_i,P)|\right]$ because the regret average over d instances is always not greater than the worst instance. Next we continue to lower bound $\mathcal{R}_1$ by specifying a π_S and selecting a subset of MNL models $\widetilde{\Theta} \subset \Theta$ for the worst case regret. Consider the assortment set $\mathbb{S} := \{\{1\},\ldots,\{d\}\}$. We define π_S as $\pi_S(s) = 1/d$ if $s \in \mathbb{S}$ otherwise $\pi_S(s) = 0$. In other words, the probability mass is uniformly distributed on all assortments in $\mathbb{S}$. Let $\widetilde{\Theta} := \{\delta\cdot v : v \in \{-1,1\}^d\}$ where $\delta \in \mathbb{R}$ is to be chosen later. Then $\mathcal{R} \geq \mathcal{R}_1 \geq \mathcal{R}_2 = \inf_{\widehat{s}}\sup_{\theta\in\widetilde{\Theta}}\mathbb{E}\left[\sum_{i=1}^d \frac{1}{d}|\mathcal{V}(s^\star;\theta) - \mathcal{V}(\widehat{s};\theta)|\right]$.

In the subsequent proof, for $v \in \{-1,1\}^d$, we use P_v to denote the distribution $P_v(s,a) = \pi_S(s)p(a|s;\delta\cdot v)$ and θ_v for $\theta_v = \delta\cdot v$. We write $v \sim_i v'$ if v and v' differ in only the i-th coordinate. we use $v^+(\theta^+)$ and $v^-(\theta^-)$ to denote a pair of v where only 1 element is different. Note that if δ is small enough, then $|\theta^T x_i| \leq 1$ for all i. To satisfy this requirement, one sufficient condition is $|\theta^T x_i| \leq ||\theta||\cdot||x_i|| \leq \frac{\delta}{\varepsilon} \leq 1 \Rightarrow \delta \leq \varepsilon$.

For any v and v' such that $v \sim_i v'$, we have

$$\begin{aligned} |\mathcal{V}(s^\star; I_i, P_v) - \mathcal{V}(s^\star; I_i, P_{v'})| &= \frac{|\exp(x_i^T \delta v) - \exp(x_i^T \delta v')|}{\big(1+\exp(x_i^T \delta v)\big)\big(1+\exp(x_i^T \delta v')\big)} \\ \geq \frac{\exp(\delta/\varepsilon) - \exp(-\delta/\varepsilon)}{2+\exp(\delta/\varepsilon)+\exp(-\delta/\varepsilon)} &\geq \frac{\exp(\delta/\varepsilon)-1}{3+e} \geq \frac{\delta}{(3+e)\varepsilon}, \end{aligned} \tag{36}$$

where for the second to last inequality in (36) we use the fact that $|\theta^T x_i| \leq 1$ hence $\exp(-\delta/\varepsilon) \leq 1$ and $\exp(\delta/\varepsilon) \leq e$; for the last inequality in (36) we use the fact that $\exp(x) - 1 \geq x$ for $0 \leq x \leq 1$.

Thus, we have proved that, for any v and v' such that $v \sim_i v'$, $|\mathcal{V}(s^\star; I_i, P_v) - \mathcal{V}(s^\star; I_i, P_{v'})| \geq \delta/(3+e)\varepsilon$. Let n denote the number of samples and P^n denote the product distribution of P. Then, from Assouad's Lemma [56], we have

$$\mathcal{R}_2 \geq \frac{\delta}{(3+e)\varepsilon} \min_{(v,v')} \{\|P_v^n \wedge P_{v'}^n\|\}, \tag{37}$$

where $v, v' \in \{-1,1\}^d$ and only differ in 1 coordinate. We further have $\min_{v,v'}\{\|P_v^n \wedge P_{v'}^n\|\} \geq \min_{v,v'}\{\frac{1}{2}\exp(-\mathrm{KL}(P_v^n \| P_{v'}^n)) = \frac{1}{2}\min_{v,v'}\{\exp(-n\mathbb{E}_{s\sim\pi_s}[\mathrm{KL}(p(\cdot|s,\theta_v)\|p(\cdot|s,\theta_{v'}))])\}\}$ where the first inequality is due to fact that $\|p \wedge p'\| \geq 1/2\exp(-\mathrm{KL}(p||p'))$ and the equality follows from the chain rule of KL divergence. With Lemma E.2, for any $v \sim_j v'$, when $s = \{j\}$, we have $\mathrm{KL}(p(\cdot|s;\theta_v)||p(\cdot|s;\theta_{v'})) \leq \frac{4\delta^2}{\varepsilon^2}$; when $s = \{i\}$ and $i \neq j$, we have $\mathrm{KL}(p(\cdot|s;\theta_v)||p(\cdot|s;\theta_{v'})) \leq (0\cdot\delta - 0\cdot\delta)^2 = 0$. Then for any pair of θ_v and $\theta_{v'}$ such that v and v' only differ in one coordinate, we have $\mathbb{E}_{s\sim\pi_s}[\mathrm{KL}(p(\cdot|s,\theta_v)||p(\cdot|s,\theta_{v'}))] \leq \frac{1}{d}\frac{4\delta^2}{\varepsilon^2}$. With these results and (37), we have $\mathcal{R}_2 \geq \frac{\delta}{44\varepsilon}\exp(-\frac{4n\delta^2}{d\varepsilon^2})$. Choose $\varepsilon = \sqrt{n}$, $\delta = \sqrt{d}$, we finally have $\mathcal{R} \geq \mathcal{R}_2 \geq c\sqrt{d/n}$ where c is an absolute constant. Note that we need $\delta \leq \varepsilon$ which is equivalent to $n \geq d$. In the case where $d > N$, with the same proof, we can have $\mathcal{R}(\mathcal{P}, \mathcal{I}) \geq c \cdot \sqrt{N/n}$. Hence, we have established that $\mathcal{R} \geq c \cdot \sqrt{\min(N,d)/n}$, when $n \geq \min(N,d)$. This concludes the proof. □

E Technical Lemmas

Lemma E.1. Suppose $A > 0$ and $C > 1$. Then for $0 \leq a \leq e^{-\frac{C}{2(C-1)}}A$, we have $\int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u \leq Ca\sqrt{\log\frac{A}{a}}$. As a special case, if $C = 2$, then for $0 \leq a \leq e^{-1}A$, we have $\int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u \leq 2a\sqrt{\log\frac{A}{a}}$.

Proof of Lemma E.1. Define

$$f(a) := \begin{cases} Ca\sqrt{\log\frac{A}{a}} - \int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u, & a > 0; \\ 0; & a = 0. \end{cases}$$

Then f is continuous at 0. Moreover, for $a > 0$, we have

$$f'(a) = (C-1)\sqrt{\log\frac{A}{a}} - \frac{C}{2\sqrt{\log A/a}},$$

which is nonnegative if and only if

$$\log(A/a) \geq \frac{C}{2(C-1)} \Leftrightarrow a \leq e^{-\frac{C}{2(C-1)}}.$$

As $a \to 0^+$, we further have

$$\frac{f(a)}{a} = C\sqrt{\log\frac{A}{a}} - \frac{1}{a}\int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u = (C-1)\sqrt{\log\frac{A}{a}} - \frac{1}{2a}\int_0^a \frac{1}{\sqrt{\log\frac{A}{u}}}\mathrm{d}u$$

$$\geq (C-1)\sqrt{\log\frac{A}{a}} - \frac{1}{2\sqrt{\log A/a}} \geq (C-1)\sqrt{\log\frac{A}{a}} - \frac{1}{2\sqrt{\log\frac{A}{a}}}.$$

This implies that $\liminf_{a\to 0^+}\frac{f(a)}{a} \geq +\infty$. Therefore, for any $0 \leq a \leq e^{-\frac{C}{2(C-1)}}A$, we have $f(a) \geq 0$, which concludes the proof. □

Lemma E.2. Let $p(a|s;\theta)$ be an MNL model as defined in (10). For any singleton assortment $s = \{i\}$, it holds that $\mathrm{KL}(p(\cdot|s;\theta_v)||p(\cdot|s;\theta_{v'})) \leq (x_i^T\theta_v - x_i^T\theta_{v'})^2$ where v, v' are defined in Appendix D.4.

Proof of Lemma E.2. For any $v, v' \in \{1,-1\}^d$ that differ only in 1 coordinate, let $p_v = \frac{1}{1+\exp(-x_i^T\theta_v)}$ and $p_{v'} = \frac{1}{1+\exp(-x_i^T\theta_{v'})}$ where $\theta_v = \delta \cdot v$ for some $\delta > 0$. Then we have

$$\mathrm{KL}(p(\cdot|s,\theta_v)\|p(\cdot|s,\theta_{v'})) = (p_v - p_{v'})\log(\frac{p_v}{1-p_v}\frac{1-p_{v'}}{p_{v'}}).$$

With the fact that $p_v/(1-p_v) = \exp(x_i^T\theta_v)$ and $(1-p_{v'})/p_{v'} = \exp(-x_i^T\theta_{v'})$, we have

$$\mathrm{KL}(p(\cdot|s,\theta_v)||p(\cdot|s,\theta_{v'})) \leq \left(\frac{\exp(-x_i^T\theta_{v'}) - \exp(-x_i^T\theta_v)}{\exp(-x_i^T\theta_{v'})}\right)(x_i^T\theta_v - x_i^T\theta_{v'}).$$

To continue, assume without loss of generality assume that $x_i^T\theta_v \geq x_i^T\theta_{v'}$, then

$$\begin{aligned}
&\mathrm{KL}(p(\cdot|s,\theta_v)||p(\cdot|s,\theta_{v'})) \leq \frac{\exp(-x_i^T\theta_{v'}) - \exp(-x_i^T\theta_v)}{\exp(-x_i^T\theta_{v'})}(x_i^T\theta_v - x_i^T\theta_{v'}) \\
&= (1 - \exp(-x_i^T\theta_v + x_i^T\theta_{v'}))(x_i^T\theta_v - x_i^T\theta_{v'}) \\
&\leq (1 - (1 + (-x_i^T\theta_v + x_i^T\theta_{v'})))(x_i^T\theta_v - x_i^T\theta_{v'}) = (x_i^T\theta_v - x_i^T\theta_{v'})^2,
\end{aligned}$$

where in the last inequality we use the fact that $\exp(x) \geq 1 + x$. □

Lemma E.3. Let $C_{s^\star} := 1/\pi_S(s^\star)$ and $r_{s^\star} := \max_{j\in s^\star} r(s^\star, j)$, then the following inequality holds: for any $p_1, p_2 \in \mathcal{P}$, $|\mathcal{V}(s^\star;p_1) - \mathcal{V}(s^\star;p_2)| \leq r_{s^\star}\sqrt{C_{s^\star}\mathbb{E}_S[\|p_1(\cdot|S) - p_2(\cdot|S)\|_1^2]}$.

Proof of Lemma E.3. With change of measure and by definition of $\mathcal{V}(s;p)$, we have $\mathcal{V}(s^\star;p_1) - \mathcal{V}(s^\star;p_2) = \mathbb{E}_S[f(S)g(S,p_1,p_2)]$, where $f(S) = \mathbb{I}(S = s^\star)/\pi_S(S)$ and $g(S,p_1,p_2) = \mathbb{I}(S = s^\star)\sum_{j\in S\cup\{0\}} r(S,j)\,(p_1(j|S) - p_2(j|S))$. Thus, we have

$$|\mathcal{V}(s^\star;p_1) - \mathcal{V}(s^\star;p_2)| \leq r_{s^\star}\mathbb{E}_S\left[f(S)\|p_1(\cdot|S) - p_2(\cdot|S)\|_1\right], \tag{38}$$

where $\|\cdot\|_1$ is the L_1 norm. By Hölder's inequality, we also have $\mathbb{E}_S[f(S)\|p_1(\cdot|S)-p_2(\cdot|S)\|_1] \le \sqrt{\mathbb{E}_S[f^2(S)]\mathbb{E}_S[\|p_1(\cdot|S)-p_2(\cdot|S)\|_1^2]}$. Moreover, we have

$$\mathbb{E}_S[f^2(S)] = \mathbb{E}_S[\mathbb{I}(S=s^\star)/\pi_S^2(S)] = 1/\pi_S(s^\star),$$

as $\frac{\mathbb{I}(S=s^\star)}{\pi_S(S)}$ has the value zero everywhere except at $s=s^\star$. Combining this with the upper bound for $|\mathcal{V}(s^\star;p_1)-\mathcal{V}(s^\star;p_2)|$ in (38), we have $|\mathcal{V}(s^\star;p_1)-\mathcal{V}(s^\star;p_2)| \le r_{s^\star}\sqrt{C_{s^\star}\mathbb{E}_S[\|p_1(\cdot|S)-p_2(\cdot|S)\|_1^2]}$. □

Lemma E.4. For any $p_1, p_2 \in \mathcal{P}$, $\mathbb{E}_S\left[\left(\sum_{j\in S\cup\{0\}}\left|p_1(j|S)-p_2(j|S)\right|\right)^2\right] \le 8H^2(p_1,p_2)$.

Proof of Lemma E.4. We use the fact that the total variation distance is less than a constant multiple of the Hellinger distance [46], i.e., for any $s\in\mathbb{S}$, $\frac{1}{2}\sum_{j\in s\cup\{0\}}\left|p_1(j|s)-p_2(j|s)\right| \le \sqrt{2}h(p_1(\cdot|s),p_2(\cdot|s))$, where h is the Hellinger distance. This implies that for any $s\in\mathbb{S}$, $\left(\sum_{j\in s\cup\{0\}}\left|p_1(j|s)-p_2(j|s)\right|\right)^2 \le 8h^2(p_1(\cdot|s),p_2(\cdot|s))$. Taking expectation with respect to S for both sides of this inequality concludes the proof. □

Lemma E.5 (Bernstein's Exponential Condition)**.** Consider a correctly specified family $\mathcal{P} = \{p_\theta(y|x):\theta\in\Theta\}$, and the corresponding family of log-likelihood ratios $\mathcal{F} := \left\{\log\frac{p_{\theta^*}(Y|X)}{p_\theta(Y|X)} : \theta\in\Theta\right\}$. Define

$$\kappa(u) := \begin{cases} \frac{2\left(e^{|u|/2}-1-|u|/2\right)}{(1-e^{-u/2})^2}, & u\neq 0; \\ 1, & u=0. \end{cases}$$

Assume that for some $\tau>0$, we have $\log\frac{p_{\theta^*}(Y|X)}{p_\theta(Y|X)} \le \tau$. Then for any $f_\theta\in\mathcal{F}$, we have $bbE\left(e^{|f_\theta|/2}-1-\frac{|f_\theta|}{2}\right) \le \kappa(\tau)H^2(\theta,\theta^*)$. By Taylor's expansion $e^u-1-u=\sum_{k=2}^\infty\frac{u^k}{k!}$, it follows that, for any $f_\theta\in\mathcal{F}$ and $k\ge 2$, we have $\mathbb{E}|f_\theta|^k \le \frac{k!}{2}2^{k-2}\times 8\kappa(\tau)H^2(\theta,\theta^*)$. In particular, $\mathbb{E}(f_\theta^2)\le 8\kappa(\tau)H^2(\theta,\theta^*)$.

Proof of Lemma E.5. Consider $f_p = f_p(Y|X) = \log\frac{p_0(Y|X)}{p(Y|X)} \in \mathcal{F}$. In particular, $f_p\le\tau$. On the event $\{p_0(Y|X)>0\}$, f_p is finite. By Lemma E.6, $\kappa(f_p)\le\kappa(\tau)$, that is,

$$\begin{aligned} e^{|f_p|/2}-1-\frac{|f_p|}{2} &\le \kappa(\tau)\cdot\frac{1}{2}(1-e^{-f_p/2})^2 \\ &= \kappa(\tau)\cdot\frac{1}{2}\left\{1-\exp\left[-\frac{1}{2}\log\frac{p_0(Y|X)}{p(Y|X)}\right]\right\}^2 \\ &= \kappa(\tau)\cdot\frac{1}{2p_0(Y|X)}\left(\sqrt{p_0(Y|X)}-\sqrt{p(Y|X)}\right)^2. \end{aligned}$$

Therefore, $\mathbb{E}\left(e^{|f_p|/2}-1-\frac{|f_p|}{2}\right) \le \kappa(\tau)\mathbb{E}\left\{\frac{1}{2}\int_{\mathcal{Y}}\left(\sqrt{p(y|X)}-\sqrt{p_0(y|X)}\right)^2\mathrm{d}y\right\} = \kappa(\tau)H^2(p,p_0)$. □

Lemma E.6. Consider $\kappa(u)$ in Lemma E.5. Then $\kappa(\cdot)$ is a non-decreasing and continuous function.

Proof of Lemma E.6. For $x > -1$, consider the change of variable $u = 2\log\frac{1}{1+x}$ such that $1 - e^{-u/2} = -x$. Then $u \geq 0$ if and only if $-1 < x < 0$. We further have $e^{|u|/2} - 1 - \frac{|u|}{2} = \big(\log(1+x) - \frac{x}{1+x}\big)\mathbf{1}_{\{-1<x<0\}} + \big(x - \log(1+x)\big)\mathbf{1}_{\{x>0\}}$. Based on the above transformation, we define

$$f(x) := \begin{cases} -\frac{2}{(1+x)x} + \frac{2}{x^2}\log(1+x), & -1 < x < 0; \\ 1, & x = 0; \\ \frac{2}{x} - \frac{2}{x^2}\log(1+x), & x > 0. \end{cases}$$

Then for $u \neq 0$, we have $\kappa(u) = f(e^{-u/2} - 1)$. It suffices to show that f is non-increasing and continuous. Consider $g(x) := -x + (1+x)\log(1+x)$; $x > -1$. Then $g(0) = g'(0) = 0$ while $g''(0) = 1$. As $x \to 0^-$, we have $\lim_{x\to 0^-}\left\{f(x) = \frac{g(x)}{(1+x)x^2/2}\right\} = g''(0) = 1$. As $x \to 0^+$, we have $\lim_{x\to 0^+}\left\{f(x) = \frac{x-\log(1+x)}{x^2/2}\right\} = 1$. This proves that f is a continuous function on $(-1, +\infty)$. For $-1 < x < 0$, we have $f'(x) = 2\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{g(x)}{(1+x)x^2}\right\} = \frac{2}{(1+x)^2x^4}\left\{g'(x)(1+x)x^2 - g(x)(2x+3x^2)\right\} = \frac{2}{(1+x)^2x^3}\left\{2x + 3x^2 - 2(1+x)^2\log(1+x)\right\}$. For $-1 < x \leq 0$, consider $h(x) := -2x - 3x^2 + 2(1+x)^2\log(1+x)$. Then $h'(x) = 4g(x) \geq 0$; $-1 < x \leq 0$. We have $h(x) \leq h(0) = 0$ for $-1 < x \leq 0$, and hence $f'(x) = -\frac{2h(x)}{(1+x)^2x^3} \leq 0$; $-1 < x < 0$. This proves that f is non-increasing on $(-1, 0)$. Moreover, $\lim_{x\to 0^-} f'(x) = -2\lim_{x\to 0^-}\left\{\frac{1}{(1+x)^2}\cdot\frac{h(x)}{x^3}\right\}$. By $h(0) = 0$, $h'(0) = 4g(0) = 0$, $h''(0) = 4g'(0) = 0$, and $h'''(0) = 4g''(0) = 4$, we have $\lim_{x\to 0} h(x)/x^3 = h'''(0)/6 = 2/3$. Then $\lim_{x\to 0^-} f'(x) = -4/3 < 0$. For $x > 0$, we have $f'(x) = 2\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{x-\log(1+x)}{x^2}\right\} = \frac{2}{x^4}\left\{\left(1 - \frac{1}{1+x}\right)x^2 - 2x[x - \log(1+x)]\right\} = \frac{2}{(1+x)x^3}\left\{-x^2 - 2x + 2(1+x)\log(1+x)\right\}$. For $x \geq 0$, consider $l(x) := x^2 + 2x - 2(1+x)\log(1+x)$. Then $l'(x) = 2[x - \log(1+x)] \geq 0$; $\forall x \geq 0$. We have $l(x) \geq l(0) = 0$ for $x \geq 0$, and hence $f'(x) = -\frac{2l(x)}{(1+x)x^3} \leq 0$; $x > 0$. This proves that f is non-increasing on $(0, +\infty)$. Moreover, $\lim_{x\to 0^+} f'(x) = -2\lim_{x\to 0^+}\left\{\frac{1}{1+x}\cdot\frac{l(x)}{x^3}\right\}$. Note that for $x \geq 0$, we have $l''(x) = 2\left(1 - \frac{1}{1+x}\right)$; $l'''(x) = \frac{2}{(1+x)^2}$. Thus we have $l(0) = l'(0) = l''(0) = 0$ and $l'''(0) = 2$. Then $\lim_{x\to 0} l(x)/x^3 = l'''(0)/6 = 1/3$, and $\lim_{x\to 0^+} f'(x) = -2/3 < 0$. Combining the above, $f(\cdot)$ is non-increasing and continuous on $(-1, +\infty)$. □

Lemma E.7. For fixed $\tau > 0$, $\alpha \in [0, +\infty]$ and any $u \geq 0$, we have $\mathcal{N}_{[]}(u, \mathcal{F}_\tau^{\mathrm{loc}}(\alpha), \mathcal{L}^2(\mathbb{P})) \leq \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}_\tau, H\right)$. Moreover, $\mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}_\tau, H\right) \leq \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}, H\right)$.

Proof of Lemma E.7. Let $\mathcal{F}_\tau := \mathcal{F}_\tau^{\mathrm{loc}}(+\infty)$. Then $\mathcal{N}_{[]}(u, \mathcal{F}_\tau^{\mathrm{loc}}(\alpha), \mathcal{L}^2(\mathbb{P})) \leq \mathcal{N}_{[]}(u, \mathcal{F}_\tau, \mathcal{L}^2(\mathbb{P}))$. For $i = 1, 2$, consider $p_i(y|x) \in \mathcal{P}$ and let $\widetilde{p}_i := (1 - e^{-\tau})p_i + e^{-\tau}p_0 \in \mathcal{P}_\tau$. In particular, $\mathcal{F}_\tau \ni \log(p_0/\widetilde{p}_i) \leq \tau$, that is, $\widetilde{p}_i \geq e^{-\tau}p_0$. Then, for some $\lambda \in [0, 1]$, we have $\left|\widetilde{p}_1^{1/2} - \widetilde{p}_2^{1/2}\right| = \left|\exp\left(\frac{1}{2}\log\widetilde{p}_1\right) - \exp\left(\frac{1}{2}\log\widetilde{p}_2\right)\right| = \frac{\widetilde{p}_1^{\lambda/2}\widetilde{p}_2^{(1-\lambda)/2}}{2}\left|\log\widetilde{p}_1 - \log\widetilde{p}_2\right| \geq \frac{e^{-\tau/2}}{2}p_0^{1/2}\left|\log\widetilde{p}_1 - \log\widetilde{p}_2\right|$, where the second equality is due to mean value theorem. Therefore, $H^2(\widetilde{p}_1, \widetilde{p}_2) \geq \frac{e^{-\tau}}{8}\mathbb{E}\left\{\int_{\mathcal{Y}} p_0(y|X)\left(\log\frac{\widetilde{p}_1(y|X)}{\widetilde{p}_2(y|X)}\right)^2\mathrm{d}y\right\} = \frac{1}{8e^\tau}\mathbb{E}_{p_0}\left(\log\frac{\widetilde{p}_1(Y|X)}{\widetilde{p}_2(Y|X)}\right)^2$. This proves that $cN_{[]}(u, \mathcal{F}_\tau, \mathcal{L}^2(\mathbb{P})) \leq \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}_\tau, H\right)$. Also note that $\left|\widetilde{p}_{p_1}^{1/2} - \widetilde{p}_{p_2}^{1/2}\right| = \frac{\left|\widetilde{p}_{p_1} - \widetilde{p}_{p_2}\right|}{\widetilde{p}_{p_1}^{1/2} + \widetilde{p}_{p_2}^{1/2}} = \frac{|p_{p_1} - p_{p_2}|}{\left(p_{p_1} + \frac{p_0}{e^\tau - 1}\right)^{1/2} + \left(p_{p_2} + \frac{p_0}{e^\tau - 1}\right)^{1/2}} \leq \frac{|p_{p_1} - p_{p_2}|}{p_{p_1}^{1/2} + p_{p_2}^{1/2}} = \left|p_{p_1}^{1/2} - p_{p_2}^{1/2}\right|$.

This further proves that $H^2(\widetilde{p}_{p_1}, \widetilde{p}_{p_2}) \leq H^2(p_{p_1}, p_{p_2})$, and hence $\mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}_\tau, H\right) \leq \mathcal{N}_{[]}\left(\frac{u}{2\sqrt{2}e^{\tau/2}}, \mathcal{P}, H\right)$. □

Lemma E.8. Consider two constants $s_0, \sigma_0 \in \mathbb{R}$. If $s = s_0$ and $\sigma = \sigma_0$ satisfy the entropy integral condition (30), then for $j \geq 0$, $s \leq s_0 + j/5$ and $\sigma = 2^{j/2}\sigma_0$ also satisfy the entropy integral condition.

Proof of Lemma E.8. Note that $s = s_0$ and $\sigma = \sigma_0$ satisfying (30) implies that $\frac{1}{\sigma_0}\int_0^{\sigma_0}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{2^{12+5s_0/2}}\sqrt{n}\sigma_0 + \frac{1}{\sigma_0}\int_0^{\sigma_0^2/2^{7+s_0/2}}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u$. By $u \mapsto \sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}$ is non-increasing, this implies that for any $j \geq 0$, $s = s_0 + j/5$ and $\sigma = 2^{j/2}\sigma_0$, we have $\frac{1}{\sigma}\int_0^{\sigma}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{\sigma_0}\int_0^{\sigma_0}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u$. Moreover, it is straightforward to show that $\frac{1}{\sigma_0}\int_0^{\sigma_0}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{\sigma}\int_0^{\sigma^2/2^{7+s/2}}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u$. Combining these gives $\frac{1}{\sigma}\int_0^{\sigma}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{2^{12+5s_0/2}}\sqrt{n}\sigma_0 + \frac{1}{\sigma}\int_0^{\sigma^2/2^{7+s/2}}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u$. That is, $\int_{\sigma^2/2^{7+s/2}}^{\sigma}\sqrt{\log \mathcal{N}_{[]}(u, \mathcal{F}, \mathcal{L}^2(\mathbb{P}))}\mathrm{d}u \leq \frac{1}{2^{12+5s/2}}\sqrt{n}\sigma^2$. This proves that $s = s_0 + j/5$ and $\sigma = 2^{j/2}\sigma_0$ also satisfy the entropy condition (30). Finally, the left hand side of (30) is non-decreasing in s while the right hand side is non-increasing in s. Therefore, it also holds for all $s \leq s_0 + j/5$. □

Lemma E.9. Consider $L(p), \widehat{L}_n(p), L_\tau(p), \widehat{L}_{\tau,n}(p)$ in (22), (14) and (23). Then we have $L(p_0) = L_\tau(p_0) = \min_{p\in p} L_\tau(p)$ and $\widehat{L}_{\tau,n}(p_0) = \widehat{L}_n(p_0)$. Moreover, for any $p \in p$, we have

$$\widehat{L}_n(p_0) - \widehat{L}_n(p) \leq \frac{1}{1-e^{-\tau}}\left\{\widehat{L}_{\tau,n}(p_0) - \widehat{L}_{\tau,n}(p)\right\}.$$

Proof of Lemma E.9. It can be established that $L(p_0) = L_\tau(p_0)$ and $\widehat{L}_{\tau,n}(p_0) = \widehat{L}_n(p_0)$ from direct derivation. For ease of notation, denote $\lambda = e^{-\tau} \in (0,1)$. First, we aim to establish that $\min_{p\in p} L_\tau(p) = L_\tau(p_0)$. For any $p \in \mathcal{P}$, denote $p_{\lambda,p} := (1-\lambda)p + \lambda p_0$, and we have

$$L_\tau(p) - L_\tau(p_0) = \mathbb{E}\log\frac{p_0(Y|X)}{p_{\lambda,p}(Y|X)} = \mathbb{E}\left\{\int_{\mathcal{Y}} p_0(y|X)\log\frac{p_0(y|X)}{p_{\lambda,p}(y|X)}\mathrm{d}y\right\} \geq 0,$$

with equality if and only if $p_0(\cdot|X) = p_{\lambda,p}(\cdot|X)$ almost everywhere on $\mathcal{Y}$, almost surely in X. This concludes that $\min_{p\in p} L_\tau(p) = L_\tau(p_0)$. Next, for any $p \in \mathcal{P}$, we have by convexity of $u \mapsto -\log(u)$ that $\widehat{L}_{\tau,n}(p) = \mathbb{E}_n\Big\{-\log[(1-\lambda)p(Y|X) + \lambda p_0(Y|X)]\Big\} \leq \mathbb{E}_n\Big\{(1-\lambda)[-\log p(Y|X)] + \lambda[-\log p_0(Y|X)]\Big\} = (1-\lambda)\widehat{L}_n(p) + \lambda\widehat{L}_n(p_0)$. By $\widehat{L}_n(p_0) = \widehat{L}_{\tau,n}(p_0)$, we have $\widehat{L}_{\tau,n}(p) - \widehat{L}_{\tau,n}(p_0) \leq (1-\lambda)\widehat{L}_n(p) + \lambda\widehat{L}_n(p_0) - \widehat{L}_{\tau,n}(p_0) \leq (1-\lambda)[\widehat{L}_n(p) - \widehat{L}_n(p_0)]$. Rearranging terms on both sides, we also have $\widehat{L}_n(p_0) - \widehat{L}_n(p) \leq \frac{1}{1-\lambda}[\widehat{L}_{\tau,n}(p_0) - \widehat{L}_{\tau,n}(p)]$. □

Lemma E.10. For $\lambda \in [0,1]$, and two conditional PDFs p_1, p_2 of $Y|X$, we have $\frac{(1-\lambda)^2}{4}H^2(p_1,p_2) \leq H^2\big((1-\lambda)p_1 + \lambda p_2, \lambda p_2\big) \leq (1-\lambda)H^2(p_1,p_2)$.

Proof of Lemma E.10. Suppose $\lambda < 1$, $p_1, p_2 \geq 0$ and p_1, p_2 are not zero simultaneously. The second inequality follows from

$$\left|\sqrt{(1-\lambda)p_1 + \lambda p_2} - \sqrt{p_2}\right| \leq \sqrt{1-\lambda}\frac{|p_1 - p_2|}{\sqrt{p_1} + \sqrt{p_2}} = \sqrt{1-\lambda}|\sqrt{p_1} - \sqrt{p_2}|.$$

The first inequality follows from

$$|\sqrt{p_1} - \sqrt{p_2}| = \frac{\sqrt{(1-\lambda)p_1 + \lambda p_2} + \sqrt{p_2}}{(1-\lambda)\left(\sqrt{p_1} + \sqrt{p_2}\right)}\left|\sqrt{(1-\lambda)p_1 + \lambda p_2} - \sqrt{p_2}\right|.$$

Note that $\sqrt{(1-\lambda)p_1 + \lambda p_2} \leq \max\left\{\sqrt{p_1}, \sqrt{p_2}\right\} \leq \sqrt{p_1} + \sqrt{p_2}$, and hence $\sqrt{(1-\lambda)p_1 + \lambda p_2} + \sqrt{p_2} \leq 2\left(\sqrt{p_1} + \sqrt{p_2}\right)$. Then we have $\left|\sqrt{p_1} - \sqrt{p_2}\right| \leq \frac{2}{1-\lambda}\left|\sqrt{(1-\lambda)p_1 + \lambda p_2} - \sqrt{p_2}\right|$. Finally, if $\lambda = 1$ or $p_1 = p_2 = 0$, then the equalities hold trivially. □

Lemma E.11. Let $p_0(y|x)$ be the true conditional density. Then for any conditional density function $p(y|x)$, $2H^2(p, p_0) \leq L(p) - L(p_0)$.

Proof of Lemma E.11.

$$L(p) - L(p_0) = \mathbb{E}\left\{\int_{\mathcal{Y}} p_0(y|X)\log\frac{p_0(y|X)}{p(y|X)}\mathrm{d}y\right\} \geq -2\mathbb{E}\left\{\int_{\mathcal{Y}} p_0(y|X)\left[\sqrt{\frac{p(y|X)}{p_0(y|X)}} - 1\right]\mathrm{d}y\right\}$$
$$= \mathbb{E}\left\{2 - \int_{\mathcal{Y}}\sqrt{p_0(y|X)p(y|X)}\mathrm{d}y\right\} = \mathbb{E}\left\{\int_{\mathcal{Y}}\left(\sqrt{p(y|X)} - \sqrt{p_0(y|X)}\right)^2\mathrm{d}y\right\} = 2H^2(p, p_0),$$

where the inequality is due to $\log(1+u) \leq u$. □

Lemma E.12 (Yang and Barron [54, Lemma 4])**.** Let $p_0(y|x)$ be the true conditional density. If for some $+\infty > \tau > 0$, $\log\frac{p_0(Y|X)}{p(Y|X)} \leq \tau$, then for any p, $L(p) - L(p_0) \leq 2(2+\tau)H^2(p, p_0)$.

References

[1] Alexander, K. S. (1984). Probability Inequalities for Empirical Processes and a Law of the Iterated Logarithm. *The Annals of Probability 12*(4), 1041–1067.

[2] Aouad, A., V. Farias, and R. Levi (2021). Assortment optimization under consider-then-choose choice models. *Management Science 67*(6), 3368–3386.

[3] Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science 59*(2), 341–357.

[4] Bertsimas, D., V. Gupta, and N. Kallus (2018). Robust sample average approximation. *Mathematical Programming 171*(1), 217–282.

[5] Bertsimas, D. and V. V. Mišic (2015). Data-driven assortment optimization. *Management Science 1*, 1–35.

[6] Bilodeau, B., D. J. Foster, and D. M. Roy (2023). Minimax rates for conditional density estimation via empirical entropy. *The Annals of Statistics 51*(2), 762–790.

[7] Cao, J. and W. Sun (2024). Tiered assortment: Optimization and online learning. *Management Science 70*(8), 5481–5501.

[8] Caro, F. and J. Gallien (2007). Dynamic assortment with demand learning for seasonal consumer goods. *Management Science 53*(2), 276–292.

[9] Chen, X., C. Shi, Y. Wang, and Y. Zhou (2021). Dynamic assortment planning under nested logit models. *Production and Operations Management 30*(1), 85–102.

[10] Chen, X., Y. Wang, and Y. Zhou (2020). Dynamic assortment optimization with changing contextual information. *Journal of Machine Learning Research*.

[11] Chen, X., Y. Wang, and Y. Zhou (2021). Optimal policy for dynamic assortment planning under multinomial logit models. *Mathematics of Operations Research 46*(4), 1639–1657.

[12] Davis, J. M., G. Gallego, and H. Topaloglu (2013). Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working Paper, Cornell University, Ithaca, NY.

[13] Davis, J. M., G. Gallego, and H. Topaloglu (2014). Assortment optimization under variants of the nested logit model. *Operations Research 62*(2), 250–273.

[14] Désir, A., V. Goyal, B. Jiang, T. Xie, and J. Zhang (2024). Robust assortment optimization under the markov chain choice model. *Operations Research 72*(4), 1595–1614.

[15] Désir, A., V. Goyal, D. Segev, and C. Ye (2020). Constrained assortment optimization under the Markov chain–based choice model. *Management Science 66*(2), 698–721.

[16] Ditchfield, H. (2021). Ethical challenges in collecting and analysing online interactions. In *Analysing digital interaction*, pp. 23–40. Springer.

[17] Duchi, J. C., P. W. Glynn, and H. Namkoong (2021). Statistics of robust optimization: A generalized empirical likelihood approach. *Mathematics of Operations Research 46*(3), 946–969.

[18] Dzyabura, D. and S. Jagabathula (2018). Offline assortment optimization in the presence of an online channel. *Management Science 64*(6), 2767–2786.

[19] Feldman, J., A. Paul, and H. Topaloglu (2019). Assortment optimization with small consideration sets. *Operations Research 67*(5), 1283–1299.

[20] Feldman, J. and H. Topaloglu (2015). Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management 24*(10), 1598–1620.

[21] Feng, Q., J. G. Shanthikumar, and M. Xue (2022). Consumer choice models and estimation: A review and extension. *Production and Operations Management 31*(2), 847–867.

[22] Flores, A., G. Berbeglia, and P. Van Hentenryck (2019). Assortment optimization under the sequential multinomial logit model. *European Journal of Operational Research 273*(3), 1052–1064.

[23] Fu, Z., Z. Qi, Z. Wang, Z. Yang, Y. Xu, and M. R. Kosorok (2022). Offline reinforcement learning with instrumental variables in confounded markov decision processes. Working Paper, Northwestern University, Evanston, IL.

[24] Gallego, G. and H. Topaloglu (2014). Constrained assortment optimization for the nested logit model. *Management Science 60*(10), 2583–2601.

[25] Han, Y., H. Zhong, M. Lu, J. Blanchet, and Z. Zhou (2025). Learning an optimal assortment policy under observational data. *arXiv preprint arXiv:2502.06777*.

[26] Hensher, D. A., J. M. Rose, and W. H. Greene (2015). *Latent class models*, pp. 706–741. Cambridge University Press.

[27] Jin, Y., Z. Yang, and Z. Wang (2021). Is pessimism provably efficient for offline rl? In *International Conference on Machine Learning*, pp. 5084–5096. PMLR.

[28] Kallus, N. and M. Udell (2020). Dynamic assortment personalization in high dimensions. *Operations Research 68*(4), 1020–1037.

[29] Kidambi, R., A. Rajeswaran, P. Netrapalli, and T. Joachims (2020). MOReL: Model-based offline reinforcement learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 21810–21823. Curran Associates, Inc.

[30] Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616.

[31] Kök, A. G., M. L. Fisher, and R. Vaidyanathan (2009). Assortment planning: Review of literature and industry practice. *Retail supply chain management: Quantitative models and empirical studies*, 99–153.

[32] Kumar, A., A. Zhou, G. Tucker, and S. Levine (2020). Conservative q-learning for offline reinforcement learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates Inc.

[33] Liu, N., Y. Ma, and H. Topaloglu (2020). Assortment optimization under the multinomial logit model with sequential offerings. *INFORMS Journal on Computing 32*(3), 835–853.

[34] Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*(2), 519–546.

[35] McFadden, D. (1973). *Conditional logit analysis of qualitative choice behavior*. Institute of Urban and Regional Development, Berkeley, CA.

[36] McFadden, D. (1981). Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications 198272*.

[37] Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

[38] Qian, M. and S. A. Murphy (2011). Performance guarantees for individualized treatment rules. *The Annals of statistics 39*(2), 1180.

[39] Ray, P. (1973). Independence of irrelevant alternatives. *Econometrica: Journal of the Econometric Society*, 987–991.

[40] Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *The Annals of Statistics 9*(1), 225–228.

[41] Rusmevichientong, P., Z.-J. M. Shen, and D. B. Shmoys (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research 58*(6), 1666–1680.

[42] Rusmevichientong, P. and H. Topaloglu (2012). Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations research 60*(4), 865–882.

[43] Shi, L., G. Li, Y. Wei, Y. Chen, and Y. Chi (2022). Pessimistic q-learning for offline reinforcement learning: Towards optimal sample complexity. In *International conference on machine learning*, pp. 19967–20025. PMLR.

[44] Sturt, B. (2025). The value of robust assortment optimization under ranking-based choice models. *Management Science 71*(5), 4246–4265.

[45] Talluri, K. and G. van Ryzin (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science 50*(1), 15–33.

[46] Tsybakov, A. B. (2008). *Introduction to Nonparametric Estimation* (1st ed.). Springer Publishing Company, Incorporated.

[47] van der Vaart, A. W. and J. A. Wellner (2023). *Weak Convergence and Empirical Processes: with Applications to Statistics* (Second ed.). Springer Series in Statistics. Springer Cham.

[48] Wang, M., H. Zhang, P. Rusmevichientong, and M. Shen (2024). Optimizing offline product design and online assortment policy: Measuring the relative impact of each decision. *Management Science*.

[49] Wang, R. (2012). Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters 40*(6), 492–497.

[50] Wang, R. (2021). Consumer choice and market expansion: Modeling, optimization, and estimation. *Operations Research 69*(4), 1044–1056.

[51] Wang, Z., P. W. Glynn, and Y. Ye (2016). Likelihood robust optimization for data-driven problems. *Computational Management Science 13*(2), 241–261.

[52] Wen, C.-H. and F. S. Koppelman (2001). The generalized nested logit model. *Transportation Research Part B: Methodological 35*(7), 627–641.

[53] Wong, W. H. and X. Shen (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *The Annals of Statistics 23*(2), 339–362.

[54] Yang, Y. and A. R. Barron (1998). An asymptotic property of model selection criteria. *IEEE Transactions on Information Theory 44*(1), 95–116.

[55] Yang, Y. and A. R. Barron (1999). Information-theoretic determination of minimax rates of convergence. *The Annals of Statistics 27*(5), 1564–1599.

[56] Yu, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pp. 423–435. Springer.

[57] Yu, T., G. Thomas, L. Yu, S. Ermon, J. Y. Zou, S. Levine, C. Finn, and T. Ma (2020). MOPO: Model-based offline policy optimization. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in Neural Information Processing Systems*, Volume 33, pp. 14129–14142. Curran Associates, Inc.