arXiv:2302.03821v1 [cs.LG] 8 Feb 2023

PASTA：悲观型商品组合优化

Juncheng Dong [*1] **Weibin Mo** [*2] **Zhengling Qi** [3] **Cong Shi** [4] **Ethan X. Fang** [5] **Vahid Tarokh** [1]

摘要

我们研究一类离线数据驱动环境下的商品组合优化问题。企业不知道底层的顾客选择模型，但可以访问一个离线数据集，其中包含历史上提供的产品组合集合、顾客选择和收入。我们的目标是利用离线数据集找到最优产品组合。由于商品组合优化具有组合性质，离线数据集中很可能出现数据覆盖不足的问题。因此，设计一种可证明高效的离线学习算法是一项重大挑战。为此，我们提出了一种基于悲观原则设计的算法，称为悲观型商品组合优化（简称 PASTA）。在一般设置下，该算法只要求离线数据覆盖最优产品组合，即可正确识别出最优产品组合。具体而言，我们针对著名的多项逻辑回归模型，建立了离线产品组合优化问题的遗憾界。我们还提出了一种高效的计算程序，用于求解我们的悲观型商品组合优化问题。数值研究表明，所提方法优于现有的基准方法。

1. 引言

卖方所面临的最关键问题之一，是为潜在买家挑选要展示的产品。无论是在实体零售还是线上零售中，卖方通常都面临陈列空间有限和仓储成本等问题，因此需要从所有可供选择的海量产品中，谨慎挑选一组产品展示给顾客。在此背景下，为了最大化卖方收入，从所有可供选择的产品中选出一组产品（即产品组合）的问题，便是商品组合优化问题。显然，顾客的选择行为（McFadden，1981）对商品组合优化问题至关重要。不失一般性，我们假设每位顾客的选择都可以用偏好向量 θ^* 来描述。这涵盖了经典的多项Logit模型（MNL）（McFadden，1973）；该模型无疑是产品组合优化文献中研究最广泛、应用最普遍的模型之一（更多详情请参见第6节）（Talluri & van Ryzin，2004；Caro & Gallien，2007年；Rusmevichientong 等，2010；Davis 等人，2013；Chen 等，2021年a；Aouad 等，2022年）。

在实践中，θ^* 通常未知，需要进行估计。假设没有顾客的历史数据，动态产品组合优化通过 θ^* 试错的方式进行自适应学习：依次更新产品组合并观察顾客随后做出的选择（Caro & Gallien，2007年；Chen 等，2020年；Rusmevichientong 等，2020年；Chen 等，2021年b；Li et al.，2022年）。与此同时，在大数据时代，企业通常会收集大量顾客数据。因此，对企业而言，利用现有（可能规模庞大的）离线数据集进行学习，往往比从头开始更有利。此外，离线学习也有益处，因为在线探索有时成本高昂，甚至不可行。因此，我们首次尝试正式研究每个卖方都会面临的以下重要问题。

研究问题：给定一个预先收集的离线数据集，其中包含历史上提供的产品组合、客户选择和收入，如何在不对离线数据集作出不切实际假设的情况下，找到一种高效且具有理论依据的离线算法来估计最优产品组合集合？

如果数据集不是通过自适应方式收集的，就经常会遇到数据覆盖不足的问题。为了使最大似然估计量（MLE）等估计量能够准确地近似 θ^*，离线数据集必须

[*]共同贡献 [1]美国北卡罗来纳州达勒姆杜克大学电气与计算机工程系，邮编 27705。[2]印第安纳州西拉斐特普渡大学克兰纳特管理学院 [3]华盛顿特区乔治·华盛顿大学决策科学系 [4]密歇根州安娜堡密歇根大学工业与运筹工程系 [5]美国北卡罗来纳州达勒姆杜克大学生物统计学与生物信息学系，邮编 27705。。通信联系人：Zhengling Qi <qizhengling@gwu.edu>。

包含足够多的产品组合和顾客选择。换句话说，数据收集过程需要充分探索不同的产品组合（由卖方提供）和不同的选择（由顾客作出）。离线数据集不太可能做到这一点，因为卖方不会选择期望收入明显次优的不合理产品组合，而顾客也不会选择违背自身偏好的产品。

主要贡献。 本文的主要贡献有两点。首先，基于悲观原则，我们提出了 悲观型产品组合优化 （简称 PASTA）框架，该框架能够正确识别最优产品组合。具体而言，我们的框架只要求离线数据集覆盖最优产品组合集合，而不要求覆盖所有可能的（数量呈组合爆炸式增长的）产品组合集合。其次，我们在多项Logit模型（MNL）下的离线产品组合优化中推导出了首个有限样本遗憾界（请参见第 6 节）。多项Logit模型是用于刻画顾客选择的最广泛使用的模型之一。随后，我们提出了一种同样名为 *PASTA* 的算法，能够高效求解悲观型产品组合优化问题。在模拟数据集上的实验（因此 θ^* 已知）验证了悲观型产品组合优化的有效性。

论文结构。 我们在第 2 节简要回顾相关工作，并在第 3 节介绍预备知识。我们在第 4 节提出悲观型产品组合优化。在第 5 节，我们给出理论结果。在第 6 节，我们以 MNL 模型为具体示例，研究该模型下的悲观型产品组合优化。在第 7 节，我们提出一种能够高效求解该问题的算法。我们在第 8 节给出实验结果，最后进行总结。

2. 相关工作

商品组合优化。在多项Logit模型下不含任何约束的商品组合优化问题，最早由 Talluri & van Ryzin（2004年）研究。随后，研究者考虑了包含各种约束的更复杂的商品组合优化问题，例如空间要求（Rusmevientchong 等，2009年）和基数约束（Rusmevientchong 等，2010年）。Davis 等人（2013年）提出了商品组合优化问题的线性规划（线性规划）形式，其中包括若干先前研究作为特例，这些特例对应于线性规划形式中的不同约束。这一系列研究假设顾客模型的真实参数已知（或至少可以准确估计）（Gallego与Topaloglu，2014年；Feldman与Topaloglu，2015年；Flores等人，2019年；Désir等人，2020年；Liu等人，2020年；Aouad 等人，2021年）。另一项密切相关的研究方向是动态产品组合优化（Caro & Gallien，2007年）。在动态产品组合优化的设定中，卖方在没有任何顾客先验信息的情况下，拥有有限销售期；在此期间，卖方观察顾客的选择，并基于观察到的行为，以自适应的试错方式优化产品组合（Sauré与Zeevi，2013年；Wang等人，2018年；Chen 等人，2021年a；Rusmevientchong 等，2020年；Chen 等人，2020年；2021年b）。与动态产品组合优化采用的在线设定相比，我们的工作不同于现有文献之处在于，我们关注离线设定：卖方只有已收集的数据集，无法控制数据收集过程。

离线学习中的悲观主义。 悲观原则已成功应用于强化学习（RL），利用预先收集的数据集寻找最优策略。从实证角度来看，它有助于提升离线设定下基于模型的方法和基于价值的方法的性能（例如，Yu 等人，2020年；Kidambi 等人，2020年；Kumar 等人，2020年）。悲观主义的重要性已在强化学习设定中得到分析和理论验证（Jin 等人，2021年；Fu 等人，2022年）。我们的主要贡献是将悲观方法用于商品组合优化问题，并展示其在实证和理论方面的价值。此外，与上述研究不同，我们聚焦于一个拥有指数级多个选项的决策问题。

3. 预备知识

令 $[N] \doteq \{1、2、\ldots\ldots N\}$ 表示由 N 个不同商品构成的集合。对于每件商品 i，都有一个可用的特征向量 $x_i \in \mathbb{R}^d$。假设 $\{x_i\}_{i\in[N]}$ 是固定向量。将所有待考虑的产品组合构成的集合记为 $\mathbb{S} \subseteq 2^{[N]} \backslash \{\varnothing\}$。对于离线数据，我们针对每位顾客定义一个随机向量 $(S、A、R)$，其中 $S \subseteq [N]$ 表示向顾客展示的产品组合，$A \in S \cup \{0\}$ 表示顾客在 $A \in S$ 时购买的商品（$A = 0$ 表示未购买任何商品），而 R 表示相应的收入。商品组合优化的最终目标是为所有顾客寻找一组最优商品 $s^* \in \mathbb{S}$，以最大化期望收入。本研究的具体目标是探究如何利用离线数据来学习最优产品组合；这些离线数据由随机三元组 $(S、A、R)$ 的独立同分布样本构成。

对于基于离线数据的商品组合优化，一个根本问题是估计尚未探索的产品组合的期望收入 $s \in \mathbb{S}$。这相当于探究产品组合与收入之间的因果关系。在著名的潜在结果框架（Rubin，1974）中，令随机变量 $R(s)$ 表示在以下干预下的潜在收入：将产品组合

设定为 $s \in \mathbb{S}$。我们的目标是找到一个最优产品组合

$$s^* \in \arg\max_{s\in\mathbb{S}} \mathbb{E}[R(s)].$$

请注意，在反事实世界中定义的预期潜在收入 $\mathbb{E}[R(s)]$ 可能无法仅凭观测数据识别。本文始终采用因果推断中的以下标准一致性和无混杂性假设。

假设3.1。 [一致性] 以概率1，观测收入与观测产品组合的潜在收入一致。即， $R = R(S)$ 几乎必然成立。

假设3.2。 [无混杂性] 潜在收入是观测产品组合的独立变量，即， $\{R(s)\}_{s\in\mathbb{S}} \perp\!\!\!\perp S$。

假设 3.1 确保观测收入与在观测产品组合 $A\ (\neq 0)$下购买商品 $A = 0$的潜在收入一致；若 S，则表示未购买。假设 3.2排除了可能混淆产品组合对收入因果效应的未观测因素[1]。

记 $\pi_S(s) \doteq \mathbb{P}(S = s)$ 为在离线数据中观察到产品组合 s 的概率。为对每个$\mathbb{E}[R(s)]$进行非参数识别 $s \in \mathbb{S}$，我们还需要以下正值性假设（Imbens 与 Rubin，2015年）。

假设 3.3。 [处处为正] 对于所有 $s \in \mathbb{S}$，观察到产品组合 s 的概率 $\pi(s)$ 为正（即$\pi(s) > 0$）。

假设 3.3 要求在离线数据中观察到每种产品组合的概率都大于零。 这是一个较强的假设，稍后将在假设 *5.1(I)* 中放宽，即只要求最优解处满足正值性。 根据假设3.1–3.3，我们可以通过逆倾向评分加权识别产品组合集合的效应（Rosenbaum 与 Rubin，1983）：对于任意 $s \in \mathbb{S}$,

$$\mathbb{E}[R(s)] = \mathbb{E}\left\{\frac{\mathbb{I}(S = s)R}{\pi_S(s)}\right\}, \tag{1}$$

其中，右侧的期望是相对于 (S的 A数据分布 R)而言的。然而，当可能的产品组合数量 $|\mathbb{S}|$ 随 N呈指数级增长时，假设 3.3 在实践中很少能对所有 $s \in \mathbb{S}$ 成立，尤其是在 N 较大时，离线数据可能有限。此外，如果产品组合 s 对应的期望收入 $\mathbb{E}[R(s)]$较低，卖方可能根本不会考虑该组合。因此，观察到此类产品组合 $\pi_S(s)$ 的概率为零。这可能使我们无法对每种产品组合 $s \in \mathbb{S}$估计（1）。

我们可能会尝试采用以下识别策略：

$$\begin{aligned}\mathbb{E}[R(s)] &= \mathbb{E}(R|S = s) \quad \text{(by Assumptions 3.1-3.2)}\\ &= \mathbb{E}[\mathbb{E}(R|S = s, A)] = \sum_{i\in s\cup\{0\}} \pi_A(i|s; \boldsymbol{x})r_{s,i},\end{aligned} \tag{2}$$

其中， $\boldsymbol{x} \doteq \{x_j\}_{j\in[N]}$ 表示各商品的特征， $\pi_A(i|s;\ \boldsymbol{x}) \doteq \mathbb{P}(A = i|S = s)$ 表示给定产品组合 s时购买第 i个商品的顾客选择概率（McFadden，1973）， $r_{s,\ i} \doteq \mathbb{E}(R|S = s,\ A = i)$ 表示给定产品组合 s 且购买第 i个商品时的条件期望收入。为简化记法，在不会引起混淆时，我们省略 π_A 中的特征 $\boldsymbol{x}$ 。如上所述，识别 $\mathbb{E}[R(s)]$ 需要了解 $\pi_A(i|s)$ 和$r_{s,\ i}$，这些量可以从数据中学习。尽管这种识别方法并不显式依赖于 $\pi_S(s)$，但要完全识别 $\pi_A(i|s)$ 和 $r_{s,\ i}$，就需要如上所假设的，对每个 $s \in \mathbb{S}$， $\pi_S(s)$ 均为正。

尽管上述产品组合覆盖不足这一挑战依然存在，我们认为，寻找最优产品组合s* 可能不一定需要 $\pi_S(s) > 0$ 在所有地方都满足，而只需在最优产品组合 s*处满足。具体而言，根据（2），在计算

$$s^* \in \arg\max_{s\in\mathbb{S}} \sum_{i\in s\cup\{0\}} \pi_A(i|s)r_{s,i}, \tag{3}$$

[1]在具有观测特征 $\{x_j\}_{j\in[N]}$的情况下，可以将假设3.2 放宽为一个更合理的条件：给定观测特征后，各变量相互独立。然而，为简化符号，且不失一般性，我们考虑假设3.2。

只要在优化过程中能够稳妥地排除 $\pi_A(i|s)$ 和$r_{s,\ i}$就未必需要准确估计 $s \neq s$*。我们的见解是，对数据中较少出现的产品组合，其对应的 $\pi_A(i|s)$ 和$r_{s,\ i}$的估计 s 往往会产生较大的误差。将估计误差纳入考量、采取悲观主义策略，可以排除这些产品组合（Jin 等人，2021年）；而标准的先预测后优化方法（Bertsimas 与 Kallus，2020年）或经验最大化方法（Zhao 等人，2012）则可能会高估 $\mathbb{E}[R(s)]$。因此，在我们提出的悲观型商品组合优化框架中，只要求最优点处的正性 $\pi_S(s^*) > 0$，这比假设3.3弱得多。

本文重点处理来自 $\pi_A(i|s)$ 的估计误差，同时假设$r_{s,\ i}$已知。这是商品组合优化文献中的一个典型假设（Talluri & van Ryzin，2004；Davis等人，2013年；Flores等人，2019年；Aouad 等，2021年）。我们的框架可以自然地扩展到需要估计$r_{s,\ i}$的情形。为使优化易于求解，我们进一步假设$r_{s,\ i} = r_i$ 期望收入仅取决于购买商品，而不取决于其所属的产品组合。在许多应用中，这一假设是合理的，因为收入是购买商品的确定性结果。在我们的悲观主义框架下，这一设定也很容易扩展，但可能会导致更复杂的商品组合优化问题。

下文不失一般性地假设，对于 $i \in [N]$，$r_i \geqslant 0$；而 $r_0 = 0$（不购买时收入为零）。对于任意向量 x，令 $x^\top$ 和 $||x||_2$ 分别表示 x的转置和 ℓ_2-范数。对于任意集合 A，令 $|A|$ 表示 A的基数。对于任意两个序列 $\{\varpi(n)\}_{n\geqslant 1}$ 和 $\{\gamma(n)\}_{n\geqslant 1}$，若存在常数 $c_1 > 0$（分别为 $c_2 > 0$），使得 $\varpi(n) \geqslant c_1\gamma(n)$（分别为 $\varpi(n) \leqslant c_2\gamma(n)$），则记作 $\varpi(n) \gtrsim \gamma(n)$（分别为 $\varpi(N) \lesssim \gamma(n)$）。此外，当 $\varpi(n) \gtrsim \gamma(n)$ 且 $\varpi(n) \lesssim \gamma(n)$时，我们记作$\varpi(n) \eqsim \gamma(n)$。

4. 悲观离线 ассортимент优化

在本节中，我们介绍悲观离线产品组合优化框架。为此，基于公式（2），我们首先根据离线数据估计选择概率 $\pi_A(i|s)$。随后，我们使用 $\pi_A(\cdot)$ 的代入估计量，计算优化问题（3）中的优化值。考虑模型的一般形式 $\pi_A(a|s;\theta^*,\ \boldsymbol{x})$ 其中未知的真实参数为 θ^*。同样，为便于记号表示，在不引起混淆时，我们省略 $\boldsymbol{x}$ 中的特征 π_A。需要指出的是，θ^*可以是有限维的，也可以是无限维的。给定离线数据集 $\mathcal{D} = \{S_i,\ A_i,\ R_i\}_{i=1}^n$，其中 n 为样本量，可以通过最大似然估计量（MLE）估计模型参数 θ^*。具体而言，将 $\hat{}$基于似然的损失函数 $L_n(\theta)$ 定义为

$$\widehat{L}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log \pi_A(A_i|S_i;\theta).$$

$\hat{}$ $\hat{}$那么，未知参数 θ^* 的最大似然估计量为 $\theta_{\mathrm{ML},\ n} \in \arg\min_{\theta\in\Theta} L_n(\theta)$，其中 Θ 为预先指定的参数空间。令

$$\mathcal{V}(s;\widehat{\theta}_{\mathrm{ML},n}) \doteq \sum_{i\in s} \pi_A(i|s;\widehat{\theta}_{\mathrm{ML},n})r_i.$$

在此，我们定义 $\mathcal{V}(s;\ \theta^*)$ 为价值函数，该函数对应于 s 在 π_A 的顾客选择模型，并取决于参数 θ^*。根据（3）得到的最优产品组合代入估计量为

$$\widehat{s}_{\mathrm{ML},n} \in \arg\max_{s\in\mathbb{S}} \left\{\mathcal{V}(s;\widehat{\theta}_{\mathrm{ML},n})\right\}.$$

基于最大似然估计的方法首先代入 θ^*的最大似然估计，然后直接优化相应的估计价值函数。

$\hat{}$如前所述，上述先估计后优化方法的一个缺点是，由数据覆盖不足造成的$\theta_{\mathrm{ML},\ n}$ 估计误差可能导致对 $\mathcal{V}(s;\theta^*)$的高估，并进一步影响下游优化。另一种方法是通过考虑以下基于似然比检验的置信区域来量化估计不确定性（Owen，1990）：

$$\Omega_n(\alpha_n) \doteq \{\theta \in \Theta : \widehat{L}_n(\theta) - \widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n}) \leqslant \alpha_n\},$$

其中 $\alpha_n > 0$ 是预先指定的。随后，我们将多项Logit模型作为一个特殊情形进行分析（请参见第 6节）。当 α_n 选为 $\mathcal{O}(d/n)$, 时，我们在定理 6.1 中证明，$\theta^* \in \Omega_n(\alpha_n)$ 以高概率成立。此类保证不需要对产品组合作出任何数据覆盖假设。

今后，为简便起见，我们省略 α_n，并在不会引起歧义时，将 Ω_n 写作 $\Omega_n(\alpha_n)$.

为了使商品组合优化能够稳健应对代入估计误差，我们通过将 3 中的估计不确定性 Ω_n 纳入考虑，研究其悲观版本。具体而言，我们提出 **悲观型商品组合优化（PASTA）**，其求解问题为

$$\hat{s}_{\mathrm{PASTA},n} \in \arg\max_{s\in\mathbb{S}} \min_{\theta\in\Omega_n} \mathcal{V}(s;\theta). \tag{4}$$

这里，对于固定的产品组合 $s \in \mathbb{S}$，内层最小化会在置信集 Ω_n内所有可能的模型^参数 θ 中计算最坏情况下的价值。特别地，如果由于数据覆盖不足，估计价值 $\mathcal{V}(s;\theta_{\mathrm{ML},\ n})$ 对于 s 具有很高的不确定性，那么最坏情况下的价值 $\min_{\theta\in\Omega_n} \mathcal{V}(s;\theta)$ 很可能远小于 $\mathcal{V}(s;\theta^*)$。在这种情况下，(4)的外层可能会选择另一个最坏情况下价值相对较高的产品组合。通过这种方式，(4)的内层会排除那些在离线数据中出现频率较低的产品组合。因此，这种策略的一个重要优势是能够避免高估价值函数。换言之，采用代入法时，以不可忽略的概率，^估计价值 $\mathcal{V}(s;\theta_{\mathrm{ML},\ n})$ 可能远高于真实值 $\mathcal{V}(s;\ \theta^*)$，进而导致一个可能并非最优、但却由基于最大似然估计的方法优化得到的产品组合。相比之下，PASTA 能够意识到数据覆盖不足，因此对于那些价值估计高度不确定的产品组合更加悲观。下一节将从理论上分析 PASTA 方法的优势。

5. 理论结果

在本节中，我们证明，在4中的 PASTA 方法满足一个一般性的遗憾保证，只需对最优点处的正性作出较弱的假设，即 $\pi_S(s^*) > 0$。具体而言，给定$\hat{s}_{\mathrm{PASTA},\ n}$ 在（4）中，我们采用以下遗憾作为性能指标来评估 PASTA 的表现

$$\mathcal{R}(\hat{s}_{\mathrm{PASTA},n}) = \mathcal{V}(s^*;\theta^*) - \mathcal{V}(\hat{s}_{\mathrm{PASTA},n};\theta^*).$$

我们旨在在一般条件下推导$\mathcal{R}(\hat{s}_{\mathrm{PASTA},\ n})$ 的遗憾界。记 $L(\theta) = \mathbb{E}[-\log\ \pi_A(A|S;\ \theta)]$ 为总体损失函数。所有详细证明见附录9。

我们首先证明，只要 $\theta^* \in \Omega_n$ 表明置信区域覆盖真实参数，PASTA 方法的遗憾就可以通过最优产品组合 s^*处价值函数在 $\theta \in \Omega_n$ 上的最坏情况估计误差来校准。

引理5.1。令$\hat{s}_{\mathrm{PASTA},\ n}$ 为 *PASTA* 方法按照（4）定义得到的解。若 $\theta^* \in \Omega_n$，则

$$\mathcal{R}(\hat{s}_{\mathrm{PASTA},n}) \leqslant \max_{\theta\in\Omega_n} \big\{\mathcal{V}(s^*;\theta^*) - \mathcal{V}(s^*;\theta)\big\}.$$

引理 *5.1*的证明。

$$\begin{aligned}
&\mathcal{V}(s^*;\theta^*) - \mathcal{V}(\hat{s}_{\mathrm{PASTA},n};\theta^*) \\
&\leqslant \mathcal{V}(s^*;\theta^*) - \min_{\theta\in\Omega_n} \mathcal{V}(\hat{s}_{\mathrm{PASTA},n};\theta) && (\text{由 } \theta^* \in \Omega_n) \\
&\leqslant \mathcal{V}(s^*;\theta^*) - \min_{\theta\in\Omega_n} \mathcal{V}(s^*;\theta) && (\text{由 } \hat{s}_{\mathrm{PASTA},\ n} \text{ 求解（4）}) \\
&= \max_{\theta\in\Omega_n} \big\{\mathcal{V}(s^*;\theta^*) - \mathcal{V}(s^*;\theta)\big\}. && \square
\end{aligned}$$

引理 5.1凸显了悲观优化的优势。特别是，对于先估计后优化方法的遗憾保证，需要在所有 $s \in \mathbb{S}$上均匀控制估计误差，即 上确界$_{s\in\mathbb{S}} \left|\mathcal{V}(s;\ \hat{\theta}_{\mathrm{ML},\ n}) - \mathcal{V}(s;\ \theta^*)\right|$，以控制贪心策略所造成的误差，即$\left|\mathcal{V}(\hat{s}_{\mathrm{ML},\ n},\ \hat{\theta}_{\mathrm{ML},\ n}) - \mathcal{V}(\hat{s}_{\mathrm{ML},\ n},\ \theta^*)\right|$。这通常可能意味着，需要在 $s \in \mathbb{S}$上对价值函数进行均匀且良好的估计。这进一步解释了为什么先估计后优化方法需要对所有 3.3 满足正值性假设 $s \in \mathbb{S}$。相比之下，我们的引理 5.1 表明，只需控制 s^* 处的估计误差就可能足够。因此，对于我们的 PASTA 方法而言，只要求最优点处满足正值性假设即可。

接下来，我们提出以下假设，以获得算法的遗憾保证。

假设 5.1。

(I) [正性 最优] 观察到最优产品组合的概率为正，即 $\pi_S(s^*) >$ 0.

（II） [基于似然的集中] 对于任意 $0<\delta<1$，以至少 $1-\delta$ 的概率，有：(1) $\theta^* \in \Omega_n$，以及 (2)

$$\sup_{\theta\in\Omega_n}\left|L(\theta)-L(\theta^*)-\left(\widehat{L}_n(\theta)-\widehat{L}_n(\theta^*)\right)\right| \leqslant \alpha_n.$$

我们强调，*PASTA* 只要求最优点处的正性。与假设3.3中对所有产品组合均要求正性相比，我们的假设5.1 (I) 弱得多，因此也更有可能满足。假设5.1(II) 是基于似然的集中所需的一般条件。我们稍后将在定理6.2中证明，在一般的多项Logit模型下，上述条件（II）确实成立。具体而言，（II）部分中的陈述（1）要求基于似然比检验的置信区域Ω_n 有效，而（II）部分中的陈述（2）要求基于似然的局部化经验过程集中（van der Vaart 和 Wellner，1996年）。

最优点处的正性与一个有限常数 $C_{s*}=1/\pi_s(s^*)$ 相关，该常数关系到学习性能。我们还将 $r_{s*} \doteq \max_{j\in s^*} r_j$ 记为 s^* 中所有商品可能获得的最高收入。注意，这两个常数 C_{s*} 和 r_{s*}仅取决于最优产品组合 s^*。在下面的引理中，我们给出最优产品组合 s^* 处的估计误差界。

引理 5.2。 在假设 *5.1*下，对于任意 $0<\delta<1$，以至少 $1-\delta$的概率，对于任意 $\theta\in\Omega_n$，

$$\mathcal{V}(s^*;\theta^*)-\mathcal{V}(s^*;\theta) \lesssim r_{s*}C_{s*}\sqrt{\alpha_n}.$$

结合引理5.1 和5.2，我们在下面的定理中总结了 PASTA 的遗憾界。

定理 5.3。 在假设 *5.1*下，对于任意 $0<\delta<1$，以 $1-\delta$的概率，我们有

$$\mathcal{R}(\hat{s}_{\mathrm{PASTA},n}) \lesssim r_{s*}C_{s*}\sqrt{\alpha_n}.$$

6. 应用：多项Logit模型

本节中，我们考虑用于顾客选择的多项Logit模型（MNL） $\pi_A(a|s)$。这是产品组合优化文献中应用最广泛的模型之一（Feng等人，2022年）。在多项Logit模型下，我们将验证假设5.1 （II），并在此情形下建立 PASTA 的遗憾界。

给定商品特定特征 $\{x_i\}_{i\in[N]}$，多项Logit模型假设，顾客对第 i 个商品的偏好与 $\exp(x_i^\top\theta^*)$ 成正比，其中 $\theta^*\in\Theta$ 是潜在的未知参数。这里，我们假设参数空间 $\Theta\subseteq\mathbb{R}^d$ 是紧集，且 $\theta_{\max} \doteq \sup_{\theta\in\Theta}\|\theta\|_2<+\infty$。给定产品组合 s，多项Logit模型下的顾客选择概率为

$$\pi_A(i|s;\theta^*)=\frac{\exp(x_i^\top\theta^*)}{1+\sum_{j\in s}\exp(x_j^\top\theta^*)}, \quad \forall i\in s. \tag{5}$$

此外，未购买的概率被归一化为 $\pi_A(0|s;\ \theta^*)=1/(1+\sum_{j\in s}\exp(x_j^\top\theta^*))$。根据（3）和多项Logit模型（5），商品组合优化的目标函数可写为

$$\mathcal{V}(s;\theta)=\frac{\sum_{i\in s}r_i\exp(x_i^\top\theta)}{1+\sum_{i\in s}\exp(x_i^\top\theta)}.$$

我们首先在多项Logit模型下证明假设5.1 的陈述（1）成立。为此，由于 Θ是紧的，存在有限常数 $C_A>0$，使得对所有 $\theta\in\Theta$，$s\in\mathbb{S}$ 以及 $i\in s$，均有 $1/\pi_A(i|s;\ \theta)\leqslant C_A$。

引理 6.1。考虑多项*Logit*模型(5)，其中集合 Θ为紧集。假设 $\theta^*\in\Theta$。对于任意 $0<\delta<1$，以概率$\geqslant(1-\delta)$，有

$$\widehat{L}_n(\theta^*)-\widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n}) \lesssim \frac{C_A d}{n}\log\frac{\theta_{\max}}{\delta}.$$

引理 6.1 表明，当 α_n 取为 $\frac{C_A d}{n}\log\frac{\theta_{\max}}{\delta}$时，我们可以保证 $\theta^*\in\Omega_n$ 以高概率成立，这证明了假设 5.1 （II）的陈述（1）。特别地，α_n 的阶为 $\mathcal{O}(d/n)$。请注意，引理 6.1 不依赖于 S的分布，这意味着不需要对观测到的产品组合作出数据覆盖假设。

假设 $\theta^* \in \Theta$ 对于 Θ 是一个紧集，这要求对于任意给定的产品组合，每种产品都有机会被数据中的顾客选中。这是一个温和的要求，因为 θ^* 始终是有限的。

接下来，我们在以下定理中论证假设5.1（II）的陈述（2）。

引理 6.2。 考虑多项*Logit*模型 (5)。假设引理 *6.1* 中的条件成立，且 $L(\theta)$ 和 $L_n(\theta)$ 均一致强凸。令 $\alpha_n \asymp \frac{C_A d}{n} \log \frac{\theta_{\max}}{\delta}$。对于 $0 < \delta < 1$，以概率 $\geqslant (1-\delta)$，有

$$\sup_{\theta \in \Omega_n} \left| L(\theta) - L(\theta^*) - \left(\widehat{L}_n(\theta) - \widehat{L}_n(\theta^*) \right) \right| \leqslant \alpha_n.$$

最后，仅在最优点满足正性假设（假设 5.1 （I））下，我们可以应用定理 5.3 来建立 PASTA 在多项Logit模型中的遗憾界。

定理 6.3。 考虑多项*Logit*模型（由方程(5)给出）。假设引理*6.2*中的条件成立，且$\pi_S(s^*) > 0$。固定一个 $\delta \in (0, 1)$。设$\widehat{s}_{\mathrm{PASTA},n}$ 是 *PASTA* 在 $\alpha_n \asymp \frac{C_A d}{n} \log \frac{\theta_{\max}}{\delta}$下的输出。则以概率$\geqslant (1-\delta)$，有

$$\mathcal{R}(\widehat{s}_{\mathrm{PASTA},n}) \lesssim r_{s*} C_{s*} \sqrt{\frac{C_A d}{n} \log \frac{\theta_{\max}}{\delta}}.$$

我们指出，在多项Logit模型（由方程（5）给出）下，遗憾的阶为 $\mathcal{O}(\sqrt{d/n})$。这是由于引理6.1中多项Logit模型经验似然比的集中速率所致。这样的遗憾界速率与参数模型假设下文献中的结果相符（Qian & Murphy, 2011；Mo & Liu, 2022）。然而，现有文献要求在每个 $s \in \mathbb{S}$处均满足正性条件 $\pi_S(s) > 0$。相比之下，定理 6.3仅要求在最优产品组合 s^*处满足正性条件 $\pi_S(s^*) > 0$。此外，我们可以证明，对于任意 $\theta \in \Theta$，$\min_{i \in N} \pi_A(i \mid S = [N];\ \theta) \leqslant 1/N$，这意味着$C_A \geqslant N$。因此，我们的遗憾的阶至少为 $\sqrt{N}$，其中 N是可用商品的总数。关于如何根据 N、 n、 d 以及 s^* 的基数建立离线产品组合优化的极小极大下界，这是一个有趣的问题。我们将在后续工作中研究这一问题。

7.*PASTA* 算法

本节针对多项Logit模型，提出一种高效算法，用于求解优化问题（4）中给出的最大最小问题。具体而言，令

$$\mathcal{V}(s;\theta) = \sum_{i \in s} \frac{r_i \exp(x_i^\top \theta)}{1 + \sum_{j \in s} \exp(x_j^\top \theta)}$$

并给定置信集 Ω_n，我们希望求解

$$\max_{s \in \mathbb{S}} \min_{\theta \in \Omega_n} \mathcal{V}(s;\theta).$$

所提出的迭代算法最多执行 T 次迭代。在第 t次迭代中，给定上一轮迭代得到的 s_t 和 θ_t，我们依次执行以下两个步骤:

- 步骤 1：计算最优产品组合 s_{t+1}，给定 θ_t （参见第 7.1节）。
- 步骤 2：计算最优 θ_{t+1}，使用 s_{t+1} （参见第 7.2节）。

相应的伪代码见下方的算法 1 。

7.1. 最优产品组合计算

给定多项Logit模型参数 θ_t，计算使期望收入最大化的产品组合 s_{t+1} 可以表述为一个线性规划（LP）问题。

假设产品组合 s 可以用一个 N 维二进制向量 $\gamma \in \{0,1\}^N$ 表示，其中 $\gamma_j = 1$ 当且仅当 $j \in s$。假设 $s \in \mathbb{S}$ 对应于以下针对 γ 的可行集，其中包含 M 个线性不等式约束:

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

算法 1PASTA

输入：离线数据集 $\{(S_i, A_i, R_i)\}_{i=1}^n$；$\alpha_n$；$\{r_i\}_{i=1}^N$；$\{x_i\}_{i=1}^N$；最大迭代次数 T输出：悲观型商品组合优化的解 $\widehat{L}_n(\theta) \doteq -\frac{1}{n}\sum_{i=1}^n \quad \pi_A(A_i|S_i\ \theta) \qquad \log$；

$\widehat{\ } \qquad \widehat{\ }\theta_{\mathrm{ML},n} \leftarrow \arg\min_{\theta\in\Theta} L_n(\theta)\widehat{\ } \qquad \widehat{\ } \quad \widehat{\ }\Omega_n \leftarrow \{\theta \in \Theta :$ $L_n(\theta) - L_n(\theta_{\mathrm{ML},n}) \leqslant \alpha_n\}\widehat{\ } \qquad \widehat{\ }t \leftarrow 0;\theta_t \leftarrow \theta_{\mathrm{ML},n}$ /* 初始化 θ_0 为 $\theta_{\mathrm{ML},n}$ */ for $t = 1$ 到 T do$s_t \leftarrow$ SolveLP$(\theta_{t-1}, \{r_i\}_{i=1}^N, \{x_i\}_{i=1}^N)$/* 第 7.1 节 */$\theta_t \leftarrow$ SolveGD$(s_t, \Omega_n, \{r_i\}_{i=1}^N, \{x_i\}_{i=1}^N)$/* 第 7.2 节 */ end for$\hat{s} \leftarrow s_T$

其中，约束系数矩阵 $[a_{ij}]_{i\in[M],\ j\in[N]}$ 是一个完全单模矩阵（Pang，2017年）。换言之，根据 s 与 γ之间的一一对应关系，我们有 $s \in \mathbb{S}$当且仅当 $\gamma \in \Gamma$。

接下来，我们将 $v_i = \exp(x_i^\top \theta_t)$ 记为第 i个商品的偏好得分。在多项Logit模型（5）下，顾客选择概率变为 $\pi_A(i|s) = \frac{v_i}{1+\sum_{j\in s} v_j}$。针对 s_{t+1}的优化可表述为

$$\max_{\gamma\in\Gamma} \frac{\sum_{i\in[N]} r_i v_i \gamma_i}{1 + \sum_{i\in[N]} v_i \gamma_i}, \tag{6}$$

这等价于以下线性规划问题（Davis等人，2013年）：

$$\begin{aligned} \max_{w_j : j\in[N]\cup\{0\}} \quad & \sum_{j\in[N]} r_j w_j \\ \text{subject to} \quad & \sum_{j\in[N]} w_j + w_0 = 1 \\ & \sum_{j\in[N]} a_{ij} \frac{w_j}{v_j} \leqslant b_i w_0 \quad \forall i \in [M] \\ & 0 \leqslant \frac{w_j}{v_j} \leqslant w_0 \quad \forall j \in [N]. \end{aligned} \tag{7}$$

具体而言，我们可以利用问题（6）的最优解（记为 γ^*）以及问题（7）的最优解（记为 w^*），通过以下公式恢复问题（6）的最优解：

$$\gamma_j^* = \frac{w_j^*}{v_j w_0^*} \quad \forall j \in [N]. \tag{8}$$

综上，在第 t次迭代中，为了针对给定的 s_{t+1} 计算最优产品组合 θ_t，我们首先求解（7）中的线性规划问题，得到 w^*。然后通过（ γ^* 8）恢复。最后，当且仅当 s_{t+1} 时，更新后的产品组合 $i \in s_{t+1}$由对应关系$\gamma_j^* = 1$得到。

7.2. 模型参数计算

给定第 s_{t+1} 7.1 节中的优化产品组合7.1后，我们旨在从置信集 θ_{t+1} 中搜索最坏情况下的多项Logit模型参数Ω_n，使期望收入最小化。具体而言，我们采用带线搜索的梯度下降（GDLS）方法，通过求解以下问题来计算 θ_{t+1}。

$$\min_{\theta\in\Omega_n(x_i^\top\theta)} \mathcal{V}(s_{t+1}; \theta). \tag{9}$$

这里我们指出，$\mathcal{V}(s_{t+1}; \theta) = \frac{\sum_{i\in s_{t+1}} r_i}{\exp 1 + \sum_{i\in s_{t+1}}} \quad (x_i^\top\theta)\exp$是 θ上的局部Lipschitz函数。给定一个可行初始参数$\theta^{(0)} \in \Omega_n$，我们最多执行 L 步梯度下降。假设 β_ℓ 是第 ℓ 步梯度下降的步长。在

每一步 $\ell = 1, 2, \cdots, L$，我们都会进行线搜索以保持可行性。具体而言，给定 $\theta^{(\ell-1)} \in \Omega_n$，我们首先计算梯度：$\xi_\ell = \nabla_\theta \mathcal{V}(s_{t+1}; \theta^{(\ell-1)})$。然后以预先指定的步长 $\beta_\ell = \tilde{\beta}$ 初始化 β_ℓ，并检查$\theta^{(\ell)} = \theta^{(\ell-1)} - \beta_\ell \xi_\ell$ 是否可行，即 $\theta^{(\ell)} \in \Omega_n$。如果不可行，我们令 $\beta_\ell \leftarrow c\beta_\ell$ 取某个预先指定的 $c \in (0, 1)$，并重新计算$\theta^{(\ell)} = \theta^{(\ell-1)} - \beta_\ell \xi_\ell$。重复此搜索，直到 $\theta^{(\ell)}$ 可行为止。我们在算法 2 中给出了整个过程的伪代码。注意，L, β, c 都是超参数。在所有数值研究中，我们设定 $L = 2, \beta = 0.01$ 和$c = \frac{1}{2}$，其经验表现良好。

算法 2带线搜索的梯度下降（GDLS）

输入： 产品组合 s_{t+1}；可行集 Ω_n；初始参数 $\theta^{(0)}$；初始步长 $\tilde{\beta}$；步长缩减常数 c；下降步数 L**输出：** 更新后的参数 $\theta_{t+1}\ell \leftarrow 0$**对于** $\ell = 1$ **至** L**执行**$\xi_\ell = \nabla_\theta \mathcal{V}(s_{t+1};\ \theta^{(\ell-1)})$ /* 计算梯度 */$\beta_\ell \leftarrow \beta\theta^{(\ell)} \leftarrow \theta^{(\ell-1)} - \beta_\ell \xi_\ell$**当** $\theta^{(\ell)} \notin \Omega_n$ **执行**$\beta_\ell \leftarrow c\beta_\ell$ /* 减小步长 */$\theta^{(\ell)} \leftarrow \theta^{(\ell-1)} - \beta_\ell \xi_\ell$**结束 while 循环，结束 for 循环**$\theta_{t+1} \leftarrow \theta^{(\ell)}$

8. 实验

我们将 PASTA 方法与不采用悲观主义的商品组合优化方法进行比较（下文称为基准方法）。我们在合成数据上评估我们的方法和基准方法；对于这些数据，最优产品组合 s^* 和真实参数 θ^* 已知，因此可以计算真实遗憾。下面详细介绍数据生成过程和基准方法。

8.1. 数据生成

我们考虑由 N、K、d、n 和 p描述的商品组合优化场景，其中 N 是可用产品的总数；K 是产品组合的基数约束，即 $\mathbb{S} = \{s : |s| \leqslant K\}$；$d$ 是 θ^* 和 $\{x_j\}_{j=1}^N$的维度；n 是离线数据集的样本量；p 是抽取最优产品组合 s^* 的概率。与（Chen 等人，2020）类似，我们首先将真实偏好向量 θ^* 生成为均匀随机的单位 d维向量。对于 $i \in \{1, \ldots, N\}$，我们从范围 $[0.5, 0.8]$ 中均匀抽取 r_i（产品 i的奖励），并将 x_i（产品 i的特征）生成为均匀随机的单位 d维向量，使得 $\exp(x_i^\top \theta^*) \leqslant \exp(-0.6)$，以避免最优产品组合包含过少商品的退化情形。给定这些信息，即可计算真实最优产品组合 s^*。接着，我们生成一个含有 n 个样本的离线数据集 $\mathcal{D} = \{(S_i, A_i, R_i)\}_{i=1}^n$。对于 $i \in \{1, \ldots, n\}$，我们按照分布 π_S生成 S_i，使得 $\pi_S(s^*) = p$和 $\pi_S(s) = \frac{1-p}{|\mathbb{S}|-1}$，其中 $0 < p < 1$是观测到最优产品组合 s^*的概率。抽取产品组合 S_i 后，根据 MNL 按照式（5）中基于真实参数 θ^* 计算的概率抽取顾客选择（动作）A_i。

8.2. 基准方法

在我们的实验中，我们使用梯度下降法来求得 $\hat{\theta}_{\mathrm{ML},\ n}$ 使经验负对数似然函数最小化。随后，给定 $\hat{\theta}_{\mathrm{ML},\ n}$ 基准方法通过求解（7）中的线性规划问题来解决商品组合优化问题。

8.3. 性能比较

对于给定的 $(N,\ K,\ d,\ n,\ p)$，我们重复第8.1 节中的数据生成过程，随机生成 50 个离线数据集。在这些实验中，我们记录了 PASTA 和基准方法的解。对于超参数，我们设定 $\alpha_n = 2L_{\mathrm{ML}}$，其中 $L_{\mathrm{ML}} = L_n(\hat{\theta}_{\mathrm{ML},\ n})$，且最大迭代次数为 $T = 30$。我们通过两个指标衡量性能：（1）解的

平均遗憾，用于表示解的性能与最优性能（即 s^* 的收入）之间的差距；（2）解的产品组合准确率（相对于最优产品组合 s^*）。产品组合 s 的产品组合准确率定义为正确选取的产品数量占s^*中产品数量的比例。主要结果总结如下。

样本量的影响。我们设定 $N = 40$， $K = 8$， $d = 16$ 和 $p = 0.9$。然后逐步增加样本数量n。结果如图1 所示，PASTA 的表现显著优于基准方法。尽管基准方法的表现随着样本数量增加而提升，PASTA 方法的遗憾始终低于基准方法的25%。在增加产品数量（$N = 60$， $K = 15$）后重复同一实验，结果表明 PASTA 方法的优势依然稳定，如图1所示。

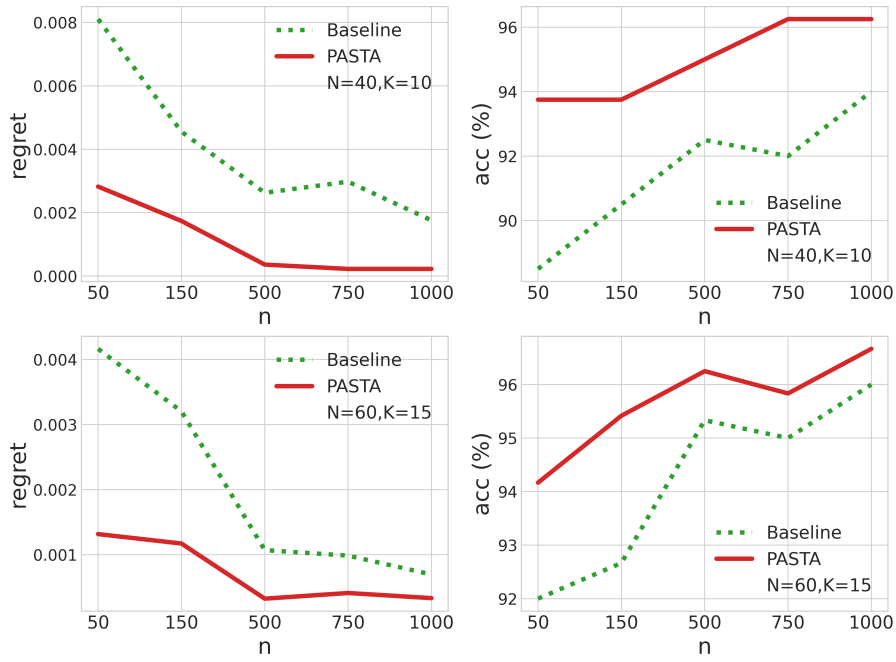


图*1*。PASTA 与基准方法在样本数量不同（n）时的性能比较。左侧为平均遗憾（越低越好），右侧为产品组合准确率（越高越好）。

离线数据中抽样得到最优产品组合的概率的影响。我们设定 $N = 40$， $K = 8$， $d = 16$， $n = 150$，并令 $p \in \{0.1, 0。3, 0。5, 0。7, 0.9\}$。我们还研究了在产品总数增加的情形（$N = 60$， $K = 15$）下 p 的影响。如图2所示，对于 p的不同取值，悲观型商品组合优化的优势始终一致且稳健。

特征维度的影响。 我们设定 $N = 20$， $K = 5$， $p = 0.9$， $n = 150$，并令 $d \in \{8, 20, 32, 64, 128\}$。为了刻画维度 d的影响，我们从均匀分布$[-1, 1]$中独立生成 d 个 θ^* 元素。结果见图 3。我们观察到，随着特征维度的增加，基准方法和悲观型商品组合优化的遗憾都会增加，但当维度 d变化时，PASTA 方法仍能保持其性能优势。

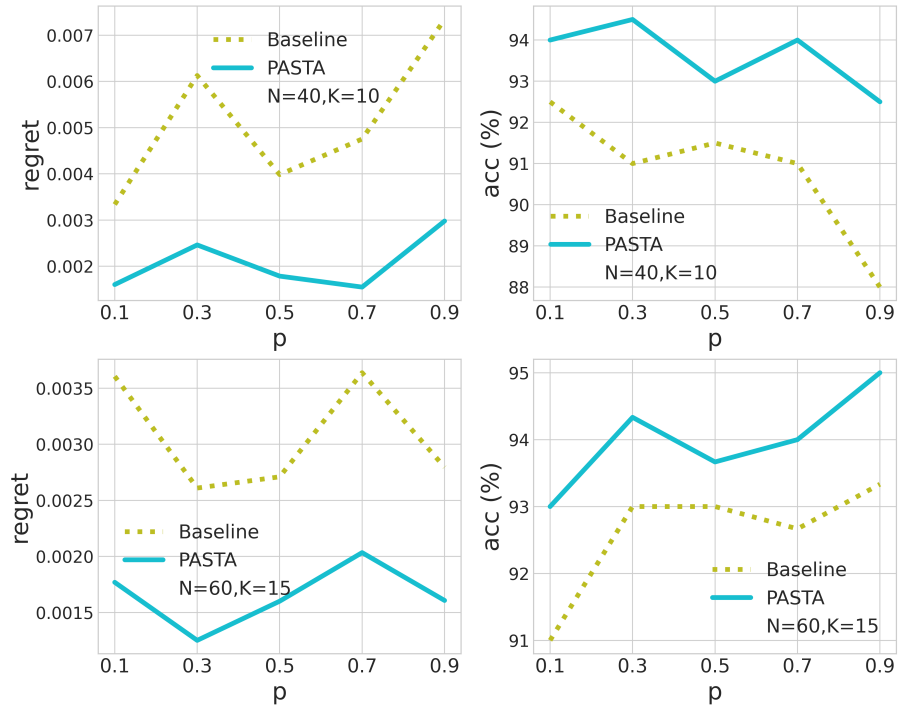


图*2*。PASTA 与基准方法在最优产品组合的概率不同（p）时的比较。上排： $N = 40$；下排： $N = 60$。

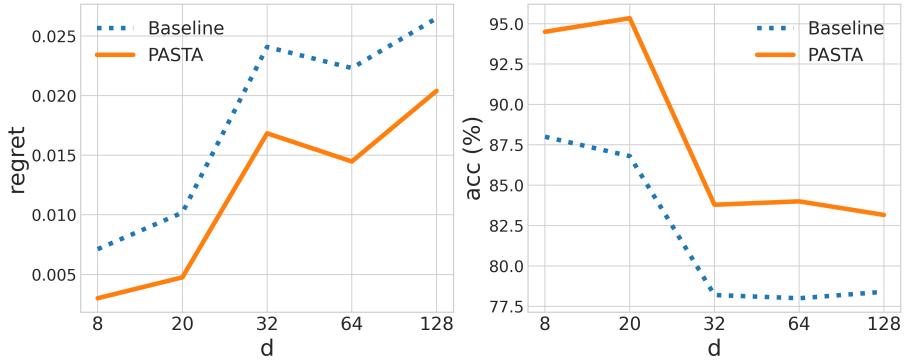


图*3*。PASTA 与基准方法在商品特征维度增加（d）时的比较。

9. 结论

本文研究了离线商品组合优化问题中数据覆盖不足的问题。随着选择数量随商品数量 N 快速增长，这一问题变得更加严峻。我们提出了悲观型商品组合优化框架，并为该方法提供了理论依据。随后，我们深入研究了多项Logit模型（MNL），并推导出针对这一常用模型的悲观型商品组合优化有限样本遗憾界。我们提出了一种高效算法，用于求解 MNL 的悲观型商品组合优化问题，并通过大量数值研究证明，我们的方法显著优于基准方法。

参考文献

Aouad, A., Farias, V. 和 Levi, R. 基于先考虑后选择选择模型的商品组合优化。管理科学，67(6):3368–3386，2021年。Aouad, A., Feldman, J. 和 Segev, D. 用于商品组合优化的指数多选择模型：多项Logit模型的替代方案？管理科学，2022年。Bertsimas, D. 和 Kallus, N. 从预测分析到处方分析。管理科学，66(3):1025–1044，2020年。Caro, F. 和 Gallien, J. 季节性消费品的动态产品组合与需求学习。管理科学，53 (2):276–292，2007年。Chen, X., Wang, Y. 和 Zhou, Y. 基于变化情境信息的动态产品组合优化。机器学习研究杂志，2020年。Chen, X., Shi, C., Wang, Y. 和 Zhou, Y. 嵌套Logit模型下的动态产品组合规划。生产与运营管理，30(1):85–102，2021年a。Chen, X., Wang, Y. 和 Zhou, Y. 多项Logit模型下动态产品组合规划的最优策略。运筹学数学，46(4):1639–1657，2021年b。Davis, J. M., Gallego, G. 和 Topaloglu, H. 具有全幺模约束结构的多项Logit模型下的产品组合规划，2013年。网址 https://people.orie.cornell.edu/jmd388/publications/ MNLConstr.pdf。工作论文，康奈尔大学，伊萨卡，纽约州。Désir, A., Goyal, V., Segev, D. 和 Ye, C. 基于马尔可夫链的选择模型下的约束产品组合优化。管理科学，66(2):698–721，2020年。Diaconis, P. 和 Saloff-Coste, L. 有限马尔可夫链的对数Sobolev不等式。应用概率年刊，6(3):695–750，1996年。Feldman, J. 和 Topaloglu, H. 多项Logit混合模型下商品组合优化的最优期望收入界限。生产与运营管理，24(10):1598–1620，2015年。Feng, Q., Shanthikumar, J. G. 和 Xue, M. 消费者选择模型与估计：综述与拓展。生产与运营管理，31(2):847–867，2022年。Flores, A., Berbeglia, G. 和 Van Hentenryck, P. 序贯多项Logit模型下的商品组合优化。欧洲运筹学杂志，273(3):1052–1064，2019年。Fu, Z., Qi, Z., Wang, Z., Yang, Z., Xu, Y. 和 Kosorok, M. R. 混杂马尔可夫决策过程中的工具变量离线强化学习，2022年。网址 https://arxiv.org/abs/2209.08666。工作论文，西北大学，埃文斯顿，伊利诺伊州。Gallego, G. 和 Topaloglu, H. 嵌套Logit模型下的约束商品组合优化。管理科学，60 (10):2583–2601，2014年。Harsha, P. 通信复杂度，2011年。网址 https://www.tifr.res.in/~ prahladh/teaching/ 2011-12/comm/lectures/l12.pdf。讲义，塔塔基础研究院（TIFR），孟买，印度。Imbens, G. W. 和 Rubin, D. B. 社会、生物医学科学中的统计因果推断。剑桥大学出版社，剑桥，英国，2015年。Jin, Y., Yang, Z. 和 Wang, Z. 悲观主义对于离线强化学习是否能被证明是高效的？载于 Meila, M. 和 Zhang, T.（编），第38届国际机器学习会议论文集，第139卷《机器学习研究论文集》，第5084–5096页。PMLR，2021年7月18日至24日。Kidambi, R., Rajeswaran, A., Netrapalli, P. 和 Joachims, T. Morel：基于模型的离线强化学习。载于 Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. 和 Lin, H.（编），神经信息处理系统进展，第33卷，第21810–21823页。Curran Associates, Inc.，2020年。

Kumar, A., Zhou, A., Tucker, G. 和 Levine, S. 用于离线强化学习的保守型Q学习。载于第34届神经信息处理系统国际会议论文集，NIPS'20，美国纽约州雷德胡克，2020年。Curran Associates公司。Li, S., Luo, Q., Huang, Z. 和 Shi, C. 马尔可夫链选择模型下约束商品组合优化的在线学习，2022年。网址 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4079753。工作论文，密歇根大学，安娜堡，密歇根州。Liu, N., Ma, Y. 和 Topaloglu, H. 具有顺序供给的多项Logit模型下的商品组合优化。INFORMS计算学报，32(3):835–853，2020年。McFadden, D. 定性选择行为的条件Logit分析。城市与区域发展研究所，伯克利，加利福尼亚州，1973年。McFadden, D. 概率选择的计量经济模型。《计量经济应用中的离散数据结构分析》，198272，1981年。Mo, W. 和 Liu, Y. 异方差或模型设定错误的无治疗效应模型中最优个体化治疗规则的高效学习。皇家统计学会期刊：B辑（统计方法论），84(2):440–472，2022年。Owen, A. 经验似然比置信域。统计学年鉴，18(1):90–120，1990年。Pang, R. 1 全幺模矩阵——斯坦福大学，2017年。网址 https://theory.stanford.edu/~jvondrak/MATH233B-2017/lec3.pdf。Qian, M. 和 Murphy, S. A. 个体化治疗规则的性能保证。统计学年鉴，39(2): 1180–1210，2011年。Rosenbaum, P. R. 和 Rubin, D. B. 倾向得分在因果效应观察性研究中的核心作用。Biometrika，70(1):41–55，1983年。Rubin, D. B. 估计随机研究和非随机研究中的治疗因果效应。《教育心理学杂志》，66(5):688，1974年。Rusmevichientong, P., Shen, M. 和 Shmoys, D. 容量约束比值和优化的多项式时间近似方案。运筹学快报，37:230–238，2009年7月。Rusmevichientong, P., Shen, Z.-J. M. 和 Shmoys, D. B. 具有容量约束的多项Logit选择模型下的动态产品组合优化。运筹学，58(6):1666–1680，2010年。Rusmevichientong, P., Sumida, M. 和 Topaloglu, H. 随机使用时长下可重复使用产品的动态产品组合优化。管理科学，66(7):2820–2844，2020年。Sauré, D. 和 Zeevi, A. 需求学习下的最优动态产品组合规划。制造与服务运营管理，15(3):387–404，2013年。Sen, B. 经验过程理论及其应用简介，2018年。网址 http://www.stat.columbia. edu/~bodhi/Talks/Emp-Proc-Lecture-Notes.pdf。讲义，哥伦比亚大学，纽约，纽约州。Talluri, K. 和 van Ryzin, G. 一般离散消费者行为选择模型下的收益管理。管理科学，50(1):15–33，2004年。van der Vaart, A. W. 和 Wellner, J. A. 弱收敛与经验过程：及其在统计学中的应用。Springer，纽约，纽约州，1996年。Wang, Y., Chen, X. 和 Zhou, Y. 动态多项Logit产品组合选择模型的近最优策略。载于 Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N. 和 Garnett, R.（编），神经信息处理系统进展，第31卷。Curran Associates, Inc.，2018年。Yu, T., Thomas, G., Yu, L., Ermon, S., Zou, J. Y., Levine, S., Finn, C. 和 Ma, T. 基于模型的离线策略优化。载于 Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. 和 Lin, H.（编），神经信息处理系统进展，第33卷，第14129–14142页。Curran Associates, Inc.，2020年。

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. 使用结果加权学习估计个体化治疗规则。美国统计学会会刊, 107(499):1106–1118，2012。

A. 理论结果的证明

在整个证明过程中，我们用 θ^* 表示真实参数，用 n 表示样本数量。我们用 L_n 表示经验负对数似然函数，即 $\widehat{L}_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log \pi_A(A_i|S_i;\theta)$ 其中 $\{(S_i,\ A_i,\ R_i)\}_{i=1}^{n}$ 是离线数据集。我们分别用 L 和 $\theta_{\mathrm{ML},\ n}$ 表示负对数似然函数，即 $L(\theta) = -\mathbb{E}[\log \pi_A(A|S;\theta)]$，以及 θ^*的最大似然估计，即 $\widehat{\theta}_{\mathrm{ML},\ n} \in \arg\min_{\theta\in\Theta}\left\{\widehat{L}_n(\theta)\right\}$。置信区域 Ω_n 定义为

$$\widehat{\Omega}_n \doteq \{\theta \in \Theta : L_n(\theta) - L_n(\theta_{\mathrm{ML},\ n}) \leqslant \alpha_n\}.$$

对于一对一般的随机变量 $(X,\ Y)$，假设 Y 给定 X 的条件概率密度函数由 $p(y|x$ 参数化建模；$\theta)$ 对于参数 θ。出于技术原因，我们将考虑以下距离。

定义 A.1（平方海灵格距离）。

$$h^2(p(\cdot|x;\theta_1), p(\cdot|x;\theta_2)) = \frac{1}{2}\int \left(\sqrt{p_1(y|x;\theta_1)} - \sqrt{p_2(y|x;\theta_2)}\right)^2 dy. \tag{10}$$

定义 A.2（海林格距离）。

$$h(p(\cdot|x;\theta_1), p(\cdot|x;\theta_2)) = \sqrt{h^2(p(\cdot|x;\theta_1), p(\cdot|x;\theta_2))}. \tag{11}$$

定义 A.3（广义平方海灵格距离）。

$$H^2(\theta_1,\theta_2) = \mathbb{E}_X\Big[h^2(p(\cdot|X;\theta_1), p(\cdot|X;\theta_2))\Big]. \tag{12}$$

定义 A.4（广义海灵格距离）。

$$H(\theta_1,\theta_2) = \mathbb{E}_X\Big[\sqrt{h^2(p(\cdot|X;\theta_1), p(\cdot|X;\theta_2))}\Big]. \tag{13}$$

在我们的理论结果中，我们特别考虑 $p(y|x;\ \theta) = \pi_A(a|s;$ 作为 A 给定 S 的条件密度（下文记为 $A|S)$。

A.1. 引理的证明 5.2

Und在假设 5.1下，对于任意 $0 < \delta < 1$，以至少 $1-\delta$的概率，对于任意 $\theta \in \Omega_n$,

$$\mathcal{V}(s^*;\theta^*) - \mathcal{V}(s^*;\theta) \lesssim\ r_{s*} C_{s*}\sqrt{\alpha_n}, \tag{14}$$

其中 $r_{s*} \doteq \max_{j\in s*} r_j$ 是最优产品组合中所有商品可能取得的最大收入。

引理的证明 5.2。对于任意满足 θ 的$L_n(\theta) - L_n(\theta_{\mathrm{ML},\ n}) \leqslant \alpha_n$，即 $\theta \in \Omega_n$，我们有

$$\begin{aligned}
&\mathcal{V}(s^*;\theta^*) - \mathcal{V}(s^*;\theta) \leqslant \left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\theta^*)\right| \\
&\leqslant \left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n}) + \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n}) - \mathcal{V}(s^*;\theta^*)\right| \\
&\leqslant \left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n})\right| + \left|\mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n}) - \mathcal{V}(s^*;\theta^*)\right| \\
&\leqslant 2\max_{\theta\in\Omega_n}\left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n})\right| \qquad \text{(By Assumption 5.1 (II) } \theta^* \in \Omega_n).
\end{aligned}$$

根据引理 B.2，对于任意 $\theta \in \Theta$,

$$\left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n})\right| \leqslant r_{s*} C_{s*}\mathbb{E}_S\Big[||\pi_A(\cdot|S;\theta) - \pi_A(\cdot|S;\widehat{\theta}_{\mathrm{ML},n})||_1\Big]$$

其中 $||\cdot||_1$ 是 ℓ_1-范数， $r_{s*} \doteq \max_{j\in s*} r_j$ 是所有商品中可能的最大收入，且 $C_{s*} = 1/\pi_S(s^*)$。在引理 B.3中，我们证明了

$$\mathbb{E}_S\left[||\pi_A(\cdot|S;\theta) - \pi_A(\cdot|S;\widehat{\theta}_{\mathrm{ML},n})||_1\right] \leqslant 2\sqrt{2}\sqrt{H^2(\theta,\widehat{\theta}_{\mathrm{ML},n})},$$

其中， H^2 是式（12）中定义的广义平方海灵格距离，其中 $p(y|x;\ \theta) = \pi_A(a|s;\ \theta)$ 是 $A|S$的条件密度。

结合上述两个不等式可得，对于任意 $\theta \in \Theta$,

$$\left|\mathcal{V}(s^*;\theta) - \mathcal{V}(s^*;\widehat{\theta}_{\mathrm{ML},n})\right| \lesssim r_{s*}C_{s*}\sqrt{H^2(\theta,\widehat{\theta}_{\mathrm{ML},n})}. \tag{15}$$

接下来，我们利用对任意 $x \geqslant 0$ 均有 $\log x \leqslant 2(\sqrt{x}-1)$ 这一事实，证明对于任意 $s \in \mathbb{S}$ 和任意 θ:

$$\begin{aligned}
&-\int \pi_A(a|s;\theta^*)\log\frac{\pi_A(a|s;\theta)}{\pi_A(a|s;\theta^*)}da \\
&\geqslant -2\int \pi_A(a|s;\theta^*)\left(\sqrt{\frac{\pi_A(a|s;\theta)}{\pi_A(a|s;\theta^*)}}-1\right)da \\
&= \int\left(\pi_A(a|s;\theta^*)+\pi_A(a|s;\theta)-2\sqrt{\pi_A(a|s;\theta)\pi_A(a|s;\theta^*)}\right)da \\
&= \int\left(\sqrt{\pi_A(a|s;\theta^*)}+\sqrt{\pi_A(a|s;\theta)}\right)^2 da \\
&\geqslant \int\left(\sqrt{\pi_A(a|s;\theta^*)}-\sqrt{\pi_A(a|s;\theta)}\right)^2 da,
\end{aligned}$$

这意味着

$$L(\theta) - L(\theta^*) \geqslant 2H^2(\theta;\theta^*). \tag{16}$$

根据引理 B.4，对于任意 $\theta \in \Omega_n$,

$$H^2(\theta,\widehat{\theta}_{\mathrm{ML},n}) \leqslant 2H^2(\theta^*,\theta) + 2H^2(\theta^*,\widehat{\theta}_{\mathrm{ML},n}) \leqslant L(\widehat{\theta}_{\mathrm{ML},n}) - L(\theta^*) + \{L(\theta) - L(\theta^*)\}. \tag{17}$$

根据假设 5.1，我们有，以至少 $1-\delta$ 的概率，对于任意 $\theta \in \Omega_n$,

$$\left|L(\theta) - L(\theta^*) - \left(\widehat{L}_n(\theta) - \widehat{L}_n(\theta^*)\right)\right| \leqslant \alpha_n.$$

换言之，在假设 5.1 下，以至少 $1-\delta$ 的概率，对于任意 $\theta \in \Omega_n$,

$$L(\theta) - L(\theta^*) \leqslant \left|\widehat{L}_n(\theta) - \widehat{L}_n(\theta^*)\right| + \alpha_n \leqslant 2\alpha_n. \tag{18}$$

将式（18）代入式（17），我们有，以至少 $1-\delta$ 的概率，对于任意 $\theta \in \Omega_n$,

$$H^2(\theta,\widehat{\theta}_{\mathrm{ML},n}) \leqslant 4\alpha_n. \tag{19}$$

结合上述不等式和式（15），我们有，以至少 $1-\delta$ 的概率， $\left|\mathcal{V}(s^*;\ \theta) - \mathcal{V}(s^*;\ \widehat{\theta}_{\mathrm{ML}})\right| \lesssim r_{s*}C_{s*}\sqrt{\alpha_n}$ 对于所有 $\theta \in \Omega_n$。证明完毕。 □

A.2. 引理 6.1 的证明

考虑紧集 Θ上的多项Logit模型（5）。假设 $\theta^* \in \Theta$。对于 $0 < \delta < 1$，其概率至少为 $1-\delta$，我们有

$$\widehat{L}_n(\theta^*) - \widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n}) \lesssim \frac{C_A d}{n}\log\frac{\theta_{\max}}{\delta}. \tag{20}$$

证明 引理 6.1。固定 $0<\delta<1$。设 $\alpha_n \asymp \frac{C_A d}{n}$ 对数 $\frac{2\theta_{\max}}{\delta}$。定义预言机置信集如下:

$$\widetilde{\Omega}_n \doteq \{\theta \in \Theta : L(\theta) - L(\theta^*) \leqslant \alpha_n\}.$$

特别地, $\theta^* \in \Omega_n$。根据引理 B.1 和 A.2，我们还有，以至少 $1-\delta/2$,

$$L(\widehat{\theta}_{\mathrm{ML},n}) - L(\theta^*) \leqslant 2C_A H^2(\widehat{\theta}_{\mathrm{ML},n}, \theta^*) \lesssim \alpha_n,$$

也就是说，$\theta_{\mathrm{ML},n} \in \Omega_n$。

定义 $\mathcal{F}_n \doteq \left\{\log \frac{\pi_A(A|S\,\theta^*)}{\pi_A(A|S\,\theta)} : \theta \in \Omega_n\right\}$。特别地，对于任意 $f \in \mathcal{F}_n$，我们有 $\|f\|_\infty \leqslant 2\log C_A$。令 $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_n} \doteq \sup_{f\in\mathcal{F}_n} |(\mathbb{P}_n - \mathbb{P})(f)|$ 为包络函数。根据 Talagrand 不等式，以至少 $1-\delta/2$, 的概率，对于任意$f \in \mathcal{F}_n$，我们有

$$(\mathbb{P}_n - \mathbb{P})(f) \lesssim \mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_n} + \sqrt{\frac{1}{n}\left\{(\log C_A)\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_n} + \sup_{f\in\mathcal{F}_n}\mathbb{E}(f-\mathbb{E}f)^2\right\}\log\frac{2}{\delta}} + \frac{\log C_A}{n}\log\frac{2}{\delta}. \tag{21}$$

对于方差项，我们有

$$\begin{aligned}
\sigma^2_{\mathcal{F}_n} &\doteq \sup_{f\in\mathcal{F}_n}\mathbb{E}(f-\mathbb{E}f)^2 \leqslant \sup_{\theta\in\widetilde{\Omega}_n}\mathbb{E}\left(\log\frac{\pi_A(A|S;\theta^*)}{\pi_A(A|S;\theta)}\right)^2 \\
&\leqslant \sup_{\theta\in\widetilde{\Omega}_n}\mathbb{E}\left\{2C_A h^2\big(\pi_A(\cdot|S;\theta^*),\pi_A(\cdot|S;\theta)\big)\right\} && \text{(by Lemma B.5)} \\
&= 2C_A \sup_{\theta\in\widetilde{\Omega}_n} H^2(\theta^*,\theta) \\
&\leqslant 2C_A \sup_{\theta\in\widetilde{\Omega}_n}[L(\theta) - L(\theta^*)] && \text{(by (17))} \\
&= 2C_A\alpha_n. && \text{(by definition of } \widetilde{\Omega}_n)
\end{aligned}$$

对于期望包络函数，我们下面的目标是应用 Sen (2018，定理 7.13）（见定理 B.6）。考虑覆盖数 $\mathcal{N}(\epsilon,\ \mathcal{F}_n,\ L^2(\mathbb{Q}))$ 对于任意给定的 $\epsilon>0$ 和有限支撑概率测度 $\mathbb{Q}$。根据引理 A.1，基于多项Logit模型（5），对于某个 $\mathsf{L}<+\infty$，$\mathcal{F}_n$ 是关于指标空间 $(\Theta,\ \|\cdot\|_2)$的一类 L-Lipschitz 函数。接着，就括号数 $\mathcal{N}_{[]}$ 和覆盖数 $\mathcal{N}$而言，对于任意 $\epsilon>0$ 和概率测度 $\mathbb{Q}$，有

$$\mathcal{N}(\epsilon\mathsf{L},\mathcal{F}_n,L^2(\mathbb{Q})) \leqslant \mathcal{N}_{[]}(2\epsilon\mathsf{L},\mathcal{F}_n,L^2(\mathbb{Q})) \leqslant \mathcal{N}(\epsilon,\Theta,\|\cdot\|_2).$$

由 $\Theta\subseteq\mathbb{R}^d$和 Θ是紧的，我们进一步得到 $\mathcal{N}(\epsilon,\ \Theta,\ \|\cdot\|_2) \lesssim \left(\frac{1}{\epsilon}\right)^d$。因此,

$$\mathcal{N}(\epsilon,\mathcal{F}_n,L^2(\mathbb{Q})) \lesssim \left(\frac{\mathsf{L}}{\epsilon}\right)^d.$$

根据定理B.6，我们进一步得到

$$\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}_n} \lesssim \sqrt{\frac{d}{n}\sigma^2_{\mathcal{F}_n}\log\frac{\mathsf{L}}{\sigma_{\mathcal{F}_n}}} \vee \left\{\frac{d}{n}\times 2\log(C_A)\log\frac{\mathsf{L}}{\sigma_{\mathcal{F}_n}}\right\} \lesssim \sqrt{\frac{C_A d}{n}\alpha_n} \asymp \frac{C_A d}{n}\sqrt{\log\frac{2\theta_{\max}}{\delta}} \lesssim \alpha_n.$$

Talagrand 不等式（21）变为

$$(\mathbb{P}_n - \mathbb{P})(f) \lesssim \alpha_n.$$

特别地，当 $\theta_{\mathrm{ML},n}\in\Omega_n$ 对应于 $f(\theta_{\mathrm{ML},n})\in\mathcal{F}_n$ 时，我们有

$$\widehat{L}_n(\theta^*) - \widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n}) = -\mathbb{P}_n f(\widehat{\theta}_{\mathrm{ML},n}) \lesssim -\mathbb{P}f(\widehat{\theta}_{\mathrm{ML},n}) + \alpha_n = \underbrace{L(\theta^*) - L(\widehat{\theta}_{\mathrm{ML},n})}_{\leqslant 0} + \alpha_n \leqslant \alpha_n.$$

至此，证明完成。 □

引理 A.1。 考虑多项*Logit*模型：

$$\pi_A(i|s;\theta)=\frac{\exp(x_i^\top\theta)}{1+\sum_{j\in s}\exp(x_j^\top\theta)};\quad \pi_A(0|s;\theta)=\frac{1}{1+\sum_{j\in s}\exp(x_j^\top\theta)};\quad \forall i\in s,\ s\in\mathbb{S},\ \theta\in\Theta.$$

令 $\theta_{\max}\doteq\max_{\theta\in\Theta}\|\theta\|_2$, $x_{\max}\doteq\max_{j\in[N]}\|x_j\|_2$。若 Θ 是紧的，即 $\theta_{\max}<+\infty$，则对数似然比$\log\frac{\pi_A(A|S\,\theta^*)}{\pi_A(A|S\,\theta)}$;; 是关于~~$\theta\in\Theta$~~的均匀 *Lipschitz* 函数。

引理 *A.1* 的证明。

$$\begin{aligned}\left\|\frac{\partial}{\partial\theta}\log\frac{\pi_A(A|S;\theta^*)}{\pi_A(A|S;\theta)}\right\|_2&=\frac{1}{\pi_A(A|S;\theta)}\left\|\frac{\partial\pi(A|S;\theta)}{\partial\theta}\right\|_2\\&=[1-\pi_A(A|S;\theta)]\left\|x_A+\sum_{j\in S}\exp(x_j^\top\theta)(x_A-x_j)\right\|_2\\&\leqslant x_{\max}+2Nx_{\max}\exp(x_{\max}\theta_{\max})\doteq\mathsf{L}<+\infty.\end{aligned}\tag{22}$$

也就是说，$\theta\mapsto\log\frac{\pi_A(A|S\,\theta^*)}{\pi_A(A|S\,\theta)}$;;是 L-Lipschitz 的。 □

引理 A.2 （参数化极大似然估计的海林格距离集中性）。 考虑多项*Logit*模型 (5)。对于 $0<\delta<1$，其概率至少为 $1-\delta$，有

$$H^2(\widehat{\theta}_{\mathrm{ML},n},\theta^*)\lesssim\frac{d}{n}\log\frac{\theta_{\max}}{\delta}.$$

引理 *A.2* 的证明。 我们将其视为 Fu 等人 (2022年，推论 2)的一个特例；与一般的马尔可夫链不同，我们的数据是独立同分布生成的。 □

A.3. 引理的证明 6.2

考虑多项Logit模型（5）。假设引理 6.1 中的条件成立，且 $L(\theta)$ 和 $L_n(\theta)$ 均一致强凸。令 $\alpha_n\asymp\frac{C_A d}{n}\log\frac{\theta_{\max}}{\delta}$。对于 $0<\delta<1$，以至少 $1-\delta$的概率，有

$$\sup_{\theta\in\Omega_n}\left|L(\theta)-L(\theta^*)-\left(\widehat{L}_n(\theta)-\widehat{L}_n(\theta^*)\right)\right|\leqslant\alpha_n.$$

引理的证明 *6.2*。 固定 $0<\delta<1$。根据 $L(\theta)$ 和 $L_n(\theta)$的强凸性假设，存在常数 $\mu>0$，使得对于任意 $\theta\in\Theta$,

$$\mu\|\theta-\theta^*\|_2^2\leqslant L(\theta)-L(\theta^*);\quad \mu\|\theta-\widehat{\theta}_{\mathrm{ML},n}\|_2^2\leqslant\widehat{L}_n(\theta)-\widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n}).$$

根据引理 6.1，以至少 $1-\delta/2$, 的概率，有 $\theta_{\mathrm{ML},\,n}\in\Omega_n$。因此，对于任意 $\theta\in\Omega_n$，有

$$\begin{aligned}&\|\theta-\theta^*\|_2\leqslant\|\theta-\widehat{\theta}_{\mathrm{ML},n}\|_2+\|\widehat{\theta}_{\mathrm{ML},n}-\theta^*\|_2 && \text{(by triangular inequality)}\\&\leqslant\frac{1}{\sqrt{\mu}}\sqrt{\widehat{L}_n(\theta)-\widehat{L}_n(\widehat{\theta}_{\mathrm{ML},n})}+\frac{1}{\sqrt{\mu}}\sqrt{L(\widehat{\theta}_{\mathrm{ML},n})-L(\theta^*)} && \text{(by strong convexity)}\\&\lesssim\sqrt{\alpha_n} && \text{(by } \theta\in\Omega_n \text{ and } \widehat{\theta}_{\mathrm{ML},n}\in\widetilde{\Omega}_n \text{ respectively)}.\end{aligned}$$

上述结果意味着，以至少 $1-\delta/2$, 的概率，我们有 $\Omega_n\subseteq\bar{\Omega}_n$，其中 $\bar{\Omega}_n$ 是一个以 θ^* 为中心、半径为 $\sqrt{\alpha_n}$ 的球：

$$\bar{\Omega}_n\doteq\left\{\theta\in\Theta:\|\theta-\theta^*\|_2\leqslant\sqrt{\alpha}_n\right\}.$$

定义 $\bar{\mathcal{F}}_n \doteq \left\{\log \frac{\pi_A(A|S\,\theta^*)}{\pi_A(A|S\,\theta)}:\ \theta \in \bar{\Omega}_n\right\}$。令 $\|\mathbb{P}_n - \mathbb{P}\|_{\bar{\mathcal{F}}_n} \doteq \sup_{f\in\bar{\mathcal{F}}_n} |(\mathbb{P}_n - \mathbb{P})(f)|$ 为包络。根据塔拉格朗德不等式，以至少 $1-\delta/2$, 的概率，对于任意 $f \in \mathcal{F}_n$，我们有：

$$|(\mathbb{P}_n - \mathbb{P})(f)| \lesssim \mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\bar{\mathcal{F}}_n} + \sqrt{\frac{1}{n}\left\{(\log C_A)\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\bar{\mathcal{F}}_n} + \sup_{f\in\bar{\mathcal{F}}_n}\mathbb{E}(f-\mathbb{E}f)^2\right\}\log\frac{2}{\delta}} + \frac{\log C_A}{n}\log\frac{2}{\delta}. \tag{23}$$

对于方差项，我们有

$$\begin{aligned}
\sigma^2_{\bar{\mathcal{F}}_n} &\doteq \sup_{f\in\bar{\mathcal{F}}_n}\mathbb{E}(f-\mathbb{E}f)^2 \leqslant \sup_{\theta\in\bar{\Omega}_n}\mathbb{E}\left(\log\frac{\pi_A(A|S;\theta^*)}{\pi_A(A|S;\theta)}\right)^2 \\
&\lesssim \sup_{\theta\in\bar{\Omega}_n}\|\theta-\theta^*\|_2^2 && \text{(by Lipschitzness in Lemma A.1)} \\
&\leqslant \alpha_n && \text{(by definition of } \bar{\Omega}_n).
\end{aligned}$$

对于期望包络函数，根据定理 B.6，我们进一步有

$$\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\bar{\mathcal{F}}_n} \lesssim \sqrt{\frac{d}{n}\sigma^2_{\bar{\mathcal{F}}_n}\log\frac{\mathsf{L}}{\sigma_{\bar{\mathcal{F}}_n}}} \vee \left\{\frac{d}{n}\times 2\log(C_A)\log\frac{\mathsf{L}}{\sigma_{\bar{\mathcal{F}}_n}}\right\} \lesssim \sqrt{\frac{d}{n}\alpha_n} \lesssim \alpha_n.$$

因此，Talagrand 不等式（23）表明，对于任意 $f \in \mathcal{F}_n$

$$|(\mathbb{P}_n - \mathbb{P})(f)| \lesssim \alpha_n.$$

换言之 以至少 $1-\delta$ 的概率，$\Omega_n \subseteq \Omega_n$ 对应于 $\{f(\theta)\}_{\theta\in\Omega_n} \subseteq \mathcal{F}_n$，我们有

$$\sup_{\theta\in\Omega_n}\left|L(\theta) - L(\theta^*) - \left(\hat{L}_n(\theta) - \hat{L}_n(\theta^*)\right)\right| \leqslant \sup_{f\in\bar{\mathcal{F}}_n}|(\mathbb{P}_n-\mathbb{P})(f)| \lesssim \alpha_n.$$

□

B. 技术引理

引理 B. **1.** 假设 $C_A > 2$ 和 $C_A \geqslant 1/\pi_A(i|s;\theta)$ 对所有 $\theta\in\Theta$、$s\in\mathbb{S}$ 和 $i\in s$均成立。那么对于任意 $\theta\in\Theta$:

$$|L(\theta) - L(\theta^*)| \leqslant 2C_A H^2(\theta,\theta^*), \tag{24}$$

引理 *B.1*的证明。 根据定义，

$$|L(\theta^*) - L(\theta)| = \left|\mathbb{E}\left\{\mathbb{E}_A\left[\log\frac{\pi_A(A|S;\theta^*)}{\pi_A(A|S;\theta)}\middle|S\right]\right\}\right|.$$

特别地，对于固定的 $s\in[N]$，我们有

$$\begin{aligned}
&\mathbb{E}\left[\log\frac{\pi_A(A|S;\theta^*)}{\pi_A(A|S;\theta)}\middle|S=s\right] = \mathrm{KL}\Big(\pi_A(\cdot|s;\theta^*)\Big\|\pi_A(\cdot|s;\theta)\Big) \\
&\leqslant \frac{\log(C_A-1)}{1-2/C_A}\Big\{1-\big[1-h^2(\pi_A(\cdot|s;\theta^*),\pi_A(\cdot|s;\theta))\big]^2\Big\} \\
&\quad \text{(by log-Sobolev inequality (Diaconis \& Saloff-Coste, 1996, Theorem A.1))} \\
&= \frac{C_A\log(C_A-2+1)}{C_A-2}\{2h^2-(h^2)^2\}\ \Big(\text{with } h^2 = h^2\big(\pi_A(\cdot|s;\theta^*),\pi_A(\cdot|s;\theta)\big)\Big) \\
&\leqslant 2C_A h^2(\pi_A(\cdot|s;\theta^*),\pi_A(\cdot|s;\theta)).
\end{aligned}$$

因此，

$$|L(\theta^*) - L(\theta)| \leqslant \mathbb{E}\left\{2C_A h^2(\pi_A(\cdot|S;\theta^*),\pi_A(\cdot|S;\theta))\right\} = 2C_A H^2(\theta^*,\theta).$$

□

引理 B.2. 勒 t $C_{s*} \doteq \frac{1}{\pi_S(s*)}$ 和 $r_{s*} \doteq$ 最大值$_{j \in s*}$ r_j，则以下不等式对任意 $\theta_1, \theta_2 \in \Theta$均成立：

$$\left|\mathcal{V}(s^*;\theta_1) - \mathcal{V}(s^*;\theta_2)\right| \leqslant r_{s*} C_{s*} \mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1\right]$$

其中 $||\cdot||_1$ 表示 L^1 范数。

引理的证明 *B.2*。

$$\left|\mathcal{V}(s^*;\theta_1) - \mathcal{V}(s^*;\theta_2)\right| = \left|\mathbb{E}_S\left[\frac{\mathbb{I}(S=s^*)}{\pi_S(S)} \sum_{i \in S} r_{i,S}\Big(\pi_A(i|S;\theta_1) - \pi_A(i|S;\theta_2)\Big)\right]\right| \tag{25}$$

$$\leqslant \mathbb{E}_S\left[\left|\frac{\mathbb{I}(S=s^*)}{\pi_S(S)} \sum_{i \in S} r_{i,S}\Big(\pi_A(i|S;\theta_1) - \pi_A(i|S;\theta_2)\Big)\right|\right] \tag{26}$$

$$\leqslant \left\|\frac{\mathbb{I}(S=s^*)}{\pi_S(S)}\right\|_\infty \mathbb{E}_S\left[\mathbb{I}(S=s^*)\left|\sum_{i \in S} r_{i,S}\Big(\pi_A(i|S;\theta_1) - \pi_A(i|S;\theta_2)\Big)\right|\right] \tag{27}$$

$$\leqslant C_{s*} \mathbb{E}_S\left[\mathbb{I}(S=s^*) \sum_{i \in S}\left|r_{i,S}\Big(\pi_A(i|S;\theta_1) - \pi_A(i|S;\theta_2)\Big)\right|\right] \tag{28}$$

$$\leqslant r_{s*} C_{s*} \mathbb{E}_S\left[\sum_{i \in S}\left|\Big(\pi_A(i|S;\theta_1) - \pi_A(i|S;\theta_2)\Big)\right|\right] \tag{29}$$

$$= r_{s*} C_{s*} \mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1\right], \tag{30}$$

其中，式（25）来自对 $\mathbb{E}[R(s)]$（式（1））的基于样本的估计；式（27）来自霍尔德不等式；式（28）来自以下事实：$\left\|\frac{\mathbb{I}(S=s^*)}{\pi_S(S)}\right\|_\infty = C_{s*}$ 因为$\frac{\mathbb{I}(S=s^*)}{\pi_S(S)}$ 除在 $s = s$*处外处处为零。最后一个等式由 L^1 范数的定义得出。

□

引理 B.3. 对于任意 θ_1，$\theta_2 \in \Theta$,

$$\mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1\right] \leqslant 2\sqrt{2}\sqrt{H^2(\theta_1,\theta_2)}.$$

引理的证明 *B.3*。 首先利用以下事实：**(1)** $L^1 = \frac{1}{2}$ TV, –其中 TV 是全变差距离；**(2)** TV $\leqslant\sqrt{2}h$ (Harsha，2011 年)，其中 h 是海林格距离，从而对任意 $s \in \mathbb{S}$,

$$||\pi_A(\cdot|s;\theta_1) - \pi_A(\cdot|s;\theta_2)||_1 \leqslant 2\sqrt{2}h\big(\pi_A(\cdot|s;\theta_1), \pi_A(\cdot|s;\theta_2)\big). \tag{31}$$

由（31），可得

$$||\pi_A(\cdot|s;\theta_1) - \pi_A(\cdot|s;\theta_2)||_1 \leqslant 2\sqrt{2}h\big(\pi_A(\cdot|s;\theta_1), \pi_A(\cdot|s;\theta_2)\big),$$
$$||\pi_A(\cdot|s;\theta_1) - \pi_A(\cdot|s;\theta_2)||_1^2 \leqslant 8h^2\big(\pi_A(\cdot|s;\theta_1), \pi_A(\cdot|s;\theta_2)\big).$$

对两边关于 S 取期望，可得

$$\mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1^2\right] \leqslant 8\mathbb{E}_S\left[h^2(\pi_A(\cdot|S;\theta_1), \pi_A(\cdot|S;\theta_2))\right],$$
$$\mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1^2\right] \leqslant 8H^2(\theta_1,\theta_2).$$

根据詹森不等式，我们有

$$\left[\mathbb{E}_S||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1\right]^2 \leqslant \mathbb{E}_S\left[||\pi_A(\cdot|S;\theta_1) - \pi_A(\cdot|S;\theta_2)||_1^2\right].$$

这意味着

$$\mathbb{E}_S\Big[||\pi_A(\cdot|S;\theta_1)-\pi_A(\cdot|S;\theta_2)||_1\Big]\leqslant 2\sqrt{2}\sqrt{H^2(\theta_1,\theta_2)}.$$

□

L引理 B.4 （ H 和 H^2的性质。）对于任意 θ_1、 θ_2、 $\theta_3\in\Theta$，以下不等式成立：

$$\begin{aligned} H(\theta_1,\theta_2) &\leqslant H(\theta_1,\theta_3)+H(\theta_2,p_3);\\ \big(H(\theta_1,\theta_2)\big)^2 &\leqslant H^2(\theta_1,\theta_2)\leqslant H(\theta_1,\theta_2);\\ H^2(\theta_1,\theta_2) &\leqslant 2H^2(\theta_1,\theta_3)+2H^2(\theta_2,\theta_3). \end{aligned}$$

引理*B.4*的证明。为方便记号，对于 $i=1$、2、3，我们用 p_i 表示 π_A 以 θ_i为参数的 $p_i(a|s)=\pi_A(a|s;\ \theta_i)$。

(1) 注意，对于任意 $s\in\mathbb{S}$、 $\sqrt{h^2(p_1(\cdot|s),\ p_2(\cdot|s))}$ 就是满足三角不等式的常规海林格距离。因此，我们有

$$\sqrt{h^2(p_1(\cdot|s),p_2(\cdot|s))}\leqslant\sqrt{h^2(p_1(\cdot|s),p_3(\cdot|s))}+\sqrt{h^2(p_2(\cdot|s),p_3(\cdot|s))}. \tag{32}$$

对式（32关于 S两边取期望，可得

$$\mathbb{E}_S\Big[\sqrt{h^2(p_1(\cdot|S),p_2(\cdot|S))}\Big]\leqslant\mathbb{E}_S\Big[\sqrt{h^2(p_1(\cdot|S),p_3(\cdot|S))}\Big]+\mathbb{E}_S\Big[\sqrt{h^2(p_2(\cdot|S),p_3(\cdot|S))}\Big].$$

根据 H 的定义，这意味着

$$H(\theta_1,\theta_2)\leqslant H(\theta_1,\theta_3)+H(\theta_2,\theta_3).$$

（2）对于第一个不等式，应用詹森不等式可得

$$\left(\mathbb{E}\Big[\sqrt{h^2(p_1(\cdot|S),p_2(\cdot|S))}\Big]\right)^2\leqslant\mathbb{E}\Big[\left(\sqrt{h^2(p_1(\cdot|S),p_2(\cdot|S))}\right)^2\Big]. \tag{33}$$

于是该不等式得证。

对于第二个不等式，我们有

$$\begin{aligned} H(\theta_1,\theta_2)-H^2(\theta_1,\theta_2)=&\mathbb{E}_S\Big[h(p_1(\cdot|S),p_2(\cdot|S))-h^2(p_1(\cdot|S),p_2(\cdot|S))\Big] &(34)\\ =&\mathbb{E}_S\Big[\Big(1-h(p_1(\cdot|S),p_2(\cdot|S))\Big)h(p_1(\cdot|S),p_2(\cdot|S))\Big]. &(35)\end{aligned}$$

注意，对于任意 s， $\Big(1-h(p_1(\cdot|s),\ p_2(\cdot|s))\Big)$ 是一个非负函数，因为海林格距离不大于 1，且 $h(p_1(\cdot|s),$ $p_2(\cdot|s))$ 也是一个正函数，因为它是度量。我们有式（35）是非负函数的期望，因此有

$$\mathbb{E}_S\Big[\Big(1-h(p_1(\cdot|S),p_2(\cdot|S))\Big)h(p_1(\cdot|S),p_2(\cdot|S))\Big]\geqslant 0.$$

于是该不等式得证。

(3) 注意，对于任意 a,b,c，我们都有 $(a-b)^2\leqslant 2(a-c)^2+2(b-c)^2$。由此可知，对于任意 s,

$$\begin{aligned} h^2(p_1(\cdot|s),p_2(\cdot|s))&=\frac{1}{2}\int\Big(\sqrt{p_1(a|s)}-\sqrt{p_2(a|s)}\Big)^2da &(36)\\ &\leqslant\frac{1}{2}\int 2\Big(\sqrt{p_1(a|s)}-\sqrt{p_3(a|s)}\Big)^2+2\Big(\sqrt{p_2(a|s)}-\sqrt{p_3(a|s)}\Big)^2da &(37)\\ &=2\cdot\frac{1}{2}\int\Big(\sqrt{p_1(a|s)}-\sqrt{p_3(a|s)}\Big)^2da+2\cdot\frac{1}{2}\int\Big(\sqrt{p_2(a|s)}-\sqrt{p_3(a|s)}\Big)^2da &(38)\\ &=2h^2(p_1(\cdot|s),p_3(\cdot|s))+2h^2(p_2(\cdot|s),p_3(\cdot|s)). &(39)\end{aligned}$$

这意味着

$$H^2(\theta_1,\theta_2) \leqslant 2H^2(\theta_1,\theta_3) + 2H^2(\theta_2,\theta_3). \tag{40}$$

□

引理 B.5（对数密度比方差界）。 假设 $X \sim p$ 是一个取值于 $\mathbb{R}$ 的随机变量，其概率密度函数为 p，且 p_1、p_2 是 X 的另外两个概率密度函数，使得在 p 的支撑集上， p_1 和 p_2 均有下界 C^{-1}。则有

$$\mathbb{E}_{X\sim p}\left(\log\frac{p_1(X)}{p_2(X)}\right)^2 \leqslant 2Ch^2(p_1,p_2),$$

其中 h^2 是式（10）中的平方海灵格距离。

引理 *B.5* 的证明。 根据 对数$(x) \leqslant 2(\sqrt{x}-1)$，对于任意 $x \geqslant 0$，有

$$\begin{aligned}
&\mathbb{E}_{X\sim p}\left(\log\frac{p_1(X)}{p_2(X)}\right)^2 = \int\left(\log\frac{p_1(X)}{p_2(X)}\right)^2 p(x)\mathrm{d}x \\
&\leqslant 4\int\max\left\{\left(\sqrt{\frac{p_1(x)}{p_2(x)}}-1\right)^2, \left(\sqrt{\frac{p_2(x)}{p_1(x)}}-1\right)^2\right\}p(x)\mathrm{d}x \\
&= 4\int\max\left\{\frac{1}{p_2(x)}\left(\sqrt{p_1(x)}-\sqrt{p_2(x)}\right)^2, \frac{1}{p_1(x)}\left(\sqrt{p_2(x)}-\sqrt{p_1(x)}\right)^2\right\}p(x)\mathrm{d}x \\
&\leqslant 4C\int\left(\sqrt{p_1(x)}-\sqrt{p_2(x)}\right)^2\mathrm{d}x \\
&= 2Ch^2(p_1,p_2).
\end{aligned}$$

□

定理 B.6（Sen（2018，定理 7.13））。 设 $\mathcal{F}$ 是一个可测函数类，满足 上确界$_{f\in\mathcal{F}}\|f\|_\infty \leqslant f_{\max}$，其中某个常数为 $f_{\max} < +\infty$。假设对于 $A \geqslant ef_{\max}$， $d \geqslant 2$， $0 \leqslant \epsilon \leqslant f_{\max}$，以及每个有限支撑的概率测度$\mathbb{Q}$，覆盖数（*Sen，2018*）满足：

$$\mathcal{N}(\epsilon,\mathcal{F},L^2(\mathbb{Q})) \lesssim \left(\frac{A}{\epsilon}\right)^d. \tag{41}$$

令 $\sigma_{\mathcal{F}}^2 \doteq$ 上确界$_{f\in\mathcal{F}}\mathbb{E}(f-\mathbb{E}f)^2$。于是有

$$\mathbb{E}\|\mathbb{P}_n-\mathbb{P}\|_{\mathcal{F}} \lesssim \sqrt{\frac{d}{n}\sigma_{\mathcal{F}}^2\log\frac{A}{\sigma_{\mathcal{F}}}} \vee \left\{\frac{d}{n}f_{\max}\log\frac{A}{\sigma_{\mathcal{F}}}\right\}.$$

定理 *B.6* 的证明。 在此证明中，我们记 X 为底层随机变量， $\{X_i\}_{i=1}^n$ 是 n 的独立同分布副本， X，且对任意 $f \in \mathcal{F}$， $\mathbb{P}_n(f) \doteq \frac{1}{n}\sum_{i=1}^{n-} f(X_i)$， $\mathbb{P}(f) \doteq \mathbb{E}[f(X)]$。不失一般性，假设 $0 \in \mathcal{F}$，并且对任意$f \in \mathcal{F}$， $\mathbb{P}(f) = 0$。令 $\{\epsilon_i\}_{i=1}^n$ 为与 $\{X_i\}_{i=1}^n$ 独立的拉德马赫随机变量。通过对称化可得

$$\mathbb{E}\|\mathbb{P}_n-\mathbb{P}\|_{\mathcal{F}} \leqslant 2\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n\epsilon_i f(X_i)\right|. \tag{42}$$

给定 $\{X_i\}_{i=1}^n$，根据 Dudley 熵界，我们有

$$\mathbb{E}_\epsilon\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n\epsilon_i\frac{f(X_i)}{\sqrt{n}}\right| \leqslant \int_0^{\sigma_{\mathcal{F},n}}\sqrt{\log\mathcal{N}(u,\mathcal{F},L^2(\mathbb{P}_n))}\mathrm{d}u, \tag{43}$$

其中，我们将 $L^2(\mathbb{P}_n)$ 视为 $\mathcal{F}$ 上的度量，也就是说，对于任意 $f \in \mathcal{F}$，$\|f\|^2_{L^2(\mathbb{P}_n)} = \frac{1}{n}\sum_{i=1}^n f(X_i)^2$。我们还记 $\sigma^2_{\mathcal{F},n} \doteq \sup_{f\in\mathcal{F}} \|f\|^2_{L^2(\mathbb{P}_n)}$，并且 $\mathbb{E}_\epsilon$ 以强调期望是关于拉德马赫随机变量 $\{\epsilon_i\}_{i=1}^n$ 计算的，同时将 $\{X_i\}_{i=1}^n$ 视为固定值。由式（41），取 $\mathbb{Q}$ 为 $\mathbb{P}_n$，我们有

$$\begin{aligned}(43) &\lesssim \sqrt{d}\int_0^{\sigma_{\mathcal{F},n}} \sqrt{\log\frac{A}{\delta}}\mathrm{d}\delta \\ &\leqslant 2\sqrt{d}\sigma_{\mathcal{F},n}\sqrt{\log\frac{A}{\sigma_{\mathcal{F},n}}} \qquad \text{(by Lemma B.7)}.\end{aligned} \tag{44}$$

特别是，$\log(A/\sigma_{\mathcal{F},n}) \geqslant \log(A/f_{\max}) \geqslant 1$ 根据假设，这满足引理 B.7 的条件。结合（42）、（43）和（44），我们有

$$\begin{aligned}\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} &\lesssim \sqrt{\frac{d}{n}} \times \mathbb{E}\sqrt{\sigma^2_{\mathcal{F},n}\log\frac{A}{\sigma_{\mathcal{F},n}}} \\ &\leqslant \sqrt{\frac{d}{n}} \times \sqrt{\frac{1}{2}\mathbb{E}(\sigma^2_{\mathcal{F},n})\log\frac{A^2}{\mathbb{E}(\sigma^2_{\mathcal{F},n})}} \qquad \left(\text{by the concavity of } u \mapsto \sqrt{u\log\frac{A^2}{u}}\right),\end{aligned} \tag{45}$$

其中，$\mathbb{E}$表示关于 $\{X_i\}_{i=1}^n$取期望。注意到

$$\mathbb{E}(\sigma^2_{\mathcal{F},n}) = \mathbb{E}\sup_{f\in\mathcal{F}}\mathbb{P}_n(f^2) \leqslant \mathbb{E}\sup_{f\in\mathcal{F}}\left\{|(\mathbb{P}_n-\mathbb{P})(f^2)| + \mathbb{P}(f^2)\right\} \leqslant \mathbb{E}\|\mathbb{P}_n-\mathbb{P}\|_{\mathcal{F}^2} + \sigma^2_{\mathcal{F}},$$

其中我们定义 $\mathcal{F}^2 \doteq \{f^2 : f\in\mathcal{F}\}$。我们的目标是将（42）、（43）、（44）、（45）应用于 $\mathcal{F}^2$。注意，$\sup_{f^2\in\mathcal{F}^2}\|f^2\|_\infty \leqslant f^2_{\max}$，$\sigma^2_{\mathcal{F}^2,n} \doteq \sup_{f^2\in\mathcal{F}^2}\|f^2\|^2_{L^2(\mathbb{P}_n)} \leqslant f^2_{\max}\sup_{f\in\mathcal{F}}\|f\|^2_{L^2(\mathbb{P}_n)} = f^2_{\max}\sigma^2_{\mathcal{F},n}$。根据 $\|f^2-g^2\|_{L^2(\mathbb{P}_n)} = \sqrt{\mathbb{P}_n[(f+g)^2(f-g)^2]} \leqslant 2f_{\max}\sqrt{\mathbb{P}_n(f-g)^2} = 2f_{\max}\|f-g\|_{L^2(\mathbb{P}_n)}$，对于任意 f，$g\in\mathcal{F}$，我们进一步有

$$\mathcal{N}(2f_{\max}\epsilon, \mathcal{F}^2, L^2(\mathbb{P}_n)) \leqslant \mathcal{N}(\epsilon,\mathcal{F},L^2(\mathbb{P}_n)) \lesssim \left(\frac{A}{\epsilon}\right)^d.$$

因此，将（42）、（43）、（44）和（45）应用于 $\mathcal{F}^2$，我们得到

$$\mathbb{E}\|\mathbb{P}_n-\mathbb{P}\|_{\mathcal{F}^2} \lesssim \sqrt{\frac{d}{n}} \times \sqrt{\frac{1}{2}f^2_{\max}\mathbb{E}(\sigma^2_{\mathcal{F},n})\log\frac{4A^2}{\mathbb{E}(\sigma^2_{\mathcal{F},n})}}.$$

定义 $B \doteq \sqrt{\frac{1}{2}\mathbb{E}(\sigma^2_{\mathcal{F},n})\log\frac{A^2}{\mathbb{E}(\sigma^2_{\mathcal{F},n})}}$于是我们得到

$$\mathbb{E}(\sigma^2_{\mathcal{F},n}) - \sigma^2_{\mathcal{F}} \lesssim \sqrt{\frac{d}{n}}f_{\max}B.$$

根据 $u\mapsto u\log\frac{A^2}{u}$ 在 $u\in(0,\ A^2/e]$ 上单调不减，在 $u\in[A^2/e,+\infty)$，上单调不增，因此我们得到

$$B^2 = \frac{1}{2}\mathbb{E}(\sigma^2_{\mathcal{F},n})\log\frac{A^2}{\mathbb{E}(\sigma^2_{\mathcal{F},n})} \lesssim \frac{1}{2}\left\{\left(\sigma^2_{\mathcal{F}} + \sqrt{\frac{d}{n}}f_{\max}B\right)\wedge\frac{A^2}{e}\right\}\log\frac{A^2}{\left(\sigma^2_{\mathcal{F}}+\sqrt{\frac{d}{n}}f_{\max}B\right)\wedge\frac{A^2}{e}}.$$

特别地，$B \leqslant \sqrt{\frac{A^2}{2e}}$，$\sigma^2_{\mathcal{F}} \leqslant f^2_{\max} \leqslant A^2/e^2 < A^2/e$。于是，渐近地，帽函数 A^2/e不起作用。因此，

$$B^2 \lesssim \left(\sigma^2_{\mathcal{F}} + \sqrt{\frac{d}{n}}f_{\max}B\right)\log\frac{A}{\sigma_{\mathcal{F}}}.$$

具体而言，B 受相应二次方程的两个根共同限定：

$$B \lesssim \frac{1}{2}\left\{\sqrt{\frac{d}{n}}f_{\max}\log\frac{A}{\sigma_{\mathcal{F}}} + \sqrt{\frac{d}{n}f^2_{\max}\left(\log\frac{A}{\sigma_{\mathcal{F}}}\right)^2 + 4\sigma^2_{\mathcal{F}}\log\frac{A}{\sigma_{\mathcal{F}}}}\right\} \lesssim \left\{\sqrt{\frac{d}{n}}f_{\max}\log\frac{A}{\sigma_{\mathcal{F}}}\right\} \vee \sqrt{\sigma^2_{\mathcal{F}}\log\frac{A}{\sigma_{\mathcal{F}}}}.$$

结合（45），我们进一步得到

$$\mathbb{E}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \lesssim \sqrt{\frac{d}{n}} B \lesssim \sqrt{\frac{d}{n}\sigma_{\mathcal{F}}^2 \log\frac{A}{\sigma_{\mathcal{F}}}} \vee \left\{\frac{d}{n} f_{\max} \log\frac{A}{\sigma_{\mathcal{F}}}\right\}.$$

□

引理 B.7。 假设 a、 $A > 0$ 使得 对数$(A/a) \geqslant 1$。则有

$$\int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u \leqslant 2a\sqrt{\log\frac{A}{a}}.$$

引理 *B.7*的证明。 定义

$$f(a) \doteq \begin{cases} 2a\sqrt{\log\frac{A}{a}} - \int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u, & a > 0; \\ 0, & a = 0. \end{cases}$$

则 f 在 0 处连续。此外，对于 $a > 0$，有

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

当 $\log(A/a) \geqslant 1$ 时，该式为非负。由于 $a \to 0^+$，我们进一步有

$$\begin{aligned} &\frac{f(a)}{a} = 2\sqrt{\log\frac{A}{a}} - \frac{1}{a}\int_0^a \sqrt{\log\frac{A}{u}}\mathrm{d}u \\ &= \sqrt{\log\frac{A}{a}} - \frac{1}{2a}\int_0^a \frac{1}{\sqrt{\log\frac{A}{u}}}\mathrm{d}u \qquad \text{(by integration-by-part)} \\ &\geqslant \sqrt{\log\frac{A}{a}} - \frac{1}{2\sqrt{\log\frac{A}{a}}}; \\ &\liminf_{a\to 0^+} \frac{f(a)}{a} \geqslant +\infty. \end{aligned}$$

因此，对于任意 $a \geqslant 0$，有 $f(a) \geqslant 0$，证明完毕。 □