Appendices

A Concentration Inequalities

We introduce the concentration inequalities used in this paper. We begin with the Hoeffding's celebrated inequality for the sum of bounded variables (Hoeffding, 1963).

Lemma A.1 (Chernoff-Hoeffding's inequality)**.** *Consider n independent bounded random variables $X_1, \dots, X_n \in [0, 1]$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[\bar{X}]$. We have $\mathbb{P}(|\bar{X} - \mu| \geq \sqrt{\frac{\log(2/\delta)}{2n}}) \leq \delta$.*

Next we state the multiplicative Chernoff inequalities for the geometric random variables (Agrawal et al., 2019). We say a random variable is *geometric* if $\mathbb{P}(X = m) = p(1-p)^m$.

Lemma A.2 (Agrawal et al. (2019), Corollary D.1)**.** *Consider n i.i.d. geometric random variables $X_1, \dots, X_n$ with expectation $\mathbb{E}[X_i] = \mu \leq 1$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. We have*

(a) $\mathbb{P}(|\bar{X} - \mu| > \sqrt{\frac{48\bar{X}\log(\sqrt{N}\ell+1)}{n}} + \frac{48\log(\sqrt{N}\ell+1)}{n}) \leq \frac{6}{N\ell^2}$,

(b) $\mathbb{P}(|\bar{X} - \mu| > \sqrt{\frac{24\mu\log(\sqrt{N}\ell+1)}{n}} + \frac{48\log(\sqrt{N}\ell+1)}{n}) \leq \frac{4}{N\ell^2}$,

(c) $\mathbb{P}(\bar{X} \geq \frac{3\mu}{2} + \frac{48\log(\sqrt{N}\ell+1)}{n}) \leq \frac{3}{N\ell^2}$.

We rephrase the above lemma into the below form. Lemma A.3 can be proved by following Appendix D in (Agrawal et al., 2019). Similar inequalities with constants smaller than 48 were shown in (Jin et al., 2019; Janson, 2018).

Lemma A.3. *Consider n i.i.d. geometric random variables $X_1, \dots, X_n$ with expectation $\mathbb{E}[X_i] = \mu \leq 1$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. We have*

(a) $\mathbb{P}(|\bar{X} - \mu| > \sqrt{\frac{48\bar{X}\log(2/\delta)}{n}} + \frac{48\log(2/\delta)}{n}) \leq 6\delta$,

(b) $\mathbb{P}(|\bar{X} - \mu| > \sqrt{\frac{24\mu\log(2/\delta)}{n}} + \frac{48\log(2/\delta)}{n}) \leq 4\delta$,

(c) $\mathbb{P}(\bar{X} \geq \frac{3\mu}{2} + \frac{48\log(2/\delta)}{n}) \leq 3\delta$.

The following lemma is a direct corollary of Lemma A.3. It can be proved by following the proof of Lemma 4.1 in (Agrawal et al., 2019, Appendix A). Here "$\wedge$" means logical and.

Lemma A.4. *Consider n i.i.d. geometric random variables $X_1, \dots, X_n$ with expectation $\mathbb{E}[X_i] = \mu \leq 1$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\epsilon = \sqrt{\frac{48\bar{X}\log(2/\delta)}{n}} + \frac{48\log(2/\delta)}{n}$. Then we have*

$$\mathbb{P}(\{\bar{X} - \epsilon \leq \mu \leq \bar{X} + \epsilon\} \wedge \{\epsilon \leq \sqrt{\frac{196\mu\log(2/\delta)}{n}} + \frac{196\log(2/\delta)}{n}\}) \geq 1 - 13\delta.$$

We state another concentration inequality to the geometric random variables. The following inequality is focused on the upper tail of the geometric random variables.

Lemma A.5 (Janson (2018), Theorem 2.1)**.** *Consider n independent "shifted" geometric random variables $X_1, \dots, X_n$ that $\mathbb{P}(X_i = k) = p_i(1-p_i)^{k-1}$. Let $p_* = \min_{1 \leq i \leq n} p_i > 0, X = \sum_{i=1}^{n} X_i, \mu = \mathbb{E}[X]$. We have $\mathbb{P}(X \geq \lambda\mu) \leq e^{-p_*\mu(\lambda - 1 - \ln\lambda)}$.*

B Proofs for Section 3.1

B.1 Proof of Proposition 1

Proof of Proposition 1. In (Rusmevichientong et al., 2010, Section 2.1), it was shown that the optimal revenue is

$$\theta^* = \max\{\theta \in \mathbb{R} : \max_{S \subseteq [N]:|S| \leq K} \sum_{i \in S} v_i(r_i - \theta) \geq \theta\}.$$

Let $S = \arg\max_{S\subseteq[N]:|S|\le K}\{\sum_{i\in S} v_i(r_i-\theta^*)\}$. By the above equation, we have

$$\sum_{i\in S} u_i = \sum_{i\in S} v_i(r_i-\theta^*) \ge \theta^*.$$

Next we show the above "$\ge$" is actually "$=$". Suppose instead, it is "$>$", then we have

$$\begin{aligned}\sum_{i\in S} v_i(r_i-\theta^*) &> \theta^*,\\ \sum_{i\in S} v_i r_i &> (1+\sum_{i\in S} v_i)\theta^*,\\ \frac{\sum_{i\in S} v_i r_i}{1+\sum_{i\in S} v_i} &> \theta^*,\end{aligned}$$

which implies that $R(S,\boldsymbol{v}) > \theta^*$ and contradicts to that θ^* is the optimal revenue. As a result, we have $\theta^* = \sum_{i\in S} u_i$. Note that when "$=$" holds, by repeating the above argument, we have $R(S,\boldsymbol{v}) = \theta^*$ and thus $S = S^*$ by Assumption 1. Therefore,

$$S^* = \operatorname*{arg\,max}_{S\subseteq[N]:|S|\le K}\{\sum_{i\in S} u_i\}.$$

It is clear that

$$\operatorname*{arg\,max}_{S\subseteq[N]:|S|\le K}\{\sum_{i\in S} u_i\} = \mathcal{F}([N],K,\boldsymbol{u}),$$

because to maximize the sum of scores under the capacity constraint, it suffices to pick all items with positive and top-K scores. □

B.2 Proof of Lemma 3.1

We prove Lemma 3.1 to show the sample complexity guarantee of Algorithm 1. We first reveal the relation between the gap of advantage score and the suboptimality gap of each item.

Lemma B.1 (Relation between Δ_i and u_i)**.** *For items $i,j\in[N]$, we have the following statements.*

(a) If $i\in S^*, j\notin S^*$, then $\Delta_i \le u_i-u_j$. In addition, $\Delta_i\le u_i$.
(b) If $i\notin S^*, j\in S^*$, then $\Delta_i\le u_j-u_i$. If in addition $|S^*|<K$, then $\Delta_i\le -u_i$.

Proof. For (a), let $S=(S^*\setminus\{i\})\cup\{j\}$. Note that $\Delta_i \le R(S^*,\boldsymbol{v})-R(S,\boldsymbol{v})$, so by Lemma B.2, we have

$$\Delta_i \le (1+\sum_{l\in S} v_l)(R(S^*,\boldsymbol{v})-R(S,\boldsymbol{v})) = u_i-u_j.$$

Let $S=S^*\setminus\{i\}$ and repeat the previous argument, we have $\Delta_i\le u_i$. For (b), let $S=(S^*\setminus\{j\})\cup\{i\}$. Similarly, we have

$$\Delta_i \le (1+\sum_{l\in S} v_l)(R(S^*,\boldsymbol{v})-R(S,\boldsymbol{v})) = u_j-u_i.$$

When $|S^*|<K$, we let $S=S^*\setminus\{j\}$ and repeat the previous argument to obtain $\Delta_i\le -u_i$. □

Lemma B.2 (Revenue Comparison Lemma)**.** *Let $S\subseteq[N]$ be an assortment. Then we have $(1+\sum_{i\in S} v_i)(\theta^*-R(S,\boldsymbol{v})) = \sum_{i\in S^*\setminus S} u_i - \sum_{i\in S\setminus S^*} u_i$.*

Proof. We have

$$\begin{aligned}(1+\sum_{i\in S} v_i)(\theta^*-R(S,\boldsymbol{v})) &= (1+\sum_{i\in S} v_i)\theta^* - \sum_{i\in S} v_i r_i\\ &= \theta^* - \sum_{i\in S} v_i(r_i-\theta^*)\end{aligned}$$

$$\begin{aligned}(\text{Proposition 1}) &= \sum_{i\in S^*} u_i - \sum_{i\in S} u_i \\ &= \sum_{i\in S^*\setminus S} u_i - \sum_{i\in S\setminus S^*} u_i. \end{aligned}$$ □

Next we prove Lemma 3.1 in twofold. First, we analyze the guarantees of the accept-reject stage at Lines 6-10 in Algorithm 1.

Lemma B.3 (Accept-Reject)**.** *In phase k, before Line 6, we have $A^{(k-1)} \subseteq S^* \subseteq A^{(k-1)} \sqcup B^{(k-1)}$ and $u_i \in [\check{\xi}_i, \hat{\xi}_i], \hat{\xi}_i - \check{\xi}_i \le \frac{\epsilon_k}{2}$ for $i \in B^{(k-1)}$. Then after Line 10, we have $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$ and $B^{(k)} \subseteq \{i \in [N] : \Delta_i \le \epsilon_k\}$.*

To facilitate readability, we divide the lemma into two lemmas and prove them separately.

Proof of Lemma B.3. We combine Lemmas B.4 and B.5. □

Lemma B.4. *Under the context of Lemma B.3, we have $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$.*

Lemma B.5. *Under the context of Lemma B.3, we have $B^{(k)} \subseteq \{i \in [N] : \Delta_i \le \epsilon_k\}$.*

We prove these two lemmas. Let $B^1_{\text{acc}} = \{b \in B^{(k-1)} : \check{\xi}_b > 0\}, B^1_{\text{rej}} = \{b \in B^{(k-1)} : \hat{\xi}_b < 0\}$. If $|B^{(k-1)}| > M$, we let $B^2_{\text{acc}} = \{b \in B^{(k-1)} : \check{\xi}_b > \beta\}, B^2_{\text{rej}} = \{b \in B^{(k-1)} : \hat{\xi}_b < \alpha\}$, where M, α, β are defined in Algorithm 1.

Proof of Lemma B.4. We recall that the notion "$\sqcup$" requires $A^{(k)} \cap B^{(k)} = \emptyset$, so we show this first. This follows directly from $A^{(k)} = A^{(k-1)} \cup B_{\text{acc}}, B^{(k)} \subseteq B^{(k-1)} \setminus B_{\text{acc}}$, and $A^{(k-1)} \cap B^{(k-1)} = \emptyset$.

Next, we show $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$. It suffices to show $i \in S^* \setminus A^{(k-1)}$ for $i \in B_{\text{acc}}$ and $i \notin S^* \setminus A^{(k-1)}$ for $i \in B_{\text{rej}}$. Suppose $|B^{(k-1)}| \le M$. Since $u_b \ge \check{\xi}_b$, we have

$$B^1_{\text{acc}} \subseteq \{b \in B^{(k-1)} : u_b > 0\} \subseteq \mathcal{F}(B^{(k-1)}, M, \boldsymbol{u}) = S^* \setminus A^{(k-1)},$$

which implies $A^{(k)} \subseteq S^*$. We have $u_b \le \hat{\xi}_b < 0$ for $b \in B_{\text{rej}}$, which implies $b \notin S^*$ and thus $S^* \subseteq (A^{(k-1)} \sqcup B^{(k-1)}) \setminus B_{\text{rej}} = A^{(k)} \sqcup B^{(k)}$.

Now consider $|B^{(k-1)}| > M$. For every $i \in B_{\text{acc}}$, since $i \in B^1_{\text{acc}}$, we have $u_i > 0$. Since $i \in B^2_{\text{acc}}$, we have $u_i \ge \check{\xi}_i > \beta$. By the definition of β, we know that $\#\{b \in B^{(k-1)} : u_b \ge u_i\} \le M$. Therefore, u_i is positive and top-M. Thus $i \in \mathcal{F}(B^{(k-1)}, M, \boldsymbol{u}) = S^* \setminus A^{(k-1)}$.

For every $i \in B_{\text{rej}}$, if $i \in B^1_{\text{rej}}$, then $u_i \le \hat{\xi}_i < 0$ is negative, thus $i \notin S^*$. Otherwise we have $i \in B^2_{\text{rej}}$. By the definition of α, we have that $u_i \le \hat{\xi}_i < \alpha$ and thus $\#\{b \in B^{(k-1)} : u_b > u_i\} > M$. Therefore, u_i is not top-M. Thus $i \notin S^* \setminus A^{(k-1)}$. □

Proof of Lemma B.5. We show $B^{(k)} \subseteq \{i \in B^{(k-1)} : \Delta_i \le \epsilon_k\}$ by showing that $\Delta_i > \epsilon_k$ implies $i \notin B^{(k)}$. Fix $i \in B^{(k-1)}$ such that $\Delta_i > \epsilon_k$.

1. Suppose $i \in S^*$. We will show that $i \in B_{\text{acc}}$. By Lemma B.1, we have $\Delta_i \le u_i$ and thus

$$\check{\xi}_i \ge \hat{\xi}_i - \frac{\epsilon_k}{2} \ge u_i - \frac{\epsilon_k}{2} \ge \Delta_i - \frac{\epsilon_k}{2} \ge \epsilon_k - \frac{\epsilon_k}{2} = \frac{\epsilon_k}{2} > 0,$$

which implies $i \in B^1_{\text{acc}}$. Note that when $|B^{(k-1)}| \le M$, we have $B_{\text{acc}} = B^1_{\text{acc}}$ and thus we conclude.

When $|B^{(k-1)}| > M$, it remains to show $i \in B^2_{\text{acc}}$. By the definition of β, it suffices to show $\#\{j \in B^{(k-1)} : \check{\xi}_i > \hat{\xi}_j\} \ge |B^{(k-1)}| - M$, which is equivalent to $\#\{j \in B^{(k-1)} : \check{\xi}_i \le \hat{\xi}_j\} \le M$.

For every $j \in B^{(k-1)}$, if $\hat{\xi}_j \ge \check{\xi}_i$, then we have $u_j + \frac{\epsilon_k}{2} \ge \hat{\xi}_j \ge \check{\xi}_i \ge u_i - \frac{\epsilon_k}{2}$. Therefore, $u_j \ge u_i - \epsilon_k \ge u_i - \Delta_i$. In summary, we have $\{j \in B^{(k-1)} : \hat{\xi}_j > \check{\xi}_i\} \subseteq \{j \in B^{(k-1)} : \xi_j \ge \xi_i - \Delta_i\}$. By Lemma B.1, we have $\xi_i - \xi_{j'} \ge \Delta_i$ for every $j' \in [N] \setminus S^*$. Thus $\{j \in B^{(k-1)} : \xi_j \ge \xi_i - \Delta_i\} \subseteq S^*$. Recall that $A^{(k-1)} \subseteq S^* \subseteq A^{(k-1)} \sqcup B^{(k-1)}$. So $\{j \in B^{(k-1)} : \hat{\xi}_j \ge \check{\xi}_i\} \subseteq S^* \setminus A^{(k-1)}$ and thus $\#\{j \in B^{(k)} : \hat{\xi}_j \ge \check{\xi}_i\} \le K - |A^{(k-1)}| = M$, which completes the proof.

2. Suppose $i \notin S^*$. We will show that $i \in B_{\text{rej}}$. Suppose $|S^*| < K$. By Lemma B.1, we have $\Delta_i \le -\xi_i$. Therefore,

$$\hat{\xi}_i \le \xi_i + \frac{\epsilon_k}{2} < -\epsilon_k + \frac{\epsilon_k}{2} < 0,$$

which implies $i \in B^1_{\text{rej}}$.

Now consider $|S^*| = K$. Since $i \notin S^*$ and $i \in B^{(k-1)}$, we must have $|B^{(k-1)}| \ge M + 1 > M$. In the following, we show that $i \in B^2_{\text{rej}}$. By the definition of α, it suffices to show $\#\{j \in B^{(k-1)} : \check{\xi}_j > \hat{\xi}_i\} \ge M$. For every $j \in S^* \cap B^{(k-1)}$, by Lemma B.1, we have $\Delta_i \le u_j - u_i$. Therefore, we have $\check{\xi}_j \ge u_j - \frac{\epsilon_k}{2} \ge \Delta_i + u_i - \frac{\epsilon_k}{2} \ge \Delta_i - \frac{\epsilon_k}{2} + \hat{\xi}_i - \frac{\epsilon_k}{2} > \hat{\xi}_i$, which implies $\{j \in B^{(k-1)} : \check{\xi}_j > \hat{\xi}_i\} \supseteq S^* \cap B^{(k-1)}$ and thus $\#\{j \in B^{(k-1)} : \check{\xi}_j > \hat{\xi}_i\} \ge M$. □

Proof of Lemma 3.1. By a union bound, the probability that EST returns confidence intervals $u_i \in [\check{\xi}_i, \hat{\xi}_i], \hat{\xi}_i - \check{\xi}_i \le \frac{\epsilon_k}{2}$ within $C_{\mathsf{EST}} \cdot \frac{|B^{(k-1)}| \log(N/\delta^{(k)})}{\epsilon_k^2}$ time steps for every phase $k \in \mathbb{N}$ is at least

$$\prod_{k=1}^{\infty}(1 - \delta^{(k)}) = \prod_{k=1}^{\infty}(1 - \frac{\delta}{3k^2}) \ge 1 - \sum_{k=1}^{\infty}\frac{\delta}{3k^2} \ge 1 - \frac{\delta}{3}\frac{\pi^2}{6} \ge 1 - \delta.$$

We condition on the above event. Note that $A^{(0)} \subseteq S^* \subseteq A^{(0)} \sqcup B^{(0)}$. By combining Lemma B.3 with an induction over phases, we can show that $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$ and $B^{(k)} \subseteq \{i \in [N] : \Delta_i \le \epsilon_k\}$ for every phase k. Therefore, when $M = 0$, the algorithm returns the optimal assortment S^*. The sample complexity of SAR-MNL with EST is

$$\begin{aligned}
T &\lesssim \sum_{k=1}^{\infty}|B^{(k-1)}| \cdot \frac{C_{\mathsf{EST}} \log(N/\delta^{(k)})}{\epsilon_k^2} \\
&\lesssim \sum_{k=1}^{\infty}\left(\sum_{i\in[N]}\mathbb{I}\{\Delta_i \ge \epsilon_k\}\right) \cdot \frac{C_{\mathsf{EST}} \log(Nk/\delta)}{\epsilon_k^2} \\
&= \sum_{i\in[N]}\sum_{k=1}^{\infty}\frac{C_{\mathsf{EST}} \log(Nk/\delta)}{\epsilon_k^2} \cdot \mathbb{I}\{\Delta_i \ge \epsilon_k\} \\
&\lesssim \sum_{i\in[N]}\sum_{k=1}^{\lceil \log \Delta_i^{-1} \rceil}\frac{C_{\mathsf{EST}} \log(Nk/\delta)}{\epsilon_k}. \\
&\lesssim C_{\mathsf{EST}} \cdot \sum_{i\in[N]}\frac{\log N + \log\log \Delta_i^{-1} + \log \delta^{-1}}{\Delta_i^2}.
\end{aligned}$$

□

B.3 Proof of Lemma 3.2

Before proving the lemma, we first specify the skipped formulas in Algorithm 2. Let $C_0 = 196, C_2 = 1024, \delta = \frac{\delta_0}{15N}$. We define $\tau = \frac{C_2 C_0 \log(2/\delta)}{\epsilon^2}$ and

$$\check{v}_i = 0 \vee (\bar{v}_i - \sigma(v_i)), \hat{v}_i = 1 \wedge (\bar{v}_i + \sigma(v_i)), \quad \sigma(v_i) = \sqrt{\frac{48\bar{v}_i \log(2/\delta)}{T_i}} + \frac{48\log(2/\delta)}{T_i},$$

where $T_i = K\tau$ is the number of offering. For each item $i \in A \cup B$, we define $\bar{v}_i = \frac{n_i}{T_i}$, where n_i is the total number of time steps with outcome "item i". One may realize that "keep offering until no purchase" is the same as the epoch-based offering in (Agrawal et al., 2019) and that EST-NAIVE uses a simplified version by only offering singletons. We adopt the notions, calling it "epoch" and referring T_i as the number of epochs.

Next we give a proof of the sample complexity guarantee using previous results in (Agrawal et al., 2019). Our proof frequently uses the big-O notations to suppress the constants, whose exact values can be calculated by following the proofs in Appendix D.3.

Proof of Lemma 3.2. 1. We prove that EST-NAIVE returns the confidence intervals $u_i \in [\check{\xi}_i, \hat{\xi}_i]$ with high probability. By Lemma 4.1 in (Agrawal et al., 2019), we have that $v_i \in [\check{v}_i, \hat{v}_i]$ and $\hat{v}_i - \check{v}_i \le \widetilde{O}(\sqrt{v_i/T_i}) = \widetilde{O}(\frac{\epsilon v_i}{K})$ with probability $1 - O(N\delta)$. By Lemma 4.2 in (Agrawal et al., 2019), we find that $\check{\theta} \le \theta^* \le \hat{\theta}$ if $\check{v}_i \le v_i \le \hat{v}_i$ for every $i \in A \cup B$. Furthermore, for the assortment $S = \arg\max_{S \subseteq A\cup B : |S| \le K} R(S, \hat{v})$, we have

$$
\begin{aligned}
R(S,\hat{v}) - R(S,\check{v}) &= \frac{\sum_{i\in S}\hat{v}_i r_i}{1+\sum_{i\in S}\hat{v}_i} - \frac{\sum_{i\in S}\check{v}_i r_i}{1+\sum_{i\in S}\check{v}_i} \\
&\le \frac{\sum_{i\in S}\hat{v}_i r_i}{1+\sum_{i\in S}\check{v}_i} - \frac{\sum_{i\in S}\check{v}_i r_i}{1+\sum_{i\in S}\check{v}_i} \\
&\le \sum_{i\in S}(\hat{v}_i - \check{v}_i) r_i \\
&\le \sum_{i\in S}(\hat{v}_i - \check{v}_i) \\
&\le K O(\frac{\epsilon}{K}) \\
&\le O(\epsilon).
\end{aligned} \tag{7}
$$

Note that $R(S,\hat{v}) = \hat{\theta}$ and $R(S,\check{v}) \le \check{\theta}$, so we conclude that $\hat{\theta} - \check{\theta} \le O(\epsilon)$. Finally, we note that $u_i = v_i(r_i - \theta^*)$ and that $v_i \in [0,1], (r_i - \theta^*) \in [-1,1]$. Therefore, we have $\hat{\xi}_i - \check{\xi}_i \le |\hat{v}_i - \check{v}_i| + |\hat{\theta} - \check{\theta}| \le O(\epsilon)$.

2. We conclude by showing EST-NAIVE achieves $C_{\mathsf{EST}} = O(K^2)$ in Lemma 3.1. When we keep offering a singleton assortment $\{i\}$ until the outcome "no purchase" occurs, it will take us $1 + v_i \le 2$ time steps in expectation. So in expectation, EST-NAIVE uses

$$
K\tau \sum_{i\in A\cup B}(1+v_i) \le 2K|A\cup B|\tau \le 2K^2\tau
$$

time steps. Using the concentration inequalities, we can turn the expectation argument into a high probability one, showing that EST-NAIVE returns in $O(K^2\tau)$ time steps with probability at least $1 - O(\delta)$. Thus we prove that $C_{\mathsf{EST}} = O(K^2)$ for EST-NAIVE. □

Finally, we discuss two questions: why the procedure only offers singletons and why the accuracy needs to be $\frac{\epsilon}{K}$. For the first question, we discuss its optimality under the epoch-based offering framework (Agrawal et al., 2019), which is used by almost all previous MNL-bandit work. Under this framework, the accuracy of our estimation to v_i solely depends on T_i, the number of epochs that offers item i.

Let us consider that all items have $v_i = \Theta(1)$ and compare two offering schemes for an assortment S: (i) offer S for an epoch; (ii) for each item $i \in S$, offer the singleton assortment $\{i\}$ for an epoch. Both offering schemes increase T_i by 1 for every $i \in S$ and thus lead to the same accuracy. Moreover, in expectation, the number of time steps used by the first scheme is $(1 + \sum_{i\in S} v_i)$ and that used by the second scheme is $\sum_{i\in S}(1 + v_i)$. When $v_i = \Theta(1)$, we have $(1 + \sum_{i\in S} v_i) \asymp \sum_{i\in S}(1+v_i)$. As a result, both schemes use a similar number of time steps, so we do not benefit from offering an assortment with size greater than 1, i.e. offering singletons could be enough.

For the second question, we consider that all items have $v_i = \Theta(\frac{1}{K})$. We note that if we need to estimate u_i to a given accuracy ϵ, we need to estimate θ^* to such accuracy: $\hat{\theta} - \check{\theta} \le \epsilon$. We observe that when $v_i = \Theta(\frac{1}{K})$, the step in Eq. (7) is almost *tight*, because

$$
(1 + \sum_{i\in S}\check{v}_i) \asymp (1 + \sum_{i\in S} v_i) \asymp (1 + \sum_{i\in S}\frac{1}{K}) \asymp (1 + |S|\frac{1}{K}) \asymp 1.
$$

To estimate θ^* to the accuracy ϵ, by Eq. (7), we need that

$$
\hat{\theta} - \check{\theta} \le \cdots \le \sum_{i\in S}(\hat{v}_i - \check{v}_i) \le \cdots \le \epsilon. \tag{8}
$$

Since $|S|$ can be $O(K)$, we need to estimate each v_i to the accuracy $\frac{\epsilon}{K}$ in order to achieve Eq. (8), which suggests that estimating to the accuracy $\frac{\epsilon}{K}$ could be necessary.

Note that we explain these two questions under different instances, namely $v_i = \Theta(1)$ and $v_i = \Theta(\frac{1}{K})$, so it is still possibly to design an estimation procedure that adapts to these different instances. Actually, this is what we show in Section 3.3 and Appendix D.

C Proofs for Section 3.2

The following lemma shows that the maximization of the reduced revenue function is monotonic in its parameters and thus we can use Eq. (2) to compute the confidence interval of the optimal revenue.

Lemma C.1 (Monotonicity)**.** *Assume $A \subseteq S^* \subseteq A \sqcup B$ and let $M = \min\{K - |A|, |B|\}$. Suppose $\zeta \in [\check{\zeta}, \hat{\zeta}]$ and $\nu_i \in [\check{\nu}_i, \hat{\nu}_i]$ for every $i \in B$. Let $\check{\theta}, \hat{\theta}, \check{\xi}_i, \hat{\xi}_i$ be those defined in Eqs. (2) (5). Then we have $\theta^* \in [\check{\theta}, \hat{\theta}]$ and $\xi_i \in [\check{\xi}_i, \hat{\xi}_i]$ for $i \in B$.*

Proof. First, we show $\theta^* \in [\check{\theta}, \hat{\theta}]$. We will only show $\theta^* \leq \hat{\theta}$, since the proof of $\theta^* \geq \check{\theta}$ is similar. By Eq. (2), we have

$$\hat{\theta} = \max_{S \subseteq B: |S| \leq M} R(S, \hat{\nu}, \hat{\zeta}) \geq R(S^* \setminus A, \hat{\nu}, \hat{\zeta}).$$

Also we have

$$\theta^* = R(S^* \setminus A, \nu, \zeta) = \frac{\zeta + \sum_{i \in S^* \setminus A} \nu_i r_i}{1 + \sum_{i \in S^* \setminus A} \nu_i},$$
$$(1 + \sum_{i \in S^* \setminus A} \nu_i)\theta^* = \zeta + \sum_{i \in S^* \setminus A} \nu_i r_i,$$
$$\theta^* = \zeta + \sum_{i \in S^* \setminus A} \nu_i (r_i - \theta^*).$$

By Proposition 1, we have $S^* = \mathcal{F}([N], K, \boldsymbol{u})$, so $u_i \geq 0$ for $i \in S^*$, thus $r_i \geq \theta^*$ for $i \in S^*$. Therefore,

$$\hat{\zeta} + \sum_{i \in S^* \setminus A} \hat{\nu}_i (r_i - \theta^*) \geq \zeta + \sum_{i \in S^* \setminus A} \nu_i (r_i - \theta^*) = \theta^*,$$
$$\hat{\zeta} + \sum_{i \in S^* \setminus A} \hat{\nu}_i r_i \geq (1 + \sum_{i \in S^* \setminus A} \hat{\nu}_i)\theta^*,$$
$$\frac{\hat{\zeta} + \sum_{i \in S^* \setminus A} \hat{\nu}_i r_i}{1 + \sum_{i \in S^* \setminus A} \hat{\nu}_i} \geq \theta^*,$$
$$R(S^* \setminus A, \hat{\nu}, \hat{\zeta}) \geq \theta^*.$$

And we conclude that $\hat{\theta} \geq \theta^*$. Second, we show $\xi_i \in [\check{\xi}_i, \hat{\xi}_i]$. Recall that $\xi_i = \nu_i (r_i - \theta^*)$. We conclude by noting that $(r_i - \theta^*) \in [-1, 1]$ and that $\nu_i \in [0, 1]$. □

C.1 Proof of Proposition 2

Proof of Proposition 2. Statement (a) can be proved by noting that $\mathbb{P}(z = r_i) = \frac{v_i}{1 + \sum_{j \in Z} v_j}$. Statement (c) can be proved by noting that $(E_\ell - 1)$ follows a geometric distribution with parameter $p = \frac{1 + \sum_{i \in Z} v_i}{1 + \sum_{i \in Z} v_i + \sum_{i \in S} v_i}$, so it has mean $\mathbb{E}[E_\ell - 1] = \frac{1-p}{p} = \frac{\sum_{i \in S} v_i}{1 + \sum_{i \in Z} v_i} = \sum_{i \in S} \nu_i$.

Now we prove statement (b). When $Z = \emptyset$, it was the same as Corollary A.1 in (Agrawal et al., 2019). We note that $Z = \emptyset$ case implies $Z \neq \emptyset$ case, because the distribution of x_i when we offer the assortment $Z \sqcup S$ under parameter $\boldsymbol{v}$ and stop at outcomes $Z \sqcup \{0\}$ is the same as when we offer S under parameter ν and stops at outcome 0. □

The next lemma bounds the sample complexity when using the generalized epoch-based offering procedure using statement (c) in last proposition.

Lemma C.2 (Sum of Epoch Lengths). *Suppose we independently explore $L \geq \log(1/\delta)$ epochs using Algorithm 3 and the expected length of each epoch $\ell \in [L]$ is $\mathbb{E}[E_\ell] \leq 3$. Let $T = \sum_{\ell=1}^{L} E_\ell$ be the total number of used time steps. With probability at least $1-\delta$, we have $T \leq 8\mathbb{E}[T] \leq 24L$.*

Proof. Note that $\{E_\ell - 1\}_{\ell=1}^{L}$ are independent geometric random variables with mean $\mathbb{E}[E_\ell] \leq 3$. Let $\lambda = 8$. Then $\lambda - 1 - \ln\lambda \geq 3$. Let $\mu = \mathbb{E}[T]$. Since $\mu \geq L$, by Lemma A.5, we have

$$\mathbb{P}(T \geq 8\mu) \leq e^{-p_* \mu(\lambda - 1 - \ln\lambda)} \leq e^{-\frac{1}{3}L(\lambda-1-\ln\lambda)} \leq e^{-L} \leq \delta. \qquad \square$$

C.2 Enhanced Version of Lemma 3.1

We show that if we assume EST returns an estimation of the reduced advantage score ξ_i, we can still obtain a similar sample complexity guarantee as that in Lemma 3.1.

Lemma C.3 (Lemma 3.1 enhanced). *Assume $A^{(k-1)} \subseteq S^*$. Suppose with probability at least $1-\delta^{(k)}$, EST (a) returns in $C_{\mathsf{EST}} \cdot \frac{|B^{(k-1)}|\log(N/\delta^{(k)})}{\epsilon_k^2}$ time steps in phase k, and (b) $\xi_i \in [\check{\xi}_i, \hat{\xi}_i]$ and $\hat{\xi}_i - \check{\xi}_i \leq \frac{\epsilon_k}{2}$ for every $i \in B^{(k-1)}$, where $\xi_i = \frac{u_i}{1+\sum_{j\in A^{(k-1)}} v_j}$ is the reduced score. Then SAR-MNL with EST is δ-PAC with sample complexity $C_{\mathsf{EST}} \cdot O(\sum_{i\in[N]} \frac{\log N + \log\delta^{-1} + \log\log\Delta_i^{-1}}{\Delta_i^2})$.*

Proof. We replace Lemma B.3 with Lemma C.4 in the proof of Lemma 3.1. $\square$

Lemma C.4. *In phase k in Algorithm 1, suppose we have $A^{(k-1)} \subseteq S^* \subseteq A^{(k-1)} \sqcup B^{(k-1)}$, and after invoking EST in phase k, we have $\xi_i \in [\check{\xi}_i, \hat{\xi}_i], \hat{\xi}_i - \check{\xi}_i \leq \frac{\epsilon_k}{2}$ for $i \in B^{(k-1)}$. Then after Line 10 , we have $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$ and $B^{(k)} \subseteq \{i \in [N] : \Delta_i \leq \epsilon_k\}$.*

Proof. Let $Z = A^{(k-1)}$ in Lemma C.5. We replace Lemma B.1 with Lemma C.5 and replace the score u_i with the score $\xi_i = \frac{u_i}{1+\sum_{j\in A^{(k-1)}} v_j}$ in the proof of Lemma B.3 to prove the lemma. $\square$

Lemma C.5 (Relation between Δ_i and ξ_i). *For a set $Z \subseteq S^*$ and an item $i \in [N] \setminus Z$, we define the reduced advantage score $\xi_i = \frac{u_i}{1+\sum_{l\in Z} v_i}$. Then for items $i, j \in [N] \setminus Z$, we have*

(a) If $i \in S^*, j \notin S^*$, then $\Delta_i \leq \xi_i - \xi_j$. In addition, $\Delta_i \leq \xi_i$.
(b) If $i \notin S^*, j \in S^*$, then $\Delta_i \leq \xi_j - \xi_i$. If in addition $|S^*| < K$, then $\Delta_i \leq -\xi_i$.

Proof. For (a), let $S = (S^* \setminus \{i\}) \cup \{j\}$. Note that $\Delta_i \leq R(S^*, \boldsymbol{v}) - R(S, \boldsymbol{v})$, so by Lemma B.2, we have

$$(1 + \sum_{l\in Z} v_l)\Delta_i \leq (1 + \sum_{l\in S} v_l)(R(S^*, \boldsymbol{v}) - R(S, \boldsymbol{v})) = u_i - u_j.$$

Note that $\xi_i = \frac{u_i}{1+\sum_{t\in Z} v_t}$ for $i \notin Z$, so $\Delta_i \leq \xi_i - \xi_j$. Let $S = S^* \setminus \{i\}$ and repeat the previous argument, we have $\Delta_i \leq \xi_i$. For (b), let $S = (S^* \setminus \{j\}) \cup \{i\}$. Similarly, we have

$$(1 + \sum_{l\in Z} v_t)\Delta_i \leq (1 + \sum_{l\in S} v_t)(R(S^*, \boldsymbol{v}) - R(S, \boldsymbol{v})) = u_j - u_i.$$

Thus $\Delta_i \leq \xi_j - \xi_i$. When $|S^*| < K$, we let $S = S^* \setminus \{j\}$ and repeat the previous argument to obtain $\Delta_i \leq -\xi_i$. $\square$

C.3 Estimation Procedure with Generalized Epoch-based Offering

We present an estimation procedure EST-REDUCED (Algorithm 7) to demonstrate the power of the generalized epoch-based offering.

Lemma C.6. *There is a δ-PAC algorithm with sample complexity $\widetilde{O}(\sum_{i=1}^{N} \frac{K}{\Delta_i^2})$ using only techniques in Sections 3.1 and 3.2.*

Algorithm 7: EST-REDUCED$(A, B, \delta_0, \epsilon)$: Estimation of ξ_i for $i \in B$

$C_0 = 196, C_2 = 1024, \delta = \frac{\delta_0}{15N}, \tau = \frac{C_2 C_0 \log(2/\delta)}{\epsilon^2}$, $Z \leftarrow A, n_Z = T_Z = 0, \forall i \in B : n_i = T_i = 0$;
$\forall i \in B$: Explore$(\{i\})$ for $K\tau$ epochs;
Compute $\check{\zeta}, \hat{\zeta}, \check{\nu}_i, \hat{\nu}_i, \check{\theta}, \hat{\theta}, \check{\xi}_i, \hat{\xi}_i$ by Eqs. (3) (4) (2) (5) for $i \in B$, **return** $\{\check{\xi}_i, \hat{\xi}_i\}_{i \in B}$;

Proof. We claim that SAR-MNL with EST-REDUCED can serve as the algorithm in the lemma. The statement (b) in Lemma C.7 shows that EST-REDUCED can serve as the estimation procedure EST in Lemma C.3 and (a) further shows that EST-REDUCED satisfies $C_{\mathsf{EST}} = O(K)$. Thus we conclude by Lemma C.3. □

Lemma C.7 (EST-REDUCED)**.** *Assume $A \subseteq S^* \subseteq A \sqcup B$. With probability $1 - \delta_0$, (a) EST-REDUCED returns in $O(K|B|\tau)$ time steps, where $\tau = O(\frac{\log N/\delta_0}{\epsilon^2})$ as defined in Algorithm 7; (b) $\xi_i = \frac{u_i}{1 + \sum_{j \in A} v_j} \in [\check{\xi}_i, \hat{\xi}_i]$ and $\hat{\xi}_i - \check{\xi}_i \le \epsilon$ for $i \in B$.*

Proof. For (a), we note that the expected epoch length of Explore$(\{i\})$ is

$$\mathbb{E}E_\ell = 1 + v_i \le 3.$$

Whenever $B \neq \emptyset$ is not empty, the procedure EST-REDUCED explores at least $\tau \ge \log(1/\delta)$ epochs, so by Lemma C.2, with probability at least $1 - \delta$, the total number of time steps used by the procedure is

$$T \le 24L \le 24 \cdot K|B|\tau.$$

For (b), we prove it by applying the results in Lemma D.1. Note that we offer each item $i \in B$ for $K\tau \ge M\tau \ge (\frac{1}{4\nu_i} \wedge \frac{M}{2})\tau$ epochs, so we meet the conditions in Lemma D.1, whose conclusion shows that (b) holds with probability at least $1 - 14N\delta$. We apply a union bound to find that (a)(b) hold simultaneously with probability at least $1 - (\delta + 14N\delta) \ge 1 - 15N\delta \ge 1 - \delta_0$. □

D Proofs for Sections 3.3 and 3.4

D.1 Error Analysis for Estimation of Advantage Score

We analyze the error of the estimations of ν_i, ζ when we use the generalized epoch-based offering procedure and how their error propagates to θ, ξ_i. By Proposition 2 and Lemma A.4, we know that the tail bound of ν_i satisfies

$$\hat{\nu}_i - \check{\nu}_i \lesssim \sqrt{\frac{\nu_i \iota}{T_i}} + \frac{\iota}{T_i},$$

where we use $\iota = \text{polylog}(\delta^{-1}, N)$ to denote the polylogarithmic terms and T_i is the number of epochs that item i is offered. The major difference between this tail and the common $\frac{1}{\sqrt{T_i}}$-type tail bound (e.g. Lemma A.1) is the existence of the term $\sqrt{\nu_i}$. We fully exploit this term to show the exploration requirement (i.e. required number of epochs) of each item $i \in B$ in the following lemma.

Lemma D.1 (Exploration Requirement)**.** *For every item $i \in B$, if $T_i \ge T_i'\tau$, where $T_i' = (\frac{1}{4\nu_i} \wedge \frac{M}{2})$ and $\tau = O(\frac{\log(N/\delta_0)}{\epsilon^2})$ is as defined in Algorithm 5, then with probability at least $1 - 14N\delta$, we have $\xi_i \in [\check{\xi}_i, \hat{\xi}_i]$ and $\hat{\xi}_i - \check{\xi}_i \le \epsilon$ for every $i \in B$.*

Our focus is to show $\hat{\xi}_i - \check{\xi}_i \le \epsilon$, which requires us to combine the tail bound with the error propagation. In the following proof, we mainly analyze the tail bound itself and defer the error propagation analysis to Lemma D.2.

Proof of Lemma D.1. For an item $i \in B$, by Lemma A.4, with probability at least $1 - 13\delta$, we have $\nu_i \in [\check{\nu}_i, \hat{\nu}_i]$ and

$$\hat{\nu}_i - \check{\nu}_i \le 2\left(\sqrt{\frac{196\nu_i \log(2/\delta)}{T_i}} + \frac{196 \log(2/\delta)}{T_i}\right)$$

$$
\begin{aligned}
&\le 2\Big(\sqrt{\frac{196\nu_i \log(2/\delta)}{(\frac{1}{4\nu_b} \wedge \frac{M}{2})\tau}} + \frac{196\log(2/\delta)}{(\frac{1}{4\nu_i} \wedge \frac{M}{2})\tau}\Big) \\
&= 2\Big(\sqrt{\frac{196\nu_i \log(2/\delta)}{(\frac{1}{4\nu_i} \wedge \frac{M}{2})\frac{C_2 C_0 \log(2/\delta)}{\epsilon^2}}} + \frac{196\log(2/\delta)}{(\frac{1}{4\nu_i} \wedge \frac{M}{2})\frac{C_2 C_0 \log(2/\delta)}{\epsilon^2}}\Big) \\
&= 2\Big(\sqrt{\frac{\nu_i \epsilon^2}{(\frac{1}{4\nu_i} \wedge \frac{M}{2})C_2}} + \frac{\epsilon^2}{(\frac{1}{4\nu_i} \wedge \frac{M}{2})C_2}\Big) \\
&= 2\Big(\sqrt{\frac{\nu_i(4\nu_i \vee \frac{2}{M})}{C_2}}\epsilon + \frac{4\nu_i \vee \frac{2}{M}}{C_2}\epsilon^2\Big) \\
&\le 2\Big(\sqrt{\frac{\nu_i(\nu_i \vee \frac{1}{M})}{C_2/4}}\epsilon + \frac{\nu_i \vee \frac{1}{M}}{C_2/4}\epsilon^2\Big) \\
&\le 2\Big(\frac{\nu_i \vee \frac{1}{M}}{\sqrt{C_2/4}}\epsilon + \frac{\nu_i \vee \frac{1}{M}}{C_2/4}\epsilon^2\Big) \\
&\le \frac{\nu_i \vee \frac{1}{M}}{\sqrt{C_2/64}}\epsilon. \qquad (9)
\end{aligned}
$$

By Lemma A.1, with probability at least $1-\delta$, we have $\zeta \in [\check{\zeta}, \hat{\zeta}]$ and

$$
\hat{\zeta} - \check{\zeta} \le 2\sqrt{\frac{\log(2/\delta)}{2T_Z}} \le 2\sqrt{\frac{\log(2/\delta)}{2\frac{C_2 C_0 \log(2/\delta)}{\epsilon^2}}} = \frac{\epsilon}{\sqrt{C_2 C_0/2}}. \qquad (10)
$$

By a union bound, we have with probability at least $1-(\delta + 13|B|\delta) \ge 1 - 14N\delta$ that $\nu_i \in [\check{\nu}_i, \hat{\nu}_i], \zeta \in [\check{\zeta}, \hat{\zeta}]$ and Eqs. (9)(10) hold for ν_i and ζ for all $i \in B$. When the event holds, we can use Lemma C.1 to show that $\xi_i \in [\check{\xi}_i, \hat{\xi}_i]$ for all $i \in B$ and use Lemma D.2 with $\epsilon_1 = \frac{\epsilon}{\sqrt{C_2/64}}, \epsilon_3 = \frac{\epsilon}{\sqrt{C_2 C_0/2}}$ to show that $\hat{\xi}_i - \check{\xi}_i \le \frac{4\epsilon}{\sqrt{C_2/64}} \le \epsilon$ for all $i \in B$. □

Lemma D.2 (Error Propagation)**.** *Assume $A \subseteq S^* \subseteq A \sqcup B$ and let $M = \min\{K - |A|, |B|\}$. Suppose we have $0 \le \hat{\nu}_i - \check{\nu}_i \le (\nu_i \vee \frac{1}{M})\epsilon_1$ for every $i \in B$ and $0 \le \hat{\zeta} - \check{\zeta} \le \epsilon_3$. Let $\check{\theta}, \hat{\theta}, \check{\xi}_i, \hat{\xi}_i$ be those defined in Eqs. (2) (5). Then $\hat{\theta} - \check{\theta} \le 2\epsilon_1 + \epsilon_3$ and $\hat{\xi}_i - \check{\xi}_i \le 3\epsilon_1 + \epsilon_3$.*

Proof. Note that $\hat{\nu}_i \ge \check{\nu}_i$, so $\hat{\nu}_i - \check{\nu}_i \le (\hat{\nu}_i \vee \frac{1}{M})\epsilon_1$. Using Lemma D.3, we have $R(S, \hat{\nu}, \hat{\zeta}) - R(S, \check{\nu}, \check{\zeta}) \le 2\epsilon_1 + \epsilon_3$. Note that $\hat{\theta} = R(S, \hat{\nu}, \hat{\zeta})$ and $\check{\theta} \ge R(S, \check{\nu}, \check{\zeta})$, together with Lemma C.1, we prove $\hat{\theta} - \check{\theta} \le 2\epsilon_1 + \epsilon_3$.

For every $i \in B$, we have

$$
\hat{\xi}_i - \check{\xi}_i \le |\hat{\nu}_i - \check{\nu}_i| + |\hat{\theta} - \check{\theta}| \le (\nu_i \vee \frac{1}{M})\epsilon_1 + 2\epsilon_1 + \epsilon_3 \le 3\epsilon_1 + \epsilon_3. \qquad \square
$$

Lemma D.3. *Suppose $|S| \le M$. Given ζ, ζ' such that $0 \le \zeta' \le \zeta \le 1$ and ν_i, ν_i' such that $0 \le \nu_i' \le \nu_i \le 1$ for every $i \in S$. Let $\epsilon_1, \epsilon_3 \in (0, 1]$. Suppose we have $\nu_i - \nu_i' \le (\nu_i \vee \frac{1}{M})\epsilon_1$ and $\zeta - \zeta' \le \epsilon_3$. Then we have $R(S, \nu, \zeta) - R(S, \nu', \zeta') \le 2\epsilon_1 + \epsilon_3$.*

Proof. We have

$$
\begin{aligned}
R(S, \nu, \zeta) - R(S, \nu', \zeta') &= \frac{\zeta + \sum_{i \in S} \nu_i r_i}{1 + \sum_{i \in S} \nu_i} - \frac{\zeta' + \sum_{i \in S} \nu_i' r_i}{1 + \sum_{i \in S} \nu_i'} \\
&\le \frac{(\zeta - \zeta') + \sum_{i \in S}(\nu_i - \nu_i')}{1 + \sum_{i \in S} \nu_i} \\
&\le \epsilon_3 + \frac{\sum_{i \in S}(\nu_i + 1/M)\epsilon_1}{1 + \sum_{i \in S} \nu_i} \\
&\le \epsilon_3 + 2\epsilon_1. \qquad \square
\end{aligned}
$$

D.2 Proof of Lemma 3.3

Proof of Lemma 3.3. For (a), in EST-ROUGH, we independently explore $L = N\tau = 4NK \cdot 196\log(2/\delta) \geq 72\log(2/\delta)$ epochs with expected length $\mathbb{E}[E_\ell] = 1 + v_i \leq 2$. By Lemma C.2, with probability at least $1 - 4\delta$, the sample complexity is bounded by $T \leq 5L \lesssim NK \log \delta^{-1}$. For each $i \in [N]$, by Lemma A.4, with probability at least $1 - 13\delta$, we have $\tilde{v}_i = \hat{\nu}_i \geq v_i$ and

$$\tilde{v}_i - v_i \leq \sqrt{\frac{196 v_i \log(2/\delta)}{\tau}} + \frac{196\log(2/\delta)}{\tau} \leq \sqrt{\frac{v_i}{4K}} + \frac{1}{4K} \leq 2v_i \vee \frac{1}{K}.$$

Using a union bound, (a) holds with probability at least $1 - (13N + 4)\delta \geq 1 - 17N\delta \geq 1 - \delta_0$.

For (b), let $V = 1 + \sum_{i \in Z} v_i$. We have

$$1 + \sum_{i \in Z} \tilde{v}_i \in [1 + \sum_{i \in Z} v_i, 1 + \sum_{i \in Z} 2v_i + \frac{|Z|}{K}] \subseteq [V, 2V].$$

Therefore, we have

$$\frac{\tilde{v}_i}{1 + \sum_{i \in Z} \tilde{v}_i} \in [v_i, 2v_i \vee \frac{1}{K}]/[V, 2V] \subseteq [\frac{v_i}{2V}, \frac{2v_i \vee \frac{1}{K}}{V}] \subseteq [\frac{\nu_i}{2}, 2\nu_i \vee \frac{1}{K}].$$ □

D.3 Proof of Lemma 3.4

Lemma D.4. *At the end of EST-ADAPTIVE, for $b \in B$, we have $T_b \geq (\frac{1}{4\nu_b} \wedge \frac{M}{2})\tau$.*

Proof. Suppose $b \in B_i$. If $i < m$, we have $\tilde{\nu}_b \in (\frac{1}{2d_i}, \frac{1}{d_i}]$. By Lemma 3.3, we have $(2\nu_b \vee \frac{1}{K}) \geq \tilde{\nu}_b \geq \frac{1}{2d_i}$. Therefore, $d_i \geq \frac{1}{2(2\nu_b \vee \frac{1}{K})} = (\frac{1}{4\nu_b} \wedge \frac{K}{2}) \geq (\frac{1}{4\nu_b} \wedge \frac{M}{2})$.

If $i = m$, we have $d_i = M \geq \frac{M}{2} \geq (\frac{1}{4\nu_b} \wedge \frac{M}{2})$. We conclude by $T_b \geq d_i\tau$. □

Lemma D.5. *With probability at least $1 - \delta$, EST-ADAPTIVE uses $T \leq 120|B|\tau$ time steps.*

Proof. Let L be the total number of epochs. Note that for every $B_{i,j}$, the expected epoch length of Explore($B_{i,j}$) is

$$\mathbb{E}E_\ell = 1 + \sum_{b \in B_{i,j}} \nu_b \leq 1 + \sum_{b \in B_{i,j}} 2\tilde{\nu}_b \leq 1 + |B_{i,j}| \cdot 2 \cdot 2^{-i} \leq 1 + d_i \cdot 2 \cdot 2^{-i} \leq 1 + 2 = 3.$$

The total number of epochs is

$$\begin{aligned} L &= \tau \cdot \sum_{i=0}^{m} \sum_{j=1}^{c_i} d_i \\ &\leq \tau \cdot \sum_{i=0}^{m} (\sum_{j=1}^{c_{i-1}} |B_{i,j}| + d_i) \\ &\leq \tau \cdot (\sum_{i=0}^{m} \sum_{j=1}^{c_{i-1}} |B_{i,j}| + \sum_{i=0}^{m} d_i) \\ &\leq \tau \cdot (|B| + 2^{m+1}) \\ &\leq \tau \cdot (|B| + 4M) \\ &\leq \tau \cdot 5|B|. \end{aligned}$$

Assume $B \neq \emptyset$. Then $L \geq \tau \geq \log(1/\delta)$. By Lemma C.2, with probability at least $1 - \delta$, we have $T \leq 24L \leq 120|B|\tau$. □

Proof of Lemma 3.4. By Lemma D.4, we meet the exploration requirement in Lemma D.1. Using a union bound, we find that Lemmas D.1 and D.5 hold simultaneously with probability at least $1 - (\delta + 14N\delta) \geq 1 - \delta_0$. Note that Lemma D.5 implies (a) and Lemma D.1 implies (b). □

D.4 Proofs of Theorems 1 and 2

Proof of Theorem 1. By Lemma 3.3, EST-ROUGH gives a rough estimation of v_i with probability at least $1 - \frac{\delta}{2}$. Given those rough estimations, by Lemmas 3.1 and 3.4, SAR-MNL with EST-ADAPTIVE is $\frac{\delta}{2}$-PAC. So the proposed algorithm returns optimal assortment with probability at least $(1 - \frac{\delta}{2})^2 \geq 1 - \delta$ and thus it is δ-PAC. We conclude by noting that we have $C_{\mathsf{EST}} = O(1)$ in Lemma C.3 for EST-ADAPTIVE. ☐

Proof of Theorem 2. We stop the algorithm provided in the proof of Theorem 1 at the phase k when $\epsilon_{k-1} \leq \frac{\varepsilon}{3}$. Then we return the assortment S corresponding to $\hat{\theta}$. Specifically, we return $S = A^{(k-1)} \sqcup S_0$,

$$S_0 = \operatorname*{arg\,max}_{S_0 \subseteq B^{(k-1)} : |S_0| \leq M} R(S_0, \hat{\nu}, \hat{\zeta}).$$

Following the proof of Lemma C.3, we can show the desired sample complexity bound. Moreover, the returned assortment satisfies

$$\begin{aligned} \theta^* - R(S, \boldsymbol{v}) &\leq \hat{\theta} - R(S, \boldsymbol{v}) \\ &= R(S_0, \hat{\nu}, \hat{\zeta}) - R(S_0, \nu, \zeta) \\ \text{(Lemma D.3)} &\leq 3\epsilon_{k-1} \\ &\leq \varepsilon. \end{aligned}$$

☐

E Proofs for Section 4

Our algorithm is to invoke SAR-MNL with $\delta = \frac{1}{T}$ and the procedure EST-REG. Note that this algorithm could possibly return the optimal assortment S^* before the time horizon T is reached. In this case, we assume our algorithm keeps offering S^* until reaching the time horizon. Note that offering S^* incurs zero regret.

We use $\check{\xi}_i^{(k)}, \hat{\xi}_i^{(k)}$ for $i \in B^{(k-1)}$ to denote the values $\{\check{\xi}_i, \hat{\xi}_i\}_{i \in B^{(k-1)}}$ returned by EST-REG in phase k. We assume $\check{\xi}_i^{(0)} = 0$ and $\hat{\xi}_i^{(0)} = 1$. The following lemma summarizes the important guarantees of SAR-MNL that we need to show the regret bound.

Lemma E.1. *With probability at least $1 - \frac{1}{T}$, throughout the algorithm, we have that $u_i \in [\check{\xi}_i^{(k)}, \hat{\xi}_i^{(k)}]$, $\hat{\xi}_i^{(k)} - \check{\xi}_i^{(k)} \leq \frac{\epsilon_k}{2}$, $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$, and $B^{(k)} \subseteq \{i \in [N] : \Delta_i \leq \epsilon_k\}$ for every phase k.*

Proof. We claim EST-REG satisfies the condition (b) in Lemma 3.1. Then we can follow the proof of Lemma 3.1 to show the that with probability at least $1 - \frac{1}{T}$, we have $u_i \in [\check{\xi}_i^{(k)}, \hat{\xi}_i^{(k)}]$, $\hat{\xi}_i^{(k)} - \check{\xi}_i^{(k)} \leq \frac{\epsilon_k}{2}$, $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$, and $B^{(k)} \subseteq \{i \in [N] : \Delta_i \leq \epsilon_k\}$ throughout the algorithm.

To show EST-REG satisfies (b) in Lemma 3.1, we need to analyze the error of the estimations it returns. Note that EST-REG offers each item i for $T_i \geq K\tau$ epochs, which satisfies the exploration requirement in Lemma D.1. Therefore, it returns $u_i \in [\check{\xi}_i^{(k)}, \hat{\xi}_i^{(k)}]$, $\hat{\xi}_i^{(k)} - \check{\xi}_i^{(k)} \leq \frac{\epsilon_k}{2}$ with the desired probability. Thus it satisfies (b) in Lemma 3.1. ☐

Now we start to analyze the regret. The key observation is that Lemma B.2 enables us to represent the regret of offering an assortment S in terms of the score difference between S and S^*. Specifically, when $Z = \emptyset$, the regret of Explore(S) is $\sum_{i \in S^* \setminus S} u_i - \sum_{i \in S \setminus S^*} u_i$. Therefore, if we know that $A^{(k)} \subseteq S^* \subseteq A^{(k)} \sqcup B^{(k)}$ and we choose a maximum subset $B \subseteq B^{(k)}$ to construct an assortment $S = A \sqcup B$ such that $|S| = K$, then the regret of Explore(S) is bounded by

$$|B|(\max_{i \in B} u_i) - |B^*|(\min_{i \in B^*} u_i) \leq (K - |A|)(\max_{i \in B^{(k)}} u_i - \min_{i \in B^{(k)}} u_i). \tag{11}$$

In the following, Lemma E.2 bounds the right hand side of Eq. (11), based on which Lemma E.3 bounds the regret of EST-REG.

Lemma E.2. *We have $(\max_{i \in B^{(k)}} \hat{\xi}_i^{(k)}) - (\min_{i \in B^{(k)}} \check{\xi}_i^{(k)}) \leq \frac{3}{2}\epsilon_k$ and $(\max_{i \in B^{(k)}} \hat{\xi}_i^{(k)}) \geq 0$.*

Proof. The second statement $(\max_{i\in B^{(k)}}\hat{\xi}_i^{(k)})\geq 0$ follows directly from that $\textsf{EST-REG}$ rejects items with negative scores. Next we show the first statement. We write $\check{\xi}_i=\check{\xi}_i^{(k)},\hat{\xi}_i=\hat{\xi}_i^{(k)}$ and $\epsilon=\frac{\epsilon_k}{2}$. If $|B^{(k-1)}|\leq M$, we have $\check{\xi}_i\leq 0\leq\hat{\xi}_i$ for $i\in B^{(k)}$ by the definitions of $B_{\mathrm{acc}},B_{\mathrm{rej}}$ and that $B_{\mathrm{acc}},B_{\mathrm{rej}}$ are excluded from $B^{(k)}$. We conclude by

$$\hat{\xi}_i-\check{\xi}_j\leq\hat{\xi}_i-\check{\xi}_i+\hat{\xi}_j-\check{\xi}_j\leq 2\epsilon.$$

If $|B^{(k-1)}|>M$, then we have $B_{\mathrm{acc}}=\{b\in B^{(k-1)}:\check{\xi}_b>(0\vee\beta)\}$ and $B_{\mathrm{rej}}=\{b\in B^{(k-1)}:\hat{\xi}_b<(0\vee\alpha)\}$. Therefore, we have $\check{\xi}_i\leq(0\vee\beta)$ and $\hat{\xi}_i\geq(0\vee\alpha)$ for $i\in B^{(k)}$. For each $i,j\in B^{(k)}$, if $\check{\xi}_i\leq 0$, then we have $\hat{\xi}_i-\check{\xi}_j\leq 2\epsilon$ using previous equation. Otherwise, we have

$$\hat{\xi}_i-\check{\xi}_j\leq\beta+\epsilon-\check{\xi}_j\leq\beta+2\epsilon-\hat{\xi}_j\leq 2\epsilon+\beta-(0\vee\alpha).$$

It suffices to show $\beta-(0\vee\alpha)\leq\epsilon$. Assume $\beta\geq 0$. Next we show $\beta-\alpha\leq\epsilon$. Let $\hat{\xi}_{i_1},\ldots,\hat{\xi}_{i_M}$ be the M largest values of $\{\hat{\xi}_i\}_{i\in B^{(k-1)}}$. By the definition of α, we have $\alpha\geq\min_{1\leq j\leq M}\check{\xi}_{i_j}$. Suppose $\alpha\geq\check{\xi}_{i_x}$ for $x\in[M]$. We have $\beta-\alpha\leq\hat{\xi}_{i_x}-\check{\xi}_{i_x}\leq\epsilon$. □

Lemma E.3. *The regret incurred by $\textsf{EST-REG}$ in phase k is* $\mathrm{Reg}^{(k)}\lesssim|B\setminus S^*|\cdot\frac{K\log NT}{\epsilon_k}$.

Proof. We note that $B=B^{(k-1)}$ and $A=A^{(k-1)}$. Let $B^*=S^*\setminus A$. For every $j\in[m]$, let $S_j'=A^{(k-1)}\sqcup B_j'$ and $B_j^*=B^*\cap B_j'$. Note that $|B_j'\setminus B_j^*|=M\geq|B^*\setminus B_j^*|$. Since $Z=\emptyset$, by Proposition 2 and Lemma B.2, the regret incurred by $\textsf{Explore}(S_j')$ is

$$\begin{aligned}(1+\sum_{i\in S_j'}v_i)(R(S^*,\boldsymbol{v})-R(S_j',\boldsymbol{v}))&=\sum_{i\in S^*}u_i-\sum_{i\in S_j'}u_i\\&=\sum_{i\in B^*\setminus B_j^*}u_i-\sum_{i\in B_j'\setminus B_j^*}u_i\\&\leq|B^*\setminus B_j^*|\cdot\max_{i\in B}\{u_i\vee 0\}-|B_j'\setminus B_j^*|\cdot\min_{i\in B}u_i\\\text{(Lemma E.2)}&\leq\frac{3}{2}\epsilon_{k-1}|B_j'\setminus B_j^*|\\&=3\epsilon_k|B_j'\setminus B_j^*|.\end{aligned}$$

Note that for every $b\in B^{(k-1)}\setminus B^*$, there are at most two $j\in[m]$ such that $b\in B_j'\setminus B_j^*$, so we have

$$\sum_{j=1}^m|B_j'\setminus B_j^*|\leq 2|B^{(k-1)}\setminus B^*|.\tag{12}$$

Thus the regret incurred in phase k is

$$\mathrm{Reg}^{(k)}\leq K\tau\sum_{j=1}^m(1+\sum_{i\in S_j'}v_i)(R(S^*,\boldsymbol{v})-R(S_j',\boldsymbol{v}))\leq K\tau\sum_{j=1}^m|B_j'\setminus B_j^*|\leq 6\epsilon_k\tau|B^{(k-1)}\setminus B^*|.$$ □

Proof of Theorem 3. Since that the event specified in Lemma E.1 happens with probability $1-\frac{1}{T}$ and that the regret is bounded by $\mathrm{Reg}_T\leq T$, it suffices we prove the regret bound under the event, which is

$$\begin{aligned}\mathrm{Reg}_T&=\sum_{k=1}^T\mathrm{Reg}^{(\mathrm{k})}\\\text{(Lemma E.3)}&\lesssim\sum_{k=1}^\infty|B^{(k-1)}\setminus S^*|\cdot\frac{K\log(NT)}{\epsilon_k}\\&=\sum_{k=1}^\infty\frac{K\log(NT)}{\epsilon_k}\cdot\sum_{i\in[N]\setminus S^*}\mathbb{I}\{i\in B^{(k-1)}\}\end{aligned}$$

$$\begin{aligned}
(\text{Lemma E.1}) &\leq \sum_{k=1}^{\infty} K\log(NT)\cdot\sum_{i\in[N]\setminus S^*}\frac{\mathbb{I}\{\Delta_i\leq\epsilon_{k-1}\}}{\epsilon_k}\\
&= \sum_{i\in[N]\setminus S^*} K\log(NT)\cdot\sum_{k=1}^{\infty}\frac{\mathbb{I}\{\Delta_i\leq\epsilon_{k-1}\}}{\epsilon_k}\\
&\lesssim \sum_{i\in[N]\setminus S^*}\frac{K\log NT}{\Delta_i}.
\end{aligned}$$

□

Finally, we discuss why we always offer full assortments in EST-REG. Actually, this is utilized by Eq. (12). If we do not offer the full assortments, the right hand side of Eq. (12) could become $2|B^{(k-1)}|$. Thus we could end up with a regret bound that depends on S^*, as we show in Section 4.

F Lower Bounds

We recall the definition of $P_S^{\boldsymbol{v}}$, the probability distribution of assortment S under MNL choice model with preference parameter $\boldsymbol{v}$.

$$P_S^{\boldsymbol{v}}(i) = \begin{cases} \frac{v_i}{v_0+\sum_{j\in S} v_j}, & i\in S\cup\{0\},\\ 0 & \text{otherwise.}\end{cases} \tag{13}$$

We show the following lower bound under the restriction $\Delta_i \lesssim \frac{1}{K}$, which gives us enough freedom to construct a simple MNL-bandit instance to realize it, as in Lemma F.2. Note that our regret upper bound in Theorem 3 only depends on items in $[N]\setminus S^*$, so in our lower bound, we only consider the gap sequence of items in $[N]\setminus S^*$. We highlight that our lower bound is for every K.

Theorem 4. *Suppose an algorithm $\mathcal{A}$ achieves $\mathbb{E}[\mathrm{Reg}_T]\lesssim T^p$ on any MNL-bandit instance for a constant $p\in(0,1)$. For any $N\geq 2, K\leq\frac{N}{2}$, suboptimality gap sequence $\{\Delta_i\}_{i=K+1}^{N}$ such that $\max_i\Delta_i\leq\frac{1}{16K}$, there is a MNL-bandit instance $\mathcal{I}$ that realizes the gap sequence. Moreover, for this instance, we have $S^*=[K]$ and the algorithm incurs regret*

$$\liminf_{T\to\infty}\frac{\mathrm{Reg}_T}{\log T}\gtrsim\sum_{i\in[N]\setminus S^*}\frac{1}{K\Delta_i}.$$

For any assortment $S\subseteq[N]$ with $|S|\leq K$, let $\mathcal{T}_S(T)$ be the number of time steps that S is offered. For any item $i\in[N]$, let $\mathcal{T}_i(T)=\sum_{|S|\leq K: i\in S}\mathcal{T}_S(T)$ be the number of time steps that item i is offered. Next we prove Theorem 4. Our proof is inspired by the proofs of the similar lower bounds in multi-armed bandits (Lattimore and Szepesvári, 2020).

Lemma F.1 (Bretagnolle-Huber inequality)**.** *Let $\mathbb{P},\mathbb{Q}$ be two measures over the same measurable space. Let A be an event. Then*

$$\mathbb{P}(A)+\mathbb{Q}(A^{\complement})\geq\frac{1}{2}\exp(-D_{\mathrm{KL}}(\mathbb{P}\,\|\,\mathbb{Q})),$$

where $D_{\mathrm{KL}}(\cdot\,\|\,\cdot)$ is the Kullback–Leibler divergence between probability measures.

Lemma F.2. *Assume the conditions of Theorem 4. For every $i\in[N]$, we let $r_i=1$ and*

$$v_i=\begin{cases}\frac{1}{K}+\frac{1}{2K(K-1)}, & i<K,\\ \frac{1}{2K}, & i=K,\\ \frac{1}{2K}-\frac{4\Delta_i}{1+2\Delta_i}, & i>K.\end{cases}$$

Then $\mathcal{I}=(N,K,\boldsymbol{r},\boldsymbol{v})$ is a MNL-bandit instance in which Δ_i complies with Definition 1.

Proof. Note that for $K=1$ we have $\max_{i\in[N]}v_i\leq\frac{1}{2K}\leq 1$ and for $K\geq 2$ we have $\max_{i\in[N]}v_i\leq\frac{1}{K}+\frac{1}{2K(K-1)}\leq\frac{1}{2}+\frac{1}{4}\leq 1$, so we always have $\max_{i\in[N]}v_i\leq 1$. Note that by the assumption $\Delta_i\leq\frac{1}{16K}$ we have

$\frac{4\Delta_i}{1+2\Delta_i} = \frac{4}{1/\Delta_i+2} \le \frac{4}{16K+2} \le \frac{1}{4K}$, so we have $\min_{i\in[N]} v_i \ge \frac{1}{4K} > 0$. Since $v_i, r_i \in [0,1]$, we know that $\mathcal{I}$ defines a MNL-bandit instance.

Let $S^* = \{1, 2, \ldots, K\}$ be the optimal assortment in this instance. For every item $i \in [N] \setminus S^*$, let $S_i^* = \arg\max_{|S|\le K: i\in S} R(S, \boldsymbol{v})$ be the best assortment containing i. We next show $\Delta_i = R(S^*, \boldsymbol{v}) - R(S_i^*, \boldsymbol{v})$. By direct computations, we have $S_i^* = \{1, 2, \ldots, K-1, i\}$ for $i \notin S^*$ and $S_i^* = S^*$ for $i \in S^*$. Therefore, we have $R(S^*, \boldsymbol{v}) - R(S_i^*, \boldsymbol{v}) = 0 = \Delta_i$ for $i \in S^*$. Note that $\sum_{i=1}^K v_i = 1$, so $R(S^*, \boldsymbol{v}) = \frac{1}{2}$. For $i \notin S^*$ we have

$$\begin{aligned} R(S^*, \boldsymbol{v}) - R(S_i^*, \boldsymbol{v}) &= \frac{1}{2} - \frac{\sum_{i=1}^K v_i - \frac{4\Delta_i}{1+2\Delta_i}}{1 + \sum_{i=1}^K v_i - \frac{4\Delta_i}{1+2\Delta_i}} \\ &= \frac{1}{2} - \frac{1 - \frac{4\Delta_i}{1+2\Delta_i}}{2 - \frac{4\Delta_i}{1+2\Delta_i}} \\ &= \Delta_i. \end{aligned}$$

□

Lemma F.3. *Under the MNL-bandit instance $\mathcal{I}$ defined in Lemma F.2, we have*

$$\mathrm{Reg}_T \ge \frac{1}{2} \sum_{i\in[N]\setminus S^*} \mathbb{E}[\mathcal{T}_i(T)] \cdot \Delta_i.$$

Proof. Under instance $\mathcal{I}$, for any assortment $S \subseteq [N]$ with $|S| \le K$, let $B = S \setminus S^*$, we have

$$\begin{aligned} \theta^* - R(S, \boldsymbol{v}) &= \frac{1}{2} - \frac{\sum_{i\in S} v_i}{1 + \sum_{i\in S} v_i} \\ &\ge \frac{1}{2} - \frac{1 - \sum_{i\in B} \frac{4\Delta_i}{1+2\Delta_i}}{2 - \sum_{i\in B} \frac{4\Delta_i}{1+2\Delta_i}} \\ &= \frac{\sum_{i\in B} \frac{4\Delta_i}{1+2\Delta_i}}{2(2 - \sum_{i\in B} \frac{4\Delta_i}{1+2\Delta_i})} \\ &\ge \sum_{i\in B} \frac{\Delta_i}{1+2\Delta_i} \\ &\ge \sum_{i\in B} \frac{\Delta_i}{2}, \end{aligned}$$

where in the second-to-third inequality we used $\frac{\alpha}{1+\alpha} \le \frac{\alpha+\beta}{1+\alpha+\beta}$ for $\alpha, \beta > 0$ with $\alpha = \sum_{i\in S} v_i$ and $\beta = \sum_{i\in S^*} v_i - \sum_{i\in S} \max\{v_i, \frac{1}{2K}\}$, and in the fifth-to-last inequality we used $\Delta_i \le \frac{1}{16K} \le \frac{1}{2}$.

Recall that S_t is the assortment offered at time step t. We have

$$\begin{aligned} \mathbb{E}[\mathrm{Reg}_T] &= \sum_{t=1}^T \mathbb{E}[\theta^* - R(S_t, \boldsymbol{v})] \\ &= \sum_{t=1}^T \sum_{|S|\le K} \mathbb{E}[\mathbb{I}\{S_t = S\}](\theta^* - R(S, \boldsymbol{v})) \\ &= \sum_{|S|\le K} (\theta^* - R(S, \boldsymbol{v})) \cdot \left(\sum_{t=1}^T \mathbb{E}[\mathbb{I}\{S_t = S\}] \right) \\ &= \sum_{|S|\le K} \mathbb{E}[\mathcal{T}_S(T)] \cdot (\theta^* - R(S, \boldsymbol{v})) \\ &\ge \sum_{|S|\le K} \mathbb{E}[\mathcal{T}_S(T)] \cdot \sum_{i\in S\setminus S^*} \frac{\Delta_i}{2} \\ &= \sum_{i\in[N]\setminus S^*} \mathbb{E}[\mathcal{T}_i(T)] \cdot \frac{\Delta_i}{2}. \end{aligned}$$

□

Lemma F.4. *Let $S \subseteq [N]$ with $|S| \le K$ be an assortment. Let $\boldsymbol{v}, \boldsymbol{v}'$ be two preference vectors such that $v'_x \ge v_x$ and $v'_i = v_i$ for $i \ne x$. Then*

$$D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) \le \frac{(v'_x - v_x)^2}{2v_x(1 + \sum_{i \in S} v_i)}.$$

Proof. Recall the definition of $P_S^{\boldsymbol{v}}$ in Eq. (13). Let $P = P_S^{\boldsymbol{v}}$ and $Q = P_S^{\boldsymbol{v}'}$. We have

$$\begin{aligned}
D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) &= \sum_{i \in S \cup \{0\}} \frac{v_i}{1 + \sum_{j \in S} v_j} \cdot \log \frac{v_i/(1 + \sum_{j \in S} v_j)}{v'_i/(1 + \sum_{j \in S} v'_j)} \\
&= \sum_{\substack{i \in S \cup \{0\} \\ i \ne x}} \frac{v_i}{1 + \sum_{j \in S} v_j} \cdot \log \frac{1 + \sum_{j \in S} v_j}{1 + \sum_{j \in S} v'_j} + \frac{v_x}{1 + \sum_{j \in S} v_j} \cdot \log \frac{v_\ell/(1 + \sum_{j \in S} v_j)}{v'_x/(1 + \sum_{j \in S} v'_j)} \\
&= \log \frac{1 + \sum_{j \in S} v'_j}{1 + \sum_{j \in S} v_j} + \frac{v_x}{1 + \sum_{j \in S} v_j} \log \frac{v_x}{v'_x}.
\end{aligned}$$

Denote $\delta = v'_x - v_x \ge 0$ and $V = 1 + \sum_{j \in S} v_j$. We have

$$\begin{aligned}
D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) &= \log(1 + \frac{\delta}{1 + \sum_{j \in S} v_j}) - \frac{v_x}{1 + \sum_{j \in S} v_j} \log(1 + \frac{\delta}{v_x}) \\
&\le \frac{\delta}{V} - \frac{v_x}{V}(\frac{\delta}{v_x} - \frac{\delta^2}{2v_x^2}) \\
&\le \frac{\delta^2}{2v_x V},
\end{aligned}$$

where we used Taylor's formula $x - \frac{x^2}{2} \le \log(1 + x) \le x$ in the second-to-third inequality. ☐

Lemma F.5. *Let $\mathcal{I} = (N, K, \boldsymbol{r}, \boldsymbol{v}), \mathcal{I}' = (N, K, \boldsymbol{r}, \boldsymbol{v}')$ be two MNL-bandit instances and $\mathcal{A}$ be an algorithm. Let $\mathbb{P}$ be the probability measure induced by $\mathcal{A}$ and $\mathcal{I}$ and $\mathbb{P}'$ be that by $\mathcal{A}$ and $\mathcal{I}'$. We have*

$$D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{P}') = \sum_{|S| \le K} \mathbb{E}[\mathcal{T}_S(T)] D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}).$$

Proof. The lemma can be proved by following the proof of Lemma 15.1 in (Lattimore and Szepesvári, 2020). ☐

Lemma F.6. *Under the assumptions of Theorem 4 and the MNL-bandit instance defined in Lemma F.2, for algorithm $\mathcal{A}$ and any item $i \in [N] \setminus S^*$, we have*

$$\liminf_{T \to \infty} \frac{\mathbb{E}[\mathcal{T}_i(T)]}{\log T} \ge \frac{1 - p}{32K\Delta_i^2}.$$

Proof. Fix an item $i \in [N] \setminus S^*$. For instance $\mathcal{I}$, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] &= \sum_{t=1}^{T} \mathbb{E}[\theta^* - R(S_t, \boldsymbol{v})]) \\
&\ge \sum_{t=1}^{T} \mathbb{E}[\mathbb{I}\{i \in S_t\} \cdot (\theta^* - R(S_t, \boldsymbol{v}))] \\
&\ge \sum_{t=1}^{T} \mathbb{E}[\mathbb{I}\{i \in S_t\} \cdot \Delta_i] \\
&= \mathbb{E}[\mathcal{T}_i(T)] \cdot \Delta_i. \qquad (14)
\end{aligned}$$

We construct another MNL-bandit instance $\mathcal{I}'$. Let $\epsilon \in (0, \frac{1}{2K})$ be a parameter. We define a preference vector $\boldsymbol{v}'$ such that

$$v'_j = \begin{cases} v_j, & j \neq i, \\ \frac{1}{2K} + \epsilon, & j = i. \end{cases}$$

Then $\mathcal{I}' = (N, K, \boldsymbol{r}, \boldsymbol{v}')$ is an MNL-bandit instance. For any algorithm $\mathcal{A}$, let $\mathbb{P}$ be the probability measure given by $\mathcal{A}$ and $\mathcal{I}$, and $\mathbb{P}'$ be that given by $\mathcal{A}$ and $\mathcal{I}'$. From now on, we use $\mathbb{E}$ to denote the expectation under $\mathbb{P}$, and $\mathbb{E}'$ to denote that under $\mathbb{P}'$.

For instance $\mathcal{I}'$, direct computations give that

$$\max_{|S|\leq K} R(S, \boldsymbol{v}') = R(\{1, \ldots, K-1, i\}, \boldsymbol{v}') = \frac{1+\epsilon}{2+\epsilon},$$
$$\max_{|S|\leq K : i \notin S} R(S, \boldsymbol{v}') = R(\{1, \ldots, K\}, \boldsymbol{v}') = \frac{1}{2}.$$

Thus we have

$$\begin{aligned} \mathbb{E}'[\mathrm{Reg}_T] &= \sum_{t=1}^{T} \mathbb{E}[\max_{|S|\leq K} R(S, \boldsymbol{v}') - R(S_t, \boldsymbol{v})] \\ &\geq \sum_{t=1}^{T} \mathbb{E}'[\mathbb{I}\{i \notin S_t\} \cdot (\max_{|S|\leq K} R(S, \boldsymbol{v}') - R(S_t, \boldsymbol{v}'))] \\ &\geq \sum_{t=1}^{T} \mathbb{E}'[\mathbb{I}\{i \notin S_t\}] \cdot (\frac{1+\epsilon}{2+\epsilon} - \frac{1}{2}) \\ &\geq \sum_{t=1}^{T} \mathbb{E}'[\mathbb{I}\{i \notin S_t\}] \cdot \frac{\epsilon}{4}. \end{aligned} \tag{15}$$

Recall that in the proof of Lemma F.2, we showed $v_i \geq \frac{1}{4K}$. For any assortment S, by Lemma F.4, we have

$$\begin{aligned} D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) &\leq \frac{(\epsilon + \frac{4\Delta_i}{1+2\Delta_i})^2}{2v_i(1 + \sum_{i\in S} v_i)} \\ &\leq \frac{(\epsilon + 4\Delta_i)^2}{2v_i} \\ &\leq 2K(4\Delta_i + \epsilon)^2. \end{aligned}$$

By Lemma F.5, we have

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{P}') &= \sum_{|S|\leq K} \mathbb{E}[\mathcal{T}_S(T)] D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) \\ &= \sum_{|S|\leq K : i \in S} \mathbb{E}[\mathcal{T}_S(T)] D_{\mathrm{KL}}(P_S^{\boldsymbol{v}} \parallel P_S^{\boldsymbol{v}'}) \\ &\leq \sum_{|S|\leq K : i \in S} \mathbb{E}[\mathcal{T}_S(T)] \cdot 2K(4\Delta_i + \epsilon)^2 \\ &= 2K(4\Delta_i + \epsilon)^2 \cdot \mathbb{E}[\mathcal{T}_i(T)]. \end{aligned}$$

Let $A = \{\mathcal{T}_i(T) > \frac{T}{2}\}$ be an event. By Lemma F.1, we have

$$\begin{aligned} \mathbb{P}(A) + \mathbb{P}'(A^{\complement}) &\geq \frac{1}{2} \exp(-D_{\mathrm{KL}}(\mathbb{P} \parallel \mathbb{P}')) \\ &\geq \frac{1}{2} \exp(-2K(4\Delta_i + \epsilon)^2 \cdot \mathbb{E}[\mathcal{T}_i(T)]). \end{aligned}$$

By Markov's inequality, we have $\mathbb{E}[\mathcal{T}_i(T)] \geq \mathbb{P}(A) \cdot \frac{T}{2}$ and $\sum_{t=1}^{T} \mathbb{E}'[\mathbb{I}\{i \notin S_t\}] \geq \mathbb{P}'(A^{\complement}) \cdot \frac{T}{2}$. Together with Eqs. (14) (15), we have

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}_T] + \mathbb{E}'[\mathrm{Reg}_T] &\geq \mathbb{E}[\mathcal{T}_i(T)] \cdot \Delta_i + \sum_{t=1}^{T} \mathbb{E}'[\mathbb{I}\{i \notin S_t\}] \cdot \frac{\epsilon}{4} \\
&\geq \frac{T}{2}(\mathbb{P}(A)\Delta_i + \mathbb{P}'(A^{\complement}) \cdot \frac{\epsilon}{4}) \\
&\geq \frac{T}{2} \min\{\Delta_i, \frac{\epsilon}{4}\}(\mathbb{P}(A) + \mathbb{P}'(A^{\complement})) \\
&\geq \frac{T}{2} \min\{\Delta_i, \frac{\epsilon}{4}\} \exp(-\mathbb{E}[\mathcal{T}_i(T)] \cdot 2K(4\Delta_i + \epsilon)^2)
\end{aligned}$$

Recall that $\mathbb{E}[\mathrm{Reg}_T] + \mathbb{E}'[\mathrm{Reg}_T] \leq 2T^p$ for some $p \in (0, 1)$. As a result, we have

$$\begin{aligned}
\liminf_{T\to\infty} \frac{\mathbb{E}[\mathcal{T}_i(T)]}{\log T} &\geq \frac{1}{2K(4\Delta_i + \epsilon)^2}(1 - p - \limsup_{T\to\infty} \frac{\log(\frac{4}{\min\{\Delta_i, \frac{\epsilon}{4}\}})}{\log T}) \\
&= \frac{1 - p}{2K(4\Delta_i + \epsilon)^2}.
\end{aligned}$$

Let $\epsilon \to 0$, we have

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathcal{T}_i(T)]}{\log T} \geq \frac{1 - p}{32K\Delta_i^2}.$$

□

Proof of Theorem 4. We consider the MNL-bandit instance defined in Lemma F.2. By Lemmas F.3 and F.6, we have

$$\begin{aligned}
\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}_T]}{\log T} &\geq \frac{1}{3} \sum_{i\in[N]\setminus S^*} \liminf_{T\to\infty} \frac{\mathbb{E}[\mathcal{T}_i(T)]}{\log T} \\
&\geq \sum_{i\in[N]\setminus S^*} \frac{1 - p}{96K\Delta_i}.
\end{aligned}$$

□