A Note on Tight Lower Bound for MNL-Bandit Assortment Selection Models

Xi Chen\textsuperscript{1} and Yining Wang\textsuperscript{2}

\textsuperscript{1}Stern School of Business, New York University
\textsuperscript{2}Machine Learning Department, Carnegie Mellon University

September 20, 2017

Abstract

In this note we prove a tight lower bound for the MNL-bandit assortment selection model that matches the upper bound given in (Agrawal et al., 2016) for all parameters, up to logarithmic factors.

1 Introduction

We consider the dynamic MNL-bandit model for assortment selection (Agrawal et al., 2016), where $N$ items are present, each associated with a known revenue parameter $r_i > 0$ and an unknown preference parameter $v_i > 0$. For a total of $T$ epochs, at each epoch $t$ a retailer, based on previous purchasing experiences, selects an assortment $S_t \subseteq [N]$ of size at most $K$ (i.e., $|S_t| \leq K$); the retailer then observes a purchasing outcome $i_t \in S_t \cup \{0\}$ sampled from the following discrete distribution:

$$\Pr[i_t = j] = \frac{v_j}{1 + \sum_{j' \in S_t} v_{j'}}, \quad v_0 = 1,$$

and collects the corresponding revenue $r_{i_t}$ (if $i_t = 0$ then no revenue is collected). The objective is to find a policy $\pi$ that minimizes the worst-case expected regret

$$\text{Reg}_\pi(N, T, K) := \sup_{v, r} \mathbb{E} \left[ \sum_{t=1}^{T} R_v(S_v^*) - R_v(S_t) \right], \quad R_v(S) := \mathbb{E} [r_{i_t} | S] = \frac{\sum_{i \in S} r_i v_i}{1 + \sum_{i \in S} v_i},$$

where $R_v(S)$ is the expected revenue collected on assortment $S$ and $S_v^* := \arg \max_{S \subseteq [N]: |S| \leq K} R_v(S)$ is the optimal assortment of size at most $K$ in hindsight. It was shown in (Agrawal et al., 2016) that a UCB-based policy achieves a regret at most $O(\sqrt{NT \log T} + N \log^3 T)$, and furthermore no policy can achieve a regret smaller than $\Omega(\sqrt{NT/K})$. There is an apparent gap between the upper and lower bounds when $K$ is large.

In this note we close this gap by proving the following result:
Theorem 1. Suppose $K \leq N/4$. There exists an absolute constant $C \geq 10^{-4}$ independent of $N$, $T$ and $K$ such that
\[
\inf_{\pi} \text{Reg}_\pi(N, T, K) \geq C \cdot \sqrt{NT}.
\]

Theorem 1 matches the upper bound $O(\sqrt{NT} \log T + N \log^3 T)$ for all three parameters $N$, $T$ and $K$, except for a logarithmic factor of $T$. The proof technique is similar to the proof of (Bubeck et al., 2012, Theorem 3.5). The major difference is that for the MNL-bandit model with assortment size $K$, a “neighboring” subset $S'$ of size $K-1$ rather than the empty set is considered in the calculation of KL-divergence. This approach reduces an $O(\sqrt{1/K})$ factor in the resulting lower bound, which matches the existing upper bound in (Agrawal et al., 2016) up to poly-logarithmic factors.

2 Proof of Theorem 1

Throughout the proof we set $r_1 = \cdots = r_N = 1$ and $v_1, \cdots, v_N \in \{1/K, (1+\epsilon)/K\}$ for some parameter $\epsilon \in (0, 1/2)$ to be specified later. For any subset $S \subseteq [N]$, we use $\theta_S$ to indicate the parameterization where $v_i = (1 + \epsilon)/K$ if $i \in S$ and $v_i = 1/K$ if $i \notin S$. We use $S_k$ to denote all subsets of $[N]$ of size $K$. Clearly, $|S_k| = \binom{N}{K}$. We use $P_S$ and $E_S$ to denote the law and expectation under the parameterization $\theta_S$.

2.1 The counting argument

We first prove the following lemma that bounds the regret of any assortment selection:

Lemma 1. Fix arbitrary $S_0 \in S_K$ and let $v$ be the parameter associated with $\theta_{S_0}$; that is, $v_i = (1 + \epsilon)/K$ if $i \in S_0$ and $v_i = 1/K$ if $i \in [N] \setminus S_0$. For any $S_t \subseteq [N]$ with $|S_t| = K$, it holds that
\[
\max_{S \in S_k} \{R_v(S)\} - R_v(S_t) \geq \frac{\delta \epsilon}{9},
\]
where $\delta = |S_t \cap S_0|/K$.

Proof. By construction of $v$, it is clear that $\max_{S \in S_k} \{R_v(S)\} = R_v(S_0) = (1 + \epsilon)/(2 + \epsilon)$. On the other hand, $R_v(S_t) = (1 + (1-\delta)\epsilon)/(2 + (1-\delta)\epsilon)$. Subsequently,
\[
\max_{S \in S_k} \{R_v(S)\} - R_v(S_t) = \frac{1 + \epsilon}{2 + \epsilon} - \frac{1 + (1-\delta)\epsilon}{2 + (1-\delta)\epsilon} = \frac{\delta \epsilon}{(2 + \epsilon)(2 + (1-\delta)\epsilon)} \geq \frac{\delta \epsilon}{9},
\]
where the last inequality holds because $0 < \epsilon < 1$. \hfill \square

For each assortment selection $S_t \subseteq [N]$, let $\tilde{S}_t \supseteq S_t$ be an arbitrary subset of size $K$ that contains $S_t$. Define $\tilde{N}_t := \sum_{i=1}^T \mathbb{I}[i \in \tilde{S}_t]$. Using Lemma 1 and the fact that $\{\tilde{S}_t\}_{t=1}^T$ suffers less regret than $\{S_t\}_{t=1}^T$, we have
\[
\max_{S \in S_k} \mathbb{E}_S \left[ \sum_{t=1}^T R_v(S_t^*) - R_v(S_t) \right] \geq \max_{S \in S_k} \mathbb{E}_S \left[ \sum_{t=1}^T R_v(S_t^*) - R_v(\tilde{S}_t) \right].
\]

1If $|S_t| = K$ then $\tilde{S}_t = S_t$. 

2
\begin{equation}
\geq \frac{1}{|S_K|} \sum_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^T R_v(S_t^*) - R_v(\tilde{S}_t) \right] \tag{1}
\end{equation}

\begin{equation}
\geq \frac{1}{|S_K|} \sum_{S \in S_K} \sum_{i \notin S} \mathbb{E}_S [\tilde{N}_i] \cdot \frac{\epsilon}{9K} \tag{2}
\end{equation}

\begin{equation}
= \frac{\epsilon}{9} \left( T - \frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S [\tilde{N}_i] \right). \tag{3}
\end{equation}

Here Eq. (1) holds because the maximum regret is always lower bounded by the average regret (averaging over all parameterization \(\theta_S\) for \(S \in S_K\)), Eq. (2) follows from Lemma 1, and Eq. (3) holds because \(\sum_{i=1}^N \mathbb{E}_S [\tilde{N}_i] = \mathbb{E}_S \left[ \sum_{i=1}^N \tilde{N}_i \right] = TK\) for any \(S \subseteq [N]\). The lower bound proof is then reduced to finding the largest \(\epsilon\) such that the summation term in Eq. (3) is upper bounded by, say, \(cT\) for some constant \(c < 1\).

### 2.2 Pinsker’s inequality

The major challenge of bounding the summation term on the right-hand side of Eq. (3) is the \(\sum_{i \in S} \mathbb{E}_S [\tilde{N}_i]\) term. Ideally, we expect this term to be small (e.g., around \(K/N\) fraction of \(\sum_{i=1}^N \mathbb{E}_S [\tilde{N}_i] = KT\)) because \(S \in S_K\) is of size \(K\). However, a bandit assortment selection algorithm, with knowledge of \(S\), could potentially allocate its assortment selections so that \(\tilde{N}_i\) becomes significantly larger for \(i \in S\) than \(i \notin S\). To overcome such difficulties, we use an analysis similar to the proof of Theorem 3.5 in (Bubeck et al., 2012) to exploit the \(\sum_{i=1}^N \mathbb{E}_S [\tilde{N}_i] = NK\) property and Pinsker’s inequality (Tsybakov, 2009) to bound the discrepancy in expectations under different parameterization.

Let \(S_{K-1}^{(i)} = S_{K-1} \cap \{S \subseteq [N] : i \notin S\}\) be all subsets of size \(K - 1\) that do not include \(i\). Re-arranging summation order we have

\[
\frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S [\tilde{N}_i] = \frac{1}{K} \sum_{i=1}^N \frac{1}{|S_K|} \sum_{S \in S_K, i \in S} \mathbb{E}_S [\tilde{N}_i] = \frac{1}{K} \sum_{i=1}^N \frac{1}{|S_K|} \sum_{S' \in S_{K-1}^{(i)}} \mathbb{E}_{S' \cup \{i\}} [\tilde{N}_i].
\]

Denote \(P = P_{S'}\) and \(Q = P_{S_{K-1}^{(i)}}\). Also note that \(0 \leq \tilde{N}_i \leq T\) almost surely under both \(P\) and \(Q\). Using Pinsker’s inequality we have that

\[
|\mathbb{E}_P [\tilde{N}_i] - \mathbb{E}_Q [\tilde{N}_i]| \leq \sum_{j=0}^T j \cdot |P[\tilde{N}_i = j] - Q[\tilde{N}_i = j]| \\
\leq T \cdot \sum_{j=0}^T |P[\tilde{N}_i = j] - Q[\tilde{N}_i = j]| \\
\leq T \cdot \|P - Q\|_{TV} \leq T \cdot \sqrt{\frac{1}{2} \text{KL}(P \| Q)}.
\]

Here \(\|P - Q\|_{TV} = \sup_A |P(A) - Q(A)|\) and \(\text{KL}(P \| Q) = \int (\log dP/dQ) dP\) are the total variation...
and the Kullback-Leibler (KL) divergence between $P$ and $Q$, respectively. Subsequently,

$$\frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{|S|} \sum_{i \in S} \mathbb{E}_S[\tilde{N}_i] \leq \frac{1}{K} \sum_{i=1}^N \frac{1}{|S_i^{(i)}|} \sum_{S' \in S_{K-1}^{(i)}} \left( \mathbb{E}_{S'}[\tilde{N}_i] + T \cdot \sqrt{\frac{1}{2} \text{KL}(P_{S'}||P_{S' \cup \{i\}})} \right). \quad (4)$$

The first term on the right-hand side of Eq. (4) is easily bounded:

$$\frac{1}{|S_K|} \sum_{i=1}^N \frac{1}{|S_i^{(i)}|} \sum_{S' \in S_{K-1}^{(i)}} \mathbb{E}_{S'}[\tilde{N}_i] = \frac{1}{|S_K|} \sum_{i=1}^N \frac{1}{|S_i^{(i)}|} \sum_{S' \in S_{K-1}^{(i)}} \mathbb{E}_{S'}[\tilde{N}_i]$$

$$\leq \frac{1}{|S_K|} \sum_{S' \in S_{K-1}^{(i)}} \frac{1}{|S_i^{(i)}|} \sum_{i=1}^N \mathbb{E}_{S'}[\tilde{N}_i]$$

$$= \frac{|S_{K-1}^{(i)}|}{K|S_K|} \cdot TK = \frac{(N-1)}{K(N)} \cdot TK = \frac{TK}{N-K+1} \leq \frac{T}{10}.$$

Here the last inequality holds because $K \leq N/4$. Combining all inequalities we have that

$$\max_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^T R_v(S_t^* - R_v(S_t)) \right] \geq \frac{\epsilon}{9} \left( \frac{9T}{10} - \frac{T}{|S_K|} \sum_{S' \in S_{K-1}^{(i)}} \frac{1}{K} \sum_{i \notin S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'}||P_{S' \cup \{i\}})} \right). \quad (5)$$

It remains to bound the KL divergence between two “neighboring” parameterization $\theta_{S'}$ and $\theta_{S' \cup \{i\}}$ for all $S' \in S_{K-1}$ and $i \notin S'$, which we elaborate in the next section.

### 2.3 KL-divergence between assortment selections

Define $N_i := \sum_{t=1}^T [i \in S_t]$. Note that because $S_t \subseteq \tilde{S}_t$, we have $N_i \leq \tilde{N}_i$ almost surely and hence $\sum_{i=1}^N \mathbb{E}_S[N_i] \leq \sum_{i=1}^N \mathbb{E}_S[\tilde{N}_i] = TK$ for all $S \subseteq [N]$.

**Lemma 2.** For any $S' \in S_{K-1}$ and $i \notin S'$, $\text{KL}(P_{S'}||P_{S' \cup \{i\}}) \leq \mathbb{E}_{S'}[N_i] \cdot 63 \epsilon^2 / K$.

Before proving Lemma 2 we first prove an upper bound on KL-divergence between categorical distributions.

**Lemma 3.** Suppose $P$ is a categorical distribution with parameters $p_0, \ldots, p_J$ and $Q$ is a categorical distribution with parameters $q_0, \ldots, q_J$. Suppose $p_j = q_j + \epsilon_j$ for all $j = 0, \ldots, J$. Then

$$\text{KL}(P||Q) \leq \sum_{j=0}^J \frac{\epsilon_j^2}{q_j}.$$

\(^2\text{Meaning that } P(X = j) = p_j \text{ for } j = 0, \ldots, J.\)
Proof. We have that
\[
\text{KL}(P \parallel Q) = \sum_{j=0}^{J} (q_j + \varepsilon_j) \log \frac{q_j + \varepsilon_j}{q_j} \leq \sum_{j=0}^{J} \left( q_j + \varepsilon_j \right) \frac{\varepsilon_j}{q_j} = \sum_{j=0}^{J} \frac{\varepsilon_j^2}{q_j}.
\]
Here (a) holds because \( \log(1 + x) \leq x \) for all \( x > -1 \) and (b) holds because \( \sum_{j=0}^{J} \varepsilon_j = 0 \).

We are now ready to prove Lemma 2.

Proof. It is clear that for any \( S_t \subseteq [N] \) such that \( i \notin S_t \), \( \text{KL}(P_{S'}(\cdot \mid S_t) \parallel P_{S' \cup \{i\}}(\cdot \mid S_t)) = 0 \). Therefore, we shall focus only on those \( S_t \subseteq [N] \) with \( i \in S_t \), which happens for \( \mathbb{E}_{S'}[N_i] \) epochs in expectation. Define \( K' := |S_t| \leq K \) and \( J := |S_t| \cap S'|. \) Re-write the probability of \( i_t = j \) as \( p_j = v_j/(a + Je/K) \) and \( q_j = v_j/(a + (J + 1)e/K) \) under \( P_{S'} \) and \( P_{S' \cup \{i\}} \), respectively, where \( a = 1 + K'/K \in (1, 2] \). We then have that
\[
|p_j - q_j| = \left| \frac{1}{a + Je/K} - \frac{1}{a + (J + 1)e/K} \right| \leq \frac{e}{K};
\]
\[
|p_j - q_j| \leq \frac{1 + e}{K^2} \left| \frac{1}{a + Je/K} - \frac{1}{a + (J + 1)e/K} \right| \leq \frac{2e}{K^2}, \quad \text{if } 1 \leq j \leq N, j \neq i;
\]
\[
|p_j - q_j| \leq \frac{2e}{K^2} + \frac{e}{K} \left| \frac{1}{a + Je/K} + \frac{1}{a + (J + 1)e/K} \right| \leq \frac{2e}{K^2} + \frac{2e}{K} \leq \frac{4e}{K}, \quad \text{if } j = i,
\]
Note that \( q_0 \geq 1/3 \) and \( q_j \geq 1/3K \) for \( j \geq 1 \). Invoking Lemma 3 we have that
\[
\text{KL}(P_{S'}(\cdot \mid S_t) \parallel P_{S' \cup \{i\}}(\cdot \mid S_t)) \leq \frac{3e^2}{K^2} + 3K \cdot \frac{4e^2}{K^4} + 3K \cdot \frac{16e^2}{K^2} \leq \frac{63e^2}{K}.
\]

2.4 Putting everything together

Using Hölder’s inequality, we have that
\[
\frac{T}{|S_K|} \sum_{S' \in S_{K-1}} \frac{1}{K} \sum_{i \in S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}})} \leq \frac{T|S_{K-1}|}{K|S_K|} \max_{S' \in S_{K-1}} \sum_{i \in S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}})} = \max_{S' \in S_{K-1}} \frac{T}{N - K + 1} \sum_{i \in S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}})}.
\]
By Jensen’s inequality and the convexity of the square root, we have
\[
\frac{1}{N - K + 1} \sum_{i \in S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}})} \leq \sqrt{\frac{1}{2(N - K + 1)} \sum_{i \in S'} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}})}.
\]
Invoking Lemma 2, we obtain
\[
\frac{1}{N - K + 1} \sum_{i \notin S'} \text{KL}(P_{S'} \parallel P_{S' \cup \{i\}}) \leq \frac{1}{N - K + 1} \sum_{i \notin S'} \mathbb{E}_{S'}[N_i] \cdot \frac{63\epsilon^2}{K}
\]
\[
\leq \frac{63\epsilon^2}{K(N - K + 1)} \sum_{i=1}^{N} \mathbb{E}_{S'}[N_i]
\]
\[
\leq \frac{126\epsilon^2}{NK} \cdot TK = \frac{126\epsilon^2}{N}.
\]

Subsequently, setting \(\epsilon = 0.01 \cdot \sqrt{N/T}\) the term inside the bracket on the right-hand side of Eq. (5) can be lower bounded by \(T/10\). The overall regret is thus lower bounded by \(\epsilon T/90 \geq 0.0001 \sqrt{NT}\). Theorem 1 is thus proved.

**References**

Agrawal, S., Avadhanula, V., Goyal, V., & Zeevi, A. (2016). A near-optimal exploration-exploitation approach for assortment selection. In *Proceedings of the 2016 ACM Conference on Economics and Computation (EC)*, (pp. 599–600). ACM.

Bubeck, S., Cesa-Bianchi, N., et al. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1), 1–122.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York.