

---

# Efficient and Accurate Top- $K$ Recovery from Choice Data (Supplementary Materials)

---

Duc Nguyen<sup>1</sup>

<sup>1</sup>Department of Computer and Information Science, University of Pennsylvania.

## 1 PROOFS

In this section, we will prove all the theoretical results of the paper. The first Section 1.1, however, details some of the important properties of IID-RUMs that are useful in proving the main theorems and lemmas. Many of these results can be of independent interest such as the *Borda consistency* property which is shared by a very broad class of IID-RUMs. The reader may also skip to Section 1.2 for the proofs of the main results stated earlier in the paper.

### 1.1 A CLOSER LOOK AT IID-RUMS

All IID-RUMs are in a family of choice models known as Fechnerian models Becker et al. [1963].

**Definition 1.** A choice model  $\rho$  is a Fechnerian model if there exists a function  $F : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$  that is increasing in the first argument and decreasing in the second argument such that

$$\rho(i|\{i, j\}) = F(U_i, U_j).$$

It is well known that Fechnerian models satisfy strongly stochastic transitivity (SST). That is, if  $\rho(i|\{i, j\}) > \frac{1}{2}$  then  $\rho(i|\{i, k\}) \geq \rho(j|\{j, k\})$  for all  $k \neq i, j$ . A weaker notion of stochastic transitivity is weak stochastic transitivity (WST): if  $\rho(i|\{i, j\}) > \frac{1}{2}, \rho(j|\{j, k\}) > \frac{1}{2}$  then  $\rho(i|\{i, k\}) > \frac{1}{2}$ . The following lemma establishes that all IID-RUMs are Fechnerian models.

**Lemma 1.1.** All IID-RUMs are Fechnerian models.

*Proof.* The proof is simply showing the existence of the function  $F$  in the definition of Fechnerian models. Consider any two items  $i$  and  $j$

$$\begin{aligned} \rho(i|\{i, j\}) &= \mathbb{P}(X_i > X_j) = \mathbb{P}(U_i + \epsilon_i > U_j + \epsilon_j) \\ &= \mathbb{P}(\epsilon_j - \epsilon_i < U_i - U_j) \\ &= F_{ij}(U_i - U_j). \end{aligned}$$

where  $F_{ij}$  is the CDF of the random variable  $\epsilon_j - \epsilon_i$ . Note that all the  $\epsilon_i$ 's are identically distributed and thus  $F = F_{ij}$  for all pairs  $(i, j)$ . This completes the proof. ■

As the class of SST models are a special subclass of WST models, there exists a universal ordering among the items, induced by pairwise preference. Under IID-RUMs, this ordering is the same as that of the partworth parameters. It's also notable that while IID-RUMs satisfy SST, RUMs with independent but not identically distributed noise (Independent RUMs) don't in general. In fact, Independent RUMs may not even satisfy WST. Next, we introduce a property of random utility models which we term the *order preserving* property.

**Definition 2.** A choice model  $\rho$  satisfies the order preserving property if

$$\rho(i|\{i, j\}) > \frac{1}{2} \Rightarrow \rho(i|S) > \rho(j|S) \forall S : i, j \in S.$$

The following lemma states that a broad class of IID-RUMs which include the MNL and Probit model satisfy the order preserving property.

**Lemma 1.2.** All IID-RUMs whose noise distribution has absolutely continuous density function and support on the real line satisfy the order preserving property.

*Proof.* Consider a pair of item  $i, j$  such that  $U_i > U_j$ . The proof for the case when  $S = \{i, j\}$  is trivial. We immediately have  $\rho(i|\{i, j\}) > \frac{1}{2}$ . Now consider any choice set  $S$  containing both  $i$  and  $j$  where  $|S| \geq 3$ . Let  $f_{\epsilon_i}$  denote the density function of the item-specific noise distribution of item  $i$ . Of course, under IID-RUM,  $f_{\epsilon_i} = f_{\epsilon_j} = f$  for all  $i \neq j$ . We have

$$\begin{aligned} \rho(i|S) &= \mathbb{P}(X_i > X_k, \forall k \in S \setminus \{i\}) \\ &= \mathbb{P}(U_i + \epsilon_i > U_k + \epsilon_k, \forall k \in S \setminus \{i\}) \\ &= \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon_i, \forall k \in S \setminus \{i\}) \\ &= \int_{\epsilon} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon_i, \forall k \in S \setminus \{i\}, \epsilon_i = \epsilon) d\epsilon \\ &= \int_{\epsilon} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon_i, \forall k \in S \setminus \{i\} | \epsilon_i = \epsilon) f_{\epsilon_i}(\epsilon) d\epsilon \\ &= \int_{\epsilon} \left( \prod_{k \in S \setminus \{i\}} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) \right) f(\epsilon) d\epsilon. \end{aligned}$$

In the last equality, we have used the independence assumption of IID-RUMs. Further breaking down the product gives

$$\begin{aligned} &\int_{\epsilon} \left( \prod_{k \in S \setminus \{i\}} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) \right) f(\epsilon) d\epsilon \\ &= \int_{\epsilon} \left( \prod_{k \in S \setminus \{i, j\}} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) \right) \mathbb{P}(\epsilon_j < U_i - U_j + \epsilon) f(\epsilon) d\epsilon. \end{aligned} \tag{1}$$

By our assumption of identicality and that the noise distribution has continuous density with support on the real line,  $\mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) < \mathbb{P}(\epsilon_k < U_j - U_k + \epsilon)$  for all  $k \in S \setminus \{i, j\}$ . On the other hand, for a random variable  $\epsilon'$  distributed identically to the  $\epsilon_i$ 's,

$$\mathbb{P}(\epsilon_j < U_i - U_j + \epsilon) > \mathbb{P}(\epsilon' < U_j - U_i + \epsilon) \quad \forall \epsilon.$$

Putting these two inequalities back into (1) gives

$$\begin{aligned} \rho(i|S) &= \int_{\epsilon} \left( \prod_{k \in S \setminus \{i, j\}} \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) \right) \mathbb{P}(\epsilon_j < U_i - U_j + \epsilon) f(\epsilon) d\epsilon \\ &> \int_{\epsilon} \left( \prod_{k \in S \setminus \{i, j\}} \mathbb{P}(\epsilon_k < U_j - U_k + \epsilon) \right) \mathbb{P}(\epsilon' < U_j - U_i + \epsilon) f(\epsilon) d\epsilon \\ &= \rho(j|S). \end{aligned}$$

This completes the proof. ■

While the class of IID-RUMs whose noise distribution has continuous density function and support on the real line includes many commonly used IID-RUMs such as the MNL and Probit model, there are other IID-RUMs outside this class that also satisfy the order preserving property. An example is IID-Exponential RUM.

**Lemma 1.3.** IID-Exponential RUM satisfies the order preserving property.

*Proof.* Since the pairwise case ( $m = 2$ ) is trivial, we will focus on proving Lemma 1.3 for  $m \geq 3$ . Consider a menu  $S \in \mathcal{C}^{(m)} : i, j \in S$ , let  $S' = S \setminus \{i, j\}$ . For this proof, let us consider a ‘copy’ of  $i$  and call it  $i'$ . That is,  $X_{i'}$  is distributed identically to  $X_i$ . We have

$$\begin{aligned}
\rho(i|S) &= \mathbb{P}(X_i > \max\{\{X_k\}_{k \in S'} \cup \{j\}\}) \\
&= \int_0^\infty \mathbb{P}(X_k < U_i + \epsilon \forall k \in S', X_j < U_i + \epsilon) \cdot f(\epsilon) d\epsilon \\
&= \int_0^\infty \mathbb{P}(X_k < U_i + \epsilon \forall k \in S') \cdot \mathbb{P}(X_j < U_i + \epsilon) \cdot f(\epsilon) d\epsilon \\
&> \int_0^\infty \mathbb{P}(X_k < U_i + \epsilon \forall k \in S') \cdot \mathbb{P}(X_{i'} < U_i + \epsilon) \cdot f(\epsilon) d\epsilon \\
&> \int_0^\infty \mathbb{P}(X_k < U_j + \epsilon \forall k \in S') \cdot \mathbb{P}(X_{i'} < U_j + \epsilon) \cdot f(\epsilon) d\epsilon \\
&= \int_0^\infty \mathbb{P}(X_k < U_j + \epsilon \forall k \in S', X_{i'} < U_j + \epsilon) \cdot f(\epsilon) d\epsilon \\
&= \mathbb{P}(X_j > \max\{\{X_k\}_{k \in S'} \cup \{i'\}\}) \\
&= \rho(j|S).
\end{aligned}$$

The first inequality holds because for any  $\epsilon > U_j - U_i$ ,  $\mathbb{P}(X_j < U_i + \epsilon) > \mathbb{P}(X_{i'} < U_i + \epsilon)$ . The second inequality holds because for any  $\epsilon > \max\{0, \max\{U_k - U_j\}_{k \in S' \cup \{j\}}\}$ ,  $\mathbb{P}(X_k < U_i + \epsilon) > \mathbb{P}(X_k < U_j + \epsilon)$  for all  $k \in S'$  and  $\mathbb{P}(X_{i'} < U_j + \epsilon) > \mathbb{P}(X_{i'} < U_i + \epsilon)$ . This completes the proof. ■

Recall the Borda Consistency property introduced in Section 4.2. One can show that a broad class of IID-RUMs which contains many commonly used RUMs such as the MNL and the Probit model satisfies Borda Consistency, the key property that ensures correctness of Generalized Borda Count.

**Theorem 1.4.** *Any IID-RUM  $\rho$  whose noise distribution has continuous density function and support on the real line satisfies Borda consistency.*

*Proof.* Note that  $|\mathcal{C}_i^{(m)}| = |\mathcal{C}_j^{(j)}| = \binom{n-1}{m-1}$  for any  $i, j$  and  $m$ . The proof for  $m = 2$  is trivial so we will focus on  $m \geq 3$ . By definition,

$$\begin{aligned}
\tau_i^{(m)} &= \frac{1}{|\mathcal{C}_i^{(m)}|} \sum_{S \in \mathcal{C}_i^{(m)}} \rho(i|S) \\
&= \frac{1}{|\mathcal{C}_i^{(m)}|} \left( \sum_{S: |S|=m, i, j \in S} \rho(i|S) + \sum_{S: |S|=m, i \in S, j \notin S} \rho(i|S) \right).
\end{aligned}$$

By the ordering preserving property of IID-RUMs, for any  $S : i, j \in S$ ,  $\rho(i|S) > \rho(j|S)$ . To prove that  $\tau_i^{(m)} > \tau_j^{(m)}$ , we only need to prove that for any  $S' : |S'| = m - 1; i, j \notin S'$ ,

$$\rho(i|S' \cup \{i\}) > \rho(j|S' \cup \{j\}).$$

Let  $f_\epsilon$  denote the density function of the noise distribution, we have

$$\begin{aligned}
\rho(i|S' \cup \{i\}) &= \mathbb{P}(X_i > X_k \forall k \in S') \\
&= \mathbb{P}(U_i + \epsilon_i > U_k + \epsilon_k \forall k \in S') \\
&= \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon_i \forall k \in S') \\
&= \int_{-\infty}^\infty \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon_i, \epsilon_i = \epsilon) \\
&= \int_{-\infty}^\infty \mathbb{P}(\epsilon_k < U_i - U_k + \epsilon) f(\epsilon) d\epsilon \\
&> \int_{-\infty}^\infty \mathbb{P}(\epsilon_k < U_j - U_k + \epsilon) f(\epsilon) d\epsilon \\
&= \rho(j|S' \cup \{j\}).
\end{aligned}$$

This completes the proof. ■

We also have a more general theorem that characterizes the class of IID-RUMs, beyond those whose noise distribution has continuous density function and support on the real line, that satisfy Borda consistency

**Theorem 1.5.** *Consider an IID-RUM  $\rho$ . Fix a menu size  $m$ . If, for any pair of item  $i, j$  where  $U_i > U_j$ ,*

- *There exists a menu  $S \in \mathcal{C}^{(m)} : i, j \in S$  such that  $\rho(i|S) \neq \rho(j|S)$ , or*
- *$m \geq 3$  and there exists a menu  $S' \in \mathcal{C}^{(m-1)} : i, j \notin S'$  such that  $\rho(i|S' \cup \{i\}) \neq \rho(j|S' \cup \{j\})$ ,*

*then  $\rho$  satisfies Borda consistency.*

*Proof.* Consider an IID-RUM  $\rho$  and a fixed menu size  $m$ . For any pair of items  $i, j$  where  $U_i > U_j$ , one can easily check that

$$\rho(i|S) \geq \rho(j|S) \quad \forall S \in \mathcal{C}^{(m)} : i, j \in S$$

and

$$\rho(i|S' \cup \{i\}) \geq \rho(j|S' \cup \{j\}) \quad \forall S' \in \mathcal{C}^{(m-1)} : i, j \notin S'.$$

If there exists a menu  $S \in \mathcal{C}^{(m)} : i, j \in S$  where  $\rho(i|S) \neq \rho(j|S)$ , then  $\rho(i|S) > \rho(j|S)$ . Similarly, if there exists a menu  $S' \in \mathcal{C}^{(m-1)} : i, j \notin S'$  where  $\rho(i|S' \cup \{i\}) \neq \rho(j|S' \cup \{j\})$  then  $\rho(i|S' \cup \{i\}) > \rho(j|S' \cup \{j\})$ . If either of these cases hold, we have

$$\begin{aligned} \tau_i^{(m)} &= \frac{1}{|\mathcal{C}_i^{(m)}|} \sum_{S \in \mathcal{C}_i^{(m)}} \rho(i|S) \\ &= \frac{1}{|\mathcal{C}_i^{(m)}|} \left( \sum_{S:|S|=m, i, j \in S} \rho(i|S) + \sum_{S:|S|=m, i \in S, j \notin S} \rho(i|S) \right) \\ &> \frac{1}{|\mathcal{C}_j^{(m)}|} \left( \sum_{S:|S|=m, i, j \in S} \rho(j|S) + \sum_{S:|S|=m, j \in S, i \notin S} \rho(j|S) \right) = \tau_j^{(m)}. \end{aligned}$$

This completes the proof. ■

## 1.2 SAMPLE COMPLEXITY FOR EXACT TOP- $K$ RECOVERY

In this section we prove all theorems stated in Section 4 as well as the sample complexity for approximate top- $K$  ranking of the choice based Borda count algorithm.

**Lemma 4.2.** *Consider an IID-RUM that satisfies Borda consistency per Definition 1. Assume input choice data with menu size  $m$  is generated according to the sampling model described in Section 4.1. For any two items  $i$  and  $j$  where  $\tau_i^{(m)} > \tau_j^{(m)}$ , the choice-based Borda count algorithm satisfies*

$$\mathbb{P}(\hat{W}_j > \hat{W}_i) \leq \exp \left( \frac{-3pR \binom{n}{m} m (\tau_i^{(m)} - \tau_j^{(m)})^2}{8n(\tau_i^{(m)} + \tau_j^{(m)})} \right).$$

*Proof.* Let  $\mathcal{E}$  denote the event  $\{i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*\}$ . We wish to prove that if  $\tau_i - \tau_j > \Delta_K^{(m)}$ , then with sufficiently large  $p$  and  $R$ ,  $\hat{W}_i$  is smaller than  $\hat{W}_j$  with very small probability. The main external concentration inequality used in this proof is Bernstein's inequality (cf. Theorem 2.8.4 Vershynin [2018]). We start our proof by expanding on the probability of

misranking  $i$  and  $j$ .

$$\begin{aligned}
& \mathbb{P}\left(\hat{W}_j > \hat{W}_i \mid \mathcal{E}\right) \\
&= \mathbb{P}\left(\sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} \underbrace{\mathbf{1}[y_S^{(r)} = j]}_{X_S^{j,r}} - \sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} \underbrace{\mathbf{1}[y_S^{(r)} = i]}_{X_S^{i,r}} > 0 \mid \mathcal{E}\right) \\
&= \mathbb{P}\left(\sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} X_S^{j,r} - \mathbb{E}[X_S^{j,r}] + \mathbb{E}[X_S^{j,r}] - \sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} X_S^{i,r} - \mathbb{E}[X_S^{i,r}] + \mathbb{E}[X_S^{i,r}] > 0 \mid \mathcal{E}\right) \quad (2) \\
&= \mathbb{P}\left(\sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} X_S^{j,r} - \mathbb{E}[X_S^{j,r}] - \sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} X_S^{i,r} - \mathbb{E}[X_S^{i,r}] > \sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} \mathbb{E}[X_S^{i,r}] - \sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} \mathbb{E}[X_S^{j,r}] \mid \mathcal{E}\right) \\
&= \mathbb{P}\left(\sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} X_S^{j,r} - \mathbb{E}[X_S^{j,r}] - \sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} X_S^{i,r} - \mathbb{E}[X_S^{i,r}] > pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})\right).
\end{aligned}$$

In the last equality, we have used the fact that  $\sum_{r=1}^R \sum_{S \in \mathcal{C}_i^{(m)}} \mathbb{E}[X_S^{i,r}] - \sum_{r=1}^R \sum_{S \in \mathcal{C}_j^{(m)}} \mathbb{E}[X_S^{j,r}] = pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)}) - \tau_j^{(m)} \geq pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})$ . Now, we will expand the two terms on the LHS of the inequality as

$$\begin{aligned}
& \mathbb{P}\left(\hat{W}_j > \hat{W}_i \mid \mathcal{E}\right) \\
&= \mathbb{P}\left(\sum_{S \in \mathcal{C}_j^{(m)} : i \notin S} \sum_{r=1}^R [X_S^{j,r} - \mathbb{E}[X_S^{j,r}]] - \sum_{S \in \mathcal{C}_i^{(m)} : j \notin S} \sum_{r=1}^R [X_S^{i,r} - \mathbb{E}[X_S^{i,r}]] \right. \\
&\quad \left. + \sum_{S \in \mathcal{C}^{(m)} : i, j \in S} \sum_{r=1}^R [(X_S^{j,r} - X_S^{i,r}) - (\mathbb{E}[X_S^{j,r}] - \mathbb{E}[X_S^{i,r}])] > pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})\right). \quad (3)
\end{aligned}$$

Note that with the above decomposition, the terms  $\{X_S^{j,r}\}_{S \in \mathcal{C}_j^{(m)} : i \notin S}$ ,  $\{X_S^{i,r}\}_{S \in \mathcal{C}_i^{(m)} : j \notin S}$ ,  $\{X_S^{j,r} - X_S^{i,r}\}_{S \in \mathcal{C}^{(m)} : i, j \in S}$  are all mutually independent. To apply Bernstein's inequality, we need to evaluate (or upper bound) the following.

$$\sum_{S \in \mathcal{C}_j^{(m)} : i \notin S} \sum_{r=1}^R \mathbb{E}[(X_S^{j,r})^2] + \sum_{S \in \mathcal{C}_i^{(m)} : j \notin S} \sum_{r=1}^R \mathbb{E}[(X_S^{i,r})^2] + \sum_{S \in \mathcal{C}^{(m)} : i, j \in S} \sum_{r=1}^R \mathbb{E}[(X_S^{j,r} - X_S^{i,r})^2].$$

We can easily see that for  $S \in \mathcal{C}_j^{(m)} : i \notin S$ ,

$$X_S^{j,r} = \begin{cases} 1 & \text{with probability } p \cdot \rho(j|S) \\ 0 & \text{with probability } 1 - p \cdot \rho(j|S) \end{cases}$$

and for  $S \in \mathcal{C}^{(m)} : i, j \in S$ ,

$$X_S^{j,r} - X_S^{i,r} = \begin{cases} 1 & \text{with probability } p \cdot \rho(j|S) \\ -1 & \text{with probability } p \cdot \rho(i|S) \\ 0 & \text{with probability } 1 - p \cdot \rho(j|S) - p \cdot \rho(i|S) \end{cases}.$$

As such  $\mathbb{E}[(X_S^{j,r})^2] = p \cdot \rho(j|S)$  and  $\mathbb{E}[(X_S^{j,r} - X_S^{i,r})^2] = p \cdot (\rho(j|S) + \rho(i|S))$ . We have

$$\begin{aligned}
& \sum_{S \in \mathcal{C}_j^{(m)} : i \notin S} \sum_{r=1}^R \mathbb{E}[(X_S^{j,r})^2] + \sum_{S \in \mathcal{C}_i^{(m)} : j \notin S} \sum_{r=1}^R \mathbb{E}[(X_S^{i,r})^2] + \sum_{S \in \mathcal{C}^{(m)} : i, j \in S} \sum_{r=1}^R \mathbb{E}[(X_S^{j,r} - X_S^{i,r})^2] \\
&= \sum_{S \in \mathcal{C}_j^{(m)} : i \notin S} \sum_{r=1}^R p \cdot \rho(j|S) + \sum_{S \in \mathcal{C}_i^{(m)} : j \notin S} \sum_{r=1}^R p \cdot \rho(i|S) + \sum_{S \in \mathcal{C}^{(m)} : i, j \in S} \sum_{r=1}^R p \cdot (\rho(j|S) + \rho(i|S)) \\
&= \sum_{S \in \mathcal{C}_j^{(m)}} \sum_{r=1}^R p \cdot \rho(j|S) + \sum_{S \in \mathcal{C}_i^{(m)}} \sum_{r=1}^R p \cdot \rho(i|S) \\
&= pR \binom{n-1}{m-1} \tau_i^{(m)} + pR \binom{n-1}{m-1} \tau_j^{(m)} \\
&= pR \binom{n-1}{m-1} (\tau_i^{(m)} + \tau_j^{(m)}).
\end{aligned}$$

Now, applying Bernstein's inequality to (3) directly yields:

$$\begin{aligned}
& \mathbb{P}(\hat{W}_j > \hat{W}_i | \mathcal{E}) \\
& \leq \exp\left(-\frac{(pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)}))^2}{2(pR \binom{n-1}{m-1} (\tau_i^{(m)} + \tau_j^{(m)})) + \frac{pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})}{3}}\right) \\
& = \exp\left(-\frac{pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})^2}{2(\tau_i^{(m)} + \tau_j^{(m)}) + \frac{2(\tau_i^{(m)} - \tau_j^{(m)})}{3}}\right) \\
& \leq \exp\left(-\frac{pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})^2}{8/3(\tau_i^{(m)} + \tau_j^{(m)})}\right).
\end{aligned}$$

This completes the proof. ■

Next we present the proof of Lemma 4.3.

**Lemma 4.3.** Consider an IID-RUM that satisfies Borda consistency per definition (1). Fix a  $K$ , we have

$$\frac{\tau_K^{(m)} + \tau_{K+1}^{(m)}}{\Delta_K^{(m)2}} = \max_{i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*} \left\{ \frac{\tau_i^{(m)} + \tau_j^{(m)}}{(\tau_i^{(m)} - \tau_j^{(m)})^2} \right\}.$$

*Proof.* Firstly, it is easy to see that, fixing an  $i \in \mathcal{S}_K^*$ , for all  $j \notin \mathcal{S}_K^*$ ,

$$\frac{\tau_i^{(m)} + \tau_{K+1}^{(m)}}{(\tau_i^{(m)} - \tau_{K+1}^{(m)})^2} \geq \frac{\tau_i^{(m)} + \tau_j^{(m)}}{(\tau_i^{(m)} - \tau_j^{(m)})^2}.$$

It remains to prove that for any  $i \in \mathcal{S}_K^*$ ,

$$\frac{\tau_K^{(m)} + \tau_{K+1}^{(m)}}{(\tau_K^{(m)} - \tau_{K+1}^{(m)})^2} \geq \frac{\tau_i^{(m)} + \tau_{K+1}^{(m)}}{(\tau_i^{(m)} - \tau_{K+1}^{(m)})^2}.$$

To declutter the notation, we'll remove the superscript  $m$ . Let  $\Delta_{iK} = \tau_i - \tau_K$  for some  $i \in \mathcal{S}_K^*$ . For any  $i \in \mathcal{S}_K^* \setminus \{K\}$ , we

have

$$\begin{aligned}
& \frac{\tau_K + \tau_{K+1}}{\Delta_K^2} \geq \frac{\tau_i + \tau_{K+1}}{(\tau_i - \tau_{K+1})^2} \\
& \Leftrightarrow \frac{\tau_K + \tau_{K+1}}{\Delta_K^2} \geq \frac{\tau_K + \Delta_{iK} + \tau_{K+1}}{(\Delta_{iK} + \Delta_K)^2} \\
& \Leftrightarrow \frac{\tau_K + \tau_{K+1}}{\Delta_K^2} \geq \frac{\tau_K + \Delta_{iK} + \tau_{K+1}}{\Delta_{iK}^2 + \Delta_K^2 + 2\Delta_{iK}\Delta_K} \\
& \Leftrightarrow \tau_K \Delta_{iK}^2 + \tau_K \Delta_K^2 + 2\tau_K \Delta_{iK} \Delta_K + \tau_{K+1} \Delta_{iK}^2 + \\
& \quad \tau_{K+1} \Delta_K^2 + 2\tau_{K+1} \Delta_{iK} \Delta_K \geq \tau_K \Delta_K^2 + \tau_{K+1} \Delta_K^2 + \Delta_{iK} \Delta_K^2 \\
& \Leftrightarrow \tau_K \Delta_{iK}^2 + 2\tau_K \Delta_{iK} \Delta_K + \tau_{K+1} \Delta_{iK}^2 + 2\tau_{K+1} \Delta_{iK} \Delta_K \geq \Delta_{iK} \Delta_K^2.
\end{aligned}$$

In deriving the last statement, the first two terms on the RHS get canceled out. Note that  $2\tau_K \Delta_{iK} \Delta_K \geq 2\Delta_K \Delta_{iK} \Delta_K \geq \Delta_{iK} \Delta_K^2$ . This completes the proof. ■

**Theorem 4.4.** *Assume the conditions of Lemma 4.2. Given sufficiently large  $p, R$  such that  $pR \binom{n}{m} \geq \frac{8n \log n}{m \Delta_K^{(m)^2}} \cdot (\Delta_K^{(m)} + 2\tau_{K+1}^{(m)})$ , the choice-based Borda count algorithm correctly identifies all of the top  $K$  items with probability at least  $1 - O(\frac{K}{n^2})$ .*

*Proof.* The exponential bound in Lemma 4.2 holds simultaneously for all pairs  $i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*$  given sufficiently large  $p, R$  such that

$$pR \binom{n-1}{m-1} \geq \max_{i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*} \frac{8 \log n}{(\tau_i^{(m)} - \tau_j^{(m)})^2} \cdot (\tau_i^{(m)} + \tau_j^{(m)}).$$

From Lemma (4.3), we have

$$\max_{i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*} \frac{\tau_i^{(m)} + \tau_j^{(m)}}{(\tau_i^{(m)} - \tau_j^{(m)})^2} = \frac{\tau_K^{(m)} + \tau_{K+1}^{(m)}}{\Delta_K^{(m)^2}}.$$

This means that if

$$pR \binom{n-1}{m-1} \geq 8 \log n \cdot \frac{\tau_K^{(m)} + \tau_{K+1}^{(m)}}{\Delta_K^{(m)^2}}$$

then

$$\mathbb{P}(\hat{W}_i < \hat{W}_j) \leq \frac{1}{n^3}.$$

Apply union bound over all pairs  $i, j$  such that  $i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*$ . Then the event

$$\hat{W}_i > \hat{W}_j \quad \forall i \in \mathcal{S}_K^*, j \notin \mathcal{S}_K^*$$

happens with probability at least  $1 - O(\frac{K}{n^2})$ . ■

### 1.3 SAMPLE COMPLEXITY FOR APPROXIMATE TOP- $K$ RECOVERY

In practice, some error in the top  $K$  estimate could be tolerable. This is known as approximate top- $K$  ranking. The metric of interest in approximate top- $K$  ranking is the edit distance between the estimate  $\hat{\mathcal{S}}_K$  and the true top  $K$  items  $\mathcal{S}_K^*$ :

$$D_{01}(\hat{\mathcal{S}}_K, \mathcal{S}_K^*) = K - |\hat{\mathcal{S}}_K \cup \mathcal{S}_K|.$$

One can see that correctly separating the top  $K - h$  items and the bottom  $n - K - h$  items, i.e., unable to identify at most  $h$  of the top  $K$  items, is sufficient to guarantee that the approximate loss is bounded by  $h$ . Considering this, the fundamental quantity that determines the hardness of approximate top- $K$  ranking is the gap between the generalized Borda Score of the  $K - h$  item and that of the  $K + h + 1$  item. For convenience, let us denote such quantity as

$$\Delta_{K,h}^{(m)} = \tau_{K-h-1}^{(m)} - \tau_{K+h+1}^{(m)}.$$

Clearly, this quantity generalizes  $\Delta_K^{(m)}$  as  $\Delta_{K,0}^{(m)} = \Delta_K^{(m)}$ . We are interested in bounding the probability that, given an error threshold  $h$ , Generalized Borda Count fails to output an estimate with loss bounded by  $h$ . Namely,

$$\mathbb{P}(D_{01}(\hat{\mathcal{S}}_K, \mathcal{S}_K^*) \leq h).$$

Building on Lemma 4.2, we obtain the following sample complexity of Generalized Borda Count for approximate top- $K$  ranking.

**Theorem 1.6.** *Assume the conditions of Lemma 4.2. Fix an error threshold  $h$ . Given sufficiently large  $p, R$  such that  $pR \binom{n}{m} \geq \frac{8n \log n}{m \Delta_{K,h}^{(m)}} \cdot (1 + \frac{\tau_{K+h+1}^{(m)}}{\Delta_{K,h}^{(m)}})$ , the choice based Borda count algorithm outputs a set  $\hat{\mathcal{S}}_K$  that satisfies*

$$D_H(\mathcal{S}_K^*, \hat{\mathcal{S}}_K) \leq h$$

with probability at least  $1 - O(\frac{K}{n^2})$ .

*Proof.* Applying Lemma 4.2, we have for every pair  $i, j$  where  $i \in \mathcal{S}_{K-h}^*, j \notin \mathcal{S}_{K+h}^*$ ,

$$\mathbb{P}(\hat{W}_j > \hat{W}_i) \leq \exp\left(-\frac{3pR \binom{n-1}{m-1} (\tau_i^{(m)} - \tau_j^{(m)})^2}{8(\tau_i^{(m)} + \tau_j^{(m)})}\right).$$

Following the same argument as in the proof of Theorem 4.4, one can show that

$$\frac{\tau_{K-h} + \tau_{K+h+1}}{\Delta_{K,h}^2} = \max_{i \in \mathcal{S}_{K-h}^*, j \notin \mathcal{S}_{K+h}^*} \frac{\tau_i^{(m)} + \tau_j^{(m)}}{(\tau_i^{(m)} - \tau_j^{(m)})^2}.$$

The rest of the proof follows by showing that given the condition on  $p, R$  in the theorem statement, the probability that Generalized Borda Count wrongly flips the relative order of any two items  $i \in \mathcal{S}_{K-h}^*, j \notin \mathcal{S}_{K+h}^*$  is upper bounded by  $O(\frac{1}{n^3})$ . Applying union bound over all such pairs of  $i, j$  completes the proof. ■

In short, Generalized Borda Count has the following sample complexity for approximate top- $K$  ranking

$$O\left(\frac{n \log n}{m \Delta_{K,h}^{(m)}} \cdot (1 + \frac{\tau_{K+h+1}^{(m)}}{\Delta_{K,h}^{(m)}})\right).$$

## 1.4 LOWER BOUND ON THE SAMPLE COMPLEXITY OF TOP- $K$ RECOVERY

### 1.4.1 Preliminaries and notations

We first restate a version of Fano's inequality Cover [1999] which will be useful to the construction of our lower bound:

**Lemma 1.7** (Fano's inequality). *Consider a set of  $L$  distributions  $\{\mathbb{P}^1, \dots, \mathbb{P}^L\}$ . Suppose that we observe a random variable (or a set of random variables)  $Y$  that was generated by first picking an index  $A \in \{1, \dots, L\}$  uniformly at random and then  $Y \sim \mathbb{P}^A$ . Fano's inequality states that any hypothesis test  $\phi$  for this problem has an error probability lower bounded as*

$$\mathbb{P}[\phi(Y) \neq A] \geq 1 - \frac{\max_{a,b \in [L], a \neq b} \mathbb{D}_{KL}(\mathbb{P}^a(Y) \parallel \mathbb{P}^b(Y)) + \log 2}{\log L}.$$

To obtain a lower bound in the form of Theorem 5.1, we need to construct an IID-RUM such that any method requires  $\Omega(n \log n)$  examples in order to accurately recover the top  $K$  items. For this purpose, we look for a model within the Multinomial Logit family. It is well known that the MNL model is an instance of IID-RUM where the noise distribution is the standard Gumbel distribution. While the MNL model has a random utility characterization, it is often more convenient to describe the MNL model in terms of weighted probabilities. Specifically, an MNL model over  $n$  items can be parameterized by a set of  $n$  positive real numbers  $w_1, \dots, w_n$  called weights. There is also a well defined relation between the weighted



probabilities representation and the utility partworth representations. Specifically, the resulting choice probability is defined to be:

$$\rho(i|S) = \frac{w_i}{\sum_{k \in S} w_k} = \frac{e^{U_i}}{\sum_{k \in S} e^{U_k}}$$

Now consider a special sub-class of MNL models that is defined by 3 parameters  $(v, \delta, \mathcal{S}_K^*)$  for some  $0 < \delta < v$  and a set  $\mathcal{S}_K^*$  of size  $K$ . The item-specific weights are then defined as follows.

$$w_i = \begin{cases} v + \delta & \text{if } i \in \mathcal{S}_K^* \\ v & \text{otherwise} \end{cases}.$$

That is, all the top items have the same weights and all the bottom items have the same (but smaller) weight. For any  $A \in [K, \dots, n]$  let  $\mathcal{M}^A$  be a special MNL model parametrized by  $(v, \delta, \{1, \dots, K-1, A\})$  as described above. One can see that Fano's inequality can be used to lower bound the error probability of any hypothesis test for the identity of  $A$ . That is, to lower bound the probability that any estimator fails to correctly identify the  $K$ -th item.

Recall our uniform sampling model described in Section 4.1, let us use  $\{Y_S^r\}_{S \in \mathcal{C}^{(m)}, r=1:R}$  to denote the observed choice data where:

$$Y_S^r = \begin{cases} y & \text{for some } y \in S \text{ if menu } S \text{ is offered in round } r \\ 0 & \text{if menu } S \text{ is not offered in round } r \end{cases}.$$

As a short hand, denote  $\rho^a$  as the choice rule corresponding to  $\mathcal{M}^a$  and  $\rho^b$  corresponding to  $\mathcal{M}^b$ . Let  $\mathbb{P}^a(\{Y_S^r\})$  denote the likelihood of choice data  $\{Y_S^r\}$  under choice model  $\mathcal{M}^a$  (define analogously for  $\mathcal{M}^b$ ).

#### 1.4.2 Exact top- $K$ recovery

The following pair of lemma and theorem establish the lower bound for exact top- $K$  recovery via showing an upper bound on the KL divergence between any two models  $\mathcal{M}^a, \mathcal{M}^b$  and then applying Fano's inequality.

**Lemma 1.8.** *Assuming the same sampling model as in Theorem 4.2, we have, for any  $a, b \in [K, \dots, n]$ ,*

$$\mathbb{D}_{KL}(\mathbb{P}^a(\{Y_S^r\}) \parallel \mathbb{P}^b(\{Y_S^r\})) \leq 4pR \binom{n-1}{m-1} \frac{\Delta_K^{(m)2}}{\tau_K^{(m)}}.$$

**Theorem 5.1.** *Consider the sampling model described in Section 4.1. There exists a choice model  $\mathcal{M}^*$  within the class of Multinomial Logit Models (MNLs) such that for  $n \geq 20$ , if  $pR \binom{n}{m} \leq \frac{n \log n}{8} \cdot \frac{\tau_K^{(m)}}{m \Delta_K^{(m)2}}$  then any statistical estimator fails to correctly identify all of the top  $K$  items with probability at least  $\frac{1}{12}$ .*

Once Lemma 1.8 has been established, we can prove Theorem 5.1 as follows. For any  $n \geq 20$ , if  $\Delta_K^{(m)2} \leq \frac{\tau_K^{(m)} \log n}{8pR \binom{n-1}{m-1}}$ , Fano's inequality gives us a lower bound on the error of any estimator for finding the top  $K$  items under  $\mathcal{M}(v, \delta, \mathcal{S}_K^*)$ :

$$\mathbb{P}(\hat{\mathcal{S}}_K \neq \mathcal{S}_K^*) \geq 1 - \frac{0.5 \log n + \log 2}{\log(n - K + 1)} \geq 1 - \frac{0.5 \log n + \log 2}{\log(n/2 + 1)} \geq 0.086.$$

The rest of the proof of Theorem 5.1 involves simple algebra, noting that  $\binom{n}{m} = \binom{n-1}{m-1} \cdot \frac{n}{m}$ . We now move on to the proof of Lemma 1.8.

*Proof.* (Proof of Lemma 1.8). Let  $\hat{\mathcal{C}}^{(m,r)}$  denote the set of menus of size  $m$  being picked in round  $r$  under the sampling

model described in Section 4.1. Let us first decompose  $\mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y_S^r\})\|\mathbb{P}^b(\{Y_S^r\}))$  into a more manageable form.

$$\begin{aligned}
& \mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y_S^r\})\|\mathbb{P}^b(\{Y_S^r\})) \\
& \text{[thanks to mutual independence among } \{Y_S^r\}] \\
& = \sum_{r=1}^R \sum_{S \in \mathcal{C}^{(m)}} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^r)\|\mathbb{P}^b(Y_S^r)) \\
& = \sum_{r=1}^R \sum_{S \in \mathcal{C}^{(m)}} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^r|S \in \hat{\mathcal{C}}_r^{(m,r)})\mathbb{P}^a(S \in \hat{\mathcal{C}}^{(m)})\|\mathbb{P}^b(Y_S^r|S \in \hat{\mathcal{C}}^{(m,r)})\mathbb{P}^b(S \in \hat{\mathcal{C}}^{(m)})) \\
& = \sum_{r=1}^R \sum_{S \in \mathcal{C}^{(m)}} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^r|S \in \hat{\mathcal{C}}^{(m,r)})\mathbb{P}^a(S \in \hat{\mathcal{C}}^{(m,r)})\|\mathbb{P}^b(Y_S^r|S \in \hat{\mathcal{C}}^{(m,r)})\mathbb{P}^b(S \in \hat{\mathcal{C}}^{(m,r)})) \\
& \text{[Using the fact that } \mathbb{P}^a(S \in \hat{\mathcal{C}}^{(m,r)}) = \mathbb{P}^b(S \in \hat{\mathcal{C}}^{(m,r)}) = p] \\
& = pR \sum_{S \in \mathcal{C}^{(m)}} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})\|\mathbb{P}^b(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})).
\end{aligned}$$

To handle the above sum, partition  $\mathcal{C}^{(m)}$  into 4 subsets  $\{S \in \mathcal{C}^{(m)} : a, b \notin S\}$ ,  $\{S \in \mathcal{C}^{(m)} : a \in S, b \notin S\}$ ,  $\{S \in \mathcal{C}^{(m)} : a \notin S, b \in S\}$ ,  $\{S \in \mathcal{C}^{(m)} : a, b \in S\}$ . It is easy to see that for  $S : a, b \notin S$ , the corresponding KL divergence terms are all 0. Hence, we only have to work with other 3 subsets. Before going into the details, let us define some shorthand notations: let  $w^a(S)$  denote the sum of the weights of the items in menu  $S$  under  $\mathcal{M}^a$  and  $S + a$  denote set union.

**Subset 1:**  $\{S \in \mathcal{C}^{(m)} : a, b \in S\}$

$$\begin{aligned}
& \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})\|\mathbb{P}^b(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})) \\
& \text{[Note that } \rho^a(x|S) = \rho^b(x|S) \quad \forall x \neq a, b \text{ and } w^a(S) = w^b(S) \text{ for all } S : a, b \in S] \\
& = \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \rho^a(a|S) \log \frac{\rho^a(a|S)}{\rho^b(a|S)} + \rho^a(b|S) \log \frac{\rho^a(b|S)}{\rho^b(b|S)} \\
& = \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{v + \delta}{w^a(S)} \log \frac{v + \delta}{v} + \frac{v}{w^a(S)} \log \frac{v}{v + \delta} \\
& = \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{\delta}{w^a(S)} \log \frac{v + \delta}{v}.
\end{aligned}$$

**Subset 2:**  $\{S \in \mathcal{C}^{(m)} : a \in S, b \notin S\}$ . Firstly, it is easy to verify the following useful identities:  $w^a(S' + a) = w^b(S' + b)$ ,  $w^a(S' + b) = w^b(S' + a) \forall S' : a, b \notin S'$ . Furthermore,  $w^a(x) = w^b(x) \forall x \neq a, b$ . Now we have

$$\begin{aligned}
& \sum_{S \in \mathcal{C}^{(m)}: a \in S, b \notin S} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})\|\mathbb{P}^b(Y_S^1|S \in \hat{\mathcal{C}}^{(m,1)})) \\
& = \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \rho^a(x|S' + a) \log \frac{\rho^a(x|S' + a)}{\rho^b(x|S' + a)} \right) + \rho^a(a|S' + a) \log \frac{\rho^a(a|S' + a)}{\rho^b(a|S' + a)} \right] \\
& = \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \frac{w^a(x)}{w^a(S' + a)} \log \frac{w^b(S' + a)}{w^a(S' + a)} \right) + \frac{v + \delta}{w^a(S' + a)} \log \frac{v + \delta}{v} \cdot \frac{w^b(S' + a)}{w^a(S' + a)} \right] \\
& = \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \frac{w^a(x)}{w^a(S' + a)} \log \frac{w^a(S' + b)}{w^a(S' + a)} \right) + \frac{v + \delta}{w^a(S' + a)} \log \frac{v + \delta}{v} \cdot \frac{w^a(S' + b)}{w^a(S' + a)} \right].
\end{aligned}$$

**Subset 3:**  $\{S \in \mathcal{C}^{(m)} : a \in S, b \notin S\}$ . Following a very similar procedure to the 2nd subset, we obtain:

$$\begin{aligned} & \sum_{S \in \mathcal{C}^{(m)}: a \notin S, b \in S} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)}) \| \mathbb{P}^b(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)})) \\ &= \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \frac{w^a(x)}{w^a(S'+b)} \log \frac{w^a(S'+a)}{w^a(S'+b)} \right) + \frac{v}{w^a(S'+b)} \log \frac{v}{v+\delta} \cdot \frac{w^a(S'+a)}{w^a(S'+b)} \right]. \end{aligned}$$

Focusing on the terms from the 2nd and 3rd subset and grouping them in an intelligent way gives us

$$\begin{aligned} & \sum_{S \in \mathcal{C}^{(m)}: a \notin S, b \in S} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)}) \| \mathbb{P}^b(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)})) \\ &+ \sum_{S \in \mathcal{C}^{(m)}: a \in S, b \notin S} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)}) \| \mathbb{P}^b(Y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)})) \\ &= \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \frac{w^a(x)}{w^a(S'+a)} \log \frac{w^a(S'+b)}{w^a(S'+a)} \right) + \frac{v+\delta}{w^a(S'+a)} \log \frac{v+\delta}{v} \cdot \frac{w^a(S'+b)}{w^a(S'+a)} \right] \\ &+ \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \sum_{x \in S'} \frac{w^a(x)}{w^a(S'+b)} \log \frac{w^a(S'+a)}{w^a(S'+b)} \right) + \frac{v}{w^a(S'+b)} \log \frac{v}{v+\delta} \cdot \frac{w^a(S'+a)}{w^a(S'+b)} \right] \\ &= \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \frac{v+\delta}{w^a(S'+a)} - \frac{v}{w^a(S'+b)} \right) \log \frac{v+\delta}{v} \right] \\ &+ \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \log \frac{w^a(S'+a)}{w^a(S'+b)} \cdot \left[ \sum_{x \in S'} \left( \frac{w^a(x)}{w^a(S'+b)} - \frac{w^a(x)}{w^a(S'+a)} \right) + \frac{v}{w^a(S'+b)} - \frac{v+\delta}{w^a(S'+a)} \right] \\ &= \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \left( \frac{v+\delta}{w^a(S'+a)} - \frac{v}{w^a(S'+b)} \right) \log \frac{v+\delta}{v} \right]. \end{aligned}$$

The last equality comes from recognizing that  $\sum_{x \in S'} \frac{w^a(x)}{w^a(S'+b)} + \frac{v}{w^a(S'+b)} = \frac{w^a(S'+b)}{w^a(S'+b)} = 1$  and similarly  $\sum_{x \in S'} \frac{w^a(x)}{w^a(S'+a)} + \frac{v+\delta}{w^a(S'+a)} = \frac{w^a(S'+a)}{w^a(S'+a)} = 1$  so they cancel out and the second term becomes 0. We now have the following much more compact identity.

$$\begin{aligned} & \mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y_S^r\}) \| \mathbb{P}^b(\{Y_S^r\})) \\ &= pR \cdot \log \frac{v+\delta}{v} \cdot \left( \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{\delta}{w^a(S)} + \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \frac{v+\delta}{w^a(S'+a)} - \frac{v}{w^a(S'+b)} \right] \right). \end{aligned}$$

To make a connection between this and  $\Delta_K^{(m)}$ , recognize that under  $\mathcal{M}^a$ ,  $\Delta_K^{(m)} = \tau_a^{(m)} - \tau_b^{(m)}$ . We therefore have

$$\begin{aligned} \Delta_K^{(m)} &= \tau_a^{(m)} - \tau_b^{(m)} \\ &= \frac{1}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}^{(m)}: a \in S} \rho^a(a|S) - \sum_{S \in \mathcal{C}^{(m)}: b \in S} \rho^a(b|S) \right) \\ &= \frac{1}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}^{(m)}: a \in S} \frac{v+\delta}{w^a(S)} - \sum_{S \in \mathcal{C}^{(m)}: b \in S} \frac{v}{w^a(S)} \right) \\ &= \frac{1}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{v}{w^a(S)} + \sum_{S \in \mathcal{C}^{(m)}: a \in S, b \notin S} \frac{v+\delta}{w^a(S)} - \sum_{S \in \mathcal{C}^{(m)}: a \notin S, b \in S} \frac{v}{w^a(S)} \right) \\ &= \frac{1}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{\delta}{w^a(S)} + \sum_{S' \in \mathcal{C}^{(m-1)}: a, b \notin S'} \left[ \frac{v+\delta}{w^a(S'+a)} - \frac{v}{w^a(S'+b)} \right] \right). \end{aligned}$$

We thus have

$$\mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y_S^r\}) \| \mathbb{P}^b(\{Y_S^r\})) = pR \log \frac{v+\delta}{v} \binom{n-1}{m-1} \Delta_K^{(m)}. \quad (4)$$

We will now bound  $\log \frac{v+\delta}{v}$  in terms of  $\frac{\Delta_K^{(m)}}{\tau_K^{(m)}}$ .

$$\begin{aligned}
\frac{\Delta_K^{(m)}}{\tau_K^{(m)}} &= \frac{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{w^a(a) - w^a(b)}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{w^a(a)}{w^a(S) + v + \delta} - \frac{w^a(b)}{w^a(S) + v}}{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{w^a(a)}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{w^a(a)}{w^a(S) + v + \delta}} \\
&= \frac{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{\delta}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{\delta(w^a(S))}{(w^a(S) + v + \delta)(w^a(S) + v)}}{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{v + \delta}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{v + \delta}{w^a(S) + v + \delta}} \\
&= \frac{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{1}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{(w^a(S))}{(w^a(S) + v + \delta)(w^a(S) + v)}}{\sum_{S \in \mathcal{C}^{(m)}: a, b \in S} \frac{1}{w^a(S) + 2v + \delta} + \sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{1}{w^a(S) + v + \delta}} \cdot \frac{\delta}{v + \delta} \\
&\geq \frac{\sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{1}{w^a(S) + v + \delta} \cdot \frac{w^a(S)}{w^a(S) + v}}{\sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{1}{w^a(S) + v + \delta}} \cdot \frac{\delta}{v + \delta} \\
&\geq \frac{\sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{1}{w^a(S) + v + \delta} \cdot \frac{(m-1)v}{(m-1)v + v}}{\sum_{S \in \mathcal{C}^{(m-1)}: a, b \notin S} \frac{1}{w^a(S) + v + \delta}} \cdot \frac{\delta}{v + \delta} \\
&= \frac{(m-1)v}{mv} \cdot \frac{\delta}{v + \delta} \\
&= \frac{m-1}{m} \frac{\delta}{v + \delta} \geq \frac{1}{2} \frac{\delta}{v + \delta}.
\end{aligned}$$

We can now bound  $\log \frac{v+\delta}{v}$  as

$$\log \frac{v + \delta}{v} \leq \frac{\delta}{v} \leq \frac{2\delta}{v + \delta} \leq \frac{4\Delta_K^{(m)}}{\tau_K^{(m)}}.$$

Combining with equation (4) completes the proof. ■

### 1.4.3 Approximate top- $K$ recovery

To obtain a lower bound on the sample complexity for approximate top- $K$  recovery, we require a more nuanced construction than that in the proof of Theorem 5.1. At a high level, we need to construct a multiset of  $K$ -size subsets  $\{\mathcal{S}^1, \dots, \mathcal{S}^L\}$ . Each  $K$ -size subset  $\mathcal{S}^l$  corresponds to an MNL model whose top  $K$  items are exactly  $\mathcal{S}^l$ . We need to carefully design this multiset such that the pairwise edit distance between any two subsets is larger than  $2h$ . This automatically ensures that any top- $K$  estimate is close in edit distance to at most one of the  $L$  subsets. At the same time, we also want the distance between the distributions induced by any two models to be small in KL divergence sense. This would then allow us to invoke Fano's lemma to lower bound the probability that any statistical estimator outputs a top- $K$  estimate set with small edit distance error. Lastly, we also want  $L$  to be sufficiently large in order to obtain a good lower bound using Fano's lemma.

We first present a reworded version of Lemma 9 of Shah and Wainwright [2017]. The original lemma, in turn, is based on a result due to Levenshtein [1971] for fixed weighted binary codes.

**Lemma 1.9.** [Shah and Wainwright, 2017, lemma 9] Consider the regime where  $h < \frac{2}{3} \min\{K, \sqrt{n}, n - K\}$ . For sufficiently large  $n$ , there exists a multiset  $\{s^1, \dots, s^L\}$  with cardinality  $L \geq \exp\left(\frac{9}{20}h \log n\right)$  and  $s^l \subseteq [\frac{n}{2}, n]$  for  $l \in [L]$  such that

$$|s^l| = \frac{3h}{2} \quad \forall l \in [L] \quad \text{and} \quad D_{01}(s^j, s^l) = 2h + 1 \quad \text{for all } j \neq l \in [L].$$

By the above lemma, there exists a multiset of cardinality  $\exp\left(\frac{9}{20}h \log n\right)$ , consisting of sets of size exactly  $\frac{3h}{2}$ , with elements from  $[\frac{n}{2}, n]$ . For  $s^l$ , let us construct a MNL model as follows: let  $u = \{1, \dots, K - \frac{3h}{2}\}$  and  $\mathcal{S}_K^l = u \cup s^l$  be the set of the top  $K$  items for model  $l \in [L]$ . Note that this construction is valid since  $h < \frac{2K}{3}$ . Following closely the description of the special MNL model used in subsection (1.4.2), we assign  $v$  as weight for all items in  $[n] \setminus \mathcal{S}_K^l$ ,  $v + \delta/2$  for all items in  $u$  and  $v + \delta$  for all items in  $s^l$  for some  $v > \delta > 0$ . In other words,  $v^l$  is the set of the top  $\frac{3h}{2}$  items for the  $l$ -th model while  $u$  is the set composing of the  $\frac{3h}{2} + 1$ -th,  $\dots$ ,  $K$ -th best item for model  $l$ .

We emphasize that by design,  $D_{01}(\mathcal{S}_K^l, \mathcal{S}_K^j) > 2h$  for any  $l \neq j \in [L]$ . This also means that for any estimate  $\hat{\mathcal{S}}_K$ , there exists at most one index  $a \in [L]$  such that  $D_{01}(\hat{\mathcal{S}}_K, \mathcal{S}_K^a) \leq h$ . In other words, outputting an estimate with low edit distance error is equivalent to exactly identifying the set of the top  $K$  items. All that remains is to prove that it is hard to distinguish between any two models as their distributions over observed choice data have small KL divergence. Such a result is summarized in the lemma below:

**Lemma 1.10.** *Consider the construction described above, for any two models indexed by  $a, b \in [L]$  with  $a \neq b$ . Let  $\mathbb{P}^a$  and  $\mathbb{P}^b$  be the distributions parametrized by  $\mathcal{S}_K^a$  and  $\mathcal{S}_K^b$  as described above, respectively. Under the sampling model described in Section 4.1, we have*

$$\mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y\}) \parallel \mathbb{P}^b(\{Y\})) \leq 12hpR \binom{n-1}{m-1} \frac{\Delta_{K,h}^{(m)2}}{\tau_{K-h}^{(m)}}$$

Before proving Lemma 1.10, let us state and prove the sample complexity lower bound for approximate top- $K$  ranking, the proof of which directly makes use of the lemma.

**Theorem 1.11.** *Consider the sampling model described in Section 4.1. There exists a choice model  $\mathcal{M}^*$  within the class of Multinomial Logit Models (MNLs) such that for  $n \geq 20$ , if  $pR \binom{n}{m} \leq \frac{1}{96} \frac{n \log n}{m \Delta_{K,h}^{(m)}} \cdot \frac{\tau_{K-h}^{(m)}}{\Delta_{K,h}^{(m)}}$  then any statistical estimator outputs an estimate  $\hat{\mathcal{S}}_K$  and*

$$D_{01}(\hat{\mathcal{S}}_K, \mathcal{S}_K^*) > h$$

with probability at least  $\frac{1}{5}$ .

*Proof.* Consider the regime when  $p, R$  are small enough such that

$$pR \binom{n-1}{m-1} \leq \frac{\log n}{96 \Delta_{K,h}^{(m)2}} \cdot \tau_{K-h}^{(m)} = \frac{\log n}{96 \Delta_{K,h}^{(m)2}} \cdot (\tau_{K-h}^{(m)} + \Delta_{K,h}^{(m)})$$

and the multiset construction as described above. Invoking Fano's inequality, we have the probability that any statistical estimator failing to output an estimate  $\hat{\mathcal{S}}_K$  such that  $D_{01}(\hat{\mathcal{S}}_K, \mathcal{S}_K^*) \leq h$ , is lower bounded as

$$1 - \frac{\frac{1}{8}h \log n + \log 2}{\frac{9}{20}h \log n} \geq \frac{1}{5} \quad \text{for } n \geq 20.$$

This finishes the proof. ■

We will now proceed to prove Lemma 1.10.

*Proof.* Following the same notation and argument as in the proof of Lemma 1.8, we have

$$\mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y\}) \parallel \mathbb{P}^b(\{Y\})) = pR \sum_{S \in \mathcal{C}^{(m)}} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)}) \parallel \mathbb{P}^b(y_S^1 | S \in \hat{\mathcal{C}}^{(m,1)})).$$

Similarly to the proof of Lemma 1.8, let us define some shorthand notations: let  $w^a(S)$  denote the sum of the weights of the items in menu  $S$  and  $w^a(i)$  the weight of item  $i$  under the special MNL model whose top  $K$  items is  $\mathcal{S}_K^a$  (similarly defined for  $b \neq a$ ). Expanding on the sum of the KL divergence terms and rearranging the summation order gives us

$$\begin{aligned}
& \frac{1}{pR} \mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y\}) \parallel \mathbb{P}^b(\{Y\})) \\
&= \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^a(i) \cdot w^b(S)}{w^b(i) \cdot w^a(S)} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^a(i) \cdot w^b(S)}{w^b(i) \cdot w^a(S)} \\
&+ \sum_{i \in \mathcal{S}_K^a \cap \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^a(i) w^b(S)}{w^b(i) w^a(S)} + \sum_{i \in [n] \setminus (\mathcal{S}_K^a \cup \mathcal{S}_K^b)} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^a(i) w^b(S)}{w^b(i) w^a(S)} \\
&= \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{v + \delta}{w^a(S)} \log \frac{(v + \delta) \cdot w^b(S)}{v \cdot w^a(S)} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{v}{w^a(S)} \log \frac{v \cdot w^b(S)}{(v + \delta) \cdot w^a(S)} \\
&+ \sum_{i \in \mathcal{S}_K^a \cap \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} + \sum_{i \in [n] \setminus (\mathcal{S}_K^a \cup \mathcal{S}_K^b)} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \\
&= \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{v + \delta}{w^a(S)} \log \frac{v + \delta}{v} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{v}{w^a(S)} \log \frac{v}{v + \delta} \\
&+ \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{v + \delta}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{v}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \quad (*) \\
&+ \sum_{i \in \mathcal{S}_K^a \cap \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} + \sum_{i \in [n] \setminus (\mathcal{S}_K^a \cup \mathcal{S}_K^b)} \sum_{S \in C_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)}.
\end{aligned}$$

In the second equality, we have made use of the fact that for any item  $i$  in  $\mathcal{S}_K^a \cup \mathcal{S}_K^b$  and  $\mathcal{S}_K^a \cap \mathcal{S}_K^b$ ,  $w^a(i) = w^b(i)$ . Focusing on the first two summations,

$$\begin{aligned}
& \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{v + \delta}{w^a(S)} \log \frac{v + \delta}{v} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{v}{w^a(S)} \log \frac{v}{v + \delta} \\
&= \left( \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in C_i^{(m)}} \frac{v + \delta}{w^a(S)} - \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in C_i^{(m)}} \frac{v}{w^a(S)} \right) \log \frac{v + \delta}{v} \\
&= (2h + 1) \binom{n-1}{m-1} \Delta_{K,h}^{(m)} \log \frac{\delta + v}{v} < 3h \binom{n-1}{m-1} \Delta_{K,h}^{(m)} \log \frac{\delta + v}{v}.
\end{aligned}$$

The last equality comes from recognizing that  $|\mathcal{S}_K^a \setminus \mathcal{S}_K^b| = |\mathcal{S}_K^b \setminus \mathcal{S}_K^a| = D_{01}(\mathcal{S}_K^a, \mathcal{S}_K^b) = 2h + 1$ . We now simplify the

remaining four summation terms of (\*).

$$\begin{aligned}
& \sum_{i \in \mathcal{S}_K^a \setminus \mathcal{S}_K^b} \sum_{S \in \mathcal{C}_i^{(m)}} \frac{v + \delta}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} + \sum_{i \in \mathcal{S}_K^b \setminus \mathcal{S}_K^a} \sum_{S \in \mathcal{C}_i^{(m)}} \frac{v}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \\
& + \sum_{i \in \mathcal{S}_K^a \cap \mathcal{S}_K^b} \sum_{S \in \mathcal{C}_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} + \sum_{i \in [n] \setminus (\mathcal{S}_K^a \cup \mathcal{S}_K^b)} \sum_{S \in \mathcal{C}_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \\
& = \sum_{i \in [n]} \sum_{S \in \mathcal{C}_i^{(m)}} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \\
& = \sum_{S \in \mathcal{C}^{(m)}} \sum_{i \in S} \frac{w^a(i)}{w^a(S)} \log \frac{w^b(S)}{w^a(S)} \\
& = \sum_{S \in \mathcal{C}^{(m)}} \log \frac{w^b(S)}{w^a(S)} \\
& = \log \frac{\prod_{S \in \mathcal{C}^{(m)}} w^b(S)}{\prod_{S \in \mathcal{C}^{(m)}} w^a(S)} \\
& = \log 1 = 0.
\end{aligned}$$

The last equality comes from recognizing that the two models  $a$  and  $b$  only differ by the identities of the items while the weights are the same. Therefore, the two products in the log term are identical. We thus obtain

$$\mathbb{D}_{\text{KL}}(\mathbb{P}^a(\{Y\}) \parallel \mathbb{P}^b(\{Y\})) < pR3h \binom{n-1}{m-1} \Delta_{K,h}^{(m)} \log \frac{\delta + v}{v}. \quad (**)$$

Following the same argument as in the proof of Lemma 1.8, we can bound

$$\log \frac{\delta + v}{v} \leq 4 \frac{\Delta_{K,h}^{(m)}}{\tau_{K-h}^{(m)}}.$$

Substituting this into (\*\*) completes the proof. ■

In short, the (matching) lower bound on the sample complexity for approximate top- $K$  ranking is

$$\Omega\left(\frac{n \log n}{m \Delta_{K,h}^{(m)}} \cdot \left(1 + \frac{\tau_{K+h+1}^{(m)}}{\Delta_{K,h}^{(m)}}\right)\right).$$

## 1.5 EFFECT OF THE MENU SIZE ON THE SAMPLE COMPLEXITY

**Theorem 6.1.** Consider an MNL model with  $n \geq 2$  items and fix a  $K$ , we have

$$\begin{aligned}
\frac{1}{m \Delta_K^{(m)}} &= \theta\left(\frac{1}{e^{U_K} - e^{U_{K+1}}} \cdot \left(1 + \frac{1}{m-1}\right)^2\right), \\
\frac{\tau_{K+1}^{(m)}}{\Delta_K^{(m)}} &= \theta\left(\frac{e^{U_{K+1}}}{e^{U_K} - e^{U_{K+1}}} \cdot \left(1 + \frac{1}{m-1}\right)\right).
\end{aligned}$$

*Proof. Part 1.* We will prove that there exists a positive valued function  $l(m)$  such that  $m \Delta_K^{(m)} < l(m)$  and that  $l(m)$  monotonically increases with  $m$ , and at diminishing rate. Once we have obtained  $l(m)$ , we can set  $f_1(m) = \frac{1}{l(m)}$ . In fact, we will prove a slightly more general result concerning any pair of items  $i, j$  where  $w_i := e^{U_i} > e^{U_j} := w_j$ . Let

$\Delta_{ij}^{(m)} = \tau_i^{(m)} - \tau_j^{(m)}$ . Additionally, let  $w_n := e^{U_n} = e^{U_{\min}}$  and  $w_1 := e^{U_1} = e^{U_{\max}}$  be two model-dependent constants. We have

$$\begin{aligned}
m\Delta_{ij}^{(m)} &= \frac{m}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i}{w_i + w_j + \sum_{k \in S} w_k} - \frac{w_j}{w_i + w_j + \sum_{k \in S} w_k} \right. \\
&\quad \left. + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \left( \frac{w_i}{w_i + \sum_{k \in S'} w_k} - \frac{w_j}{w_j + \sum_{k \in S'} w_k} \right) \right) \\
&= \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{w_i + w_j + \sum_{k \in S} w_k} + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \frac{1}{w_j + \sum_{k \in S'} w_k} \cdot \frac{\sum_{k \in S'} w_k}{w_i + \sum_{k \in S'} w_k} \right) \\
&\leq \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \binom{n-2}{m-2} \cdot \frac{1}{w_i + w_j + (m-2)w_n} + \binom{n-2}{m-1} \frac{1}{w_j + (m-1)w_n} \cdot \frac{(m-1)w_1}{w_i + (m-1)w_1} \right) \\
&= \frac{m(w_i - w_j)}{n-1} \left( (m-1) \cdot \frac{1}{w_i + w_j + (m-2)w_n} + (n-m) \frac{1}{w_j + (m-1)w_n} \cdot \frac{(m-1)w_1}{w_i + (m-1)w_1} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \frac{m}{[w_j + (m-1)w_n]} \frac{(m-1)w_1}{[w_i + (m-1)w_1]} \right. \\
&\quad \left. + m(m-1) \cdot \left[ \frac{1}{w_i + w_j + (m-2)w_n} - \frac{mw_1}{[w_i + (m-1)w_1][w_j + (m-1)w_n]} \right] \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \frac{m}{[mw_n + w_j - w_n]} \frac{(m-1)w_1}{[w_i + (m-1)w_1]} \right. \\
&\quad \left. + m(m-1) \cdot \frac{(w_1 - w_i)(w_n - w_j)}{[(m-2)w_n + w_i + w_j][(m-1)w_n + w_j][(m-1)w_1 + w_i]} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \frac{m}{[mw_n + w_j - w_n]} \frac{(m-1)w_1}{[w_i + (m-1)w_1]} \right. \\
&\quad \left. + \underbrace{\frac{(w_1 - w_i)(w_n - w_j)}{(m-2)w_n + w_i + w_j} \cdot \frac{m}{mw_n + w_j - w_n} \cdot \frac{(m-1)}{(m-1)w_1 + w_i}}_{\leq 0} \right) \\
&\leq (w_i - w_j) \cdot \frac{n}{n-1} \cdot \frac{m}{[mw_n + w_j - w_n]} \frac{(m-1)w_1}{[w_i + (m-1)w_1]} \\
&= (w_i - w_j) \cdot \frac{n}{n-1} \cdot \left( 1 - \frac{w_j/w_n - 1}{m-1 + w_j/w_n} \right) \cdot \left( 1 - \frac{w_i/w_1}{m-1 + w_i/w_1} \right).
\end{aligned}$$

It is easy to see that all of the terms containing  $m$  are positive and *increase* with  $m$ . Additionally, it is easy to show that all of these terms increase with *decreasing* rate with respect to  $m$ . Though complicated, the above lower bound can somewhat be simplified using big- $O$  notations.

$$m\Delta_{ij}^{(m)} = \Omega\left((w_i - w_j) \cdot \left(1 - \frac{1}{m}\right)^2\right).$$

By modifying the above argument, we can also show a variational lower bound on  $m\Delta_{ij}^{(m)}$ .



$$\begin{aligned}
m\Delta_{ij}^{(m)} &= \frac{m}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i}{w_i + w_j + \sum_{k \in S} w_k} - \frac{w_j}{w_i + w_j + \sum_{k \in S} w_k} \right. \\
&\quad \left. + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \left( \frac{w_i}{w_i + \sum_{k \in S'} w_k} - \frac{w_j}{w_j + \sum_{k \in S'} w_k} \right) \right) \\
&= \frac{m}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i - w_j}{w_i + w_j + \sum_{k \in S} w_k} + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \left( \frac{w_i}{w_i + \sum_{k \in S'} w_k} - \frac{w_j}{w_j + \sum_{k \in S'} w_k} \right) \right) \\
&= \frac{m}{\binom{n-1}{m-1}} \cdot \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i - w_j}{w_i + w_j + \sum_{k \in S} w_k} + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \frac{(w_i - w_j)(\sum_{k \in S'} w_k)}{(w_i + \sum_{k \in S'} w_k)(w_j + \sum_{k \in S'} w_k)} \right) \\
&= \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{w_i + w_j + \sum_{k \in S} w_k} + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \frac{1}{w_i + \sum_{k \in S'} w_k} \cdot \frac{\sum_{k \in S'} w_k}{w_j + \sum_{k \in S'} w_k} \right) \\
&\text{[Using the fact that } \frac{\sum_{k \in S} w_k}{w_j + \sum_{k \in S} w_k} \geq \frac{\sum_{k \in S} w_n}{w_j + \sum_{k \in S} w_n} \forall S : j \in S] \\
&\geq \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{w_i + w_j + \sum_{k \in S} w_k} + \sum_{S' \in \mathcal{C}^{(m-1)}: i, j \notin S'} \frac{1}{w_i + \sum_{k \in S'} w_k} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right) \\
&\geq \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \binom{n-2}{m-2} \frac{1}{w_i + w_j + (m-2)\bar{w}} + \binom{n-2}{m-1} \frac{1}{w_i + (m-1)\bar{w}} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right).
\end{aligned}$$

where  $\bar{w} = \frac{1}{n-2} \sum_{k \neq i, j} w_k$ . The last inequality comes from applying Jensen's inequality. More precisely, one could treat the sum  $\sum_{k \in S} w_k$  as a random variable and the summation as the unnormalized 'expectation' over the uniform distribution of the menus. As  $\frac{1}{x}$  is a convex function in  $x$  for positive  $x$ , Jensen's inequality applies. Furthermore, without loss of generality, one can assume that  $\bar{w} = 1$  (via scaling of the weights). Consequently,  $w_n \leq 1$ . Continuing with the expansion gives us

$$\begin{aligned}
m\Delta_{ij}^{(m)} &\geq \frac{m(w_i - w_j)}{\binom{n-1}{m-1}} \left( \binom{n-2}{m-2} \frac{1}{w_i + w_j + (m-2)\bar{w}} + \binom{n-2}{m-1} \frac{1}{w_i + (m-1)\bar{w}} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right) \\
&= \frac{m(w_i - w_j)}{n-1} \left( \frac{m-1}{w_i + w_j + (m-2)} + \frac{n-m}{w_i + (m-1)} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( \frac{m(m-1)}{w_i + w_j + (m-2)} + \frac{m(n-m)}{w_i + (m-1)} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right) \\
&\geq \frac{(w_i - w_j)}{n-1} \left( \frac{m(m-1)}{w_i + w_j + (m-2)} + \frac{(m-1)(n-m)}{w_i + (m-1)} \cdot \frac{(m-1)w_n}{w_j + (m-1)w_n} \right) \\
&\geq \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} + \frac{m(m-1)}{w_i + w_j + (m-2)} - \frac{m(m-1)(m-1)w_n}{(w_i + m-1)(w_j + (m-1)w_n)} \right) \\
&\geq \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} \right. \\
&\quad \left. + m(m-1) \left( \frac{1}{w_i + w_j + (m-2)} - \frac{(m-1)w_n}{(w_i + m-1)(w_j + (m-1)w_n)} \right) \right)
\end{aligned}$$

[Expanding on the difference term]

$$\begin{aligned}
&\geq \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} \right. \\
&\quad \left. + m(m-1) \left( \frac{1}{w_i + w_j + (m-1)} - \frac{(m-1)w_n}{(w_i + m-1)(w_j + (m-1)w_n)} \right) \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} \right. \\
&\quad \left. + m(m-1) \cdot \frac{(w_i w_j + w_i(m-1)w_n + m w_j + m(m-1)w_n - w_j - (m-1)w_n - w_i(m-1)w_n - w_j(m-1)w_n - m(m-1)w_n + 1(m-1)w_n)}{((w_i + w_j + (m-1))(w_i + m-1)(w_j + (m-1)w_n))} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} \right. \\
&\quad \left. + m(m-1) \cdot \frac{m w_j + w_i w_j - w_j - w_j(m-1)w_n}{(w_i + w_j + (m-1))(w_i + m-1)(w_j + (m-1)w_n)} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} \right. \\
&\quad \left. + m(m-1)w_j \cdot \frac{m + w_i - 1 - (m-1)w_n}{(w_i + w_j + (m-1))(w_i + m-1)(w_j + (m-1)w_n)} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} + \frac{m(m-1)w_j((m-1)(1-w_n) + w_i)}{(w_i + w_j + (m-1))(w_i + m-1)(w_j + (m-1)w_n)} \right) \\
&\geq \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} + \frac{m(m-1)w_j((m-1)(1-w_n) + w_i)}{(w_i + w_j + m)(w_i + m-1)(w_j + (m-1)w_n)} \right) \\
&= \frac{(w_i - w_j)}{n-1} \left( n \cdot \frac{m-1}{m-1+w_i} \cdot \frac{(m-1)w_n}{(m-1)w_n + w_j} + \underbrace{w_j \cdot \frac{m}{w_i + w_j + m} \cdot \frac{m-1}{w_j + (m-1)w_n} \cdot \frac{w_i + (m-1)(1-w_n)}{w_i + (m-1)}}_{\geq 0} \right) \\
&\geq \frac{(w_i - w_j)}{n-1} \left( n \cdot \left( 1 - \frac{w_i}{m-1+w_i} \right) \cdot \left( 1 - \frac{w_j}{(m-1)w_n + w_j} \right) \right) \\
&= (w_i - w_j) \cdot \frac{n}{n-1} \cdot \left( 1 - \frac{w_i}{m-1+w_i} \right) \cdot \left( 1 - \frac{w_j/w_n}{m-1+w_j/w_n} \right).
\end{aligned}$$

We thus have  $m\Delta_{ij} = O((w_i - w_j) \cdot (1 - \frac{1}{m})^2)$ . Combining with the upper bound shown earlier, we get

$$m\Delta_{ij} = \theta\left((w_i - w_j) \cdot (1 - \frac{1}{m})^2\right).$$

**Part 2.** We will prove that there exists a positive valued function  $l(m)$  such that  $\frac{\Delta_{ij}^{(m)}}{\tau_j^{(m)}} \geq l(m)$  and that  $l(m_2) > l(m_1)$  for  $m_2 > m_1 \geq 2$ . Consider for any two items  $i, j$  such that  $w_i > w_j$ . By definition

$$\begin{aligned} \frac{\Delta_{ij}^{(m)}}{\tau_j^{(m)}} &= \frac{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i - w_j}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_i}{w_i + \sum_{k \in S} w_k} - \frac{w_j}{w_j + \sum_{k \in S} w_k}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_j}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{w_j + \sum_{k \in S} w_k}} \\ &= \frac{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_i - w_j}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{(w_i - w_j)(\sum_{k \in S} w_k)}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{w_j}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{w_j + \sum_{k \in S} w_k}} \\ &= \frac{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{\sum_{k \in S} w_k}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{1}{w_j + \sum_{k \in S} w_k}} \cdot \frac{w_i - w_j}{w_j} \\ &= \left(1 - \frac{\sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{1}{w_j + \sum_{k \in S} w_k}}\right) \cdot \frac{w_i - w_j}{w_j}. \end{aligned}$$

The goal is to upper (and lower) bound the quantity

$$\frac{\sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{1}{w_j + \sum_{k \in S} w_k}}.$$

For a lower bound, we can show that

$$\begin{aligned} &\frac{\sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{1}{w_j + \sum_{k \in S} w_k}} \\ &\geq \frac{\binom{n-2}{m-1} \frac{1}{(m-1)w_1 + w_i} \frac{w_j}{w_j + (m-1)w_1}}{\binom{n-2}{m-2} \cdot \frac{1}{(m-2)w_n + w_i + w_j} + \binom{n-2}{m-1} \cdot \frac{1}{(m-1)w_n + w_j}} \\ &= \frac{(n-m) \cdot \frac{1}{(m-1)w_1 + w_i} \frac{w_j}{w_j + (m-1)w_1}}{(m-1) \cdot \frac{1}{(m-2)w_n + w_i + w_j} + (n-m) \cdot \frac{1}{(m-1)w_n + w_j}} \\ &[\text{Noting that } (m-2)w_n + w_i + w_j = (m-1)w_n + w_j + (w_i - w_n) \geq (m-1)w_n + w_j] \\ &\geq \frac{(n-m) \cdot \frac{1}{(m-1)w_1 + w_i} \frac{w_j}{w_j + (m-1)w_1}}{(m-1) \cdot \frac{1}{(m-1)w_n + w_j} + (n-m) \cdot \frac{1}{(m-1)w_n + w_j}} \\ &= \frac{n-m}{n-1} \cdot \frac{1}{(m-1)w_1 + w_i} \cdot \frac{(m-1)w_n + w_j}{(m-1)w_1 + w_j} \\ &\geq \frac{n-m}{n-1} \cdot \frac{1/w_1}{m-1 + w_i/w_1} \cdot \frac{w_n}{w_1}. \end{aligned}$$

For an upper bound, we can show that

$$\begin{aligned}
& \frac{\sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{w_j}{(w_i + \sum_{k \in S} w_k)(w_j + \sum_{k \in S} w_k)}}{\sum_{S \in \mathcal{C}_i^{(m)} \cap \mathcal{C}_j^{(m)}} \frac{1}{\sum_{k \in S} w_k} + \sum_{S \in \mathcal{C}^{(m-1)}: i, j \notin S} \frac{1}{w_j + \sum_{k \in S} w_k}} \\
& \leq \frac{\binom{n-2}{m-1} \frac{1}{(m-1)w_n + w_i} \frac{w_j}{w_j + (m-1)w_n}}{\binom{n-2}{m-2} \cdot \frac{1}{(m-2)w_1 + w_i + w_j} + \binom{n-2}{m-1} \cdot \frac{1}{(m-1)w_1 + w_j}} \\
& = \frac{(n-m) \cdot \frac{1}{(m-1)w_n + w_i} \frac{w_j}{w_j + (m-1)w_n}}{(m-1) \cdot \frac{1}{(m-2)w_1 + w_i + w_j} + (n-m) \cdot \frac{1}{(m-1)w_1 + w_j}} \\
& \text{[Noting that } (m-2)w_1 + w_i + w_j = (m-1)w_1 + w_j + (w_i - w_1) \leq (m-1)w_1 + w_j\text{]} \\
& \leq \frac{(n-m) \cdot \frac{1}{(m-1)w_n + w_i} \frac{w_j}{w_j + (m-1)w_n}}{(m-1) \cdot \frac{1}{(m-1)w_1 + w_j} + (n-m) \cdot \frac{1}{(m-1)w_1 + w_j}} \\
& = \frac{n-m}{n-1} \cdot \frac{1/w_n}{m-1 + w_i/w_n} \cdot \frac{w_1}{w_n}.
\end{aligned}$$

In both cases, it is straight forward to check that both of the upper and lower bounds decrease with increasing  $m$  and at a decreasing rate. Furthermore, they have the same asymptotic dependency on  $m$ . Putting both bounds together, we have

$$\frac{\Delta_{ij}^{(m)}}{\tau_j^{(m)}} = \theta \left( \left(1 - \frac{n-m}{n-1} \cdot \frac{1}{m}\right) \cdot \frac{w_i - w_j}{w_j} \right) = \theta \left( \frac{n}{n-1} \cdot \left(1 - \frac{1}{m}\right) \cdot \frac{w_i - w_j}{w_j} \right).$$

Simplifying the above expression by removing the dependency on  $n$  gives

$$\frac{\Delta_{ij}^{(m)}}{\tau_j^{(m)}} = \theta \left( \left(1 - \frac{1}{m}\right) \cdot \frac{w_i - w_j}{w_j} \right).$$

This completes the proof. ■

## 1.6 CONNECTION AMONG MNL-MLE, GENERALIZED BORDA COUNT AND SPECTRAL RANKING

In this section, we will prove the theorems about the connections among the three algorithms. Here, we provide more general results pertaining to ranking output (as opposed to just top  $K$  estimate).

**Theorem 7.1.** *Consider the sampling model described in Section 4.1, for any  $p > 0$ , in the limit as  $R \rightarrow \infty$ , MNL-MLE and choice-based Borda count will produce the same top  $K$  estimate. Moreover, this holds even if the data does not come from the MNL model or any IID-RUM.*

*Proof. Notation:* As a shorthand notation, we will use  $p(i|S)$  denote the probability of item  $i$  being chosen from menu  $S$ . Note that this probability does not necessarily follow any parametric RUM. Let  $\text{ord}$  denote the ordering function whose input is a set of real numbers and whose output is the full ordering of the indices of those numbers in increasing order. Under our sampling model, in the limit as  $R \rightarrow \infty$ , all the possible menus in  $\mathcal{C}^{(m)}$  are observed given that  $p > 0$ . Furthermore, given infinite data, the observed probability becomes exact. Consider the log likelihood function

$$\mathcal{L}(\mathbf{U}) = \sum_{i=1}^n \left( \sum_{S \in \mathcal{C}_i^{(m)}} p(i|S) \cdot \log \frac{e^{U_i}}{\sum_{j \in S} e^{U_j}} \right).$$

The derivative of the log likelihood with respect to individual partworth parameter is given as

$$\begin{aligned}
\nabla_{U_i} \mathcal{L}(\mathbf{U}) &= \sum_{S \in \mathcal{C}_i^{(m)}} \left( - \sum_{j \in S, j \neq i} p(j|S) \frac{e^{U_i}}{\sum_{k \in S} e^{U_k}} - p(i|S) \frac{e^{U_i}}{\sum_{k \in S} e^{U_k}} + p(i|S) \right) \\
&= \sum_{S \in \mathcal{C}_i^{(m)}} \left( p(i|S) - \frac{e^{U_i}}{\sum_{k \in S} e^{U_k}} \right).
\end{aligned}$$

As the log likelihood function is concave in  $U$ , the MLE estimate  $\hat{U}$  must satisfy:

$$\sum_{S \in \mathcal{C}_i^{(m)}} \frac{e^{\hat{U}_i}}{\sum_{j \in S} e^{\hat{U}_j}} = \sum_{S \in \mathcal{C}_i^{(m)}} p(i|S) \quad \forall i.$$

On the other hand it can easily be shown that for any two items  $i \neq j$ ,

$$\hat{U}_i > \hat{U}_j \Rightarrow \sum_{S \in \mathcal{C}_i^{(m)}} \frac{e^{\hat{U}_i}}{\sum_{k \in S} e^{\hat{U}_k}} > \sum_{S \in \mathcal{C}_j^{(m)}} \frac{e^{\hat{U}_j}}{\sum_{k \in S} e^{\hat{U}_k}}.$$

Hence,

$$\text{ord}(\{\hat{U}_i\}_{i=1}^n) = \text{ord}\left(\left\{\sum_{S \in \mathcal{C}_i^{(m)}} \frac{e^{\hat{U}_i}}{\sum_{k \in S} e^{\hat{U}_k}}\right\}_{i=1}^n\right) = \text{ord}\left(\left\{\sum_{S \in \mathcal{C}_i^{(m)}} p(i|S)\right\}_{i=1}^n\right).$$

Observing that the ordering given by the Borda Count algorithm is consistent with the ordering induced by  $\{\sum_{S \in \mathcal{C}_i^{(m)}} p(i|S)\}_{i=1}^n$  completes the proof. ■

**Theorem 7.2.** *Consider the sampling model described in Subsection 4.1. Assume that the underlying choice model generating the data is in the class of IID-RUMs whose noise distribution has absolutely continuous density function with support on the real line. For any  $p > 0$ , in the limit as  $R \rightarrow \infty$ , then Spectral Ranking, MNL-MLE and choice-based Borda count produce the same top  $K$  estimate.*

*On the other hand, there exists a choice model where in the limit as  $R \rightarrow \infty$ , the spectral ranking algorithm produces a different top  $K$  estimate from MNL-MLE/Borda count.*

*Proof.* Following the argument as in the proof of Theorem 7.1, given  $p > 0$  as  $R \rightarrow \infty$ , all possible menus of size  $m$  are observed and all choice probabilities are exact. For the first part of the theorem it suffices to prove that under IID-RUMs, the Spectral Ranking algorithm is consistent in recovering the true ordering among the items, as this will be the same ordering as given by Generalized Borda Count/MLE.

The Spectral Ranking algorithm ranks the items by the stationary distribution of a Markov Chain constructed using choice data. For analysis purpose, we follow the construction due to Negahban et al. [2017], Maystre and Grossglauser [2015]. Fix a menu size  $m \geq 2$  and consider the following Markov Chain where for any two items  $i, j$ ,

$$\mathbf{M}_{ij} = \begin{cases} \frac{1}{\binom{n-1}{m-1}} \sum_{S \in \mathcal{C}^{(m)} : i, j \in S} \rho(j|S) & \text{if } j \neq i \\ 1 - \sum_{k \neq i} \mathbf{M}_{ik} & \text{if } j = i \end{cases}.$$

Consider any two items  $i, j$  such that  $U_i > U_j$ . Under IID-RUMs assumptions in the theorem statement, we have

$$\rho(i|S) > \rho(j|S) \quad \forall S \in \mathcal{C}^{(m)} : i, j \in S$$

and

$$\rho(i|S' \cup \{i\}) > \rho(j|S' \cup \{j\}) \quad \forall S' \in \mathcal{C}^{(m-1)} : i, j \notin S'.$$

It is easy to show that

$$\mathbf{M}_{ki} > \mathbf{M}_{kj} \quad \forall k \neq i, j$$

and

$$\mathbf{M}_{ii} > \mathbf{M}_{jj}.$$

Let  $\pi$  denote the stationary distribution of the Markov Chain constructed above. By definition, the stationary distribution of the Markov Chain satisfies

$$\pi_i = \sum_{k \in [n]} \pi_k \cdot \mathbf{M}_{ki} > \sum_{k \in [n]} \pi_k \cdot \mathbf{M}_{kj} = \pi_j.$$

For the second half of the theorem, let us consider the pairwise comparison setting ( $m = 2$ ). We will construct a pairwise choice model such that MNL-MLE/Generalized Borda Count give a different ordering among the items from Spectral

Ranking algorithm in the limit as  $R \rightarrow \infty$ . Consider an universe of 4 items with the following pairwise choice probability. Note that  $P_{ij} = P(j|\{i, j\})$ . Define

$$P = \begin{bmatrix} 0.5 & 0.6 & 0.55 & 0.55 \\ 0.2 & 0.5 & 0.85 & 0.60 \\ 0.45 & 0.40 & 0.5 & 0.95 \\ 0.45 & 0.45 & 0.15 & 0.5 \end{bmatrix}.$$

It is easy to check that  $\tau_4 > \tau_3 > \tau_2 > \tau_1$ . In the limit of infinite data, MLE/Borda Count will output the ordering 4, 3, 2, 1 (best item first). However, the Spectral Ranking algorithm will give the ordering 4, 2, 3, 1. This completes the proof. ■

## 2 ADDITIONAL EXPERIMENTS

### 2.1 DATASET DESCRIPTIONS

Table (1) shows the characteristics of the datasets used in our experiments: number of items, number of rankings available and whether the data contains partial rankings.

| Dataset     | $n$ | Num rankings | Partial ranking? |
|-------------|-----|--------------|------------------|
| APA         | 5   | 6000         | No               |
| Sushi       | 11  | 5000         | No               |
| Irish-North | 12  | 44k          | Yes              |
| Irish-West  | 9   | 30k          | Yes              |
| Irish-Meath | 14  | 64k          | Yes              |
| F1          | 22  | 18           | Yes              |

Table 1: Characteristics of the datasets used in our experiments.

**Data preparation:** Given menu size  $m$ , to estimate choice probability from rankings and partial rankings, we first enumerate all the possible menus of size  $m$ . For each menu  $S$ , and for each  $i \in S$ , we count the number of rankings where  $i$  is ranked ahead of all the other items in  $S$ , a 'win' for  $i$ . For a partial ranking where there are  $l$  items ranked as 'equal', we count  $\frac{1}{l}$ -'win' for each item. Choice probability is obtained by normalizing the number of wins by the number of rankings.

**Experiments:** For each independent trial, we first generate all the data as specified by the maximum sample size. We then feed increasing portion of this data to the algorithm and check if the algorithm correctly identifies *all* of the top  $K$  items. For all experiments, we keep  $R = 100$  and adjust  $p$  to obtain the appropriate expected sample size.

### 2.2 ADDITIONAL EXPERIMENTAL RESULTS

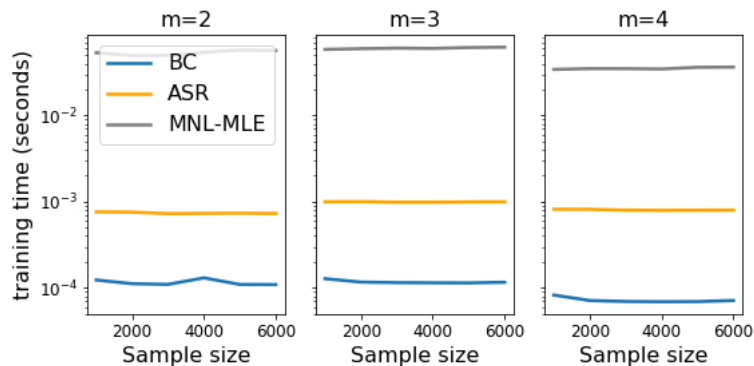


Figure 1: **APA election dataset:** Average training time (seconds) against sample size for  $m = 2, 3, 4$ .

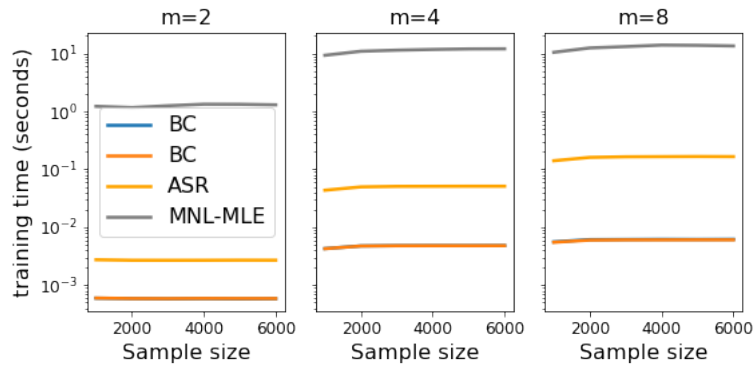


Figure 2: **Irish North dataset:** Average training time (seconds) against sample size for  $m = 2, 4, 8$ .

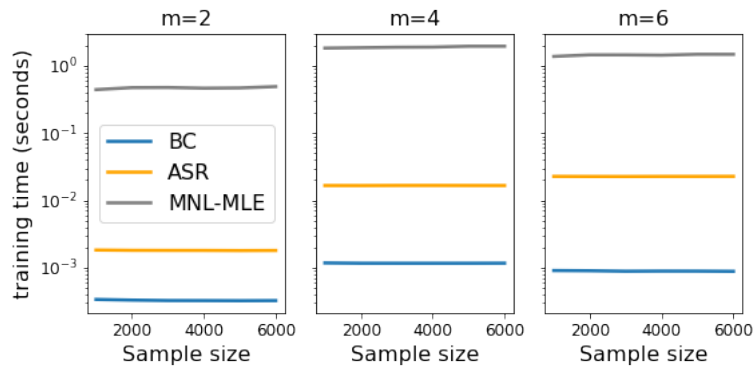


Figure 3: **Irish West dataset:** Average training time (seconds) against sample size for  $m = 2, 4, 6$ .

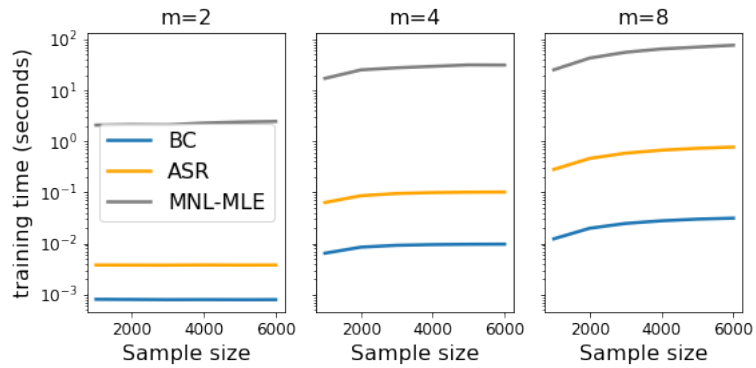


Figure 4: **Irish Meath dataset:** Average training time (seconds) against sample size for  $m = 2, 4, 8$ .



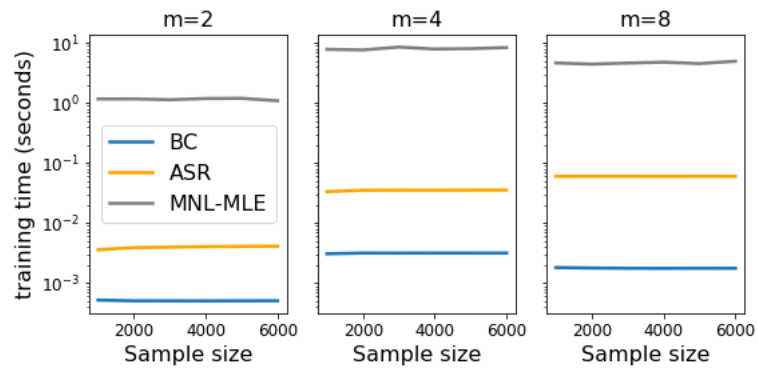


Figure 5: **SUSHI dataset:** Average training time (seconds) against sample size for  $m = 2, 4, 8$ .

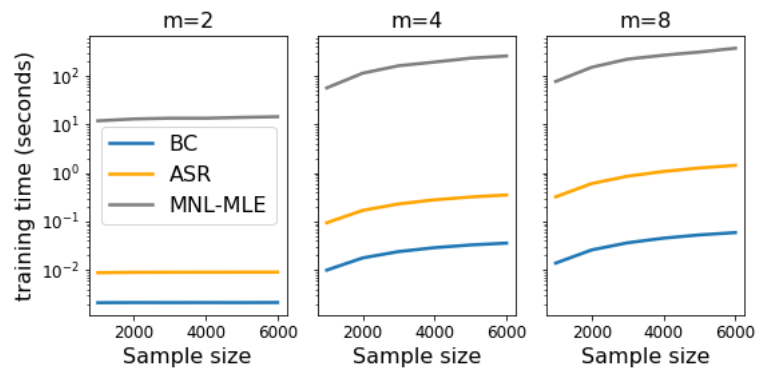


Figure 6: **F1 dataset:** Average training time (seconds) against sample size for  $m = 2, 4, 8$ .

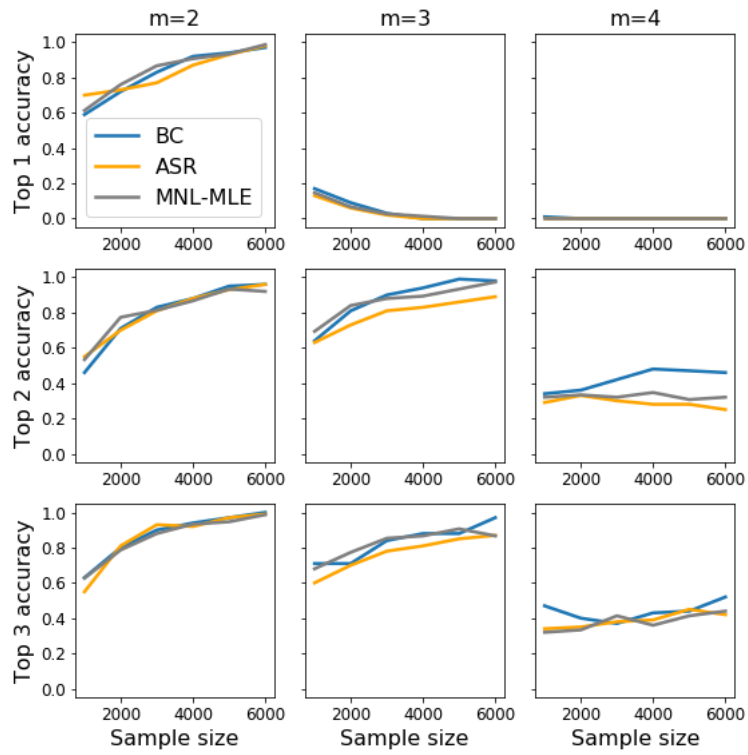


Figure 7: **APA election dataset:** Exact top  $K$  accuracy against sample size

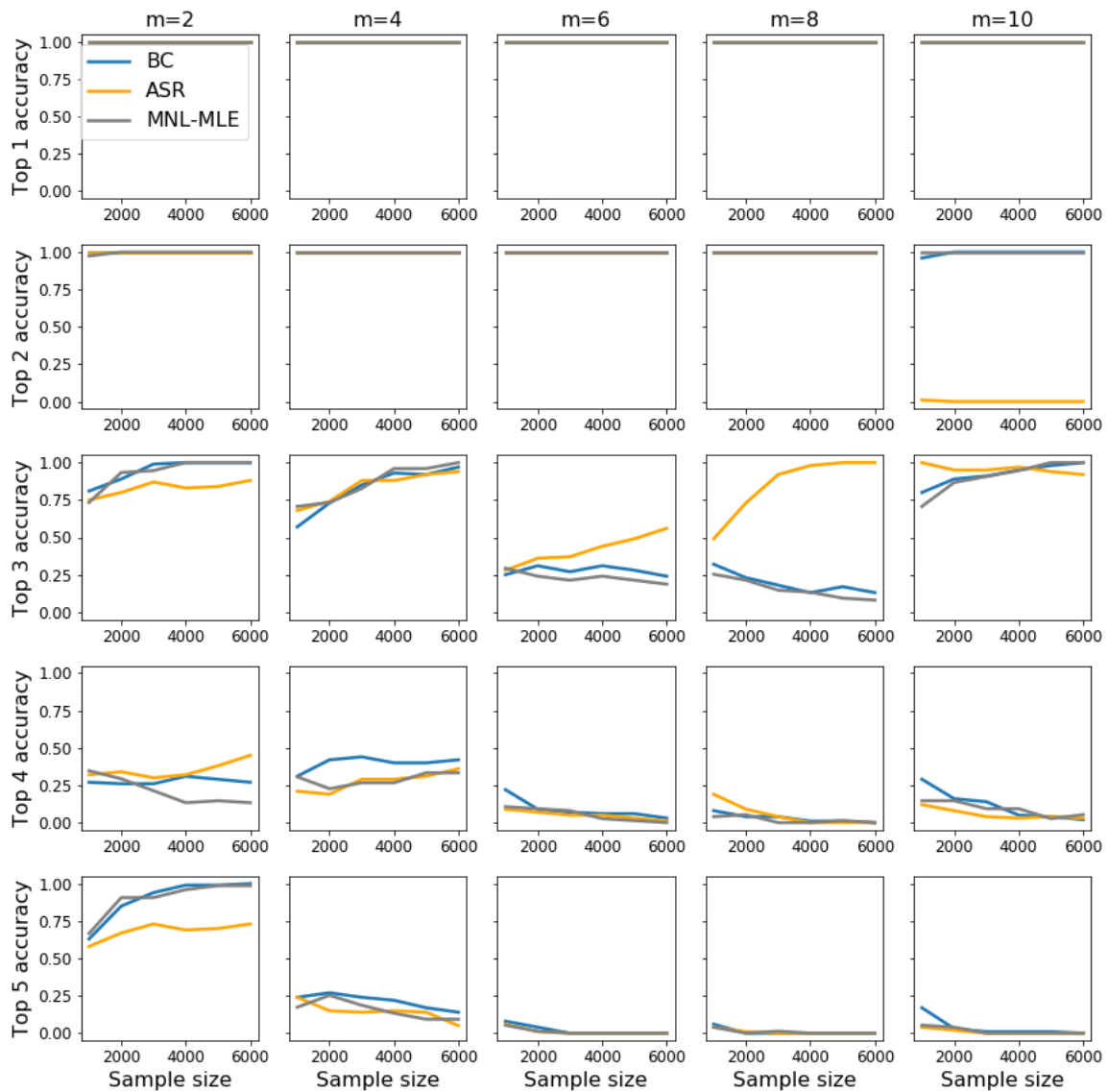


Figure 8: **SUSHI dataset**: Exact top  $K$  accuracy against sample size

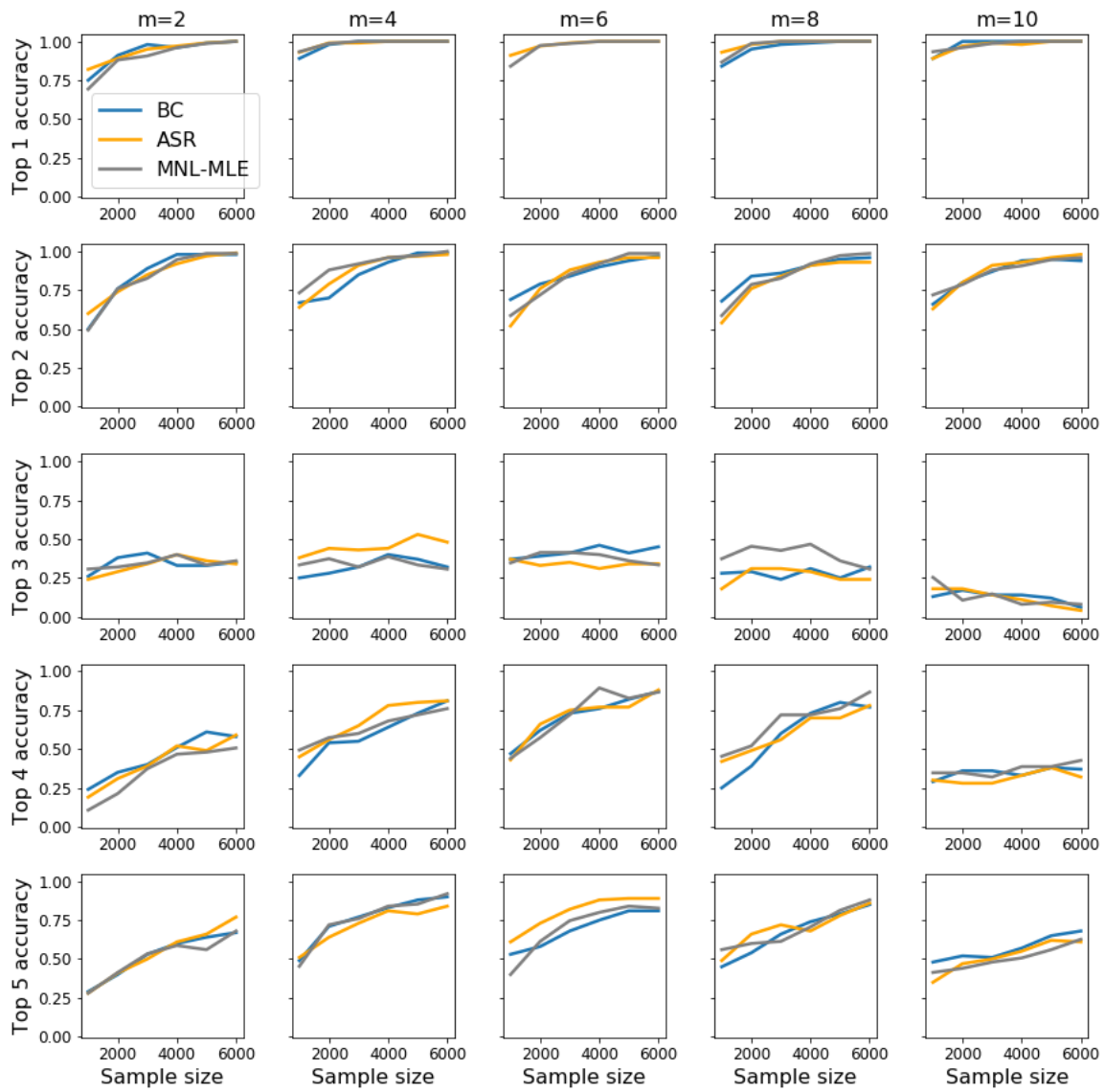


Figure 9: **Irish North dataset:** Exact top  $K$  accuracy against sample size

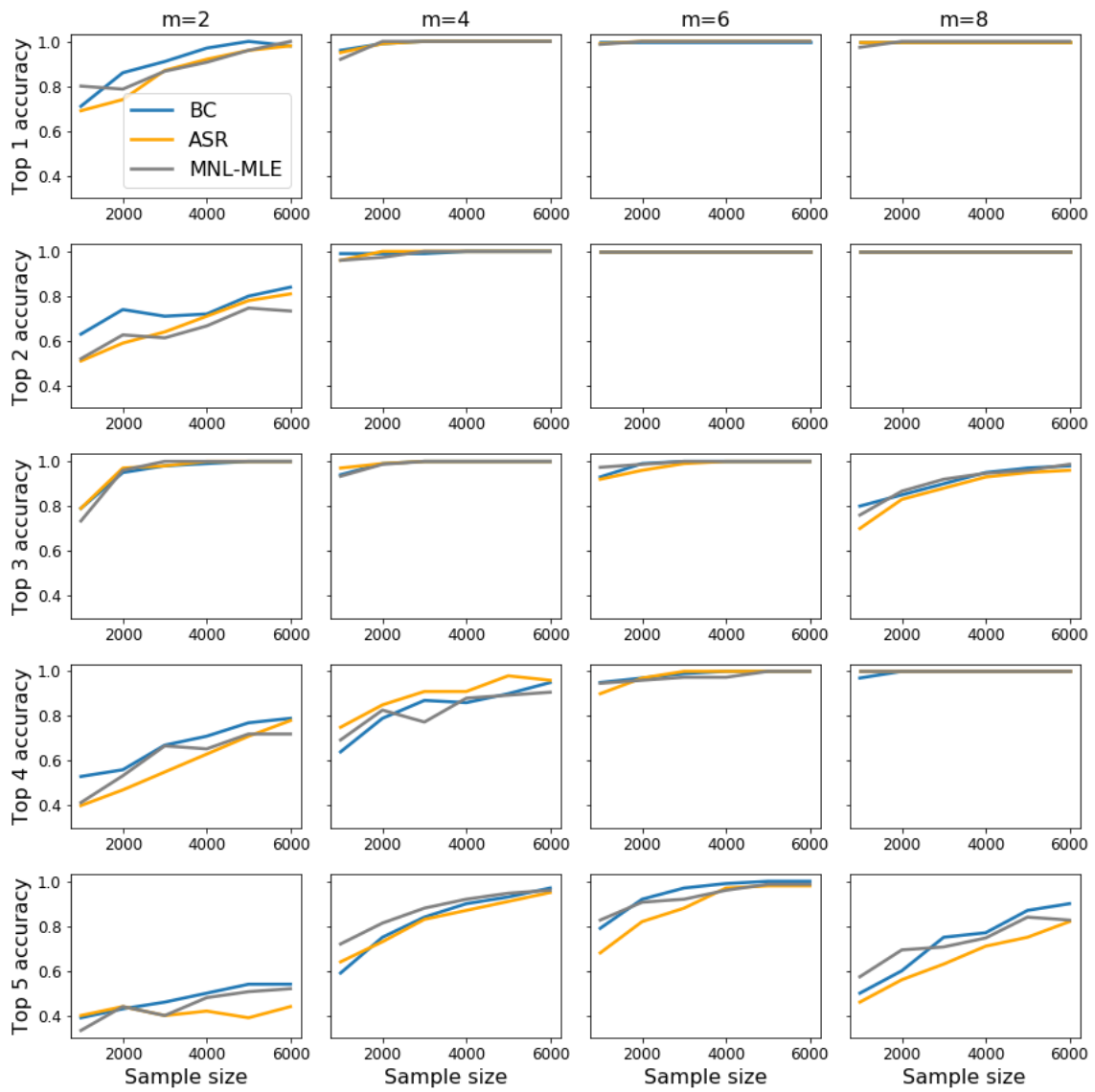


Figure 10: **Irish West dataset:** Exact top  $K$  accuracy against sample sizes

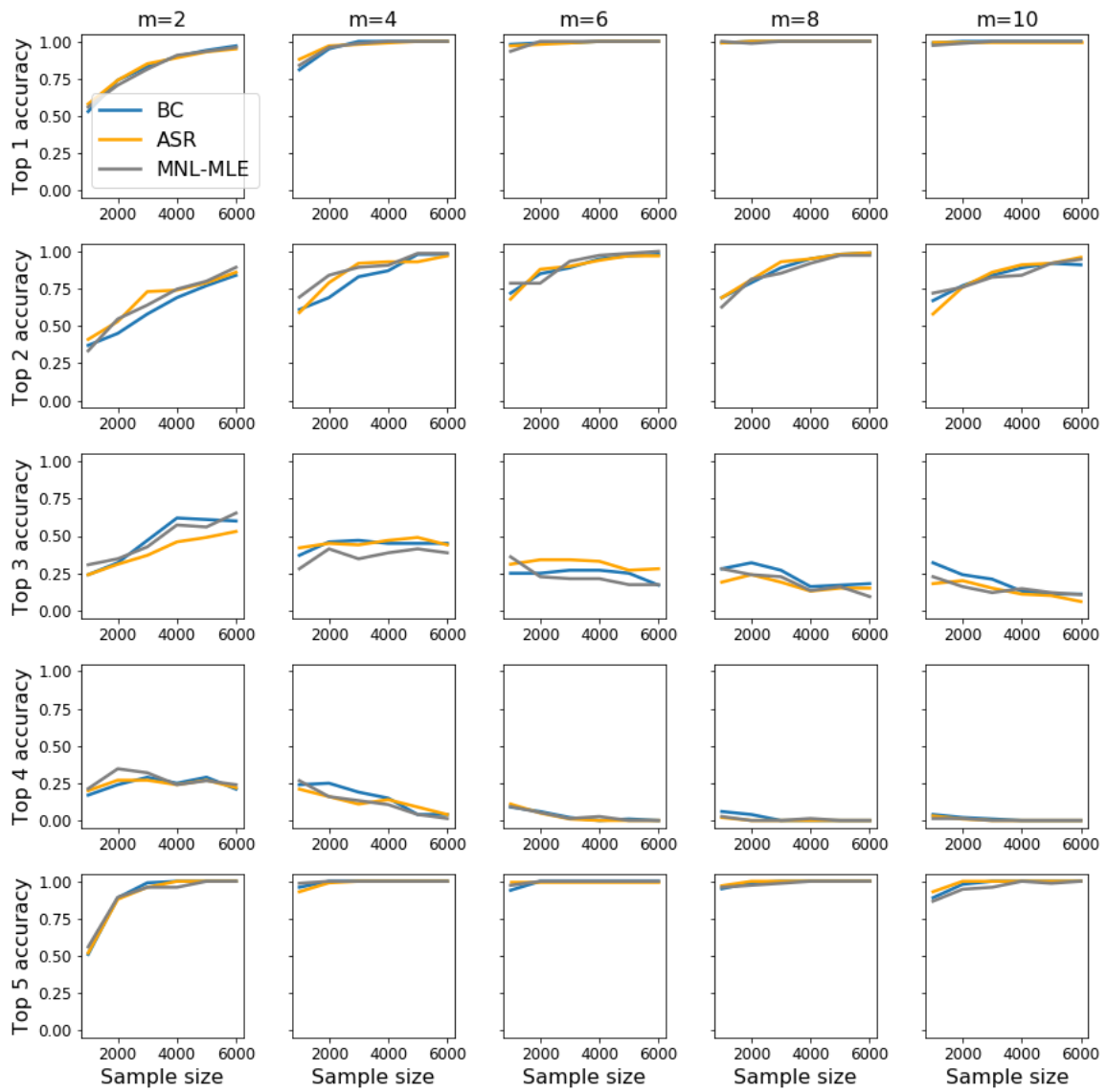


Figure 11: **Irish Meath dataset:** Exact top  $K$  accuracy against sample size

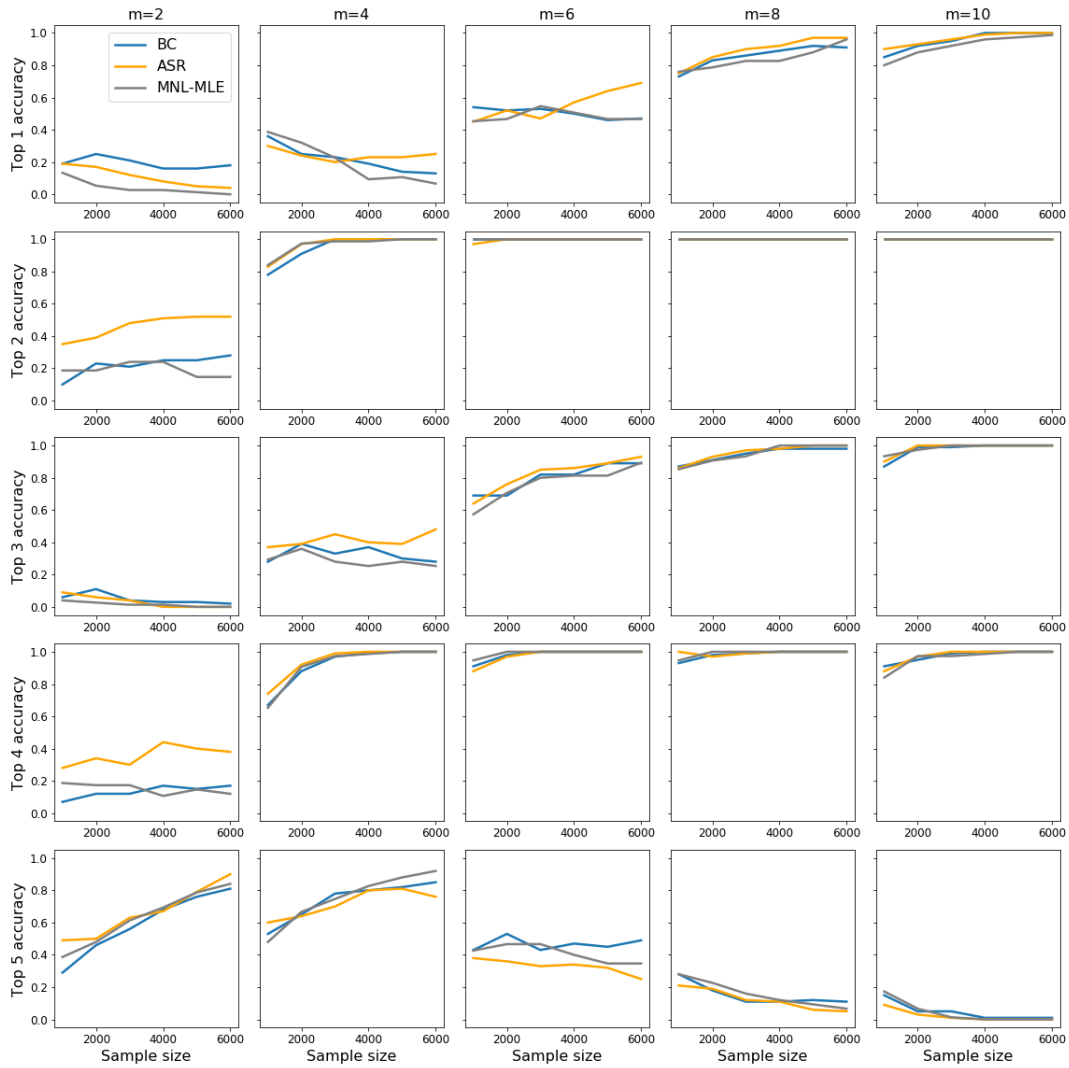


Figure 12: **F1 dataset:** Exact top  $K$  accuracy against sample size.

## References

- Gordon M Becker, Morris H DeGroot, and Jacob Marschak. Stochastic models of choice behavior. *Behavioral science*, 8(1): 41–55, 1963.
- Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- Vladimir Iosifovich Levenshtein. Upper-bound estimates for fixed-weight codes. *Problemy Peredachi Informatsii*, 7(4): 3–12, 1971.
- Lucas Maystre and Matthias Grossglauser. Fast and accurate inference of plackett–luce models. In *Advances in neural information processing systems*, pages 172–180, 2015.
- Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.
- Nihar B Shah and Martin J Wainwright. Simple, robust and optimal ranking from pairwise comparisons. *The Journal of Machine Learning Research*, 18(1):7246–7283, 2017.
- Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.