Spectral MLE: Top-K Rank Aggregation from Pairwise Comparisons
— Supplemental Materials —

Yuxin Chen*　　Changho Suh†

Abstract

This supplemental document presents details concerning analytical derivations that support the theorems made in the main text "Spectral MLE: Top-K Rank Aggregation from Pairwise Comparisons", accepted to the 32th International Conference on Machine Learning (ICML 2015). One can find here the detailed proof of Theorems 2 - 4.

1 Main Theorems

We repeat the main theorems as follows for convenience of presentation.

Theorem 2 (**Minimax Lower Bounds**). *Fix* $\epsilon \in \left(0, \frac{1}{2}\right)$*, and let* $\mathcal{G} \sim \mathcal{G}_{n,p_{\mathrm{obs}}}$*. If*

$$L \;\leq\; c\frac{(1-\epsilon)\log n - 2}{np_{\mathrm{obs}}\Delta_K^2} \tag{1}$$

holds for some absolute constant[1] $c > 0$*, then for any ranking scheme* ψ*, there exists a preference vector* $\boldsymbol{w}$ *with separation* Δ_K *such that the probability of error* $P_{\mathrm{e}}(\psi) \geq \epsilon$*.*

Theorem 3. *Let* $c_0, c_1, c_2, c_3 > 0$ *be some sufficiently large constants. Suppose that* $L = O(\mathrm{poly}(n))$*, the comparison graph* $\mathcal{G} \sim \mathcal{G}_{n,p_{\mathrm{obs}}}$ *with* $p_{\mathrm{obs}} > c_0 \log n / n$*, and assume that the separation measure satisfies*

$$\Delta_K \;>\; c_1\sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}. \tag{2}$$

Then with probability exceeding $1 - 1/n^2$*, Spectral MLE perfectly identifies the set of top-*K *ranked items, provided that the parameters obey* $T \geq c_2 \log n$ *and*

$$\xi_t := c_3\left\{\xi_{\min} + \frac{1}{2^t}\left(\xi_{\max} - \xi_{\min}\right)\right\}, \tag{3}$$

where $\xi_{\min} := \sqrt{\frac{\log n}{nLp_{\mathrm{obs}}}}$ *and* $\xi_{\max} := \sqrt{\frac{\log n}{p_{\mathrm{obs}}L}}$*.*

Theorem 4. *Suppose that* $\mathcal{G} \sim \mathcal{G}_{n,p_{\mathrm{obs}}}$ *with* $p_{\mathrm{obs}} > c_1 \log n / n$ *for some large constant* c_1*, and that there exists a score* $\hat{\boldsymbol{w}}^{\mathrm{ub}} \in [w_{\min}, w_{\max}]^n$ *independent of* $\mathcal{G}$ *satisfying*

$$\left|\hat{w}_i^{\mathrm{ub}} - w_i\right| \leq\; \xi w_{\max}, \quad \forall 1 \leq i \leq n; \tag{4}$$

$$\left\|\hat{\boldsymbol{w}}^{\mathrm{ub}} - \boldsymbol{w}\right\| \leq\; \delta\left\|\boldsymbol{w}\right\|. \tag{5}$$

*Department of Statistics, Stanford University, CA 94305, U.S.A.
†Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
[1] More precisely, $c = w_{\min}^4/(2w_{\max}^4)$.

Then with probability at least $1 - c_2 n^{-4}$ for some constant $c_2 > 0$, the coordinate-wise MLE

$$w_i^{\mathrm{mle}} := \arg \max_{\tau \in [w_{\min}, w_{\max}]} \mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) \tag{6}$$

satisfies

$$\left|w_i - w_i^{\mathrm{mle}}\right| < \frac{20\left(6 + \frac{\log L}{\log n}\right) w_{\max}^5}{w_{\min}^4} \max\left\{\delta + \frac{\xi \log n}{n p_{\mathrm{obs}}}, \quad \sqrt{\frac{\log n}{n p_{\mathrm{obs}} L}}\right\} \tag{7}$$

simultaneously for all scores $\hat{\boldsymbol{w}} \in [w_{\min}, w_{\max}]^n$ obeying

$$|\hat{w}_i - w_i| \leq \left|\hat{w}_i^{\mathrm{ub}} - w_i\right|, \quad 1 \leq i \leq n. \tag{8}$$

2 Performance Guarantees of Spectral MLE

In this section, we establish the theoretical guarantees of Spectral MLE in controlling the ranking accuracy and ℓ_∞ estimation errors, which are the subjects of Theorem 3 and Theorem 4. The proof of Theorem 3 relies heavily on the claim of Theorem 4; for this reason, we present the proofs of Theorem 3 and Theorem 4 in a reverse order. Before proceeding, we recall that the coordinate-wise log-likelihood of τ is given by

$$\frac{1}{L} \log \mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) := \sum_{j:(i,j) \in \mathcal{E}} y_{ij} \log \frac{\tau}{\tau + \hat{w}_j} + (1 - y_{ij}) \log \frac{\hat{w}_j}{\tau + \hat{w}_j}, \tag{9}$$

and we shall use $\boldsymbol{w}_{\setminus i}$ (resp. $\hat{\boldsymbol{w}}_{\setminus i}$) to denote the vector $\boldsymbol{w} = [w_1, \cdots, w_n]$ (resp. $\hat{\boldsymbol{w}} = [\hat{w}_1, \cdots, \hat{w}_n]$) excluding the entry w_i (resp. $\hat{w}_i$).

2.1 Proof of Theorem 4

To prove Theorem 4, we aim to demonstrate that for every $\tau \in [w_{\min}, w_{\max}]$ that is sufficiently separated from the ground truth w_i (or, more formally, $|\tau - w_i| \gtrsim \max\left\{\delta + \frac{\xi \log n}{n p_{\mathrm{obs}}}, \sqrt{\frac{\log n}{n p_{\mathrm{obs}} L}}\right\}$), the coordinate-wise likelihood satisfies

$$\log \mathcal{L}\left(w_i, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) > \log \mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) \tag{10}$$

and, therefore, τ cannot be the coordinate-wise MLE.

To begin with, we provide a lemma (which will be proved later) that concerns (10) for *any single* τ that is well separated from w_i.

Lemma 1. *Fix any $\gamma \geq 3$. Under the conditions of Theorem 4, for any $\tau \in [w_{\min}, w_{\max}]$ obeying*

$$|w_i - \tau| > \gamma \cdot \frac{w_{\max}^5}{w_{\min}^4} \max\left\{\frac{25}{4}\left(\delta + \frac{\xi \log n}{n p_{\mathrm{obs}}}\right), \quad 20\sqrt{\frac{\log n}{n p_{\mathrm{obs}} L}}\right\}, \tag{11}$$

one has

$$\frac{1}{L} \log \mathcal{L}\left(w_i, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L} \log \mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) > \frac{w_{\max}^6}{100 w_{\min}^6} \frac{\log n}{L}. \tag{12}$$

with probability exceeding $1 - 4n^{-\gamma} - 2n^{-10}$; this holds simultaneously for all $\hat{\boldsymbol{w}}_i \in [w_{\min}, w_{\max}]^n$ satisfying (8).

To establish Theorem 4, we still need to derive a uniform control over all τ satisfying (11). This will be accomplished via a standard covering argument. Specifically, for any small quantity $\epsilon > 0$, we construct a set $\mathcal{N}_\epsilon$ (called an ϵ-cover) within the interval $[w_{\min}, w_{\max}]$ such that for any $\tau \in [w_{\min}, w_{\max}]$, there exists an $\tau_0 \in \mathcal{N}_\epsilon$ obeying

$$|\tau - \tau_0| \leq \epsilon \quad \text{and} \quad |\tau_0 - w_i| \geq |\tau - w_i|. \tag{13}$$

It is self-evident that one can produce such a cover $\mathcal{N}_\epsilon$ with cardinality $\left\lceil \frac{w_{\max}}{\epsilon} \right\rceil + 1$. If we set $\gamma = 6 + \frac{\log L}{\log n}$ in Lemma 1, taking the union bound over $\mathcal{N}_\epsilon$ gives

$$\frac{1}{L}\log\mathcal{L}\left(w_i, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L}\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) > \frac{w_{\max}^6}{100 w_{\min}^6}\frac{\log n}{L} \tag{14}$$

simultaneously over all $\tau_0 \in \mathcal{N}_\epsilon$ obeying $|w_i - \tau_0| > \frac{\left(6+\frac{\log L}{\log n}\right) w_{\max}^5}{w_{\min}^4} \max\left\{\frac{25}{4}\left(\delta + \frac{\xi \log n}{n p_{\text{obs}}}\right),\ 20\sqrt{\frac{\log n}{n p_{\text{obs}} L}}\right\}$; this occurs with probability at least $1 - 4\left|\mathcal{N}_\epsilon\right| n^{-6-\frac{\log L}{\log n}} - 8\left|\mathcal{N}_\epsilon\right| n^{-10}$.

We then proceed by bounding the difference between $\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)$ and $\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)$. To achieve this, we first recognize that the Lipschitz constant of $\frac{1}{L}\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)$ (cf. (9)) is bounded above by

$$\begin{aligned}\frac{1}{L}\cdot\left|\frac{\partial \log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)}{\partial \tau}\right| &= \left|\sum_{j:(i,j)\in\mathcal{E}} y_{i,j}\left(\frac{1}{\tau} - \frac{1}{\tau + \hat{w}_j}\right) - \left(1 - y_{i,j}\right)\frac{1}{\tau + \hat{w}_j}\right| \\ &\overset{\text{(a)}}{\leq} \deg(i)\cdot\frac{2}{w_{\min}} \overset{\text{(b)}}{\leq} \frac{12}{5}\frac{n p_{\text{obs}}}{w_{\min}}.\end{aligned}$$

where (a) follows since

$$\left|y_{i,j}\left(\frac{1}{\tau} - \frac{1}{\tau + \tilde{w}_j}\right) - \left(1 - y_{i,j}\right)\frac{1}{\tau + \tilde{w}_j}\right| = \left|\frac{y_{i,j}}{\tau} - \frac{1}{\tau + \tilde{w}_j}\right| \leq \left|\frac{y_{i,j}}{\tau}\right| + \left|\frac{1}{\tau + \tilde{w}_j}\right| < \frac{2}{w_{\min}},$$

and (b) holds since $\deg(i) \leq 2.4 n p_{\text{obs}}$ with probability $1 - O\left(n^{-4}\right)$ as long as $p_{\text{obs}} > \frac{c_1 \log n}{n}$ for some sufficiently large $c_1 > 0$. As a result, by picking

$$\epsilon = \frac{\frac{w_{\max}^6}{100 w_{\min}^6}\frac{\log n}{L}}{\frac{12}{5}\frac{n p_{\text{obs}}}{w_{\min}}} = \frac{w_{\max}^6}{240 w_{\min}^5}\frac{\log n}{n p_{\text{obs}} L}, \tag{15}$$

one can make sure that for any $|\tau - \tau_0| \leq \epsilon$,

$$\frac{1}{L}\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L}\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) \leq \epsilon\cdot\frac{12}{5}\frac{n p_{\text{obs}}}{w_{\min}}, \tag{16}$$

$$\Rightarrow \quad \frac{1}{L}\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) < \frac{1}{L}\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) + \frac{w_{\max}^6}{100 w_{\min}^6}\frac{\log n}{L}. \tag{17}$$

In addition, with the above choice (15) of ϵ in place, the cardinality of the ϵ-cover is bounded above by

$$\left|\mathcal{N}_\epsilon\right| \leq \left\lceil \frac{w_{\max}}{\epsilon} \right\rceil + 1 = \left\lceil \frac{240 n p_{\text{obs}} L}{\log n}\cdot\frac{w_{\min}^5}{w_{\max}^5} \right\rceil + 1 \ll n^2 L$$

for any sufficiently large n.

Putting (14) and (17) together suggests that for *all* $\tau \in [w_{\min}, w_{\max}]$ sufficiently apart from the ground truth w_i, namely,

$$\forall \tau \in [w_{\min}, w_{\max}]: \quad |\tau - w_i| \geq \frac{\left(6 + \frac{\log L}{\log n}\right) w_{\max}^5}{w_{\min}^4}\max\left\{\frac{25}{4}\left(\delta + \frac{\xi \log n}{n p_{\text{obs}}}\right),\ 20\sqrt{\frac{\log n}{n p_{\text{obs}} L}}\right\}, \tag{18}$$

one necessarily has

$$\begin{aligned}&\frac{1}{L}\log\mathcal{L}\left(w_i, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L}\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) \\ &= \left\{\frac{1}{L}\log\mathcal{L}\left(w_i, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L}\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)\right\} + \left\{\frac{1}{L}\log\mathcal{L}\left(\tau_0, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right) - \frac{1}{L}\log\mathcal{L}\left(\tau, \hat{\boldsymbol{w}}_{\setminus i}; \boldsymbol{y}_i\right)\right\} \\ &> 0,\end{aligned} \tag{19}$$

with probability at least $1-4\left|\mathcal{N}_{\epsilon}\right| n^{-6-\frac{\log L}{\log n}}-O\left(n^{-4}\right) \geq 1-4n^{2}Ln^{-6-\frac{\log L}{\log n}}-O(n^{-4})=1-O(n^{-4})$. Consequently, any $\tau \in [w_{\min}, w_{\max}]$ that obeys (18) cannot be the coordinate-wise MLE, which in turn justifies the claim (7) of Theorem 4 (which is slightly weaker than what we prove here).

Proof of Lemma 1. We start by evaluating the true coordinate-wise likelihood gap

$$\log \mathcal{L}\left(w_{i}, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right)-\log \mathcal{L}\left(\tau, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right) \tag{20}$$

for any fixed $\tau \neq w_i$ independent of $\boldsymbol{y}_i$. Here, $\boldsymbol{y}_i := \{y_{i,j} \mid (i,j) \in \mathcal{E}\}$ is assumed to be generated under the BTL model parameterized by $\boldsymbol{w}$, which clearly obeys

$$\mathbb{E}\left[y_{i,j}\right]=\frac{w_{i}}{w_{i}+w_{j}} \quad \text{and} \quad \mathbf{Var}\left[y_{i,j}\right]=\frac{1}{L}\frac{w_{i}w_{j}}{\left(w_{i}+w_{j}\right)^{2}}.$$

In order to quantify the average value of (20), we rewrite the likelihood function as

$$\begin{aligned}
\frac{1}{L}\log \mathcal{L}\left(\tau, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right) &= \sum_{j:(i,j)\in\mathcal{E}}\left\{y_{i,j}\log\left(\frac{\tau}{\tau+w_{j}}\right)+\left(1-y_{i,j}\right)\log\left(\frac{w_{j}}{\tau+w_{j}}\right)\right\} && (21)\\
&= \sum_{j:(i,j)\in\mathcal{E}} y_{i,j}\log\left(\frac{\tau}{w_{j}}\right)+\sum_{j:(i,j)\in\mathcal{E}}\log\left(\frac{w_{j}}{\tau+w_{j}}\right). && (22)
\end{aligned}$$

Taking expectation w.r.t. $\boldsymbol{y}_i$ using the form (21) reveals that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{L}\log \mathcal{L}\left(w_{i}, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right)-\frac{1}{L}\log \mathcal{L}\left(\tau, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right)\middle|\,\mathcal{G}\right]=\sum_{j:(i,j)\in\mathcal{E}}\left\{\frac{w_{i}}{w_{i}+w_{j}}\log\left(\frac{\frac{w_{i}}{w_{i}+w_{j}}}{\frac{\tau}{\tau+w_{j}}}\right)+\frac{w_{j}}{w_{i}+w_{j}}\log\left(\frac{\frac{w_{j}}{w_{i}+w_{j}}}{\frac{w_{j}}{\tau+w_{j}}}\right)\right\}\\
&\quad=\sum_{j:(i,j)\in\mathcal{E}}\mathsf{KL}\left(\frac{w_{i}}{w_{i}+w_{j}}\,\Big\|\,\frac{\tau}{\tau+w_{j}}\right), \qquad (23)
\end{aligned}$$

where $\mathsf{KL}\left(p\|q\right)$ stands for the Kullback–Leibler (KL) divergence of Bernoulli (q) from Bernoulli (p). Using Pinsker's inequality [1, Theorem 2.33], that is, $\mathsf{KL}\left(p\|q\right) \geq 2\left(p-q\right)^{2}$, we arrive at the following lower bound

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{L}\log \mathcal{L}\left(w_{i}, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right)-\frac{1}{L}\log \mathcal{L}\left(\tau, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right)\middle|\,\mathcal{G}\right]\geq 2\sum_{j:(i,j)\in\mathcal{E}}\left(\frac{w_{i}}{w_{i}+w_{j}}-\frac{\tau}{\tau+w_{j}}\right)^{2}\\
&\quad=2\left(w_{i}-\tau\right)^{2}\sum_{j:(i,j)\in\mathcal{E}}\frac{w_{j}^{2}}{\left(w_{i}+w_{j}\right)^{2}\left(\tau+w_{j}\right)^{2}}. \qquad (24)
\end{aligned}$$

That being said, the true coordinate-wise likelihood of w_i strictly dominates that of τ in the mean sense.

However, when running Spectral MLE, we do not have access to the ground truth scores $\boldsymbol{w}_{\backslash i}$; what we actually compute is $\mathcal{L}(w_i, \hat{\boldsymbol{w}}_{\backslash i}; \boldsymbol{y}_i)$ (resp. $\mathcal{L}(\tau, \hat{\boldsymbol{w}}_{\backslash i}; \boldsymbol{y}_i)$) rather than $\mathcal{L}\left(\boldsymbol{w}; \boldsymbol{y}_i\right)$ (resp. $\mathcal{L}(\tau, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_i)$). Fortunately, such surrogate likelihoods are sufficiently close to the true coordinate-wise likelihoods, which we will show in the rest of the proof. For brevity, we shall denote respectively the heuristic and true log-likelihood functions by

$$\begin{cases}\hat{\ell}_{i}\left(w_{i}\right) & := \frac{1}{L}\log \mathcal{L}\left(w_{i}, \hat{\boldsymbol{w}}_{\backslash i}; \boldsymbol{y}_{i}\right),\\ \ell^{*}\left(w_{i}\right) & := \frac{1}{L}\log \mathcal{L}\left(w_{i}, \boldsymbol{w}_{\backslash i}; \boldsymbol{y}_{i}\right),\end{cases} \tag{25}$$

whenever it is clear from context. Note that $\hat{\boldsymbol{w}}_{\backslash i}$ could depend on $\boldsymbol{y}_i$.

As seen from (22), for any candidate $\tau \in [w_{\min}, w_{\max}]$, we can quantify the difference between $\hat{\ell}_i\left(\tau\right)$ and $\ell^{*}\left(\tau\right)$ as

$$\hat{\ell}_{i}\left(\tau\right)-\ell^{*}\left(\tau\right) = \sum_{j:(i,j)\in\mathcal{E}} y_{i,j}\log\left(\frac{w_{j}}{\hat{w}_{j}}\right)+\sum_{j:(i,j)\in\mathcal{E}}\left\{\log\left(\frac{\hat{w}_{j}}{\tau+\hat{w}_{j}}\right)-\log\left(\frac{w_{j}}{\tau+w_{j}}\right)\right\}. \tag{26}$$

As a consequence, the gap between the true loss $\ell^*(w_i) - \ell^*(\tau)$ and the surrogate loss $\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau)$ is given by

$$\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) - (\ell^*(w_i) - \ell^*(\tau)) = \hat{\ell}_i(w_i) - \ell^*(w_i) - \left(\hat{\ell}_i(\tau) - \ell^*(\tau)\right)$$

$$= \sum_{j:(i,j)\in\mathcal{E}} \left\{ \log\left(\frac{\hat{w}_j}{w_i + \hat{w}_j}\right) - \log\left(\frac{w_j}{w_i + w_j}\right) - \left(\log\left(\frac{\hat{w}_j}{\tau + \hat{w}_j}\right) - \log\left(\frac{w_j}{\tau + w_j}\right)\right)\right\} \tag{27}$$

$$= \sum_{j:(i,j)\in\mathcal{E}} \left\{ \log\left(\frac{\tau + \hat{w}_j}{w_i + \hat{w}_j}\right) - \log\left(\frac{\tau + w_j}{w_i + w_j}\right)\right\}. \tag{28}$$

This gap relies on the function

$$g(t) := \log\left(\frac{\tau + t}{w_i + t}\right) - \log\left(\frac{\tau + w_j}{w_i + w_j}\right), \quad t \in [w_{\min}, w_{\max}],$$

which apparently obeys the following two properties: (i) $g(w_j) = 0$; (ii)

$$\left|\frac{\partial g(t)}{\partial t}\right| = \left|\frac{1}{\tau + t} - \frac{1}{w_i + t}\right| = \frac{|\tau - w_i|}{(w_i + t)(\tau + t)} \le \frac{|\tau - w_i|}{4w_{\min}^2}, \quad \forall t \in [w_{\min}, w_{\max}].$$

Taken together these two properties demonstrate that

$$|g(t)| \le \frac{1}{4w_{\min}^2} |\tau - w_i| \, |t - w_j|, \quad \forall t \in [w_{\min}, w_{\max}].$$

Substitution into (28) gives

$$\begin{aligned}\left|\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) - (\ell^*(w_i) - \ell^*(\tau))\right| &\le \frac{1}{4w_{\min}^2} |\tau - w_i| \sum_{j:(i,j)\in\mathcal{E}} |\hat{w}_j - w_j| \\ &\le \frac{1}{4w_{\min}^2} |\tau - w_i| \sum_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right|. \end{aligned} \tag{29}$$

Notably, this is a deterministic inequality which holds for all $\hat{\boldsymbol{w}}_j$ obeying $|\hat{w}_j - w_j| \le |\hat{w}_j^{\mathrm{ub}} - w_j|$ $(1 \le j \le n)$. A desired property of the upper bound (29) is that it is independent of $\mathcal{G}$ and the data $\boldsymbol{y}_i$, due to our assumption on $\hat{\boldsymbol{w}}^{\mathrm{ub}}$.

We now move on to develop an upper bound on (29). From our assumptions on the initial estimate, we have

$$\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|^2 \le \|\hat{\boldsymbol{w}}^{\mathrm{ub}} - \boldsymbol{w}\|^2 \le \delta^2 \|\boldsymbol{w}\|^2 \le n w_{\max}^2 \delta^2.$$

Since $\mathcal{G}$ and $\hat{\boldsymbol{w}}^{\mathrm{ub}}$ are statistically independent, this inequality immediately gives rise to the following two consequences:

$$\begin{aligned}\mathbb{E}\left[\sum\nolimits_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right|\right] &= p_{\mathrm{obs}} \|\hat{\boldsymbol{w}}^{\mathrm{ub}} - \boldsymbol{w}\|_1 \le p_{\mathrm{obs}}\sqrt{n}\|\hat{\boldsymbol{w}}^{\mathrm{ub}} - \boldsymbol{w}\| \\ &\le n p_{\mathrm{obs}} w_{\max} \delta \end{aligned} \tag{30}$$

and

$$\mathbb{E}\left[\sum\nolimits_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right|^2\right] = p_{\mathrm{obs}} \|\hat{\boldsymbol{w}}^{\mathrm{ub}} - \boldsymbol{w}\|_2^2 \le n p_{\mathrm{obs}} w_{\max}^2 \delta^2. \tag{31}$$

Recall our assumption that $\max_j \left|\hat{w}_j^{\mathrm{ub}} - w_j\right| \le \xi w_{\max}$. For any fixed $\gamma \ge 3$, if $p_{\mathrm{obs}} > \frac{2\log n}{n}$, then with probability at least $1 - 2n^{-\gamma}$,

$$\begin{aligned}\sum_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right| &\overset{\text{(i)}}{\le} \mathbb{E}\left[\sum_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right|\right] + \sqrt{2\gamma \log n \cdot \mathbb{E}\left[\sum\nolimits_{j:(i,j)\in\mathcal{E}} \left|\hat{w}_j^{\mathrm{ub}} - w_j\right|^2\right]} + \frac{2\gamma}{3}\xi w_{\max} \log n \\ &\le n p_{\mathrm{obs}} w_{\max} \delta + \sqrt{2\gamma \cdot n p_{\mathrm{obs}} \log n}\, w_{\max} \delta + \frac{2\gamma}{3} \xi w_{\max} \log n \\ &\overset{\text{(ii)}}{\le} (1 + \sqrt{\gamma})\, n p_{\mathrm{obs}} w_{\max} \delta + \frac{2\gamma}{3} \xi w_{\max} \log n \\ &\overset{\text{(iii)}}{\le} \gamma n p_{\mathrm{obs}} w_{\max} \delta + \gamma \xi w_{\max} \log n, \end{aligned}$$

where (i) comes from the Bernstein inequality as given in Lemma 4, (ii) follows since $\log n < \frac{p_{\text{obs}} n}{2}$ by assumption, and (iii) arises since $1+\sqrt{\gamma} \leq \gamma$ whenever $\gamma \geq 3$. This combined with (29) allows us to control

$$\left| \hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) - (\ell^*(w_i) - \ell^*(\tau)) \right| \leq \frac{|\tau - w_i| \gamma w_{\max}}{4 w_{\min}^2} (n p_{\text{obs}} \delta + \xi \log n) \tag{32}$$

with high probability.

The above arguments basically reveal that $\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau)$ is reasonably close to $\ell^*(w_i) - \ell^*(\tau)$. Thus, to show that $\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) > 0$, it is sufficient to develop a lower bound on $\ell^*(w_i) - \ell^*(\tau)$ that exceeds the gap (32). In expectation, the preceding inequality (24) gives

$$\begin{aligned} \mathbb{E}\left[\ell^*(w_i) - \ell^*(\tau) \mid \mathcal{G}\right] &\geq 2(w_i - \tau)^2 \sum_{j:(i,j)\in\mathcal{E}} \frac{w_j^2}{(w_i + w_j)^2 (\tau + w_j)^2} \\ &\geq \frac{w_{\min}^2}{8 w_{\max}^4} (w_i - \tau)^2 \deg(i). \end{aligned} \tag{33}$$

Recognizing that $y_{i,j} = \frac{1}{L}\sum_{l=1}^{L} y_{i,j}^{(l)}$ is a sum of independent random variables $y_{i,j}^{(l)} \sim \text{Bernoulli}\left(\frac{w_i}{w_i + w_j}\right)$, we can control the conditional variance as

$$\begin{aligned} &\mathbf{Var}\left[\ell^*(w_i) - \ell^*(\tau) \mid \mathcal{G}\right] \overset{\text{(a)}}{=} \mathbf{Var}\left[\sum_{j:(i,j)\in\mathcal{E}} y_{i,j} \log\left(\frac{w_i}{\tau}\right) \,\middle|\, \mathcal{G} \right] \\ &= \log^2\left(\frac{w_i}{\tau}\right) \sum_{j:(i,j)\in\mathcal{E}} \frac{1}{L} \frac{w_i w_j}{(w_i + w_j)^2} \overset{\text{(b)}}{\leq} \frac{1}{L} \frac{(w_i - \tau)^2}{\min\{w_i^2, \tau^2\}} \sum_{j:(i,j)\in\mathcal{E}} \frac{w_{\max}^2}{4 w_{\min}^2} \\ &\leq \frac{w_{\max}^2}{4 w_{\min}^4} \cdot \frac{1}{L} (w_i - \tau)^2 \deg(i), \end{aligned} \tag{34}$$

where (a) is an immediate consequence of (22), and (b) follows since $\left|\log \frac{\beta}{\alpha}\right| \leq \frac{\beta - \alpha}{\alpha}$ for any $\beta > \alpha > 0$. Note that $0 \leq \frac{1}{L} y_{i,j}^{(l)} \leq \frac{1}{L}$. Making use of the Bernstein inequality, (33) and (34) suggests that: conditional on $\mathcal{G}$,

$$\begin{aligned} \ell^*(w_i) - \ell^*(\tau) &\geq \mathbb{E}\left[\ell^*(w_i) - \ell^*(\tau) \mid \mathcal{G}\right] - \sqrt{2\gamma \mathbf{Var}\left[\ell^*(w_i) - \ell^*(\tau) \mid \mathcal{G}\right] \log n} - \frac{2\gamma \log n \cdot \left|\log\left(\frac{w_i}{\tau}\right)\right|}{3L} \\ &\geq \frac{w_{\min}^2}{8 w_{\max}^4} (w_i - \tau)^2 \deg(i) - \frac{\sqrt{2\gamma} w_{\max} |w_i - \tau|}{2 w_{\min}^2} \sqrt{\frac{\deg(i) \log n}{L}} - \frac{2\gamma |w_i - \tau| \log n}{3 L w_{\min}} \end{aligned} \tag{35}$$

holds with probability at least $1 - 2n^{-\gamma}$, where the last inequality follows again from the inequality $\left|\log\left(\frac{\beta}{\alpha}\right)\right| \leq \frac{\beta - \alpha}{\alpha}$ for any $\beta \geq \alpha > 0$.

The above bound relies on $\deg(i)$, which is on the order of $n p_{\text{obs}}$ with high probability. More precisely, taking the Chernoff bound [2, Corollary 4.6] as well as the union bound reveals that: there exists some constant $c_1 > 1$ such that if $p_{\text{obs}} > \frac{c_1 \log n}{n}$, then

$$\frac{4}{5} n p_{\text{obs}} < \deg(i) < \frac{6}{5} n p_{\text{obs}}, \quad \forall 1 \leq i \leq n \tag{36}$$

with probability at least $\frac{2}{n^{10}}$. This taken collectively with (35) and the assumption $n p_{\text{obs}} > 2 \log n$ implies that

$$
\begin{aligned}
\ell^*(w_i) - \ell^*(\tau) &\geq \frac{w_{\min}^2}{8w_{\max}^4}(w_i-\tau)^2 \cdot \frac{4}{5}np_{\mathrm{obs}} - \sqrt{\frac{\gamma}{2}}\frac{w_{\max}|w_i-\tau|}{w_{\min}^2}\sqrt{\frac{6np_{\mathrm{obs}}\log n}{5L}} - \frac{2\gamma|w_i-\tau|\log n}{3Lw_{\min}} \\
&\geq \frac{w_{\min}^2}{10w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}} - \left(\sqrt{\frac{3\gamma}{5}} + \frac{2\gamma}{3}\frac{1}{\sqrt{2}}\right)\frac{w_{\max}|w_i-\tau|}{w_{\min}^2}\sqrt{\frac{np_{\mathrm{obs}}\log n}{L}} \\
&\geq \frac{w_{\min}^2}{10w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}} - \gamma\frac{w_{\max}|w_i-\tau|}{w_{\min}^2}\sqrt{\frac{np_{\mathrm{obs}}\log n}{L}} \qquad (37) \\
&\geq \frac{w_{\min}^2}{20w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}} \qquad (38)
\end{aligned}
$$

with probability at least $1-2n^{-\gamma}-2n^{-10}$, as long as

$$\gamma \cdot \frac{w_{\max}|w_i-\tau|}{w_{\min}^2}\sqrt{\frac{np_{\mathrm{obs}}\log n}{L}} \leq \frac{w_{\min}^2}{20w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}}$$

or, equivalently,

$$|w_i-\tau| \geq \frac{20\gamma \cdot w_{\max}^5}{w_{\min}^4}\sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}. \qquad (39)$$

Finally, we are ready to control $\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau)$ from below. Putting (32) and (38) together, we see that with high probability,

$$
\begin{aligned}
\hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) &\geq \ell^*(w_i) - \ell^*(\tau) - \frac{|\tau-w_i|\gamma w_{\max}(np_{\mathrm{obs}}\delta + \xi\log n)}{4w_{\min}^2} \\
&\geq \frac{w_{\min}^2}{20w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}} - \frac{|\tau-w_i|\gamma w_{\max}}{4w_{\min}^2}(np_{\mathrm{obs}}\delta + \xi\log n) \\
&> \frac{w_{\min}^2}{100w_{\max}^4}(w_i-\tau)^2 np_{\mathrm{obs}} \qquad (40) \\
&> \frac{w_{\max}^6}{100w_{\min}^6}\frac{\log n}{L}, \qquad (41)
\end{aligned}
$$

where (40) holds under the condition

$$|\tau - w_i| > \frac{25\gamma w_{\max}^5}{4w_{\min}^4}\left(\delta + \frac{\xi\log n}{np_{\mathrm{obs}}}\right),$$

and (41) follows from the assumption (39). This establishes the claim (12). □

2.2 Proof of Theorem 3

The accuracy of top-K identification is closely related to the ℓ_∞ error of the score estimate. In the sequel, we shall assume that $w_{\max}=1$ to simplify presentation, and our goal is to demonstrate that

$$\left\|\boldsymbol{w}^{(t)} - \boldsymbol{w}\right\|_\infty \lesssim \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}} + \frac{1}{2^t}\sqrt{\frac{\log n}{p_{\mathrm{obs}}L}} \asymp \xi_t, \quad \forall t \in \mathbb{N}, \qquad (42)$$

where

$$\xi_t := c_3\left\{\xi_{\min} + \frac{1}{2^t}(\xi_{\max} - \xi_{\min})\right\}, \quad \forall t \geq -1 \qquad (43)$$

with $\xi_{\min} = \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}$ and $\xi_{\max} = \sqrt{\frac{\log n}{p_{\mathrm{obs}}L}}$. If $T \geq c_2\log n$ for some sufficiently large $c_2>0$, then this gives

$$\left\|\boldsymbol{w}^{(T)} - \boldsymbol{w}\right\|_\infty \asymp \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}} = \xi_{\min}.$$

The key implication is the following: if $w_K - w_{K-1} \geq c_1\sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}$ for some sufficiently large $c_1 > 0$, then

$$w_i^{(T)} - w_j^{(T)} \geq w_i - w_j - \left|w_i^{(T)} - w_i\right| - \left|w_j^{(T)} - w_j\right| \geq w_K - w_{K+1} - 2\left\|\boldsymbol{w}^{(T)} - \boldsymbol{w}\right\| > 0$$

for all $1 \leq i \leq K$ and $j \geq K+1$, indicating that Spectral MLE will output the first K items as desired. The remaining proof then boils down to showing (42).

We start from $t = 0$. When the initial estimate $\boldsymbol{w}^{(0)}$ is computed by Rank Centrality, the ℓ_2 estimation error satisfies [3]

$$\frac{\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\|}{\|\boldsymbol{w}\|} \leq c_4\sqrt{\frac{\log n}{np_{\mathrm{obs}}L}} = c_4\xi_{\min} := \delta \tag{44}$$

with high probability, where $c_4 > 0$ is some universal constant independent of n, p_{obs}, L and Δ_K. A by-product of this result is an upper bound

$$\left\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\right\|_\infty \leq \left\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\right\| \leq \delta\|\boldsymbol{w}\| \leq \delta\sqrt{n} = c_4\sqrt{\frac{\log n}{p_{\mathrm{obs}}L}}, \tag{45}$$

which together with the fact $\left\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\right\|_\infty \leq w_{\max} - w_{\min} \leq 1$ give

$$\left\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\right\|_\infty \leq \min\left\{c_4\sqrt{\frac{\log n}{p_{\mathrm{obs}}L}},\ 1\right\} = \min\left\{c_4\xi_{\max}, 1\right\}. \tag{46}$$

This justifies that $\boldsymbol{w}^{(0)}$ satisfies the claim (42). Notably, $\boldsymbol{w}^{(0)}$ is independent of $\mathcal{E}^{\mathrm{iter}}$ and $\boldsymbol{y}^{\mathrm{iter}}$ and, therefore, independent of the iterative steps.

In what follows, we divide the iterative stage into two phases: (1) $t \leq T_0$ and (2) $t > T_0$, where T_0 is a threshold such that

$$\xi_t \geq c_{10}\xi_{\min} = c_{10}\sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}, \quad \text{iff} \quad t \leq T_0, \tag{47}$$

for some large constant c_{10}. As is seen from the definition of ξ_t, $T_0 \lesssim \log n$ holds as long as $L = O(\mathrm{poly}(n))$.

For the case where $t \leq T_0$, we proceed by induction on t w.r.t. the following hypotheses:

- $\mathcal{M}_t$: $\left\|\boldsymbol{w}^{(\mathrm{mle})} - \boldsymbol{w}\right\|_\infty \leq \frac{1}{2}\xi_t$ holds at the t^{th} iteration (the iteration where we compute $\boldsymbol{w}^{(t+1)}$);
- $\mathcal{B}_t$: *all* entries $w_i^{(\tau)}$ of $\boldsymbol{w}^{(\tau)}$ $(\tau \leq t-1)$ satisfying $|w_i^{(\tau)} - w_i| \geq 1.5\xi_t$ have been replaced by time t;
- $\mathcal{H}_t$: *none* of the entries $w_i^{(\tau)}$ $(\tau \leq t-1)$ satisfying $|w_i^{(\tau)} - w_i| \leq \frac{1}{2}\xi_t$ have been replaced by time t.

We note that $\mathcal{B}_t$ and $\mathcal{H}_t$ are immediate consequences of $\mathcal{M}_t$, $\mathcal{B}_{t-1}$, and $\mathcal{H}_{t-1}$. First of all, with $\mathcal{B}_{t-1}$ in mind, we only need to examine those entries $w_i^{(\tau)}$ obeying $|w_i^{(\tau)} - w_i| \geq 1.5\xi_t$ that have not been replaced by time $t-1$. To this end, we recall that Spectral MLE replaces $w_i^{(\tau)}$ iff $|w_i^{(\tau)} - w_i^{\mathrm{mle}}| > \xi_t$. With $\mathcal{M}_t$ in place, for each i obeying $|w_i^{(\tau)} - w_i| \geq 1.5\xi_t$, one has

$$|w_i^{(\tau)} - w_i^{\mathrm{mle}}| \geq |w_i^{(\tau)} - w_i| - |w_i^{\mathrm{mle}} - w_i| > 1.5\xi_t - \frac{1}{2}\xi_t = \xi_t$$

and hence it is necessarily replaced by w_i^{mle} by time t. Similarly, for any i obeying $|w_i^{(\tau)} - w_i| \leq 0.5\xi_t$, one has

$$|w_i^{(\tau)} - w_i^{\mathrm{mle}}| \leq |w_i^{(\tau)} - w_i| + |w_i^{\mathrm{mle}} - w_i| < \frac{1}{2}\xi_t + \frac{1}{2}\xi_t = \xi_t$$

and, therefore, it cannot be replaced by time t. These establish $\mathcal{B}_t$ and $\mathcal{H}_t$. As a consequence, it suffices to verify $\mathcal{M}_t$, which is achieved by induction.

When $t = 0$, applying Theorem 4 and setting $\boldsymbol{w}^{\mathrm{ub}} = \boldsymbol{w}^{(0)}$, we see that

$$\left\|\boldsymbol{w}^{\mathrm{mle}} - \boldsymbol{w}\right\|_\infty \leq c_7\xi_{\min} + c_9\frac{\log n}{np_{\mathrm{obs}}}\xi_{\max}$$

for some universal constants $c_7, c_9 > 0$, where we have made use of the properties (44) and (46). When c_{10} is sufficiently large, the definition of T_0 (cf. (47)) gives $\xi_0 \gg c_7\sqrt{\frac{\log n}{np_{\text{obs}}L}}$; additionally, $c_9 \frac{\log n}{np_{\text{obs}}}\xi_{\max} \ll \xi_{\max} \leq \xi_0$ holds as long as $\frac{\log n}{np_{\text{obs}}}$ is sufficiently small. Putting these conditions together gives

$$\left\|\boldsymbol{w}^{\text{mle}} - \boldsymbol{w}\right\|_\infty \leq c_7\xi_{\min} + c_9 c_4 \frac{\log n}{np_{\text{obs}}}\xi_{\max} < \frac{1}{2}\xi_0,$$

which verifies the property $\mathcal{M}_0$.

We now turn to extending these inductive hypotheses to the t^{th} iteration, assuming that all of them hold up to time $t-1$. Taken together $\mathcal{M}_{t-1}$ and $\mathcal{B}_{t-1}$ immediately reveal that

$$\left\|\boldsymbol{w}^{(t)} - \boldsymbol{w}\right\|_\infty \leq 1.5\xi_{t-1}. \tag{48}$$

In order to invoke Theorem 4 for the coordinate-wise MLEs, we need to construct a looser auxiliary score estimate $\boldsymbol{w}^{\text{ub}}$. With $\mathcal{B}_{t-1}$, $\mathcal{H}_{t-1}$ and (48) in mind, we propose a candidate for the t^{th} iteration as follows[2]

$$w_i^{\text{ub}} = \begin{cases} w_i + 1.5\xi_{t-1}, & \text{if } |w_i^{(0)} - w_i| > \frac{1}{2}\xi_{t-1}, \\ w_i^{(0)} & \text{else.} \end{cases} \tag{49}$$

which is clearly independent of $\mathcal{E}^{\text{iter}}$ and $\boldsymbol{y}^{\text{iter}}$. According to $\mathcal{B}_{t-1}$ and $\mathcal{H}_{t-1}$, (i) none of the entries $w_i^{(0)}$ with $|w_i^{(0)} - w_i| \leq \frac{1}{2}\xi_{t-1}$ have been replaced so far; (ii) if an entry $w_i^{(0)}$ has ever been replaced, then the error of the new iterate cannot exceed $1.5\xi_{t-1}$ (otherwise it'll be replaced by the MLE in time $t-1$ which gives an error below $0.5\xi_{t-1}$). As a result, $\boldsymbol{w}^{\text{ub}}$ clearly satisfies

$$\left|w_i^{(t)} - w_i\right| \leq \left|w_i^{\text{ub}} - w_i\right| \leq 1.5\xi_{t-1}, \tag{50}$$

$$\text{and} \quad \left\|\boldsymbol{w}^{(t)} - \boldsymbol{w}\right\| \leq \left\|\boldsymbol{w}^{(\text{ub})} - \boldsymbol{w}\right\| \overset{\text{(i)}}{\leq} \frac{1.5\xi_{t-1}}{0.5\xi_{t-1}}\left\|\boldsymbol{w}^{(0)} - \boldsymbol{w}\right\| \leq 3\delta\|\boldsymbol{w}\|. \tag{51}$$

Here, (i) arises since if $w_i^{(0)}$ is replaced, then the error $\left|w_i^{(0)} - w_i\right|$ is at least $0.5\xi_{t-1}$, whereas the replaced pointwise error is $1.5\xi_{t-1}$, which inflates the original error by no more than 3 times. With these in place, applying Theorem 4 gives

$$\left\|\boldsymbol{w}^{\text{mle}} - \boldsymbol{w}\right\|_\infty \leq c_8\xi_{\min} + 1.5c_9\frac{\log n}{np_{\text{obs}}}\xi_{t-1},$$

which relies on the fact $\delta \lesssim \sqrt{\frac{\log n}{np_{\text{obs}}L}}$. Recognize that

$$\xi_t \gg c_8\xi_{\min} \quad \text{and} \quad 1.5c_9\frac{\log n}{np_{\text{obs}}}\xi_{t-1} \ll \xi_t$$

hold in the regime where $t \leq T_0$ and $\frac{\log n}{np_{\text{obs}}} \ll 1$, which taken together give

$$\left\|\boldsymbol{w}^{\text{mle}} - \boldsymbol{w}\right\|_\infty \leq \frac{1}{2}\xi_t$$

as claimed in $\mathcal{M}_t$. Having verified these inductive hypotheses, we see from the above argument that the worst case ℓ_∞ error bound at the t^{th} iteration is at most $1.5\xi_t$, which in turn leads to the claim (42) for any $t \leq T_0$.

[2] Careful readers will note that when $|w_i^{(0)} - w_i| \geq \frac{1}{2}\Delta_{t-1}$, the resulting w_i^{ub} might exceed the range $[w_{\min}, w_{\max}]$. This can be easily addressed if we do the following: (1) change w_i^{ub} to $w_i - 1.5\Delta_{t-1}$ instead if $w_i - 1.5\Delta_{t-1} \in [w_{\min}, w_{\max}]$; (2) if it is still infeasible, set w_i^{ub} to be $w_{\max}$ if $|w_i - w_{\max}| > |w_i - w_{\min}|$ and $w_{\min}$ otherwise. For simplicity of presentation, however, we omit these boundary situations and assume $w_i + 1.5\Delta_{t-1} \leq w_{\max}$ throughout, which will not change the results anyway.

Starting from $t = T_0 + 1$, we fix the auxiliary score as follows

$$w_i^{\mathrm{ub}} = \begin{cases} w_i + 1.5\xi_{T_0}, & \text{if } |w_i^{(0)} - w_i| > \frac{1}{2}\xi_\infty, \\ w_i^{(0)} & \text{else}, \end{cases} \tag{52}$$

where we recall that $\xi_\infty = c_3 \xi_{\min}$ and $\xi_{T_0} = c_{10}\xi_{\min}$. This apparently satisfies

$$\left| w_i^{(t)} - w_i \right| \le \left| w_i^{\mathrm{ub}} - w_i \right| \le 1.5\xi_{T_0}$$

for $t = T_0 + 1$, due to the preceding analysis for $t \le T_0$. Moreover, the number of indices that satisfy $|w_i^{(0)} - w_i| > \frac{1}{2}\xi_\infty$, denoted by k, obeys

$$k \cdot \left(\frac{1}{2}\xi_\infty \right)^2 \le \left\| \boldsymbol{w} - \boldsymbol{w}^{(0)} \right\|^2 \le \delta^2 \|\boldsymbol{w}\|^2 \quad \Longleftrightarrow \quad k \cdot \le \frac{4\delta^2 \|\boldsymbol{w}\|^2}{\xi_\infty^2},$$

which further gives

$$\begin{aligned} \left\| \boldsymbol{w}^{\mathrm{ub}} - \boldsymbol{w} \right\|^2 &\le \left\| \boldsymbol{w}^{(0)} - \boldsymbol{w} \right\|^2 + \sum_{i:\ |w_i^{(0)} - w_i| > \frac{1}{2}\Delta_\infty} \left(1.5\xi_{T_0}\right)^2 \le \delta^2 \|\boldsymbol{w}\|^2 + 2.25k\xi_{T_0}^2 \\ &\le \delta^2 \|\boldsymbol{w}\|^2 \left(1 + \frac{9\xi_{T_0}^2}{\xi_\infty^2} \right). \end{aligned}$$

If we pick $\frac{c_{10}}{c_3} = \frac{\xi_{T_0}}{\xi_\infty} \le \sqrt{2}$, then the above inequality gives rise to

$$\left\| \boldsymbol{w}^{\mathrm{ub}} - \boldsymbol{w} \right\| \le \sqrt{19}\delta \|\boldsymbol{w}\|.$$

Apply Theorem 4 to deduce

$$\left\| \boldsymbol{w}^{\mathrm{mle}} - \boldsymbol{w} \right\|_\infty \lesssim \delta + \frac{\log n}{np_{\mathrm{obs}}}\xi_{T_0} + \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}} \asymp \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}} \ll \frac{1}{2}\xi_\infty,$$

as long as $\frac{\log n}{p_{\mathrm{obs}} n}$ is small and c_{10}, c_3 are sufficiently large.

The main point of the above calculation is that: for any entry $w_i^{(0)}$ satisfying $|w_i^{(0)} - w_i| < \frac{1}{2}\xi_\infty$, one must have

$$\left| w_i^{(0)} - w_i^{\mathrm{mle}} \right| \le \left| w_i^{(0)} - w_i \right| + \left| w_i^{(\mathrm{mle})} - w_i \right| < \xi_\infty < \xi_t,$$

and hence it will never be replaced. As a result, the auxiliary score (52) remains valid for all iterations that follow. Putting the above arguments together we obtain

$$\left\| \boldsymbol{w}^{(t)} - \boldsymbol{w} \right\|_\infty \le \frac{1}{2}\xi_\infty \asymp \sqrt{\frac{\log n}{np_{\mathrm{obs}}L}}, \quad t > T_0.$$

This establishes the claim (42) for $t > T_0$, and in turn finishes the proof of the theorem.

3 Minimax Lower Bound

This section establishes the minimax lower limit given in Theorem 2. To bound the minimax probability of error, we proceed by constructing a finite set of hypotheses, followed by an analysis based on classical Fano-type argument. For notational simplicity, each hypothesis is represented by a permutation σ over $[n]$, and we denote by $\sigma(i)$ and $\sigma([K])$ the index of the i^{th} ranked item and the index set of all top-K items, respectively.

We now single out a set of hypotheses and some prior to be imposed on them. Suppose that the values of $\boldsymbol{w}$ are fixed up to permutation in such a way that

$$w_{\sigma(i)} = \begin{cases} w_K, & 1 \le i \le K, \\ w_{K+1}, & K < i \le n, \end{cases}$$

where we abuse the notation w_K, w_{K+1} to represent any two values satisfying

$$\frac{w_K - w_{K+1}}{w_{\max}} = \Delta_K > 0.$$

Below we suppose that the ranking scheme is informed of the values w_K, w_{K+1}, which only makes the ranking task easier. In addition, we impose a uniform prior over a collection $\mathcal{M}$ of $M := \max\{K, n-K\} + 1$ hypotheses regarding the permutation: if $K < n/2$, then

$$\mathbb{P}\left\{\sigma\left([K]\right) = \mathcal{S}\right\} = \frac{1}{M}, \ \text{ if } \mathcal{S} = \{2, \cdots, K\} \cup \{i\}, \qquad (i = 1, K+1, \cdots, n); \tag{53}$$

if $K \ge n/2$, then

$$\mathbb{P}\left\{\sigma\left([K]\right) = \mathcal{S}\right\} = \frac{1}{M}, \ \text{ if } \mathcal{S} = \{1, \cdots, K+1\} \setminus \{i\}, \qquad (i = 1, \cdots, K+1). \tag{54}$$

In words, each alternative hypothesis is generated by swapping two indices of the hypothesis obeying $\sigma([K]) = [K]$. Denoting by $P_{\mathrm{e},M}$ the average probability of error with respect to the prior we construct, one can easily verify that the minimax probability of error is at least $P_{\mathrm{e},M}$.

This Bayesian probability of error will be bounded using classical Fano-type bounds. To accommodate partial observation, we introduce an erased version of $\boldsymbol{y}_{i,j} := (y_{i,j}^{(1)}, \cdots, y_{i,j}^{(L)})$ such that

$$\boldsymbol{z}_{i,j} = \begin{cases} \boldsymbol{y}_{i,j}, & \text{with probability } p_{\mathrm{obs}}, \\ \text{erasure}, & \text{else}, \end{cases}$$

and set $\boldsymbol{Z} := \{\boldsymbol{z}_{i,j}\}_{1 \le i \le j \le n}$. With a slight abuse of notation, we denote by σ and $\hat{\sigma}$ the ground truth permutation and the output of any ranking procedure, respectively. Making use of (53) and (54) gives

$$\begin{aligned}
\log M &= H(\sigma) = I(\sigma; \hat{\sigma}) + H(\sigma | \hat{\sigma}) \\
&\overset{(a)}{\le} I(\sigma; \boldsymbol{Z}) + 1 + P_{\mathrm{e},M} \log M \\
&\overset{(b)}{\le} \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \mathsf{KL}\left(\mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_1} \parallel \mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_2}\right) + 1 + P_{\mathrm{e},M} \log M \\
&\overset{(c)}{=} \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{i \ne j} \mathsf{KL}\left(\mathbb{P}_{\boldsymbol{z}_{i,j}|\sigma=\sigma_1} \parallel \mathbb{P}_{\boldsymbol{z}_{i,j}|\sigma=\sigma_2}\right) + 1 + P_{\mathrm{e},M} \log M \\
&= \frac{p_{\mathrm{obs}}}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{i \ne j} \mathsf{KL}\left(\mathbb{P}_{\boldsymbol{y}_{i,j}|\sigma=\sigma_1} \parallel \mathbb{P}_{\boldsymbol{y}_{i,j}|\sigma=\sigma_2}\right) + 1 + P_{\mathrm{e},M} \log M \\
&\overset{(d)}{=} \frac{p_{\mathrm{obs}} L}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{i \ne j} \mathsf{KL}\left(\mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel \mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_2}\right) + 1 + P_{\mathrm{e},M} \log M \\
&\overset{(e)}{\le} \frac{2 w_{\max}^4}{w_{\min}^4} n p_{\mathrm{obs}} L \Delta_K^2 + 1 + P_{\mathrm{e},M} \log M,
\end{aligned}$$

where $H(X)$, $I(X;Y)$, and $\mathsf{KL}(P \parallel Q)$ denote the entropy, mutual information, and Kullback–Leibler (KL) divergence, respectively. Here, (a) results from the data processing inequality and Fano's inequality [4]; (b) arises from Lemma 2 (see below); (c) follows from the independence assumption of the $\boldsymbol{z}_{i,j}$'s; (d) is a consequence of the fact that $y_{i,j}^{(\ell)}$ $(1 \le l \le L)$ are i.i.d.; and (e) follows from Lemma 3 (see below). This immediately yields

$$P_{\mathrm{e},M} \ge \frac{\log M - \frac{2w_{\max}^4}{w_{\min}^4} n p_{\mathrm{obs}} L \Delta_K^2 - 1}{\log M}.$$

Consequently, one would have $P_{\mathrm{e}} \ge P_{\mathrm{e},M} \ge \epsilon$ if

$$\frac{2w_{\max}^4}{w_{\min}^4} n p_{\mathrm{obs}} L \Delta_K^2 \le (1-\epsilon)\log M - 1.$$

Since $|\mathcal{M}| = M \ge \frac{n}{2}$, the above condition is necessarily satisfied when

$$\frac{2w_{\max}^4}{w_{\min}^4} n p_{\mathrm{obs}} L \Delta_K^2 \le (1-\epsilon)\log n - 2 \quad \Longleftrightarrow \quad L \le \frac{w_{\min}^4}{2w_{\max}^4} \cdot \frac{(1-\epsilon)\log n - 2}{n p_{\mathrm{obs}} \Delta_K^2},$$

which finishes the proof.

Lemma 2. *Under the prior* (53) *and* (54)*, one has*

$$I(\sigma; \boldsymbol{z}) \le \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \mathsf{KL}\left(\mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_1} \parallel \mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_2}\right). \tag{55}$$

Proof. It follows from the definition of mutual information that

$$\begin{aligned}
I(\sigma; \boldsymbol{z}) &= \sum_{\sigma_1 \in \mathcal{M}} \sum_{\boldsymbol{z}} \mathbb{P}\left(\sigma = \sigma_1, \boldsymbol{Z} = \boldsymbol{z}\right) \log \frac{\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_1\right)}{\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z}\right)} \\
&= \frac{1}{M} \sum_{\sigma_1 \in \mathcal{M}} \sum_{\boldsymbol{z}} \mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_1\right) \log \left\{ \frac{\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_1\right)}{\frac{1}{M} \sum_{\sigma_2 \in \mathcal{M}} \mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_2\right)} \right\} \\
&\le \frac{1}{M} \sum_{\sigma_1 \in \mathcal{M}} \sum_{\boldsymbol{z}} \mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_1\right) \left\{ \frac{1}{M} \sum_{\sigma_2 \in \mathcal{M}} \log \frac{\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_1\right)}{\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{z} \mid \sigma = \sigma_2\right)} \right\} \\
&= \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \mathsf{KL}\left(\mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_1} \parallel \mathbb{P}_{\boldsymbol{Z}|\sigma=\sigma_2}\right),
\end{aligned}$$

where the inequality is due to Jensen's inequality. □

Lemma 3. *If $w_K, w_{K+1} \in [w_{\min}, w_{\max}]$, then for any $\sigma_1, \sigma_2 \in \mathcal{M}$:*

$$\sum_{i \ne j} \mathsf{KL}\left(\mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel \mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_2}\right) \le \frac{2w_{\max}^4}{w_{\min}^4} n \Delta_K^2. \tag{56}$$

Proof. To start with, for any two measures $P \sim \mathrm{Bernoulli}\,(p)$ and $Q \sim \mathrm{Bernoulli}\,(q)$, one has [5, Eqn. (7)]

$$\mathsf{KL}\left(P \parallel Q\right) \le \chi^2\left(P \parallel Q\right) = \frac{(p-q)^2}{q} + \frac{(p-q)^2}{1-q} = \frac{(p-q)^2}{q\,(1-q)}. \tag{57}$$

where $\chi^2\left(P \parallel Q\right)$ denotes the χ^2 divergence.

Recall that given $\sigma = \sigma_1$ (resp. $\sigma = \sigma_2$), $y_{i,j}^{(1)}$ is Bernoulli distributed with mean $r_1 := \frac{w_{\sigma_1(i)}}{w_{\sigma_1(i)} + w_{\sigma_1(j)}}$ (resp. $r_2 := \frac{w_{\sigma_2(i)}}{w_{\sigma_2(i)} + w_{\sigma_2(j)}}$). If we set $\delta = r_1 - r_2$, then (57) yields

$$\mathsf{KL}\left(\mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel \mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_2}\right) \le \frac{\delta^2}{r_2\,(1-r_2)} \le \frac{4w_{\max}^2}{w_{\min}^2} \delta^2,$$

where the last inequality follows since

$$r_2(1-r_2)=\frac{w_{\sigma_2(i)}w_{\sigma_2(j)}}{\left(w_{\sigma_2(i)}+w_{\sigma_2(j)}\right)^2}\geq\frac{w_{\min}^2}{4w_{\max}^2}.$$

By construction, conditional on any hypotheses $\sigma_1,\sigma_2\in\mathcal{M}$, the resulting $\boldsymbol{y}_{i,j}$ are different over at most $2n$ locations. For each of these $O(n)$ locations, our construction of $\mathcal{M}$ ensures that

$$|\delta|=|r_2-r_1|\leq\frac{w_K}{w_K+w_{K+1}}-\frac{w_{K+1}}{w_K+w_{K+1}}=\frac{w_K-w_{K+1}}{w_K+w_{K+1}}\leq\frac{w_{\max}}{2w_{\min}}\Delta_K.$$

As a result, the total contribution is bounded above by

$$\sum_{i\neq j}\mathsf{KL}\left(\mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_1}\parallel\mathbb{P}_{y_{i,j}^{(1)}|\sigma=\sigma_2}\right)\leq 2n\cdot\left(\max_{i,j}\delta^2\right)\frac{4w_{\max}^2}{w_{\min}^2}\leq\frac{2w_{\max}^4}{w_{\min}^4}n\Delta_K^2.$$

□

A Bernstein Inequality

Our analysis relies on the Bernstein inequality. To simplify presentation, we state below a user-friendly version of Bernstein inequality.

Lemma 4. *Consider n independent random variables z_l ($1\leq l\leq n$), each satisfying $|z_l|\leq B$. Then there exists a universal constant $c_0>0$ such that for any $a\geq 2$,*

$$\left|\sum_{l=1}^n z_l-\mathbb{E}\left[\sum_{l=1}^n z_l\right]\right|\leq\sqrt{2a\log n\sum_{l=1}^n\mathbb{E}\left[z_l^2\right]}+\frac{2a}{3}B\log n \tag{58}$$

with probability at least $1-\frac{2}{n^a}$.

This is an immediate consequence of the well-known Bernstein inequality

$$\mathbb{P}\left\{\left|\sum_{l=1}^n z_l-\mathbb{E}\left[\sum_{l=1}^n z_l\right]\right|>t\right\}\leq 2\exp\left(-\frac{\frac{1}{2}t^2}{\sum_{l=1}^n\mathbb{E}\left[z_l^2\right]+\frac{1}{3}Bt}\right). \tag{59}$$

References

[1] R. W. Yeung, *Information theory and network coding.* Springer, 2008. 2.1

[2] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomized algorithms and probabilistic analysis.* Cambridge University Press, 2005. 2.1

[3] S. Negahban, S. Oh, and D. Shah, "Rank centrality: Ranking from pair-wise comparisons," 2012. [Online]. Available: http://arxiv.org/abs/1209.1688 2.2

[4] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012. 3

[5] T. van Erven and P. Harremoes, "Renyi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014. 3