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 \(\epsilon \in (0, \frac{1}{2})\), and let \(\mathcal{G} \sim \mathcal{G}_{n,p_{obs}}\). If

\[
L \leq c \frac{(1 - \epsilon) \log n - 2}{n p_{obs} \Delta_K^2}
\]

holds for some absolute constant \(c > 0\), then for any ranking scheme \(\psi\), there exists a preference vector \(w\) with separation \(\Delta_K\) such that the probability of error \(P_e(\psi) \geq \epsilon\).

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

\[
\Delta_K > c_1 \sqrt{\frac{\log n}{np_{obs} L}}.
\]

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

\[
\xi_t := c_3 \left\{ \xi_{\text{min}} + \frac{1}{2T} (\xi_{\text{max}} - \xi_{\text{min}}) \right\},
\]

where \(\xi_{\text{min}} := \sqrt{\frac{\log n}{np_{obs} L}}\) and \(\xi_{\text{max}} := \sqrt{\frac{\log n}{p_{obs} L}}\).

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

\[
|\hat{w}_{\text{ub}} - w_i| \leq \xi_{\text{max}}, \quad \forall 1 \leq i \leq n;
\]

\[
||\hat{w}_{\text{ub}} - w|| \leq \delta ||w||.
\]

\(^*\)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, \(c = w_{\text{min}}^4/(2w_{\text{max}}^4)\).
Then with probability at least \(1 - c_2 n^{-4}\) for some constant \(c_2 > 0\), the coordinate-wise MLE

\[
    w_i^\text{mle} := \arg \max_{\tau \in [w_{\min}, w_{\max}]} L (\tau, \hat{w}_i; y_i)
\]

satisfies

\[
    |w_i - w_i^\text{mle}| < \frac{20 \left(6 + \frac{\log L \log n}{\log n}\right)}{w_{\min}^4} \max \left\{ \delta + \frac{\xi \log n}{n \log L}, \sqrt{\frac{\log n}{n \log L}} \right\}
\]

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

\[
    |\hat{w}_i - w_i| \leq |\hat{w}_{i}^{\text{ub}} - w_i|, \quad 1 \leq i \leq n.
\]

2 Performance Guarantees of Spectral MLE

In this section, we establish the theoretical guarantees of Spectral MLE in controlling the ranking accuracy and \(L_\infty\) 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 \(\tau\) is given by

\[
    \frac{1}{L} \log L (\tau, \hat{w}_i; y_i) := \sum_{j: (i, j) \in \mathcal{E}} y_{ij} \log \frac{\tau}{\tau + \hat{w}_j} + (1 - y_{ij}) \log \frac{\tau + \hat{w}_j}{\tau + \hat{w}_i},
\]

and we shall use \(w_i\) (resp. \(\hat{w}_i\)) to denote the vector \(w = [w_1, \cdots, w_n]\) (resp. \(\hat{w} = [\hat{w}_1, \cdots, \hat{w}_n]\)) 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, \(|\tau - w_i| \geq \max \left\{ \delta + \frac{\xi \log n}{n \log L}, \sqrt{\frac{\log n}{n \log L}} \right\}\)), the coordinate-wise likelihood satisfies

\[
    \log L (w_i, \hat{w}_{i}; y_i) > \log L (\tau, \hat{w}_{i}; y_i)
\]

and, therefore, \(\tau\) cannot be the coordinate-wise MLE.

To begin with, we provide a lemma (which will be proved later) that concerns (10) for any single \(\tau\) 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

\[
    |w_i - \tau| > \gamma \cdot \frac{w_{\max}^5}{w_{\min}^4} \max \left\{ \frac{25}{4} \left( \delta + \frac{\xi \log n}{n \log L} \right), 20 \sqrt{\frac{\log n}{n \log L}} \right\},
\]

one has

\[
    \frac{1}{L} \log L (w_i, \hat{w}_{i}; y_i) - \frac{1}{L} \log L (\tau, \hat{w}_{i}; y_i) > \frac{w_{\max}^6}{100 w_{\min}^6} \frac{\log n}{L}.
\]

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

To establish Theorem 4, we still need to derive a uniform control over all \(\tau\) 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 \(\epsilon\)-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

\[
    |\tau - \tau_0| \leq \epsilon \quad \text{and} \quad |\tau_0 - w_i| \geq |\tau - w_i|.
\]
It is self-evident that one can produce such a cover \( \mathcal{N}_\epsilon \) with cardinality \( \left\lceil \frac{w_{\text{max}}}{\epsilon} \right\rceil + 1 \). If we set \( \gamma = 6 + \frac{\log L}{\log n} \) in Lemma \([11]\) taking the union bound over \( \mathcal{N}_\epsilon \) gives
\[
\frac{1}{L} \log \mathcal{L}(w_i, \hat{w}_i; y_i) - \frac{1}{L} \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) > \frac{w_{\text{max}}^6}{100w_{\text{min}}^6} \frac{\log n}{L},
\] (14)

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

We then proceed by bounding the difference between \( \log \mathcal{L}(\tau, \hat{w}_i; y_i) \) and \( \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) \). To achieve this, we first recognize that the Lipschitz constant of \( \frac{1}{L} \log \mathcal{L}(\tau, \hat{w}_i; y_i) \) (cf. \([10]\)) is bounded above by
\[
\frac{1}{L} \left| \frac{\partial \log \mathcal{L}(\tau, \hat{w}_i; y_i)}{\partial \tau} \right| = \left| \sum_{j: y_{i,j} \neq 0} y_{i,j} \left( \frac{1}{\tau} - \frac{1}{\tau + \hat{w}_j} \right) - (1 - y_{i,j}) \frac{1}{\tau + \hat{w}_j} \right|
\leq \deg(i) \cdot \frac{2}{w_{\text{min}}} \leq \frac{12 n_{\text{obs}}}{5} \frac{w_{\text{obs}}}{w_{\text{min}}},
\]

where (a) follows since
\[
\left| y_{i,j} \left( \frac{1}{\tau} - \frac{1}{\tau + \hat{w}_j} \right) - (1 - y_{i,j}) \frac{1}{\tau + \hat{w}_j} \right| = \left| y_{i,j} \left( \frac{1}{\tau} - \frac{1}{\tau + \hat{w}_j} \right) \right| \leq \frac{|y_{i,j}|}{\tau} + \left| \frac{1}{\tau + \hat{w}_j} \right| < \frac{2}{w_{\text{min}}},
\]
and (b) holds since \( \deg(i) \leq 2.4n_{\text{obs}} \), with probability \( 1 - O(n^{-4}) \) as long as \( p_{\text{obs}} > \frac{c_1 \log n}{n} \) for some sufficiently large \( c_1 > 0 \). As a result, by picking
\[
\epsilon = \frac{w_{\text{max}}^6}{100w_{\text{min}}^6} \frac{\log n}{L} = \frac{w_{\text{max}}^6}{240w_{\text{min}}^5} \frac{\log n}{n_{\text{obs}} L},
\] (15)

one can make sure that for any \( |\tau - \tau_0| \leq \epsilon \),
\[
\frac{1}{L} \log \mathcal{L}(\tau, \hat{w}_i; y_i) - \frac{1}{L} \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) \leq \epsilon \cdot \frac{12 n_{\text{obs}}}{5} \frac{w_{\text{obs}}}{w_{\text{min}}},
\] (16)

\[
\frac{1}{L} \log \mathcal{L}(\tau, \hat{w}_i; y_i) < \frac{1}{L} \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) + \frac{w_{\text{max}}^6}{100w_{\text{min}}^6} \frac{\log n}{L}.
\] (17)

In addition, with the above choice \([15]\) of \( \epsilon \) in place, the cardinality of the \( \epsilon \)-cover is bounded above by
\[
|\mathcal{N}_\epsilon| \leq \left\lfloor \frac{w_{\text{max}}}{\epsilon} \right\rfloor + 1 = \left\lfloor \frac{240n_{\text{obs}} L}{\log n} \frac{w_{\text{min}}^5}{w_{\text{max}}^5} \right\rfloor + 1 \ll n^2 L
\]
for any sufficiently large \( n \).

Putting \([14]\) and \([17]\) together suggests that for all \( \tau \in [w_{\text{min}}, w_{\text{max}}] \) sufficiently apart from the ground truth \( w_i \), namely,
\[
\forall \tau \in [w_{\text{min}}, w_{\text{max}}]: \ |\tau - w_i| \geq \left( \frac{6 + \frac{\log L}{\log n}}{w_{\text{min}}^4} \right) \frac{w_{\text{max}}^5}{w_{\text{min}}} \max \left\{ \frac{25}{4} \left( \delta + \frac{\xi \log n}{n_{\text{obs}}} \right), \ 20 \sqrt{\frac{\log n}{n_{\text{obs}} L}} \right\},
\] (18)

one necessarily has
\[
\frac{1}{L} \log \mathcal{L}(w_i, \hat{w}_i; y_i) - \frac{1}{L} \log \mathcal{L}(\tau, \hat{w}_i; y_i)
= \left\{ \frac{1}{L} \log \mathcal{L}(w_i, \hat{w}_i; y_i) - \frac{1}{L} \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) \right\} + \left\{ \frac{1}{L} \log \mathcal{L}(\tau_0, \hat{w}_i; y_i) - \frac{1}{L} \log \mathcal{L}(\tau, \hat{w}_i; y_i) \right\}
> 0,
\] (19)
with probability at least $1 - 4|\mathcal{N}| n^{-6-\log_2 n} - O(n^{-4}) \geq 1 - 4n^2L n^{-6-\log_2 n} - O(n^{-4}) = 1 - O(n^{-4})$. Consequently, any $\tau \in [w_{\min}, w_{\max}]$ that obeys \cite{13} cannot be the coordinate-wise MLE, which in turn justifies the claim \cite{7} of Theorem \cite{4} (which is slightly weaker than what we prove here).

**Proof of Lemma \cite{7}** We start by evaluating the true coordinate-wise likelihood gap

$$
\log \mathcal{L}(w_i, w_{\backslash i}; y_i) - \log \mathcal{L}(\tau, w_{\backslash i}; y_i)
$$

(20)

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

$$
\mathbb{E}[y_{ij}] = \frac{w_i}{w_i + w_j} \quad \text{and} \quad \text{Var}[y_{ij}] = \frac{w_i w_j}{(w_i + w_j)^2}.
$$

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

$$
\frac{1}{L} \log \mathcal{L}(\tau, w_{\backslash i}; y_i) = \sum_{j: (i, j) \in \mathcal{E}} \left\{ y_{ij} \log \left( \frac{\tau}{\tau + w_j} \right) + (1 - y_{ij}) \log \left( \frac{w_j}{\tau + w_j} \right) \right\}
$$

(21)

$$
= \sum_{j: (i, j) \in \mathcal{E}} y_{ij} \log \left( \frac{\tau}{w_j} \right) + \sum_{j: (i, j) \in \mathcal{E}} \log \left( \frac{w_j}{\tau + w_j} \right).
$$

(22)

Taking expectation w.r.t. $y_i$ using the form (21) reveals that

$$
\mathbb{E} \left[ \frac{1}{L} \log \mathcal{L}(w_i, w_{\backslash i}; y_i) - \frac{1}{L} \log \mathcal{L}(\tau, w_{\backslash i}; y_i) \bigg| G \right] = \sum_{j: (i, j) \in \mathcal{E}} \left\{ \frac{w_i}{w_i + w_j} \log \left( \frac{w_i}{w_i + w_j} \right) + \frac{w_j}{w_i + w_j} \log \left( \frac{w_j}{\tau + w_j} \right) \right\}
$$

(23)

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

$$
\mathbb{E} \left[ \frac{1}{L} \log \mathcal{L}(w_i, w_{\backslash i}; y_i) - \frac{1}{L} \log \mathcal{L}(\tau, w_{\backslash i}; y_i) \bigg| G \right] \geq 2 \sum_{j: (i, j) \in \mathcal{E}} \left( \frac{w_i}{w_i + w_j} - \frac{\tau}{\tau + w_j} \right)^2
$$

$$
= 2(w_i - \tau)^2 \sum_{j: (i, j) \in \mathcal{E}} \frac{w_j^2}{(w_i + w_j)^2(\tau + w_j)^2}.
$$

(24)

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

However, when running Spectral MLE, we do not have access to the ground truth scores $w_{\backslash i}$; what we actually compute is $\mathcal{L}(w_i, \hat{w}_{\backslash i}; y_i)$ (resp. $\mathcal{L}(\tau, \hat{w}_{\backslash i}; y_i)$) rather than $\mathcal{L}(w; y_i)$ (resp. $\mathcal{L}(\tau, w_{\backslash i}; y_i)$). 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

$$
\hat{\ell}_i(w_i) := \frac{1}{L} \log \mathcal{L}(w_i, \hat{w}_{\backslash i}; y_i),
$$

$$
\ell^*(w_i) := \frac{1}{L} \log \mathcal{L}(w_i, w_{\backslash i}; y_i),
$$

(25)

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

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

$$
\hat{\ell}_i(\tau) - \ell^*(\tau) = \sum_{j: (i, j) \in \mathcal{E}} y_{ij} \log \left( \frac{w_j}{\hat{w}_j} \right) + \sum_{j: (i, j) \in \mathcal{E}} \left\{ \log \left( \frac{\hat{w}_j}{\tau + \hat{w}_j} \right) - \log \left( \frac{\tau}{\tau + w_j} \right) \right\}.
$$

(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
\[
\hat{\ell}_i (w_i) - \hat{\ell}_i (\tau) - (\ell^* (w_i) - \ell^* (\tau)) = \hat{\ell}_i (w_i) - \ell^* (w_i) - (\hat{\ell}_i (\tau) - \ell^* (\tau))
\]
\[
= \sum_{j: (i,j) \in E} \left\{ \log \left( \frac{\hat{w}_j}{\hat{w}_j + w_i} \right) - \log \left( \frac{w_j}{w_j + w_i} \right) \right\} - \log \left( \frac{\tau + w_j}{\tau + w_i} \right) - \log \left( \frac{\tau + \hat{w}_j}{\tau + \hat{w}_j} \right)
\]
\[
= \sum_{j: (i,j) \in E} \left\{ \log \left( \frac{\tau + \hat{w}_j}{\tau + w_i} \right) - \log \left( \frac{\tau + w_j}{\tau + w_i} \right) \right\}.
\]
(27)

This gap relies on the function
\[
g (t) := \log \left( \frac{\tau + t}{w_i + t} \right) - \log \left( \frac{\tau + w_i}{w_i + w_j} \right), \quad t \in [w_{\min}, w_{\max}],
\]
which apparently obeys the following two properties: (i) $g (w_j) = 0$; (ii)
\[
\left| \frac{\partial g (t)}{\partial t} \right| = \left| \frac{1}{\tau + t} - \frac{1}{w_i + t} \right| = \frac{|\tau - w_i|}{(w_i + t)(\tau + t)} \leq \frac{|\tau - w_i|}{4w_{\min}^2}, \quad \forall t \in [w_{\min}, w_{\max}],
\]
Taken together these two properties demonstrate that
\[
|g (t)| \leq \frac{1}{4w_{\min}^2} |\tau - w_i| |t - w_j|, \quad \forall t \in [w_{\min}, w_{\max}].
\]
Substitution into (28) gives
\[
\left| \hat{\ell}_i (w_i) - \hat{\ell}_i (\tau) - (\ell^* (w_i) - \ell^* (\tau)) \right| \leq \frac{1}{4w_{\min}^2} |\tau - w_i| \sum_{j: (i,j) \in E} |\hat{w}_j - w_j|
\]
\[
\leq \frac{1}{4w_{\min}^2} |\tau - w_i| \sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j|.
\]
(29)

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

We now move on to develop an upper bound on (29). From our assumptions on the initial estimate, we have
\[
\|\hat{w} - w\|^2 \leq \|\hat{w}_{\text{ub}} - w\|^2 \leq \delta^2 \|w\|^2 \leq n w_{\max}^2 \delta^2.
\]
Since $\mathcal{G}$ and $\hat{w}_{\text{ub}}$ are statistically independent, this inequality immediately gives rise to the following two consequences:
\[
\mathbb{E} \left[ \sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j| \right] = p_{\text{obs}} \|\hat{w}_{\text{ub}} - w\|_1 \leq p_{\text{obs}} \sqrt{n} \|\hat{w}_{\text{ub}} - w\| \leq np_{\text{obs}} w_{\max} \delta
\]
(30)
and
\[
\mathbb{E} \left[ \sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j|^2 \right] = p_{\text{obs}} \|\hat{w}_{\text{ub}} - w\|^2_2 \leq np_{\text{obs}} w_{\max}^2 \delta^2.
\]
(31)
Recall our assumption that $\max_j |\hat{w}_{\text{ub}} - w_j| \leq \xi w_{\max}$. For any fixed $\gamma \geq 3$, if $p_{\text{obs}} > \frac{2 \log n}{n}$, then with probability at least $1 - 2n^{-\gamma}$,
\[
\sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j| \overset{(i)}{\leq} \mathbb{E} \left[ \sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j| \right] + \sqrt{2 \gamma \log n \cdot \mathbb{E} \left[ \sum_{j: (i,j) \in E} |\hat{w}_{\text{ub}} - w_j|^2 \right] + \frac{2\gamma}{3} \xi w_{\max} \log n}
\]
\[
\leq np_{\text{obs}} w_{\max} \delta + \sqrt{2 \gamma \cdot np_{\text{obs}} \log n w_{\max} \delta + \frac{2\gamma}{3} \xi w_{\max} \log n}
\]
\[
\overset{(ii)}{\leq} (1 + \sqrt{\gamma}) np_{\text{obs}} w_{\max} \delta + \frac{2\gamma}{3} \xi w_{\max} \log n
\]
\[
\overset{(iii)}{\leq} \gamma np_{\text{obs}} w_{\max} \delta + \xi w_{\max} \log n,
\]

5
where (i) comes from the Bernstein inequality as given in Lemma 4, (ii) follows since \( \log n < \frac{\log \gamma}{\alpha} \) by assumption, and (iii) arises since \( 1 + \sqrt{\gamma} \leq \gamma \) whenever \( \gamma \geq 3 \). This combined with (29) allows us to control

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

(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

\[
E \left[ \ell^* (w_i) - \ell^* (\tau) \mid \mathcal{G} \right] \geq 2 (w_i - \tau)^2 \sum_{j: (i, j) \in \mathcal{E}} \frac{w_j^2}{(w_i + w_j)^2 (\tau + w_j)^2} \geq \frac{w_{\text{min}}^2}{8w_{\text{max}}^4} (w_i - \tau)^2 \deg (i)
\]

(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

\[
\text{Var} \left[ \ell^* (w_i) - \ell^* (\tau) \mid \mathcal{G} \right] \stackrel{(a)}{=} \text{Var} \left[ \sum_{j: (i, j) \in \mathcal{E}} y_{i, j} \log \left( \frac{w_i}{\tau} \right) \mid \mathcal{G} \right] = \log^2 \left( \frac{w_i}{\tau} \right) \sum_{j: (i, j) \in \mathcal{E}} \frac{1}{L} \frac{w_i w_j}{(w_i + w_j)^2} \leq \frac{1}{L} \min \{ w_i^2, \tau^2 \} \sum_{j: (i, j) \in \mathcal{E}} \frac{w_{\text{max}}^2}{4w_{\text{min}}^2} \leq \frac{w_{\text{max}}^2}{4w_{\text{min}}^4} \frac{1}{L} (w_i - \tau)^2 \deg (i)
\]

(34)

where (a) is an immediate consequence of (22), and (b) follows since \( \log \frac{w_i}{\tau} \leq \frac{w_i}{\tau} \leq \frac{\beta - \alpha}{\alpha} \) 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 \( \mathcal{G} \),

\[
\ell^* (w_i) - \ell^* (\tau) \geq E \left[ \ell^* (w_i) - \ell^* (\tau) \mid \mathcal{G} \right] - \sqrt{2 \gamma \text{Var} \left[ \ell^* (w_i) - \ell^* (\tau) \mid \mathcal{G} \right] \log n} - \frac{2 \gamma \log n \cdot \log \left( \frac{w_i}{\tau} \right)}{3L}
\]

(35)

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

The above bound relies on \( \deg (i) \), which is on the order of \( np_{\text{obs}} \) 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

\[
\frac{4}{5} np_{\text{obs}} < \deg (i) < \frac{6}{5} np_{\text{obs}}, \quad \forall 1 \leq i \leq n
\]

(36)

with probability at least \( \frac{2}{n^{c_{1}}} \). This taken collectively with (35) and the assumption \( np_{\text{obs}} > 2 \log n \) implies
\[ \ell^*(w_i) - \ell^*(\tau) \geq \frac{w_{\min}^2}{8w_{\max}^2} \left( w_i - \tau \right)^2 - \frac{4}{5} np_{\text{obs}} - \frac{\sqrt{\gamma}}{2} w_{\max} \left| w_i - \tau \right| \sqrt{6np_{\text{obs}} \log n} - \frac{2\gamma}{3} w_{\min} \left| w_i - \tau \right| \sqrt{\frac{np_{\text{obs}} \log n}{L}} \]

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

\[ \gamma \cdot \frac{w_{\max} \left| w_i - \tau \right|}{w_{\min}^2} \sqrt{\frac{np_{\text{obs}} \log n}{L}} \leq \frac{w_{\min}^2}{20w_{\max}^4} \left( w_i - \tau \right)^2 np_{\text{obs}} \]

or, equivalently,

\[ |w_i - \tau| \geq \frac{20\gamma \cdot w_{\max}^5}{w_{\min}^4} \sqrt{\frac{\log n}{np_{\text{obs}} L}}. \tag{39} \]

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,

\[ \hat{\ell}_i(w_i) - \hat{\ell}_i(\tau) \geq \ell^*(w_i) - \ell^*(\tau) - \frac{|\tau - w_i| \gamma w_{\max} (np_{\text{obs}} \delta + \xi \log n)}{4w_{\min}^2} \]

\[ \geq \frac{w_{\min}^2}{20w_{\max}^4} \left( w_i - \tau \right)^2 np_{\text{obs}} - \frac{|\tau - w_i| \gamma w_{\max}}{4w_{\min}^2} (np_{\text{obs}} \delta + \xi \log n) \]

\[ > \frac{w_{\min}^2}{100w_{\max}^4} \left( w_i - \tau \right)^2 np_{\text{obs}} \]

\[ > \frac{w_{\min}^6}{100w_{\max}^4} \log n \tag{40} \]

where (40) holds under the condition

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

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

\[ \square \]

2.2 Proof of Theorem 3

The accuracy of top-\( K \) identification is closely related to the \( \ell_{\infty} \) 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

\[ \|w^{(t)} - w\|_\infty \lesssim \sqrt{\frac{\log n}{np_{\text{obs}} L}} + \frac{1}{2t} \sqrt{\frac{\log n}{p_{\text{obs}} L}} \approx \xi_t, \quad \forall t \in \mathbb{N}, \tag{42} \]

where

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

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

\[ \|w^{(T)} - w\|_\infty \lesssim \sqrt{\frac{\log n}{np_{\text{obs}} L}} = \xi_{\min}. \]

7
The key implication is the following: if \( w_K - w_{K-1} \geq c_1 \sqrt{\frac{\log n}{np_{\text{obs}} L}} \) for some sufficiently large \( c_1 > 0 \), then
\[
|w_i^{(T)} - w_j^{(T)}| \geq |w_i^{(T)} - w_j| - |w_j^{(T)} - w_j| \geq w_K - w_{K+1} - 2 \left\| w^{(T)} - w \right\| > 0
\]
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 \( w^{(0)} \) is computed by Rank Centrality, the \( \ell_2 \) estimation error satisfies
\[
\left\| w^{(0)} - w \right\| \leq c_4 \sqrt{\frac{\log n}{np_{\text{obs}} L}} = c_4 \xi_{\text{min}} := \delta
\]
with high probability, where \( c_4 > 0 \) is some universal constant independent of \( n, p_{\text{obs}}, L \) and \( \Delta_K \). A by-product of this result is an upper bound
\[
\left\| w^{(0)} - w \right\| \leq \delta \left\| w \right\| \leq \delta \sqrt{n} = c_4 \left\{ \log n \frac{1}{p_{\text{obs}} L} \right\}
\]
which together with the fact \( \left\| w^{(0)} - w \right\|_{\infty} \leq w_{\text{max}} - w_{\text{min}} \leq 1 \) give
\[
\left\| w^{(0)} - w \right\|_{\infty} \leq \min \left\{ c_4 \sqrt{\frac{\log n}{p_{\text{obs}} L}}, 1 \right\} = \min \left\{ c_4 \xi_{\text{max}}, 1 \right\}.
\]
This justifies that \( w^{(0)} \) satisfies the claim (42). Notably, \( w^{(0)} \) is independent of \( c_{\text{iter}} \) and \( y_{\text{iter}} \) 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
\[
\xi_t \geq c_{10} \xi_{\text{min}} = c_{10} \left\{ \log n \frac{1}{p_{\text{obs}} L} \right\} \quad \text{iff} \quad t \leq T_0,
\]
for some large constant \( c_{10} \). As is seen from the definition of \( \xi_t, T_0 \lesssim \log n \) holds as long as \( L = O(p_{\text{iter}}) \).

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

- \( \mathcal{M}_t \): \( \left\| w^{(\text{mle})} - w \right\|_{\infty} \leq \frac{1}{2} \xi_t \) holds at the \( t \)-th iteration (the iteration where we compute \( w^{(t+1)} \));
- \( \mathcal{B}_t \): all entries \( w^{(\text{iter})}_t \) of \( w^{(\text{iter})} (t \leq t - 1) \) satisfying \( |w^{(\text{iter})}_t - w_i| \geq 1.5 \xi_t \) have been replaced by time \( t \);
- \( \mathcal{H}_t \): none of the entries \( w^{(\text{iter})}_t \) (\( t \leq t - 1 \)) satisfying \( |w^{(\text{iter})}_t - w_i| \leq \frac{1}{2} \xi_t \) 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 \( w^{(\text{iter})}_i \) obeying \( |w^{(\text{iter})}_i - w_i| \geq 1.5 \xi_t \) that have not been replaced by time \( t - 1 \). To this end, we recall that Spectral MLE replaces \( w^{(\text{iter})}_i \) iff \( |w^{(\text{iter})}_i - w^{\text{mle}}_i| > \xi_t \). With \( \mathcal{M}_t \) in place, for each \( i \) obeying \( |w^{(\text{iter})}_i - w_i| \geq 1.5 \xi_t \), one has
\[
|w^{(\text{iter})}_i - w^{\text{mle}}_i| \geq |w^{(\text{iter})}_i - w_i| - |w^{\text{mle}}_i - w_i| > 1.5 \xi_t - \frac{1}{2} \xi_t = \xi_t
\]
and hence it is necessarily replaced by \( w^{\text{mle}}_i \) by time \( t \). Similarly, for any \( i \) obeying \( |w^{(\text{iter})}_i - w_i| \leq 0.5 \xi_t \), one has
\[
|w^{(\text{iter})}_i - w^{\text{mle}}_i| \leq |w^{(\text{iter})}_i - w_i| + |w^{\text{mle}}_i - w_i| < \frac{1}{2} \xi_t + \frac{1}{2} \xi_t = \xi_t
\]
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 3 and setting \( w^{\text{ub}} = w^{(0)} \), we see that
\[
\left\| w^{\text{mle}} - w \right\| \leq c_7 \xi_{\text{min}} + c_9 \frac{\log n}{np_{\text{obs}}} \xi_{\text{max}}
\]
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. (48)) gives \( \xi_0 \gg c_7 \sqrt{\frac{\log n}{n_{\text{obs}} L}} \), additionally, \( c_9 \frac{\log n}{n_{\text{obs}}} \xi_{\text{max}} \ll \xi_{\text{max}} \leq \xi_0 \) holds as long as \( \frac{\log n}{n_{\text{obs}}} \) is sufficiently small. Putting these conditions together gives

\[
\|\mathbf{w}^{\text{MLE}} - \mathbf{w}\|_\infty \leq c_7 \xi_{\text{min}} + c_9 c_4 \frac{\log n}{n_{\text{obs}}} \xi_{\text{max}} < \frac{1}{2} \xi_0,
\]

which verifies the property \( M_0 \).

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 \( M_{t-1} \) and \( B_{t-1} \) immediately reveal that

\[
\|\mathbf{w}^{(t)} - \mathbf{w}\|_\infty \leq 1.5 \xi_{t-1}.
\]  

(48)

In order to invoke Theorem 3 for the coordinate-wise MLEs, we need to construct a looser auxiliary score estimate \( \mathbf{w}^{\text{ub}} \). With \( B_{t-1} \), \( H_{t-1} \) and \( \text{(48)} \) in mind, we propose a candidate for the \( t^{th} \) iteration as follows:

\[
\mathbf{w}^{\text{ub}}_i = \begin{cases} 
\mathbf{w}_i + 1.5 \xi_{t-1}, & \text{if } |w_i^{(0)} - w_i| > \frac{1}{2} \xi_{t-1}, \\
\mathbf{w}_i^{(0)}, & \text{else.}
\end{cases}
\]  

(49)

which is clearly independent of \( \xi_{\text{iter}} \) and \( \mathbf{y}^{\text{iter}} \). According to \( B_{t-1} \) and \( H_{t-1} \), (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, \( \mathbf{w}^{\text{ub}} \) clearly satisfies

\[
|w_i^{(t)} - w_i| \leq |w^{\text{ub}}_i - w_i| \leq 1.5 \xi_{t-1},
\]  

(50)

and

\[
\|\mathbf{w}^{(t)} - \mathbf{w}\| \leq \|\mathbf{w}^{(\text{ub})} - \mathbf{w}\| \leq \|\mathbf{w}^{(0)} - \mathbf{w}\| + \frac{1.5 \xi_{t-1}}{0.5 \xi_{t-1}} \|\mathbf{w}^{(0)} - \mathbf{w}\| \leq 3 \|\mathbf{w}\|.
\]  

(51)

Here, (i) arises since if \( w_i^{(0)} \) is replaced, then the error \( |w_i^{(0)} - w_i| \) 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 3 gives

\[
\|\mathbf{w}^{\text{MLE}} - \mathbf{w}\|_\infty \leq c_8 \xi_{\text{min}} + 1.5 c_9 \frac{\log n}{n_{\text{obs}}} \xi_{t-1},
\]

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

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

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

\[
\|\mathbf{w}^{\text{MLE}} - \mathbf{w}\|_\infty \leq \frac{1}{2} \xi_t
\]

as claimed in \( M_1 \). Having verified these inductive hypotheses, we see from the above argument that the worst case \( \ell_\infty \) error bound at the \( t^{th} \) iteration is at most 1.5 \( \xi_t \), which in turn leads to the claim (42) for any \( t \leq T_0 \).

\[\text{Careful readers will note that when } |w_i^{(0)} - w_i| \geq \frac{1}{2} \Delta_{t-1}, \text{ the resulting } w_i^{\text{ub}} \text{ might exceed the range } [w_{\text{min}}, w_{\text{max}}]. \text{ This can be easily addressed if we do the following: (1) change } w_i^{\text{ub}} \text{ to } w_i - 1.5 \Delta_{t-1} \text{ instead if } w_i - 1.5 \Delta_{t-1} \in [w_{\text{min}}, w_{\text{max}}]; (2) if it is still infeasible, set } w_i^{\text{ub}} \text{ to } w_{\text{max}} \text{ if } |w_i - w_{\text{max}}| > |w_i - w_{\text{min}}| \text{ and } w_{\text{min}} \text{ otherwise. For simplicity of presentation, however, we omit these boundary situations and assume } w_i + 1.5 \Delta_{t-1} \leq w_{\text{max}} \text{ throughout, which will not change the results anyway.}\]
Starting from $t = T_0 + 1$, we fix the auxiliary score as follows

$$w_{i}^{ab} = \begin{cases} w_i + 1.5\xi_T, & \text{if } |w_i^{(0)} - w_i| > \frac{1}{2}\xi_{\infty}, \\ w_i^{(0)}, & \text{else,} \end{cases}$$ (52)

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

$$|w_i^{(t)} - w_i| \leq |w_i^{ab} - w_i| \leq 1.5\xi_T$$

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

$$k \cdot \left(\frac{1}{2}\xi_{\infty}\right)^2 \leq \|w - w^{(0)}\|^2 \leq \delta^2\|w\|^2 \iff k \leq \frac{4\delta^2\|w\|^2}{\xi_{\infty}^2},$$

which further gives

$$\|w^{ab} - w\|^2 \leq \|w^{(0)} - w\|^2 + \sum_{i: |w_i^{(0)} - w_i| > \frac{1}{2}\Delta_{\infty}} (1.5\xi_T)^2 \leq \delta^2\|w\|^2 + 2.25k\xi_T^2 \leq \delta^2\|w\|^2 \left(1 + \frac{9\xi_T^2}{\xi_{\infty}^2}\right).$$

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

$$\|w^{ab} - w\| \leq \sqrt{19\delta}\|w\|.$$

Apply Theorem 4 to deduce

$$\|w^{\text{mle}} - w\|_{\infty} \leq \delta + \frac{\log n}{np_{\text{obs}}} + \sqrt{\log n \frac{\log n}{np_{\text{obs}}}} \leq \sqrt{\log n \frac{\log n}{np_{\text{obs}}}} \leq \frac{1}{2}\xi_{\infty},$$

as long as $\frac{\log n}{np_{\text{obs}}}$ 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

$$|w_i^{(0)} - w_i^{\text{mle}}| \leq |w_i^{(0)} - w_i| + |w_i^{(\text{mle})} - w_i| < \xi_{\infty} < \xi_t,$$

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

$$\|w^{(t)} - w\|_{\infty} \leq \frac{1}{2}\xi_{\infty} \leq \sqrt{\log n \frac{\log n}{np_{\text{obs}}}}$$

for $t > T_0$.

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 $\sigma$ 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 \( w \) are fixed up to permutation in such a way that
\[
w_{\sigma(i)} = \begin{cases} w_K, & 1 \leq i \leq K, \\ w_{K+1}, & K < i \leq n, \end{cases}
\]
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 \( M := \max \{ K, n - K + 1 \} + 1 \) hypotheses regarding the permutation: if \( K < n/2, \) then
\[
\mathbb{P} \{ \sigma ([K]) = \mathcal{S} \} = \frac{1}{M}, \quad \text{if } \mathcal{S} = \{ 2, \cdots, K \} \cup \{ i \}, \quad (i = 1, K + 1, \cdots, n);
\]
if \( K \geq n/2, \) then
\[
\mathbb{P} \{ \sigma ([K]) = \mathcal{S} \} = \frac{1}{M}, \quad \text{if } \mathcal{S} = \{ 1, \cdots, K + 1 \} \setminus \{ i \}, \quad (i = 1, \cdots, K + 1).
\]
In words, each alternative hypothesis is generated by swapping two indices of the hypothesis obeying \( \sigma ([K]) = [K] \). Denoting by \( P_{e, M} \) 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 \( P_{e, M} \).

This Bayesian probability of error will be bounded using classical Fano-type bounds. To accommodate partial observation, we introduce an erased version of \( y_{i,j} := (y_{i,j}^{(1)}, \cdots, y_{i,j}^{(L)}) \) such that
\[
z_{i,j} = \begin{cases} y_{i,j}, & \text{with probability } p_{\text{obs}}, \\ \text{erasure}, & \text{else}, \end{cases}
\]
and set \( Z := \{ z_{i,j} \}_{1 \leq i \leq j \leq n} \). With a slight abuse of notation, we denote by \( \sigma \) and \( \hat{\sigma} \) the ground truth permutation and the output of any ranking procedure, respectively. Making use of (53) and (54) gives
\[
\log M = H (\sigma) = I (\sigma; \hat{\sigma}) + H (\sigma|\hat{\sigma})
\]
\[
\leq (a) I (\sigma; Z) + 1 + P_{e, M} \log M
\]
\[
\leq (b) \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \mathrm{KL} (\mathbb{P}_{Z|\sigma = \sigma_1} \| \mathbb{P}_{Z|\sigma = \sigma_2}) + 1 + P_{e, M} \log M
\]
\[
\leq (c) \frac{p_{\text{obs}}}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{i \neq j} \mathrm{KL} (\mathbb{P}_{y_{i,j}|\sigma = \sigma_1} \| \mathbb{P}_{y_{i,j}|\sigma = \sigma_2}) + 1 + P_{e, M} \log M
\]
\[
\leq (d) \frac{p_{\text{obs}} L}{M^2} \sum_{\sigma_1, \sigma_2 \in \mathcal{M}} \sum_{i \neq j} \mathrm{KL} (\mathbb{P}_{y_{i,j}^{(1)}|\sigma = \sigma_1} \| \mathbb{P}_{y_{i,j}^{(1)}|\sigma = \sigma_2}) + 1 + P_{e, M} \log M
\]
\[
\leq (e) \frac{2w_{\max}^2}{w_{\min}^4} n p_{\text{obs}} L \Delta_k^2 + 1 + P_{e, M} \log M,
\]
where \( H (X), I (X; Y), \) and \( \mathrm{KL} (P \| Q) \) 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 \( z_{i,j} \)’s; \( (d) \) is a
consequence of the fact that $y_{i,j}^{(t)} (1 ≤ l ≤ L)$ are i.i.d.; and (e) follows from Lemma 3 (see below). This immediately yields

$$P_{e,M} ≥ \frac{\log M - 2u_{\max}^4 n p_{\text{obs}} L \Delta_K^2}{\log M}.$$  

Consequently, one would have $P_e ≥ P_{e,M} ≥ \epsilon$ if

$$\frac{2u_{\max}^4}{u_{\min}^4} n p_{\text{obs}} L \Delta_K^2 ≤ (1 - \epsilon) \log M - 1.$$  

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

$$\frac{2u_{\max}^4}{u_{\min}^4} n p_{\text{obs}} L \Delta_K^2 ≤ (1 - \epsilon) \log n - 2 \iff L ≤ \frac{w_{\min}^4}{2u_{\max}^4} \frac{(1 - \epsilon) \log n - 2}{n p_{\text{obs}} \Delta_K^2},$$

which finishes the proof.

**Lemma 2.** Under the prior [53] and [54], one has

$$I(\sigma; z) ≤ \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 ∈ M} KL (P_{Z|\sigma=\sigma_1} \parallel P_{Z|\sigma=\sigma_2}).$$ (55)

**Proof.** It follows from the definition of mutual information that

$$I(\sigma; z) = \sum_{\sigma_1 ∈ M} \sum_{z} P(\sigma = \sigma_1, Z = z) \log \frac{P(Z = z | \sigma = \sigma_1)}{P(Z = z)}$$

$$= \frac{1}{M} \sum_{\sigma_1 ∈ M} \sum_{z} P(Z = z | \sigma = \sigma_1) \log \left\{ \frac{P(Z = z | \sigma = \sigma_1)}{\frac{1}{M} \sum_{\sigma_2 ∈ M} P(Z = z | \sigma = \sigma_2)} \right\}$$

$$≤ \frac{1}{M} \sum_{\sigma_1 ∈ M} \sum_{z} P(Z = z | \sigma = \sigma_1) \left\{ \frac{1}{M} \sum_{\sigma_2 ∈ M} \log \frac{P(Z = z | \sigma = \sigma_1)}{P(Z = z | \sigma = \sigma_2)} \right\}$$

$$= \frac{1}{M^2} \sum_{\sigma_1, \sigma_2 ∈ M} KL (P_{Z|\sigma=\sigma_1} \parallel P_{Z|\sigma=\sigma_2}),$$

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

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

$$\sum_{i ≠ j} KL (P_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel P_{y_{i,j}^{(1)}|\sigma=\sigma_2}) ≤ \frac{2w_{\max}^4}{w_{\min}^4} n \Delta_K^2.$$ (56)

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

$$KL (P \parallel Q) ≤ \chi^2 (P \parallel Q) = \frac{(p - q)^2}{q} + \frac{(p - q)^2}{1 - q} = \frac{(p - q)^2}{q (1 - q)}.$$ (57)

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

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

$$KL (P_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel P_{y_{i,j}^{(1)}|\sigma=\sigma_2}) ≤ \frac{\delta^2}{r_2 (1 - r_2)} ≤ \frac{4w_{\max}^2 \delta^2}{w_{\min}^4},$$

$$\sum_{i ≠ j} KL (P_{y_{i,j}^{(1)}|\sigma=\sigma_1} \parallel P_{y_{i,j}^{(1)}|\sigma=\sigma_2}) ≤ \frac{4w_{\max}^2 \Delta_K^2}{w_{\min}^4}. $$
where the last inequality follows since

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

By construction, conditional on any hypotheses \( \sigma_1, \sigma_2 \in \mathcal{M} \), the resulting \( y_{i,j} \) 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

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

\( \square \)

### 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 \( a \geq 2 \),

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

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

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

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

### References


