Ranking from Pairwise Comparisons

Cheng Mao

November 14, 2018

Ranking from comparisons arises in various applications, including recommender systems, social choice and sports tournament. We consider the following setup. Suppose that there are items $1, \dots, n$ associated with unknown ranks $\pi^*(1), \dots, \pi^*(n)$, where $\pi^* : [n] \to [n]$ is a permutation. Observing a set of pairwise comparisons, each of the form $i \succ j$ meaning that "item i beats item j", we aim to recover the ranking π^*.

1 Modeling pairwise comparisons

We first give an overview of common models for ranking from pairwise comparisons.

1.1 Models for probabilities of outcomes

Each pairwise comparison is a Bernoulli outcome. Let us denote the probability that the item at rank k beats the item at rank ℓ by $M_{k,\ell}$ where $M \in \mathbb{R}^{n \times n}$, so that

$$\mathbb{1}\{i \succ j\} \sim \mathsf{Ber}(M_{\pi^*(i),\pi^*(j)}).$$

In the sequel, we present several models on the matrix M of probabilities. It is vacuous to compare an item to itself, so we assume without loss of generality that $M_{i,i} = 1/2$ for $i \in [n]$. Moreover, we consider the case that there is one and only one winner in a pairwise comparison, so it always holds that $M_{k,\ell} + M_{\ell,k} = 1$.

Parametric models Parametric models assume that for $i \in [n]$, item i is associated with a strength parameter $\theta_i \in \mathbb{R}$, and

$$M_{\pi^*(i),\pi^*(j)} = F(\theta_i - \theta_j)$$

where $F : \mathbb{R} \to (0, 1)$ is a known, increasing link function. Two classical examples are the logistic function $F(x) = \frac{1}{1+e^{-x}}$ and the Gaussian cumulative density function, which correspond to the Bradley-Terry model and the Thurstone model respectively.

Noisy sorting The noisy sorting model [BM08] assumes that

$$M_{k,\ell} = \begin{cases} 1/2 + \lambda & \text{if } k > \ell, \\ 1/2 - \lambda & \text{if } k < \ell. \end{cases} \tag{1}$$

This is the model we focus on later, as it is simple yet captures important concepts and tools.

Strong stochastic transitivity Strong stochastic transitivity (SST) means that for any triplet $(k, \ell, m) \in [n]^3$ such that $k < \ell < m$, we have

$$M_{k,m} \geq M_{k,\ell} \vee M_{\ell,m}.$$

In matrix terminology, this is saying that M is bivariate isotonic (bi-isotonic) in addition to the constraint $M + M = \mathbf{1}\mathbf{1}^\top$. More precisely, all the columns of M are nonincreasing while all the rows of M are nondecreasing. Note that any parametric model, as well as the noisy sorting model, satisfies SST.

1.2 Sampling models

We consider uniform sampling. Namely, for $m \in [N]$ where N is the sample size, we observe independent outcomes

$$y_m \sim \mathsf{Ber}(M_{\pi^*(i_m),\pi^*(j_m)}), \tag{2}$$

where the random pairs (i_m, j_m) are sampled uniformly randomly with replacement from all possible pairs $\{(i,j)\}_{i \neq j}$. Here $y_m = 1$ means that $i_m \succ j_m$ and $y_m = 0$ means that $j_m \succ i_m$. We collect the outcomes of comparisons in a matrix $A \in \mathbb{R}^{n \times n}$ whose entry $A_{i,j}$ is defined to be the number of times item i beats item j.

Note that for parametric models, we have for $m \in [N]$,

$$\mathbb{E}[y_m] = F(\theta_{i_m} - \theta_{j_m}) = F\big(x_m^\top \theta\big),$$

where $x_m = e_{i_m} - e_{j_m}$ is the design point. This is simply the setup of generalized linear regression. Particularly, the Bradley-Terry model is essentially logistic regression with this special design.

2 Kendall's tau and minimax rates for noisy sorting

In general, we would like to estimate both π^* and M, but let us focus on estimating π^* under the noisy sorting model (1) for the rest of the notes. Full details of the discussion can be found in the paper [MWR18].

Consider the Kendall tau distance, i.e., the number of inversions between permutations, defined as

$$d_{\mathsf{KT}}(\pi, \sigma) = \sum_{i,j \in [n]} \mathbb{1}\big(\pi(i) > \pi(j),\, \sigma(i) < \sigma(j)\big).$$

Note that $d_{\mathsf{KT}}(\pi, \sigma) \in [0, \binom{n}{2}]$ and it is equal to the minimum number of adjacent transpositions required to change from π to σ (think of bubble sort). A closely related distance is the ℓ_1-distance, also known as Spearman's footrule, defined as

$$\|\pi - \sigma\|_1 = \sum_{i=1}^{n} |\pi(i) - \sigma(i)|.$$

It is well known [DG77] that

$$d_{\mathsf{KT}}(\pi, \sigma) \leq \|\pi - \sigma\|_1 \leq 2 d_{\mathsf{KT}}(\pi, \sigma). \tag{3}$$

Theorem 2.1. *Consider the noisy sorting model (1) with $\lambda \in (0, \frac{1}{2} - c]$ where c is a positive constant. Suppose N independent comparisons are given according to (2). Then it holds that*

$$\min_{\tilde{\pi}} \max_{\pi^*} \mathbb{E}_{\pi^*}[d_{\mathsf{KT}}(\tilde{\pi}, \pi^*)] \asymp \frac{n^3}{N\lambda^2} \wedge n^2.$$

2.1 Inversions and metric entropy

Before proving the theorem, we study the metric entropy of the set of permutations $\mathfrak{S}_n$ with respect to the Kendall tau distance d_{KT}. Let $\mathcal{B}(\pi, r) = \{\sigma \in \mathfrak{S}_n : d_{\mathsf{KT}}(\pi, \sigma) \le r\}$.

The inversion table $b_1, \dots, b_n$ of a permutation $\pi \in \mathfrak{S}_n$ is defined by

$$b_i = \sum_{j : i<j} \mathbb{1}\big(\pi(i) > \pi(j)\big).$$

Note that $b_i \in \{0, 1, \dots, n-i\}$ and $d_{\mathsf{KT}}(\pi, \mathsf{id}) = \sum_{i=1}^n b_i$. On can reconstruct a permutation using its inversion table $\{b_i\}_{i=1}^n$, so the set of inversion tables is bijective to $\mathfrak{S}_n$. (Try the permutation $(3\,5\,2\,4\,1)$ which has inversion table $(4\,2\,0\,1\,0)$.)

Lemma 2.2. *For $0 \le k \le \binom{n}{2}$, we have that*

$$n \log(k/n) - n \le \log |\mathcal{B}(\mathsf{id}, k)| \le n \log(1 + k/n) + n\,.$$

Proof. According to the discussion above, $|\mathcal{B}(\mathsf{id}, k)|$ is equal to the number of inversion tables $b_1, \dots, b_n$ such that $\sum_{i=1}^n b_i \le k$ where $b_i \in \{0, 1, \dots, n-i\}$. On the one hand, if $b_i \le \lfloor k/n \rfloor$ for all $i \in [n]$, then $\sum_{i=1}^n b_i \le k$, so a lower bound is given by

$$\begin{aligned}
|\mathcal{B}(\mathsf{id}, k)| &\ge \prod_{i=1}^n (\lfloor k/n \rfloor + 1) \wedge (n - i + 1) \\
&\ge \prod_{i=1}^{n - \lfloor k/n \rfloor} (\lfloor k/n \rfloor + 1) \prod_{i=n-\lfloor k/n \rfloor + 1}^{n} (n - i + 1) \\
&\ge (k/n)^{n - k/n} \lfloor k/n \rfloor !\,.
\end{aligned}$$

Using Stirling's approximation, we see that

$$\begin{aligned}
\log |\mathcal{B}(\mathsf{id}, k)| &\ge n \log(k/n) - (k/n) \log(k/n) + \lfloor k/n \rfloor \log \lfloor k/n \rfloor - \lfloor k/n \rfloor \\
&\ge n \log(k/n) - n\,.
\end{aligned}$$

On the other hand, if b_i is only required to be a nonnegative integer for each $i \in [n]$, then we can use a standard "stars and bars" counting argument to get an upper bound

$$|\mathcal{B}(\mathsf{id}, k)| \le \binom{n+k}{n} \le e^n (1 + k/n)^n\,.$$

Taking the logarithm finishes the proof. □

For $\varepsilon > 0$ and $S \subseteq \mathfrak{S}_n$, let $N(S, \varepsilon)$ and $D(S, \varepsilon)$ denote respectively the ε-covering number and the ε-packing number of S with respect to d_{KT}.

Proposition 2.3. *We have that for $\varepsilon \in (0, r)$,*

$$n \log \big(\frac{r}{n+\varepsilon}\big) - 2n \le \log N(\mathcal{B}(\pi, r), \varepsilon) \le \log D(\mathcal{B}(\pi, r), \varepsilon) \le n \log \big(\frac{2n+2r}{\varepsilon}\big) + 2n\,.$$

For $n \lesssim \varepsilon < r \le \binom{n}{2}$, the ε-metric entropy of $\mathcal{B}(\pi, r)$ scales as $n \log \frac{r}{\varepsilon}$. In other words, $\mathfrak{S}_n$ equipped with d_{KT} is a doubling space[1] with doubling dimension $\Theta(n)$.

Proof. The relation between the covering and the packing number is standard. We employ a volume argument for the bounds. Let $\mathcal{P}$ be a 2ε-packing of $\mathcal{B}(\pi, r)$ so that the balls $\mathcal{B}(\sigma, \varepsilon)$ are disjoint for $\sigma \in \mathcal{P}$. By the triangle inequality, $\mathcal{B}(\sigma, \varepsilon) \subseteq \mathcal{B}(\pi, r + \varepsilon)$ for each $\sigma \in \mathcal{P}$. By the invariance of the Kendall tau distance under composition, Lemma 2.2 yields

$$\begin{aligned} \log D(\mathcal{B}(\pi, r), 2\varepsilon) &\le n \log(1 + r/n) + n - n \log(\varepsilon/n) + n \\ &= n \log \big(\frac{n+r}{\varepsilon}\big) + 2n\,. \end{aligned}$$

In addition, if $\mathcal{N}$ is an ε-net of $\mathcal{B}(\pi, r)$, then the set of balls $\{\mathcal{B}(\sigma, \varepsilon)\}_{\sigma \in \mathcal{N}}$ covers $\mathcal{B}(\pi, r)$. By Lemma 2.2, we obtain

$$\begin{aligned} \log N(\mathcal{B}(\pi, r), \varepsilon) &\ge \log |\mathcal{B}(\pi, r)| - \log |\mathcal{B}(\sigma, \varepsilon)| \\ &\ge n \log(r/n) - n - n \log(1 + \varepsilon/n) - n \\ &= n \log \big(\frac{r}{n+\varepsilon}\big) - 2n\,, \end{aligned}$$

as claimed. ∎

2.2 Proof of the minimax upper bound

We only present the proof of the upper bound in Theorem 2.1 with $\lambda = 1/4$ for simplicity. The estimator we use is a sieve maximum likelihood estimator (MLE), meaning that it is the MLE over a net (called a sieve). More precisely, define $\varphi = \frac{n}{N}\binom{n}{2}$. Let $\mathcal{P}$ be a maximal φ-packing (and thus a φ-net) of $\mathfrak{S}_n$ with respect to d_{KT}. Consider the sieve MLE

$$\hat{\pi} \in \operatorname*{argmax}_{\pi \in \mathcal{P}} \sum_{\pi(i) < \pi(j)} A_{i,j}\,. \tag{4}$$

Basic setup Since $\mathcal{P}$ is a φ-net, there exists $\sigma \in \mathcal{P}$ such that $D \triangleq d_{\mathsf{KT}}(\sigma, \pi^*) \le \varphi$. By definition of $\hat{\pi}$, $\sum_{\hat{\pi}(i) < \hat{\pi}(j)} A_{i,j} \ge \sum_{\sigma(i) < \sigma(j)} A_{i,j}$. Canceling concordant pairs (i, j) under $\hat{\pi}$ and σ, we see that

$$\sum_{\hat{\pi}(i) < \hat{\pi}(j),\, \sigma(i) > \sigma(j)} A_{i,j} \ge \sum_{\hat{\pi}(i) > \hat{\pi}(j),\, \sigma(i) < \sigma(j)} A_{i,j}\,.$$

Splitting the summands according to π^* yields that

$$\sum_{\substack{\hat{\pi}(i) < \hat{\pi}(j), \\ \sigma(i) > \sigma(j), \\ \pi^*(i) < \pi^*(j)}} A_{i,j} + \sum_{\substack{\hat{\pi}(i) < \hat{\pi}(j), \\ \sigma(i) > \sigma(j), \\ \pi^*(i) > \pi^*(j)}} A_{i,j} \ge \sum_{\substack{\hat{\pi}(i) > \hat{\pi}(j), \\ \sigma(i) < \sigma(j), \\ \pi^*(i) < \pi^*(j)}} A_{i,j} + \sum_{\substack{\hat{\pi}(i) > \hat{\pi}(j), \\ \sigma(i) < \sigma(j), \\ \pi^*(i) > \pi^*(j)}} A_{i,j}\,.$$

[1] A metric space (X, d) is called a doubling space with doubling dimension $\log_2 M$, if M is the smallest number such that any ball of radius r in (X, d) can be covered with M balls of radius $r/2$.

Since $A_{i,j} \geq 0$, we may drop the rightmost term and drop the condition $\hat{\pi}(i) < \hat{\pi}(j)$ in the leftmost term to obtain that

$$\sum_{\substack{\sigma(i)>\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j} + \sum_{\substack{\hat{\pi}(i)<\hat{\pi}(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j} \geq \sum_{\substack{\hat{\pi}(i)>\hat{\pi}(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}\,. \tag{5}$$

To set up the rest of the proof, we define, for $\pi \in \mathcal{P}$,

$$\begin{aligned} L_\pi &= \left|\{(i,j) \in [n]^2 : \pi(i) < \pi(j), \sigma(i) > \sigma(j), \pi^*(i) > \pi^*(j)\}\right| \\ &= \left|\{(i,j) \in [n]^2 : \pi(i) > \pi(j), \sigma(i) < \sigma(j), \pi^*(i) < \pi^*(j)\}\right|. \end{aligned}$$

Moreover, define the random variables

$$X_\pi = \sum_{\substack{\pi(i)>\pi(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}\,, \quad Y_\pi = \sum_{\substack{\pi(i)<\pi(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j}\,, \quad \text{and} \quad Z = \sum_{\substack{\sigma(i)>\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}.$$

We show that the random process $X_\pi - Y_\pi - Z$ is positive with high probability if $d_{\mathsf{KT}}(\pi, \sigma)$ is large.

Binomial tails For a single pairwise comparison sampled uniformly from the possible $\binom{n}{2}$ pairs, the probability that

1. the chosen pair (i,j) satisfies $\pi(i) > \pi(j)$, $\sigma(i) < \sigma(j)$ and $\pi^*(i) < \pi^*(j)$, *and*
2. item i wins the comparison,

is equal to $\frac{3}{4}L_\pi\binom{n}{2}^{-1}$. By definition, X_π is the number of times the above event happens if N independent pairwise comparisons take place, so $X_\pi \sim \mathsf{Bin}\big(N, \frac{3}{4}L_\pi\binom{n}{2}^{-1}\big)$. Similarly, we have $Y_\pi \sim \mathsf{Bin}\big(N, \frac{1}{4}L_\pi\binom{n}{2}^{-1}\big)$ and $Z \sim \mathsf{Bin}\big(N, \frac{3}{4}D\binom{n}{2}^{-1}\big)$. The tails of a Binomial random variable can be bounded by the following lemma.

Lemma 2.4. *For $0 < r < p < s < 1$ and $X \sim \mathsf{Bin}(N,p)$, we have*

$$\mathbb{P}(X \leq rN) \leq \exp\Big(-N\frac{(p-r)^2}{2p(1-r)}\Big) \quad \textit{and} \quad \mathbb{P}(X \geq sN) \leq \exp\Big(-N\frac{(p-s)^2}{2s(1-p)}\Big).$$

Therefore, we obtain

1. $\mathbb{P}\left(X_\pi \leq \frac{5}{8}L_\pi N\binom{n}{2}^{-1}\right) \leq \exp\big(-L_\pi N\binom{n}{2}^{-1}/128\big)$,
2. $\mathbb{P}\left(Y_\pi \geq \frac{3}{8}L_\pi N\binom{n}{2}^{-1}\right) \leq \exp\big(-L_\pi N\binom{n}{2}^{-1}/128\big)$, and
3. $\mathbb{P}\left(Z \geq 2\varphi N\binom{n}{2}^{-1}\right) \leq \exp\big(-\varphi N\binom{n}{2}^{-1}/4\big) = \exp(-n/4)\,.$

Then we have that

$$\mathbb{P}\left(X_\pi - Y_\pi \leq \tfrac{1}{4}L_\pi N\tbinom{n}{2}^{-1}\right) \leq 2\exp\big(-L_\pi N\tbinom{n}{2}^{-1}/128\big)\,. \tag{6}$$

Peeling and union bounds For an integer $r \in [C\varphi, \binom{n}{2}]$ where C is a sufficiently large constant to be chosen, consider the slice $\mathcal{S}_r = \{\pi \in \mathcal{P} : L_\pi = r\}$. Note that if $\pi \in \mathcal{S}_r$, then

$$\begin{aligned} d_{\mathsf{KT}}(\pi, \pi^*) &= |\{(i,j) : \hat{\pi}(i) < \hat{\pi}(j), \pi^*(i) > \pi^*(j)\}| \\ &\leq |\{(i,j) : \hat{\pi}(i) < \hat{\pi}(j), \sigma(i) > \sigma(j), \pi^*(i) > \pi^*(j)\}| \\ &\quad + |\{(i,j) : \sigma(i) < \sigma(j), \pi^*(i) > \pi^*(j)\}| \\ &= L_\pi + d_{\mathsf{KT}}(\sigma, \pi^*) \leq r + \varphi\,, \end{aligned} \tag{7}$$

showing that $\mathcal{S}_r \subseteq \mathcal{B}(\pi^*, r+\varphi)$. Therefore, Proposition 2.3 gives

$$\log |\mathcal{S}_r| \leq n \log \frac{2n + 2r + 2\varphi}{\varphi} + 2n \leq n \log \frac{45r}{\varphi}\,.$$

By (6) and a union bound over $\mathcal{S}_r$, we have $\min_{\pi \in \mathcal{S}_r} (X_\pi - Y_\pi) > \frac{1}{4} r N \binom{n}{2}^{-1}$ with probability

$$1 - \exp\Big(n \log \frac{45r}{\varphi} + \log 2 - \frac{rN}{128\binom{n}{2}}\Big) \geq 1 - \exp(-2n)\,,$$

where the inequality holds by the definition of φ and the range of r. Then a union bound over integers $r \in [C\varphi, \binom{n}{2}]$ yields that

$$X_\pi - Y_\pi > \frac{C}{4} \varphi N \binom{n}{2}^{-1}$$

for all $\pi \in \mathcal{P}$ such that $L_\pi \geq C\varphi$ with probability at least $1 - e^{-n}$. This is larger than the above high probability upper bound on Z, so we conclude that with probability at least $1 - e^{-n/8}$,

$$X_\pi - Y_\pi - Z > 0$$

for all $\pi \in \mathcal{P}$ with $L_\pi \geq C\varphi$. However, (5) says that $X_{\hat{\pi}} - Y_{\hat{\pi}} - Z \leq 0$, so $L_{\hat{\pi}} \leq C\varphi$ on the above event. By (7), $d_{\mathsf{KT}}(\hat{\pi}, \pi^*) \leq L_{\hat{\pi}} + \varphi$ on the same event, which completes the proof.

3 An efficient algorithm for noisy sorting

Let us move on to present an efficient algorithm. We continue to assume $\lambda = 1/4$. To recover the underlying order of items, it is equivalent to estimate the row sums $\sum_{j=1}^n M_{\pi^*(i), \pi^*(j)}$ which we call scores of the items. Initially, for each $i \in [n]$, we estimate the score of item i by the number of wins item i has. If item i has a much higher score than item j in the first stage, then we are confident that item i is stronger than item j. Hence in the second stage, we know $M_{\pi^*(i), \pi^*(j)} = 3/4$ with high probability. For those pairs that we are not certain about, $M_{\pi^*(i), \pi^*(j)}$ is still estimated by its empirical version. The variance of each score is thus greatly reduced in the second stage, thereby yielding a more accurate order of the items. Then we iterate this process to obtain finer and finer estimates of the scores and the underlying order.

To present the T-stage sorting algorithm formally, we split the sample into T subsamples each containing N/T pairwise comparisons. For $t \in [T]$, we define a matrix $A^{(t)} \in \mathbb{R}^{n \times n}$ by setting $A^{(t)}_{i,j}$ to be the number of times item i beats item j in the t-th sample. The algorithm proceeds as follows:

1. For $i \in [n]$, define $I^{(0)}(i) = [n]$, $I_-^{(0)}(i) = \varnothing$ and $I_+^{(0)}(i) = \varnothing$. For $0 \le t \le T$, we use $I^{(t)}(i)$ to denote the set of items j whose ranking relative to i has not been determined by the algorithm at stage t.

2. At stage t, compute the score $S_i^{(t)}$ of item i:
$$S_i^{(t)} = \frac{T\binom{n}{2}}{N} \sum_{j \in I^{(t-1)}(i)} A_{i,j}^{(t)} + \frac{3}{4}\big|I_-^{(t-1)}(i)\big| + \frac{1}{4}\big|I_+^{(t-1)}(i)\big| .$$

3. Set the threshold
$$\tau_i^{(t)} \asymp n\sqrt{|I^{(t-1)}(i)| T N^{-1} \log(nT)}\,,$$
and define the sets
$$\begin{aligned} I_+^{(t)}(i) &= \{j \in [n] : S_j^{(t)} - S_i^{(t)} < -\tau_i^{(t)}\}, \\ I_-^{(t)}(i) &= \{j \in [n] : S_j^{(t)} - S_i^{(t)} > \tau_i^{(t)}\}, \text{ and} \\ I^{(t)}(i) &= [n] \setminus \big(I_-^{(t)}(i) \cup I_+^{(t)}(i)\big) . \end{aligned}$$

4. Repeat step 2 and 3 for $t = 1, \dots, T$. Output a permutation $\hat{\pi}^{\mathsf{MS}}$ by sorting the scores $S_i^{(T)}$ in nonincreasing order, i.e., $S_i^{(T)} \ge S_j^{(T)}$ if $\hat{\pi}^{\mathsf{MS}}(i) < \hat{\pi}^{\mathsf{MS}}(j)$.

We take $T = \lfloor \log\log n \rfloor$ so that the overall time complexity of the algorithm is only $O(n^2 \log\log n)$.

Theorem 3.1. *With probability at least $1 - n^{-7}$, the algorithm with $T = \lfloor \log\log n \rfloor$ stages outputs an estimator $\hat{\pi}^{\mathsf{MS}}$ that satisfies*
$$\|\hat{\pi}^{\mathsf{MS}} - \pi^*\|_\infty \lesssim \frac{n^2}{N}(\log n)\log\log n$$
and
$$d_{\mathsf{KT}}(\hat{\pi}^{\mathsf{MS}}, \pi^*) \lesssim \frac{n^3}{N}(\log n)\log\log n\,.$$

The second statement follows from the first one together with (3).

3.1 Proof (sketch) of Theorem 3.1

Assume that $\pi^* = \mathsf{id}$ without loss of generality. We define a score
$$s_i^* = \sum_{j \in [n]\setminus\{i\}} M_{i,j} = \frac{i}{2} + \frac{n}{4} - \frac{3}{4}$$
for each $i \in [n]$, which is simply the i-th row sum of M minus $1/2$.

Lemma 3.2. *Fix $t \in [T]$, $I \subseteq [n]$ and $i \in I$. Let us define*
$$S = \frac{T\binom{n}{2}}{N} \sum_{j \in I} A_{i,j}^{(t)} + \frac{3}{4}\big|\{j \in [n] \setminus I : j < i\}\big| + \frac{1}{4}\big|\{j \in [n] \setminus I : j > i\}\big| .$$
If $|I|$ is not too small, then it holds with probability at least $1 - (nT)^{-9}$ that
$$|S - s_i^*| \lesssim n\sqrt{|I| T N^{-1} \log(nT)}\,.$$

Proof. The probability that a uniform pair consists of item i and an item in $I \setminus \{i\}$, and that item i wins the comparison, is equal to $q \triangleq \big(\sum_{j \in I \setminus \{i\}} M_{i,j}\big)/\binom{n}{2}$. Thus the random variable $X \triangleq \sum_{j \in I} A^{(t)}_{i,j}$ has distribution $\mathsf{Bin}(N/T, q)$. In particular, we have $\mathbb{E}[X] = Nq/T = \frac{N}{T\binom{n}{2}} \sum_{j \in I \setminus \{i\}} M_{i,j}$, so S is an unbiased estimate of s^*_i. Moreover, we have the tail bound

$$\mathbb{P}\left(\big|X - \mathbb{E}[X]\big| \gtrsim \sqrt{qNT^{-1}\log(nT)}\right) \le (nT)^{-9},$$

from which the conclusion follows. □

We apply Lemma 3.2 inductively to each stage of the algorithm. By a union bound over all $i \in [n]$ and $t \in [T]$, all the events studied below hold with high probability. For $t \in [T]$, define

$$\mathcal{E}^{(t-1)} \triangleq \big\{j < i \text{ for all } j \in I^{(t-1)}_{-}(i) \text{ and } j > i, \text{ for all } j \in I^{(t-1)}_{+}(i)\big\}.$$

On the event $\mathcal{E}^{(t-1)}$, the score $S^{(t)}_i$ is exactly the quantity S in Lemma 3.2 with $I = I^{(t-1)}(i)$, so

$$|S^{(t)}_i - s^*_i| \lesssim n\sqrt{|I^{(t-1)}(i)|TN^{-1}\log(nT)} = \tau^{(t)}_i/2. \tag{8}$$

For any $j \in I^{(t)}_{+}(i)$, by definition $S^{(t)}_j - S^{(t)}_i < -\tau^{(t)}_i$, so we have $s^*_j < s^*_i$ and thus $j > i$. Similarly, $j < i$ for any $j \in I^{(t)}_{-}(i)$. Hence $\mathcal{E}^{(t)}$ occurs with high probability. Moreover, if $|s^*_j - s^*_i| > 2\tau^{(t)}_i$, then $|S^{(t)}_j - S^{(t)}_i| > \tau^{(t)}_i$, so $j \notin I^{(t)}(i)$. Hence if $j \in I^{(t)}(i)$, then $|j - i| \lesssim \tau^{(t)}_i$. Consequently,

$$|I^{(t)}(i)| \lesssim \tau^{(t)}_i \lesssim n\sqrt{|I^{(t-1)}(i)|TN^{-1}\log(nT)}\,. \tag{9}$$

Note that if we have $\alpha^{(0)} = n$ and the iterative relation $\alpha^{(t)} \le \beta\sqrt{\alpha^{(t-1)}}$ where $\alpha^{(t)} > 0$ and $\beta > 0$, then it is easily seen that $\alpha^{(t)} \le \beta^2 n^{2^{-t}}$. Consequently, we obtain that

$$|I^{(T-1)}(i)| \lesssim \frac{n^2 T}{N}\log(nT) n^{2^{-T+1}} \lesssim \frac{n^2}{N}(\log n)(\log\log n)$$

for $T = \lfloor \log\log n \rfloor$. Taking T to be larger does not make $|I^{(T-1)}(i)|$ smaller, because Lemma 3.2 requires a lower bound on $|I^{(T-1)}(i)|$. The details are left out. It follows from (8) that

$$|S^{(T)}_i - s^*_i| \lesssim \frac{n^2}{N}(\log n)\log\log n =: \delta\,.$$

As the permutation $\hat{\pi}^{\mathsf{MS}}$ is defined by sorting the scores $S^{(T)}_i$ in nonincreasing order, we see that $\hat{\pi}^{\mathsf{MS}}(i) < \hat{\pi}^{\mathsf{MS}}(j)$ for all pairs (i, j) with $s^*_i - s^*_j > 2\delta$, i.e., $j - i > \delta$.

Finally, suppose that $\hat{\pi}^{\mathsf{MS}}(i) - i > \delta$ for some $i \in [n]$. Then there exists $j > i + \delta$ such that $\hat{\pi}^{\mathsf{MS}}(j) < \hat{\pi}^{\mathsf{MS}}(i)$, contradicting the guarantee we have just proved. A similar argument leads to a contradiction if $\hat{\pi}^{\mathsf{MS}}(i) - i < -\delta$. Therefore, we obtain that $|\hat{\pi}^{\mathsf{MS}}(i) - i| \le \delta$, completing the proof.

References

[BM08] Mark Braverman and Elchanan Mossel. Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 268–276. ACM, New York, 2008.

[DG77] Persi Diaconis and Ronald L. Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 262–268, 1977.

[MWR18] Cheng Mao, Jonathan Weed, and Philippe Rigollet. Minimax rates and efficient algorithms for noisy sorting. *Proceedings of the 28th International Conference on Algorithmic Learning Theory*, 2018.