Supplementary Material for
“Statistical Inference for Incomplete Ranking Data: The Case of Rank-Dependent Coarsening”
1 Proofs of Theoretical Results in Section 7

Definition 1. Let $\hat{\pi}_N$ denote the ranking produced as a prediction by a ranking method on the basis of $N$ observed (pairwise) preferences. The method is consistent if $p(\hat{\pi}_N = \pi^*) \to 1$ for $N \to \infty$.

Definition 2. Consider a complete ranking $\pi \in S_K$, and let us consider two indices $i \neq j$. We define the $(i,j)$-swap ranking, $\pi_{i,j} : [K] \to [K]$, as follows: $\pi_{i,j}(k) = \pi(k)$, $\forall k \in [K] \setminus \{i,j\}$, $\pi_{i,j}(i) = \pi(j)$ and $\pi_{i,j}(j) = \pi(i)$.

Lemma 3. (Lemma 2 in the paper) Let us consider a probability measure $p_\theta$ over $S_K$. Consider $q_{i,j} = \sum_{\pi \in E(a_i \succ a_j)} p_\theta(\pi) \lambda_{\pi(i),\pi(j)}$, $\forall i \neq j$. (The model (8) in the paper, without assuming that the marginal distribution is necessarily PL). Then:

$$p_\theta(\pi) \geq p_\theta(\pi_{i,j}), \forall \pi \in E(a_i \succ a_j) \Rightarrow q_{i,j} > q_{j,i}.$$

Proof. We easily observe that:

$$q_{i,j} = \sum_{\pi \in E(a_i \succ a_j)} p_\theta(\pi) \lambda_{\pi(i),\pi(j)}$$

$$q_{j,i} = \sum_{\pi \in E(a_j \succ a_i)} p_\theta(\pi) \lambda_{\pi(j),\pi(i)}$$

Furthermore, let us notice that the set $E(a_j \succ a_i)$ coincides with $\{\pi_{i,j} : \pi \in E(a_i \succ a_j)\}$ and that $\lambda_{\pi_{i,j}(i),\pi_{i,j}(j)} = \lambda_{\pi(j),\pi(i)}$ for every $\pi \in E(a_i \succ a_j)$. Therefore, we can write:

$$q_{j,i} = \sum_{\pi \in E(a_i \succ a_j)} p_\theta(\pi_{i,j}) \lambda_{\pi(i),\pi(j)}.$$

By hypothesis, the following inequalities hold:

$$p_\theta(\pi) \geq p_\theta(\pi_{i,j}), \forall \pi \in E(a_i \succ a_j),$$

and therefore we deduce that $q_{i,j} > q_{j,i}$.

Lemma 4. Consider the PL model with $\theta_i > 0$ for all $i \in [K]$, and let $\lambda = \{\lambda_{u,v} | 1 \leq u < v \leq K\}$ be any (pairwise) coarsening such that $\lambda_{u,v}$ is the probability to select positions $u$ and $v$. Then, $q_{i,j} > 0$ for all $i, j \in [K]$, $i \neq j$. Thus, each preference $a_i \succ a_j$ has a positive probability to be observed.

Proof. Take any $\lambda_{u,v} > 0$ and fix $i, j \in [K]$, $i \neq j$. According to the PL model, if $\theta_k > 0$ for all $k \in [K]$, $\text{pl}_\theta(\pi) > 0$ for all $\pi \in S_K$. Thus, there is a probability $p > 0$ that $\pi(i) = u$ and $\pi(j) = v$. Consequently, $q_{i,j} \geq p\lambda_{u,v} > 0$.

Lemma 5. (Lemma 3 in the paper) Assume the model (8) and let $\theta_i > 0$ for all $i \in [K]$, $\theta_i \neq \theta_j$ for $i \neq j$. The coarsening (7) is order-preserving for PL in the sense that $p_{i,j} > 1/2$ if and only if $q_{i,j} > 1/2$, where $q_{i,j} = q_{i,j}/(q_{i,j} + q_{j,i})$. 

2
Proof. First, note that, according to the previous lemma, \( q_{i,j} > 0 \) for all \( i, j \in [K] \), \( i \neq j \), so all \( q'_{i,j} \) are well defined.

- Let us first prove the “only if” part. According to Lemma 3, it only remains to prove that any Plackett-Luce distribution \( pl_\theta \) satisfies the following implication:

\[
p_{i,j} > p_{j,i} \Rightarrow pl_\theta(\pi) \geq pl_\theta(\pi_{i,j}), \; \forall \pi \in E(a_i > a_j).
\]

Let us first notice that \( p_{i,j} > p_{j,i} \) if and only if \( \theta_i > \theta_j \). Let us take an arbitrary ranking \( \pi \in E(a_i > a_j) \). We can write:

\[
pl_\theta(\pi) = C_{i,j} \cdot \frac{\theta_{\pi^{-1}(\pi(i))}}{\sum_{s=\pi(i)} \theta_{\pi^{-1}(s)}} \cdot \frac{\theta_{\pi^{-1}(\pi(j))}}{\sum_{s=\pi(j)} \theta_{\pi^{-1}(s)}}
\]

\[
pl_\theta(\pi_{i,j}) = C_{i,j} \cdot \frac{\theta_{\pi_{i,j}^{-1}(\pi_{i,j}(i))}}{\sum_{s=\pi_{i,j}(i)} \theta_{\pi_{i,j}^{-1}(s)}} \cdot \frac{\theta_{\pi_{i,j}^{-1}(\pi_{i,j}(j))}}{\sum_{s=\pi_{i,j}(j)} \theta_{\pi_{i,j}^{-1}(s)}}
\]

where

\[
C_{i,j} = \prod_{r \notin \{\pi(i),\pi(j)\}} \frac{\theta_{\pi^{-1}(r)}}{\theta_{\pi(i)} + \theta_{\pi(i+1)} + \cdots + \theta_{\pi(K)}}
\]

According to the relation between \( \pi \) and \( \pi_{i,j} \), we can easily check the following equality:

\[
\sum_{s=\pi(i)} \theta_{\pi^{-1}(s)} = \sum_{s=\pi_{i,j}(i)} \theta_{\pi_{i,j}^{-1}(s)}
\]

(In fact, both \( \theta_i \) and \( \theta_j \) appear in both sums). Furthermore, we observe that:

\[
\sum_{s=\pi(j)} \theta_{\pi^{-1}(s)} - \sum_{s=\pi_{i,j}(i)} \theta_{\pi_{i,j}^{-1}(s)} = \theta_j - \theta_i,
\]

and therefore

\[
\sum_{s=\pi(j)} \theta_{\pi^{-1}(s)} < \sum_{s=\pi_{i,j}(i)} \theta_{\pi_{i,j}^{-1}(s)}.
\]

We deduce that \( pl_\theta(\pi) > pl_\theta(\pi_{i,j}) \).

- Let us now prove the “if” part. Suppose that \( q'_{i,j} > 1/2 \). Therefore, according to the “only if part”, \( p_{i,j} \) must be greater than or equal to 1/2 (as otherwise, we would get \( q'_{i,j} < 1/2 \)). Now, according to the hypotheses, all the components of the parameter \( \theta \) are different from each other, and therefore \( p_{i,j} \neq 1/2 \), so we deduce that it must be strictly greater than 1/2.
Lemma 6. (Lemma 4 in the paper) Assume the model (8), $\theta_i \neq \theta_j$ for $i \neq j$, and $\theta_i > 0$ for all $i \in [K]$. Let us take an arbitrarily small $\epsilon^* > 0$. There exists $N_0 \in \mathbb{N}$ such that $\theta_i > \theta_j$ if and only if $\hat{p}_{i,j} > 1/2$ for all $i, j \in [K]$, with probability at least $1 - \epsilon^*$, after having observed at least $N_0$ preferences.

Proof. Take an arbitrary pair $(i, j)$, and let us consider the sequence $(\hat{p}_{i,j}^{(n)})_{n \in \mathbb{N}}$, where $\hat{p}_{i,j}^{(n)} = \frac{c_{i,j}^{(n)}}{c_{i,j}^{(n)} + c_{j,i}^{(n)}}$, and $c_{i,j}^{(n)}$ denotes the number of times the pair $a_i \succ a_j$ is observed in the sample. According to the Strong Law of Large Numbers, the sequence $\hat{p}_{i,j}^{(n)} = \frac{c_{i,j}^{(n)}}{c_{i,j}^{(n)} + c_{j,i}^{(n)}}$ converges in probability to $q'_{i,j} = \frac{q_{i,j}}{q_{i,j} + q_{j,i}}$. This means that, for any pair of arbitrary $\epsilon > 0$ and $\delta > 0$, there exists $N_{\delta, i, j} \in \mathbb{N}$ such that $|\hat{p}_{i,j}^{(n)} - q'_{i,j}| < \delta$, with probability greater than $1 - \epsilon$, for every $n \geq N_{\delta, i, j}$. Now, $\theta_i < \theta_j$ if and only if $\hat{p}_{i,j} > 0.5$, which is equivalent to saying that $q'_{i,j} = q_{i,j}/(q_{i,j} + q_{j,i}) > 1/2$, according to Lemma 5. Let us now take $\delta = \min_{i,j} |q'_{i,j} - 1/2|/2$, $\epsilon = \epsilon^*/K(K-1)$, and $N_0 = \max_{i,j} N_{\delta, i, j}$. Then, thanks to the union bound, with probability at least $1 - \epsilon^*$, we can assure for every $n \geq N_0$ that $\hat{p}_{i,j}^{(n)} > 0.5$ if and only if $q'_{i,j} > 1/2$ for every pair $(i, j)$.

Theorem 7. (Theorem 5 in the paper) Copeland ranking is consistent.

Proof. It is a direct consequence of Lemma 6.

Theorem 8. (Theorem 6 in the paper) FAS, FAS(R), and FAS(B) are consistent.

Proof. Let us separately consider the three cases.

- FAS(R). Let us first notice that $\hat{p}_{i,j} + \hat{p}_{j,i} = 1$ and therefore,

$$\arg \min_{\pi \in S_k} \left( \sum_{(i,j) : \pi(i) < \pi(j)} \hat{p}_{j,i} \right) = \arg \max_{\pi \in S_k} \left( \sum_{(i,j) : \pi(i) < \pi(j)} \hat{p}_{i,j} \right).$$

Now, for an arbitrary $\pi \in S_k$, the following equality holds:

$$\sum_{(i,j) : \pi(i) < \pi(j)} \hat{p}_{i,j} = \sum_{i < j} r_{i,j}^\pi,$$

where $r_{i,j}^\pi$ is defined as follows for every $i < j$:

$$r_{i,j}^\pi = \begin{cases} \hat{p}_{i,j} & \text{if } \pi(i) < \pi(j) \\ \hat{p}_{j,i} & \text{otherwise} \end{cases}.$$
Furthermore, according to Lemmas 5 and 6, for an arbitrarily small $\epsilon^* > 0$, there exists $N_0 \in \mathbb{N}$ such that $\theta_i > \theta_j$, or equivalently, $q_{i,j} > q_{i,j}$, if and only if $\hat{p}_{i,j} > 1/2$, for every pair $(i, j)$ and every $n \geq N_0$, with probability greater than or equal to $1 - \epsilon^*$. Let us now consider the function $f : \mathcal{S}_K \to \mathbb{R}$: $f(\pi) = \sum_{(i,j) : \pi(i) < \pi(j)} \pi^\pi_{i,j}$, where

$$\pi^\pi_{i,j} = \begin{cases} q_{i,j} & \text{if } \pi(i) < \pi(j) \\ q_{j,i} & \text{otherwise.} \end{cases}$$

The argument of the maximum of this function is $\pi^* = \text{arg sort}\{\theta_1, \ldots, \theta_k\}$. Consequently, the solution to FAS, after having observed at least $N_0$ preferences, coincides with $\text{arg sort}\{\theta_1, \ldots, \theta_k\}$ with probability at least $1 - \epsilon^*$.

- **FAS.** The proof is analogous to the previous case. (Let us notice that $c_{i,j} > c_{j,i}$ if and only if $\hat{p}_{i,j} > 1/2$).

- **FAS(B).** First of all, let us take into account that $\mathbb{I}(\hat{p}_{j,i} > 1/2) + \mathbb{I}(\hat{p}_{i,j} > 1/2) = 1$, for every $(i, j)$ and therefore the FAS(B) ranking is $\hat{\pi}$ satisfying:

$$\hat{\pi} = \arg \max_{\pi \in \mathcal{S}_K} \sum_{(i,j) : \pi(i) < \pi(j)} \mathbb{I}(\hat{p}_{i,j} > 1/2) = \arg \max_{\pi \in \mathcal{S}_K} \#\{(i, j) : \pi(i) < \pi(j) \text{ and } \hat{p}_{i,j} > 1/2\}.$$  

Now, let us take an arbitrarily small $\epsilon^* > 0$. According to Lemma 6, there exists $N_0$ such that $\hat{p}_{i,j} > 1/2$ if and only if $\theta_i > \theta_j$, for every pair $(i,j)$ and for all $n \geq N_0$ with probability at least $1 - \epsilon^*$. Therefore, with probability at least $1 - \epsilon^*$ and for a sufficiently large sample, we can equivalently write that the solution to the FAS(B) algorithm is $\hat{\pi}$ satisfying:

$$\hat{\pi} = \arg \max_{\pi \in \mathcal{S}_K} \sum_{(i,j) : \pi(i) < \pi(j)} \mathbb{I}(\hat{p}_{i,j} > 1/2) = \arg \max_{\pi \in \mathcal{S}_K} \#\{(i, j) : \pi(i) < \pi(j) \text{ and } \theta_i > \theta_j\}.$$  

Clearly, the solution to this problem is $\pi^* = \text{arg sort}\{\theta_1, \ldots, \theta_k\}$ (the mode of the underlying PL distribution).

2 Experimental Results
Figure 1: Simulation results of algorithms for coarsened rankings when full rankings are generated according to PL (with $K = 3$) and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 2: Simulation results of algorithms for coarsened rankings when full rankings are generated according to PL (with $K = 4$) and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 3: Simulation results of algorithms for coarsened rankings when full rankings are generated according to PL (with $K = 5$) and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 4: Simulation results of algorithms for coarsened rankings when full rankings are generated according to PL (with $K = 7$) and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 5: Performance of algorithms averaged over all coarsening positions \((i, j)\) when rankings are generated according to PL. Each plot from left to right corresponds to the number of items \(K \in \{3, 4, 5, 7\}\), respectively. X-axis is the sample size. Y-axis shows the Kendall distance.
Figure 6: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 3$ and dispersion $\phi = 0.1$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 7: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 4$ and dispersion $\phi = 0.1$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 8: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 5$ and dispersion $\phi = 0.1$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 9: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 7$ and dispersion $\phi = 0.1$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 10: Performance of algorithms averaged over all coarsening positions \((i, j)\) when rankings are generated according to Mallows with dispersion \(\phi = 0.1\). Each plot from left to right corresponds to the number of items \(K \in \{3, 4, 5, 7\}\), respectively. X-axis is the sample size. Y-axis shows the Kendall distance.
Figure 11: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 4$ and dispersion $\phi = 0.5$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.
Figure 12: Simulation results of algorithms for coarsened rankings when full rankings are generated according to Mallows (with $K = 5$ and dispersion $\phi = 0.5$), and coarsening is degenerate ($\lambda_{i,j} = 1$ for some $1 \leq i < j \leq K$). Full breaking (extraction of all pairwise preferences) is shown as a baseline. X-axis shows the sample size. Y-axis is the Kendall distance averaged over 500 runs.