1 General Notations

- N: Number of users
- n: Number of items, $n = \begin{cases} 2m-1, & \text{if n is odd} \\ 2m, & \text{otherwise} \end{cases}$
- $\mathcal{P}_n$: the space of permutation of n items
- $\boldsymbol{R}^1, ..., \boldsymbol{R}^N$: full rankings given by the users
- $\boldsymbol{R}^j \in \mathcal{P}_n = \{R_1^j, ..., R_n^j\} \sim \text{Mallows}(\boldsymbol{\rho}^0, \alpha^0)$
- $P(\boldsymbol{\rho}|\boldsymbol{R}^1, ..., \boldsymbol{R}^N, \alpha^o)$: Mallows posterior
- $\{i_1, ..., i_n\}$: a ranking of n items that determines the sequence following which the items are to be sampled. i.e. $i_j = k$ indicates that item j is the k-th item is to be sampled
- $\{o_1, ..., o_n\}$: an ordering of n items that corresponds to $\{i_1, ..., i_n\}$ s.t. $i_{o_k} = k$. $\{o_1, ..., o_n\}$ and $\{i_1, ..., i_n\}$ have a one-to-one relationship
- $\sum_{\{i_1,...,i_n\} \in \mathcal{P}_n} q(\tilde{\boldsymbol{\rho}}|i_1, ..., i_n, \alpha_0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N) \cdot g(i_1, ..., i_n| \dots)$: pseudolikelihood that approximates the Mallows posterior
- $q(\tilde{\boldsymbol{\rho}}|i_1, ..., i_n, \alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N) = q(\tilde{\boldsymbol{\rho}}|o_1, ..., o_n, \alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N)$
  $= q(\tilde{\rho}_{o_1}|\alpha^0, o_1, R_{o_1}^1, ..., R_{o_1}^N) \cdot q(\tilde{\rho}_{o_2}|\alpha^0, o_2, \tilde{\rho}_{o_1} R_{o_2}^1, ..., R_{o_2}^N) \cdot ... \cdot$
  $q(\tilde{\rho}_{o_{n-1}}|\alpha^0, o_{n-1}, \tilde{\rho}_{o_1}, ..., \tilde{\rho}_{o_{n-2}}, R_{n-1}^1, ..., R_{n-1}^N) \cdot q(\tilde{\rho}_{o_n}|\alpha^0, o_n, \tilde{\rho}_{o_1}, ..., \tilde{\rho}_{o_{n-1}}, R_n^1, ..., R_n^N)$
  - $q(\tilde{\rho}_{o_1}|\alpha^0, o_1, R_{o_1}^1, ..., R_{o_1}^N) = \dfrac{\exp\{-\frac{\alpha_0}{n}\sum_{j=1}^N d(R_{o_1}^j, \tilde{\rho}_{o_1})\}\mathbb{1}_{\tilde{\rho}_{o_1} \in \{1,...,n\}}}{\sum_{\tilde{r}_{o_1} \in \{1,..,n\}} \exp\{-\frac{\alpha_0}{n}\sum_{j=1}^N d(R_{o_1}^j, \tilde{r}_{o_1})\}}$
  - $q(\tilde{\rho}_{o_k}|\alpha^0, o_k, \tilde{\rho}_{o_1}, ..., \tilde{\rho}_{o_{k-1}}, R_{o_k}^1, ..., R_{o_k}^N) = \dfrac{\exp\{-\frac{\alpha_0}{n}\sum_{j=1}^N d(R_{o_k}^j, \tilde{\rho}_{o_k})\}\mathbb{1}_{\tilde{\rho}_{o_k} \in \{1,..,n\} \setminus \{\tilde{\rho}_{o_1}, ..., \tilde{\rho}_{o_{k-1}}\}}}{\sum_{\tilde{r}_{o_k} \in \{1,..,n\} \setminus \{\tilde{\rho}_{o_1}, ..., \tilde{\rho}_{o_{k-1}}\}} \exp\{-\frac{\alpha_0}{n}\sum_{j=1}^N d(R_{o_k}^j, \tilde{r}_{o_k})\}}$

    for $k = 2, ..., n$

- $\boldsymbol{o}^0$: a set of ordering that corresponds to $\boldsymbol{\rho}^0$ s.t. ${\rho^0}^{-1}(m) = o^0_m$
- Define the "v-function" $f_v(\cdot)$ such that $f_v(\boldsymbol{\rho}^0) = \mathcal{V}_{\boldsymbol{\rho}^0}$, where
  - $\mathcal{V}_{\boldsymbol{\rho}^o} = \begin{cases} \{\boldsymbol{r} \in \mathcal{P}_n : r_{o^0_m} = 1, r_{o^0_{m\pm k}} \in \{2k, 2k+1\}, k = 1, ..., m-1\}, & \text{if n is odd} \\ \{\boldsymbol{r} \in \mathcal{P}_n : \{r_{o^0_{m-k}}, r_{o^0_{m+k+1}}\} \in \{2k+1, 2k+2\}, k = 0, ..., m\}, & \text{if n is even} \end{cases}$

2 Theorems and Lemmas

2.1

Lemma 2.1.1 *Given there are odd number of items, i.e. $n = 2m-1$. $\forall \alpha^0 \in (0, \infty)$,*

1. $\mathbb{E}(R_{o^0_m} | \boldsymbol{\rho}_0, \alpha^0) = \rho^0_{o_m} = m$
2. $\forall j \in [1, m-2]$, $j < \mathbb{E}[R_{o^0_j} | \boldsymbol{\rho}^0, \alpha^0] < \mathbb{E}[R_{o^0_{j+1}} | \boldsymbol{\rho}^0, \alpha^0] < m$
3. $\forall j \in [m+2, 2m-1]$, $m < \mathbb{E}[R_{o^0_{j-1}} | \boldsymbol{\rho}^0, \alpha^0] < \mathbb{E}[R_{o^0_j} | \boldsymbol{\rho}^0, \alpha^0] < j$

Similarly, if n is even, i.e. $n = 2m$, $\forall \alpha^0 \in (0, \infty)$,

1. $\forall j \in [1, m-1]$, $j < \mathbb{E}[R_{o^0_j} | \boldsymbol{\rho}^0, \alpha^0] < \mathbb{E}[R_{o^0_{j+1}} | \boldsymbol{\rho}^0, \alpha^0]$
2. $\forall j \in [m+2, 2m]$, $\mathbb{E}[R_{o^0_{j-1}} | \boldsymbol{\rho}^0, \alpha^0] < \mathbb{E}[R_{o^0_j} | \boldsymbol{\rho}^0, \alpha^0] < j$

Note that for both cases, it satisfies that $\forall 1 \leq j < k \leq n$ and $\forall \alpha > 0$, $\mathbb{E}[R_{o^0_j} | \boldsymbol{\rho}^0, \alpha^0] < \mathbb{E}[R_{o^0_k} | \boldsymbol{\rho}^0, \alpha^0]$

Lemma 2.1.2 *As $N \to \infty$, $\frac{1}{N} \sum\limits_{j=1}^{N} R^j_i \to \mathbb{E}[R_i | \boldsymbol{\rho}^0, \alpha^0]$, $\forall i = 1, ..., n$*

Definition 1 *Given a vector of length n, i.e. $\{x_1, ..., x_n\}$,the function $rank(x_1, ..., x_n)$ is defined as $rank(x_1, ..., x_n) = \{r_1, ..., r_n\}$ such that $x_{(r_k)} = x_k$, $\forall k = 1, ..., n$*

Theorem 2.1.3 *As $N \to \infty$, and $\forall \alpha > 0$,*
$rank(\frac{1}{N} \sum\limits_{j=0}^{N} R^j_1, ..., \frac{1}{N} \sum\limits_{j=0}^{N} R^j_n) \to rank(\mathbb{E}[R_1 | \boldsymbol{\rho}^0, \alpha_0], ..., \mathbb{E}[R_n | \boldsymbol{\rho}^0, \alpha_0]) = \boldsymbol{\rho}^0$

To rephrase, as N approaches infinity, the Mallows consensus parameter $\boldsymbol{\rho}^0$ can be inferred from the data by taking the marginal mean for each item and then apply the rank function to these marginal means.

2.2

Theorem 2.2.1 *For a function g defined on $\mathcal{P}_n$ which can depend on $\boldsymbol{\rho}^0$,*

$$\underset{g \in \mathcal{D}_{\boldsymbol{\rho}^0}}{\arg\min} \lim_{N\to\infty} KL(P(\boldsymbol{\rho}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N)|| \sum_{\{i_1,...,i_n\}\in\mathcal{P}_n} q(\tilde{\boldsymbol{\rho}}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N, i_1, ..., i_n) g(i_1, ..., i_n|\boldsymbol{\rho}^0)$$
$$= g^*(i_1, ..., i_n|\mathcal{V}_{\boldsymbol{\rho}^0}),$$

where

- $\mathcal{D}_{\boldsymbol{\rho}^0}$ *is the set of all distrbutions on* $\mathcal{P}_n$*, which can depend on* $\boldsymbol{\rho}^0$
- $g^*(i_1, ..., i_n|\mathcal{V}_{\boldsymbol{\rho}^0})$ *is a distribution whose density is concentrated on* $\boldsymbol{\rho}^0$*, defined as*
$$\begin{cases} g^*(i_1, ..., i_n|\mathcal{V}_{\boldsymbol{\rho}^0}) = |\mathcal{V}_{\boldsymbol{\rho}^0}|^{-1} > 0, & \text{if } \{i_1, ..., i_n\} \in \mathcal{V}_{\boldsymbol{\rho}^0} \\ g^*(i_1, ..., i_n|\mathcal{V}_{\boldsymbol{\rho}^0}) = 0, & \text{if } \{i_1, ..., i_n\} \notin \mathcal{V}_{\boldsymbol{\rho}^0} \end{cases}, \text{ where } |\mathcal{V}_{\boldsymbol{\rho}^0}| = \begin{cases} 2^{m-1}, & \text{if } n \text{ is odd} \\ 2^m, & \text{otherwise} \end{cases}$$

That is to say, for a set of distributions g, which are defined on the space of permutation of n items $\mathcal{P}_n$, the distribution g^* that minimizes the KL-divergence betweeeh the Mallows posterior and the pseudolikelihood defined above, is a uniform distribution with its density concentrated on $\mathcal{V}_{\boldsymbol{\rho}^o}$

2.3

For a given $N < \infty$, define $\hat{\boldsymbol{\rho}}^0$ as $rank(\frac{1}{N}\sum_{j=0}^{N} R_1^j, ..., \frac{1}{N}\sum_{j=0}^{N} R_n^j)$ and $\mathcal{V}_{\hat{\boldsymbol{\rho}}^0} = f_v(\hat{\boldsymbol{\rho}}^0)$.

Theorem 2.3.1 $\exists \sigma \geq 0$ *and* $g'(i_1, ..., i_n|\mathcal{V}_{\hat{\boldsymbol{\rho}}^0}, \sigma)$ *such that*

$$KL\ (P(\boldsymbol{\rho}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N)|| \sum_{\{i_1,...,i_n\}\in\mathcal{P}_n} q(\tilde{\boldsymbol{\rho}}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N, i_1, ..., i_n) g^*(i_1, ..., i_n|\mathcal{V}_{\hat{\boldsymbol{\rho}}^0)} \geq$$
$$KL\ (P(\boldsymbol{\rho}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N)|| \sum_{\{i_1,...,i_n\}\in\mathcal{P}_n} q(\tilde{\boldsymbol{\rho}}|\alpha^0, \boldsymbol{R}^1, ..., \boldsymbol{R}^N, i_1, ..., i_n) g'(i_1, ..., i_n|\mathcal{V}_{\hat{\boldsymbol{\rho}}^0}, \sigma)$$

where $g'(i_1, ..., i_n|\mathcal{V}_{\hat{\boldsymbol{\rho}}^0}, \sigma) = \sum_{\hat{\boldsymbol{v}}\in\mathcal{V}_{\hat{\boldsymbol{\rho}}^0}} \{g^*(\hat{\boldsymbol{v}}|\mathcal{V}_{\hat{\boldsymbol{\rho}}^0}) \int_{\boldsymbol{x}} \mathcal{F}_r(i_1, ..., i_n|x_1, ..., x_n) \prod_{i=1}^{n} \mathcal{N}(x_i|\hat{v}_i, \sigma) d\boldsymbol{x}\}$*, and*

- $\hat{\boldsymbol{v}} \sim g^*(\hat{\boldsymbol{v}}|\mathcal{V}_{\hat{\boldsymbol{\rho}^0}})$
- $x_i \sim \mathcal{N}(x_i|\hat{v}_i, \sigma)$ *for* $i = 1, ..., n$
- $i_1, ..., i_n \sim \mathcal{F}_r(i_1, ..., i_n|x_1, ..., x_n)$*, where* $\mathcal{F}_r = \begin{cases} 1, & \text{if } \{i_1, ..., i_n\} = rank(x_1, ..., x_n) \\ 0, & \text{otherwise} \end{cases}$

As N is limited, $\boldsymbol{\rho}^0$ and therefore, $\mathcal{V}_{\boldsymbol{\rho}^0}$ usually cannot be accurately inferred from the data. We can however, sample for $i_1, ..., i_n$ by sampling for each item i from a univariate Gaussian distribution centeredd on $\hat{v}_i$ with a fixed variance σ for all items, and then obtain a ranking using the rank function. By introducing the variance, a smaller KL divergence from the Mallows posterior can be achieved.

2.4

Theorem 2.4.1 *With the usage of $g'(i_1, ..., i_n | \mathcal{V}_{\hat{\boldsymbol{\rho}}^0}, \sigma)$, the value of σ that minimizes the KL-divergence between the Mallows posterior and the resulted pseudolikelihood is*

$$\sigma = \begin{cases} 0, & \textit{if } \delta(\alpha^0, n, N) \leq \delta^* \\ f(\alpha^0, n, N), & \textit{otherwise} \end{cases}$$

In other words, σ should be 0 when $\delta(\alpha^0, n, N) \geq \delta^*$. Beyond this point, the optimal choice of σ should be greater than 0, and it follows a function $f(\alpha^0, n, N)$.

2.5

Theorem 2.5.1 *As $N \to \infty, \sigma = 0 \ \forall \alpha > 0$ and ≥ 1*