Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Partition–Mallows Model and Its Inference for Rank Aggregation

Wanchuang Zhu, Yingkai Jiang, Jun S. Liu & Ke Deng

To cite this article: Wanchuang Zhu, Yingkai Jiang, Jun S. Liu & Ke Deng (2021): Partition–Mallows Model and Its Inference for Rank Aggregation, Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1930547

To link to this article: https://doi.org/10.1080/01621459.2021.1930547

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–17: Theory and Methods
https://doi.org/10.1080/01621459.2021.1930547

Taylor & Francis
Taylor & Francis Group

Partition–Mallows Model and Its Inference for Rank Aggregation

Wanchuang Zhu[a], Yingkai Jiang[b], Jun S. Liu[c], and Ke Deng[a]

[a]Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing, China; [b]Yau Mathematical Sciences Center & Department of Mathematical Sciences, Tsinghua University, Beijing, China; [c]Department of Statistics, Harvard University, Cambridge, MA

ABSTRACT

Learning how to aggregate ranking lists has been an active research area for many years and its advances have played a vital role in many applications ranging from bioinformatics to internet commerce. The problem of discerning reliability of rankers based only on the rank data is of great interest to many practitioners, but has received less attention from researchers. By dividing the ranked entities into two disjoint groups, that is, relevant and irrelevant/background ones, and incorporating the Mallows model for the relative ranking of relevant entities, we propose a framework for rank aggregation that can not only distinguish quality differences among the rankers but also provide the detailed ranking information for relevant entities. Theoretical properties of the proposed approach are established, and its advantages over existing approaches are demonstrated via simulation studies and real-data applications. Extensions of the proposed method to handle partial ranking lists and conduct covariate-assisted rank aggregation are also discussed.

ARTICLE HISTORY
Received November 2019
Accepted May 2021

KEYWORDS
Covariate-assisted aggregation; Heterogeneous rankers; Mallows model; Meta-analysis; Partial ranking lists

1. Introduction

Rank data arise naturally in many fields, such as web searching (Renda and Straccia 2003), design of recommendation systems (Linas, Tadas, and Francesco 2010), and genomics (Bader 2011). Many probabilistic models have been proposed for analyzing this type of data, among which the Thurstone model (Thurstone 1927), the Mallows model (Mallows 1957), and the Plackett-Luce model (Luce 1959; Plackett 1975) are the most well-known representatives. The Thurstone model assumes that each entity possesses a hidden score and all the scores come from a joint probability distribution. The Mallows model is a location model defined on the permutation space of ordered entities, in which the probability mass of a permuted order is an exponential function of its distance from the true order. The Plackett-Luce model assumes that the preference of entity E_i is associated with a weight w_i, and describes a recursive procedure for generating a random ranking list: entities are picked one by one with the probability proportional to their weights in a sequential fashion without replacement, and ranked based on their order of being selected.

Rank aggregation aims to derive a "better" aggregated ranking list $\hat{\tau}$ from multiple ranking lists $\tau_1, \tau_2, \ldots, \tau_m$. It is a classic problem and has been studied in a variety of contexts for decades. Early applications of rank aggregation can be traced back to the 18th-century France, where the idea of rank aggregation was proposed to solve the problem of political elections (Borda 1781). In the past 30 years, efficient rank aggregation algorithms have played important roles in many fields, such as web searching (Renda and Straccia 2003), information retrieval (Fagin et al. 2003), design of recommendation systems (Linas, Tadas, and Francesco 2010), social choice studies (Porello and Endriss 2012; Soufiani, Parkes, and Xia 2014), genomics (Bader 2011), and bioinformatics (Lin and Ding 2010; Chen et al. 2016).

Some popular approaches for rank aggregation are based on certain summary statistics. These methods simply calculate a summary statistics, such as the mean, median, or geometric mean, for each entity E_i based on its rankings across different ranking lists, and obtain the aggregated ranking list based on these summary statistics. Optimization-based methods obtain the aggregated ranking by minimizing a user-defined objective function, that is, let $\hat{\tau} = \arg\min_{\tau} \frac{1}{m} \sum_{i=1}^{m} d(\tau, \tau_i)$, where distance measurement $d(\cdot, \cdot)$ could be either *Spearman's footrule distance* (Diaconis and Graham 1977) or the *Kendall tau distance* (Diaconis 1988). More detailed studies on these optimization-based methods can be found in Young and Levenglick (1978),Young (1988), and Dwork et al. (2001).

In early 2000s, a novel class of Markov chain-based methods have been proposed (Dwork et al. 2001; Lin and Ding 2010; Lin 2010; Deconde et al. 2011), which first use the observed ranking lists to construct a probabilistic transition matrix among the entities and then use the magnitudes of the entities' equilibrium probabilities of the resulting Markov chain to rank them. The boosting-based method *RankBoost* (Freund et al. 2003), employs a *feedback function* $\Phi(i, j)$ to construct the final ranking, where $\Phi(i, j) > 0$ (or ≤ 0) indicates that entity E_i is (or is not) preferred to entity E_j. Some statistical methods use aforementioned probabilistic models (such as the Thurstone

CONTACT Ke Deng kdeng@tsinghua.edu.cn Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing, China.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

model) and derive the maximum likelihood estimate (MLE) of the final ranking. More recently, researchers have begun to pay attention to rank aggregation methods for pairwise comparison data (Rajkumar and Agarwal 2014; Chen and Suh 2015; Chen et al. 2019).

We note that all aforementioned methods assume that the rankers of interest are equally reliable. In practice, however, it is very common that some rankers are more reliable than the others, whereas some are nearly non-informative and may be regarded as "spam rankers." Such differences in the rankers' qualities, if ignored in analysis, may significantly corrupt the rank aggregation and lead to seriously misleading results. To the best of our knowledge, the earliest effort to address this critical issue can be traced to Aslam and Montague (2001), which derived an aggregated ranking list by calculating a weighted summation of the observed ranking lists, known as the *Borda Fuse*. Lin and Ding (2010) extended the objective function of Dwork et al. (2001) to a weighted fashion. Independently, Liu et al. (2007) proposed a supervised rank aggregation to determine weights of the rankers by training with some external data. Although assigning weights to rankers is an intuitive and simple way to handle quality differences, how to scientifically determine these weights is a critical and unsolved problem in the aforementioned works.

Recently, Deng et al. (2014) proposed BARD, a Bayesian approach to deal with quality differences among independent rankers without the need of external information. BARD introduces a Partition Model, which assumes that all involved entities can be partitioned into two groups: the relevant ones and the background ones. A rationale of the approach is that, in many applications, distinguishing relevant entities from background ones has the priority over the construction of a final ranking of all entities. Under this setting, BARD decomposes the information in a ranking list into three components: (i) the relative rankings of all background entities, which is assumed to be uniform; (ii) the relative ranking of each relevant entity among all background ones, which takes the form of a truncated power-law; and, (iii) the relative rankings of all relevant entities, which is again uniform. The parameter of the truncated power-law distribution, which is ranker-specific, naturally serves as a quality measure for each ranker, as a ranker of a higher quality means a less spread truncated power-law distribution.

Li et al. (2020) proposed a stage-wise data-generation process based on an extended Mallows model (EMM) introduced by Fligner and Verducci (1986). EMM assumes that each entity comes from a two-components mixture model involving a uniform distribution to model non-informative entities, a modified Mallows model for informative entities and a ranker-specific proportion parameter. Li, Yi, and Liu (2021) followed the Thurstone model framework to deal with available covariates for the entities as well as different qualities of the rankers. In their model, each entity is associated with a Gaussian-distributed latent score and a ranking list is determined by the ranking of these scores. The quality of each ranker is determined by the standard deviation parameter in the Gaussian model so that a larger standard deviation indicates a poorer quality ranker.

Although these recent articles have proposed different ways for learning the quality variation among rankers, they all suffer from some limitations. The BARD method (Deng et al. 2014) simplifies the problem by assuming that all relevant entities are exchangeable. In many applications, however, the observed ranking lists often have a strong ordering information for relevant entities, and simply labeling these entities as "relevant" without considering their relative rankings tends to lose too much information and oversimplify the problem. Li et al. (2020) did not explicitly measure quality differences by their extended Mallows model. Although they mentioned that some of their model parameters can indicate the rankers' qualities, it is not clear how to properly combine multiple indicators to produce an easily interpretable quality measurement. The learning framework of Li, Yi, and Liu (2021) based on Gaussian latent variables appears to be more suitable for incorporating covariates than for handling heterogeneous rankers.

In this article, we propose a *Partition–Mallows Model* (PAMA), which combines the partition modeling framework of Deng et al. (2014) with the Mallows model, to accommodate the detailed ordering information among the relevant entities. The new framework can not only quantify the quality difference of rankers and distinguish relevant entities from background entities like BARD, but also provide an explicit ranking estimate among the relevant entities in rank aggregation. In contrast to the strategy of imposing the Mallows on the full ranking lists, which tends to be sensitive to noises in low-ranking entities, the combination of the Partition and Mallows Models allows us to focus on highly ranked entities, which typically contain high-quality signals in data, and is thus more robust. Both simulation studies and real data applications show that the proposed approach is superior to existing methods, for example, BARD and EMM, for a large class of rank aggregation problems.

The rest of this article is organized as follows. A brief review of BARD and the Mallows model is presented in Section 2 as preliminaries. The proposed PAMA model is described in Section 3 with some key theoretical properties established. Statistical inference of the PAMA model, including the Bayesian inference and the pursuit of MLE, is detailed in Section 4. Performance of PAMA is evaluated and compared to existing methods via simulations in Section 5. Two real data applications are shown in Section 6 to demonstrate strength of the PAMA model in practice. Finally, we conclude the article with a short discussion in Section 7.

2. Notations and Preliminaries

Let $U = \{E_1, E_2, \ldots, E_n\}$ be the set of entities to be ranked. We use "$E_i \preceq E_j$" to represent that entity E_i is preferred to entity E_j in a ranking list τ, and denote the position of entity E_i in τ by $\tau(i)$. Note that more preferred entities always have lower rankings. Our research interest is to aggregate m observed ranking lists, $\tau_1, \ldots, \tau_m$, presumably constructed by m rankers independently into one consensus ranking list which is supposed to be "better" than each individual one.

2.1. BARD and Its Partition Model

The Partition Model in BARD (Deng et al. 2014) assumes that U can be partitioned into two non-overlapping subsets: $U = U_R \cup U_B$, with U_R representing the set of relevant entities and

U_B for the background ones. Let $I = \{I_i\}_{i\in U}$ be the vector of group indicators, where $I_i = \mathbb{I}(E_i \in U_R)$ and $\mathbb{I}(\cdot)$ is the indicator function. This formulation makes sense in many applications where people are only concerned about a fixed number of top-ranked entities. Under this formulation, the information in a ranking list τ_k can be equivalently represented by a triplet $(\tau_k^0, \tau_k^{1|0}, \tau_k^1)$, where τ_k^0 denotes relative rankings of all background entities, $\tau_k^{1|0}$ denotes relative rankings of relevant entities among the background entities and τ_k^1 denotes relative rankings of all relevant entities.

Deng et al. (2014) suggested a three-component model for τ_k by taking advantage of its equivalent decomposition:

$$\begin{aligned} P(\tau_k \mid I) &= P(\tau_k^0, \tau_k^{1|0}, \tau_k^1 \mid I) \\ &= P(\tau_k^0 \mid I) \times P(\tau_k^{1|0} \mid I) \times P(\tau_k^1 \mid \tau_k^{1|0}, I), \end{aligned} \quad (1)$$

where both $P(\tau_k^0 \mid I)$ (relative ranking of the background entities) and $P(\tau_k^1 \mid \tau_k^{1|0}, I)$ (relative ranking of the relevant entities conditional on their set of positions relative to background entities) are uniform, and the relative ranking of a relevant entity E_i among background ones follows a power-law distribution with parameter $\gamma_k > 0$, that is,

$$P(\tau_k^{1|0}(i) = t \mid I) = q(t \mid \gamma_k, n_0) \propto t^{-\gamma_k} \cdot \mathbb{I}(1 \le t \le n_0 + 1),$$

leading to the following explicit forms for the three terms in Equation (1):

$$P(\tau_k^0 \mid I) = \frac{1}{n_0!}, \quad (2)$$

$$\begin{aligned} P(\tau_k^{1|0} \mid I) &= \prod_{i\in U_R} q(\tau_k^{1|0}(i) \mid \gamma_k, I) \\ &= \frac{1}{(B_{\tau_k,I})^{\gamma_k} \times (C_{\gamma_k,n_1})^{n_1}}, \end{aligned} \quad (3)$$

$$P(\tau_k^1 \mid \tau_k^{1|0}, I) = \frac{1}{A_{\tau_k,I}} \times \mathbb{I}\big(\tau_k^1 \in \mathcal{A}_{U_R}(\tau_k^{1|0})\big), \quad (4)$$

where $n_1 = \sum_{i=1}^n I_i$ and $n_0 = n - n_1$ are the counts of relevant and background entities respectively, $B_{\tau_k,I} = \prod_{i\in U_R} \tau_k^{1|0}(i)$, $C_{\gamma_k,n_1} = \sum_{t=1}^{n_0+1} t^{-\gamma_k}$ is the normalizing constant of the power-law distribution, $\mathcal{A}_{U_R}(\tau_k^{1|0})$ is the set of τ_k^1's that are compatible with $\tau_k^{1|0}$, and $A_{\tau_k,I} = \#\{\mathcal{A}_{U_R}(\tau_k^{1|0})\} = \prod_{t=1}^{n_0+1}(n_{\tau_k,t}^{1|0}!)$ with $n_{\tau_k,t}^{1|0} = \sum_{i\in U_R} \mathbb{I}(\tau_k^{1|0}(i) = t)$.

Intuitively, this model assumes that each ranker first randomly places all background entities to generate τ_k^0, then "inserts" each relevant entity independently into the list of background entities according to a truncated power-law distribution to generate $\tau_k^{1|0}$, and finally draws τ_k^1 uniformly from $\mathcal{A}_{U_R}(\tau_k^{1|0})$. In other words, τ_k^0 serves as a baseline for modeling $\tau_k^{1|0}$ and τ_k^1. It is easy to see from the model that a more reliable ranker should possess a larger γ_k. With the assumption of independent rankers, we have the full-data likelihood:

$$\begin{aligned} P(\tau_1, \ldots, \tau_m \mid I, \boldsymbol{\gamma}) &= \prod_{k=1}^m P(\tau_k \mid I, \gamma_k) = [(n_0)!]^{-m} \\ &\times \prod_{k=1}^m \frac{\mathbb{I}\big(\tau_k^1 \in \mathcal{A}_{U_R}(\tau_k^{1|0})\big)}{A_{\tau_k,I} \times (B_{\tau_k,I})^{\gamma_k} \times \big(C_{\gamma_k,n_1}\big)^{n_1}}, \end{aligned} \quad (5)$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$. A detailed Bayesian inference procedure for $(I, \boldsymbol{\gamma})$ via Markov chain Monte Carlo can be found in Deng et al. (2014).

2.2. The Mallows Model

Mallows (1957) proposed the following probability model for a ranking list τ of n entities:

$$\pi(\tau \mid \tau_0, \phi) = \frac{1}{Z_n(\phi)} \cdot \exp\{-\phi \cdot d(\tau, \tau_0)\}, \quad (6)$$

where τ_0 denotes the true ranking list, $\phi > 0$ characterizing the reliability of τ, function $d(\cdot,\cdot)$ is a distance metric between two ranking lists, and

$$Z_n(\phi) = \sum_{\tau'} \exp\{-\phi \cdot d(\tau', \tau_0)\} = \frac{\prod_{t=2}^n (1 - e^{-t\phi})}{(1 - e^{-\phi})^{n-1}} \quad (7)$$

being the normalizing constant, whose analytic form was derived in Diaconis (1988). Clearly, a larger ϕ means that τ is more stable and concentrates in a tighter neighborhood of τ_0. A common choice of $d(\cdot,\cdot)$ is the Kendall tau distance.

The Mallows model under the Kendall tau distance can also be equivalently described by an alternative multistage model, which selects and positions entities one by one in a sequential fashion, where ϕ serves as a common parameter that governs the probabilistic behavior of each entity in the stochastic process (Mallows 1957). Later on, Fligner and Verducci (1986) extended the Mallows model by allowing ϕ to vary at different stages, that is, introducing a position-specific parameter ϕ_i for each position i, which leads to a very flexible, in many cases too flexible, framework to model rank data. To stabilize the generalized Mallows model by Fligner and Verducci (1986), Li et al. (2020) proposed to put a structural constraint on ϕ_is of the form $\phi_i = \phi \cdot (1-\alpha^i)$ with $0 < \phi < 1$ and $0 \le \alpha \le 1$. As a probabilistic model for rank data, the Mallows model enjoys great interpretability, model compactness, inference and computation efficiency. For a comprehensive review of the Mallows model and its extensions, see Irurozki et al. (2014) and Li et al. (2020).

3. The PAMA

The Partition Model employed by BARD (Deng et al. 2014) tends to oversimplify the problem for scenarios where we care about the detailed rankings of relevant entities. To further enhance the Partition Model of BARD so that it can reflect the detailed rankings of relevant entities, we describe a new PAMA in this section.

3.1. The Reverse Partition Model

To combine the Partition Model with the Mallows Model, a naive strategy is to simply replace the uniform model for the relevant entities, that is, $P(\tau_k^1 \mid \tau_k^{1|0}, I)$ in (1), by the Mallows Model, which leads to the updated Equation (4) as below:

$$P(\tau_k^1 \mid \tau_k^{1|0}, I) = \frac{\pi(\tau_k^1)}{Z_{\tau_k,I}} \times \mathbb{I}\big(\tau_k^1 \in \mathcal{A}_{U_R}(\tau_k^{1|0})\big),$$

$\mathcal{I}^+$	$\mathcal{I}^0$	I		$\mathcal{I}$	U	τ_k		τ_k^1	$\tau_k^{0\|1}$	τ_k^0		τ_k^0	$\tau_k^{1\|0}$	τ_k^1
1	-	1		1	E_1	2		2	-	-		-	1	2
2	-	1		2	E_2	6		4	-	-		-	3	4
3	-	1		3	E_3	4		3	-	-		-	2	3
4	-	1		4	E_4	1		1	-	-		-	1	1
5	-	1	$\Longleftarrow$	5	E_5	7	$\Longleftrightarrow$	5	-	-	$\Longleftrightarrow$	-	3	5
-	0	0		0	E_6	5		-	3	2		2	-	-
-	0	0		0	E_7	3		-	4	1		1	-	-
-	0	0		0	E_8	8		-	1	3		3	-	-
-	0	0		0	E_9	9		-	1	4		4	-	-
-	0	0		0	E_{10}	10		-	1	5		5	-	-

Figure 1. An illustrative example of construction of $\mathcal{I}^+, \mathcal{I}^0$ and I based on the enhanced indicator vector $\mathcal{I}$ of $n_1 = 5$ to a universe of 10 entities, and the decomposition of a ranking list τ_k into triplet $(\tau_k^1, \tau_k^{0|1}, \tau_k^0)$ and $(\tau_k^0, \tau_k^{1|0}, \tau_k^1)$, respectively, given $\mathcal{I}$.

where $\pi(\tau_k^1)$ is the Mallows density of τ_k^1 and $Z_{\tau_k,I} = \sum_{\tau \in \mathcal{A}_{U_R}(\tau_k^{1|0})} \pi(\tau)$ is the normalizing constant of the Mallows model with a constraint due to the compatibility of τ_k^1 with respect to $\mathcal{A}_{U_R}(\tau_k^{1|0})$. Apparently, the calculation of $Z_{\tau_k,I}$, which involves the summation over the whole space of $\mathcal{A}_{U_R}(\tau_k^{1|0})$, whose size is $A_{\tau_k,I} = \#\{\mathcal{A}_{U_R}(\tau_k^{1|0})\} = \prod_{t=1}^{n_0+1}(n_{\tau_k,t}^{1|0}!)$, is infeasible for most practical cases, rendering such a naive combination of the Mallows model and the Partition Model impractical.

To avoid the challenging computation caused by the constraints due to $\mathcal{A}_{U_R}(\tau_k^{1|0})$, we rewrite the Partition Model by switching the roles of τ_k^0 and τ_k^1 in the model: instead of decomposing τ_k as $(\tau_k^0, \tau_k^{1|0}, \tau_k^1)$ conditioning on the group indicators I, we decompose τ_k into an alternative triplet $(\tau_k^1, \tau_k^{0|1}, \tau_k^0)$, where $\tau_k^{0|1}$ denotes the *relative reverse rankings* of background entities among the relevant ones. Formally, we note that $\tau_k^{0|1}(i) \triangleq n_1 + 2 - \tau_{k|\{i\}\cup U_R}(i)$ for any $i \in U_R$, where $\tau_{k|\{i\}\cup U_R}(i)$ denotes the relative ranking of a background entity among the relevant ones. In this *reverse Partition Model*, we first order the relevant entities according to a certain distribution and then use them as a reference system to "insert" the background entities. Figure 1 illustrates the equivalence between τ_k and its two alternative presentations, $(\tau_k^0, \tau_k^{1|0}, \tau_k^1)$ and $(\tau_k^1, \tau_k^{0|1}, \tau_k^0)$.

Given the group indicator vector I, the reverse Partition Model based on $(\tau_k^1, \tau_k^{0|1}, \tau_k^0)$ gives rise to the following distributional form for τ_k:

$$\begin{aligned} P(\tau_k \mid I) &= P(\tau_k^1, \tau_k^{0|1}, \tau_k^0 \mid I) \\ &= P(\tau_k^1 \mid I) \times P(\tau_k^{0|1} \mid I) \times P(\tau_k^0 \mid \tau_k^{0|1}, I), \end{aligned} \tag{8}$$

which is analogous to Equation (1) for the original Partition Model in BARD. Comparing to Equation (1), however, the new form Equation (8) enables us to specify an unconstrained marginal distribution for τ_k^1. Moreover, due to the symmetry between $\tau_k^{1|0}$ and $\tau_k^{0|1}$, it is highly likely that the power-law distribution, which was shown in Deng et al. (2014) to approximate the distribution of $\tau_k^{1|0}(i)$ well for each $E_i \in U_R$, can also model $\tau_k^{0|1}(i)$ for each $E_i \in U_B$ reasonably well. Detailed numerical validations are shown in the supplementary material.

If we assume that all relevant entities are exchangeable, then all background entities are exchangeable, and the relative reserve ranking of a background entity among the relevant entities follows a power-law distribution, we have

$$P(\tau_k^1 \mid I) = \frac{1}{n_1!}, \tag{9}$$

$$\begin{aligned} P(\tau_k^{0|1} \mid I, \gamma_k) &= \prod_{i \in U_B} P(\tau_k^{0|1}(i) \mid I, \gamma_k) \\ &= \frac{1}{(B^*_{\tau_k,I})^{\gamma_k} \times (C^*_{\gamma_k,n_1})^{n_0}}, \end{aligned} \tag{10}$$

$$P(\tau_k^0 \mid \tau_k^{0|1}, I) = \frac{1}{A^*_{\tau_k,I}} \times \mathbb{I}\big(\tau_k^0 \in \mathcal{A}_{U_R}(\tau_k^{0|1})\big), \tag{11}$$

where n_1 and n_0 are numbers of relevant and background entities, respectively, $B^*_{\tau_k,I} = \prod_{i \in U_B} \tau_k^{0|1}(i)$ is the unnormalized part of the power-law, $C^*_{\gamma_k,n_1} = \sum_{t=1}^{n_1+1} t^{-\gamma_k}$ is the normalizing constant, $\mathcal{A}_{U_B}(\tau_k^{0|1})$ is the set of all τ_k^0 that are compatible with a given $\tau_k^{0|1}$, and $A^*_{\tau_k,I} = \#\{\mathcal{A}_{U_B}(\tau_k^{0|1})\} = \prod_{t=1}^{n_1+1}(n_{\tau_k,t}^{0|1}!)$ with $n_{\tau_k,t}^{0|1} = \sum_{i \in U_B} \mathbb{I}(\tau_k^{0|1}(i) = t)$. Apparently, the likelihood of this reverse-Partition Model shares the same structure as that of the original Partition Model in BARD, and thus can be inferred in a similar way.

3.2. The PAMA

The reverse Partition Model introduced in Section 3.1 allows us to freely model τ_k^1 beyond a uniform distribution, which is

infeasible for the original Partition Model in BARD. Here we employ the Mallows model for τ_k^1 due to its interpretability, compactness and computability. To achieve this, we replace the group indicator vector I in the Partition Model by a more general indicator vector $\mathcal{I} = \{\mathcal{I}_i\}_{i=1}^n$, which takes value in $\Omega_{\mathcal{I}}$, the space of all permutations of $\{1,\ldots,n_1,\underbrace{0,\ldots,0}_{n_0}\}$, with $\mathcal{I}_i = 0$ if $E_i \in U_B$, and $\mathcal{I}_i = k > 0$ if $E_i \in U_R$ and is ranked at position k among all relevant entities in U_R. Figure 1 provides an illustrative example of assigning an enhanced indicator vector $\mathcal{I}$ to a universe of 10 entities with $n_1 = 5$.

Based on the status of $\mathcal{I}$, we can define subvectors $\mathcal{I}^+$ and $\mathcal{I}^0$, where $\mathcal{I}^+$ stands for the subvector of $\mathcal{I}$ containing all positive elements in $\mathcal{I}$, and $\mathcal{I}^0$ for the remaining zero elements in $\mathcal{I}$. Figure 1 demonstrates the constructions of $\mathcal{I}$, $\mathcal{I}^+$ and $\mathcal{I}^0$, and the equivalence between τ_k, $(\tau_k^0, \tau_k^{1|0}, \tau_k^1)$, and $(\tau_k^1, \tau_k^{0|1}, \tau_k^0)$ given $\mathcal{I}$. Note that different from the Partition Model in BARD, in which we allow the number of relevant entities represented by n_1 to vary nearby its expected value, the number of relevant entities in the new model, is assumed to be fixed and known for conceptual and computational convenience. In other words, we have $|U_R| = n_1$ in the new setting.

As an analogy of Equations (1) and (8), we have the following decomposition of τ_k given the enhanced indicator vector $\mathcal{I}$

$$\begin{aligned} P(\tau_k \mid \mathcal{I}) &= P(\tau_k^1, \tau_k^{0|1}, \tau_k^0 \mid \mathcal{I}) \\ &= P(\tau_k^1 \mid \mathcal{I}) \times P(\tau_k^{0|1} \mid \mathcal{I}) \times P(\tau_k^0 \mid \tau_k^{0|1}, \mathcal{I}). \end{aligned} \quad (12)$$

Assume that $\tau_k^1 \mid \mathcal{I}$ follows the Mallows model (with parameter ϕ_k) centered at $\mathcal{I}^+$

$$P(\tau_k^1 \mid \mathcal{I}, \phi_k) = P(\tau_k^1 \mid \mathcal{I}^+, \phi_k) = \frac{\exp\{-\phi_k \cdot d_\tau(\tau_k^1, \mathcal{I}^+)\}}{Z_{n_1}(\phi_k)}, \quad (13)$$

where $d_\tau(\cdot,\cdot)$ denotes Kendall tau distance and $Z_{n_1}(\phi_k)$ is defined as in Equation (7). Clearly, a larger ϕ_k indicates that ranker τ_k is of a higher quality, as the distribution is more concentrated at the "true ranking" defined by $\mathcal{I}^+$. Since the relative ranking of background entities are of no interest to us, we still assume that they are randomly ranked. Together with the power-law assumption for $\tau_k^{0|1}(i)$, we have

$$\begin{aligned} P(\tau_k^{0|1} \mid \mathcal{I}) &= P(\tau_k^{0|1} \mid I, \gamma_k) \\ &= \frac{1}{(B^*_{\gamma_k,I})^{\gamma_k} \times (C^*_{\gamma_k,n_1})^{n-n_1}}, \end{aligned} \quad (14)$$

$$\begin{aligned} P(\tau_k^0 \mid \tau_k^{0|1}, \mathcal{I}) &= P(\tau_k^0 \mid \tau_k^{0|1}, I) \\ &= \frac{1}{A^*_{\tau_k,I}} \times \mathbb{I}\big(\tau_k^0 \in \mathcal{A}_{U_R}(\tau_k^{0|1})\big), \end{aligned} \quad (15)$$

where notations $A^*_{\tau_k,I}$, $B^*_{\tau_k,I}$ and $C^*_{\gamma_k,n_1}$ are the same as in the reverse-Partition Model. We call the resulting model the *PAMA*.

Different from the Partition and reverse-Partition models, which quantify the quality of ranker τ_k with only one parameter γ_k in the power-law distribution, the PAMA model contains two quality parameters ϕ_k and γ_k, with the former indicating the ranker's ability of ranking relevant entities and the latter reflecting the ranker's ability in differentiating relevant entities from background ones. Intuitively, ϕ_k and γ_k reflect the quality of ranker τ_k in two different aspects. However, considering that a good ranker is typically strong in both dimensions, it looks quite natural to further simplify the model by assuming

$$\phi_k = \phi \cdot \gamma_k, \quad (16)$$

with $\phi > 0$ being a common factor for all rankers. This assumption, while reducing the number of free parameters by almost half, captures the natural positive correlation between ϕ_k and γ_k and serves as a first-order (i.e., linear) approximation to the functional relationship between ϕ_k and γ_k. A wide range of numerical studies based on simulated data suggest that the linear approximation showed in Equation (16) works reasonably well for many typical scenarios for rank aggregation. In contrast, the more flexible model with both ϕ_k and γ_k as free parameters (which is referred to as PAMA*) suffers from unstable performance from time to time. Detailed evidences to support assumption (16) can be found in the supplementary material.

Plugging in Equation (16) into Equation (13), we have a simplified model for τ_k^1 given $\mathcal{I}$ as follows:

$$\begin{aligned} P(\tau_k^1 \mid \mathcal{I}, \phi, \gamma_k) &= P(\tau_k^1 \mid \mathcal{I}^+, \phi, \gamma_k) \\ &= \frac{\exp\{-\phi \cdot \gamma_k \cdot d_\tau(\tau_k^1, \mathcal{I}^+)\}}{Z_{n_1}(\phi \cdot \gamma_k)}. \end{aligned} \quad (17)$$

Combining Equations (14), (15), and (17), we get the full likelihood of τ_k

$$\begin{aligned} P(\tau_k \mid \mathcal{I}, \phi, \gamma_k) &= P(\tau_k^1 \mid \mathcal{I}, \phi, \gamma_k) \times P(\tau_k^{0|1} \mid \mathcal{I}, \gamma_k) \times P(\tau_k^0 \mid \tau_k^{0|1}, \mathcal{I}) \\ &= \frac{\mathbb{I}\big(\tau_k^0 \in \mathcal{A}_{U_R}(\tau_k^{0|1})\big)}{A^*_{\tau_k,I} \times (B^*_{\tau_k,I})^{\gamma_k} \times (C^*_{\gamma_k,n_1})^{n-n_1} \times (D^*_{\tau_k,\mathcal{I}})^{\phi\cdot\gamma_k} \times E^*_{\phi,\gamma_k}}, \end{aligned} \quad (18)$$

where $D^*_{\tau_k,\mathcal{I}} = \exp\{d_\tau(\tau_k^1, \mathcal{I}^+)\}$, $E^*_{\phi,\gamma_k} = Z_{n_1}(\phi \cdot \gamma_k) = \frac{\prod_{t=2}^{n_1}(1-e^{-t\phi\gamma_k})}{(1-e^{-\phi\gamma_k})^{n_1-1}}$, and $A^*_{\tau_k,\mathcal{I}}$, $B^*_{\tau_k,\mathcal{I}}$ and $C^*_{\tau_k,n_1}$ keep the same meaning as in the reverse-Partition model. At last, for the set of observed ranking lists $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_m)$ from m independent rankers, we have the joint likelihood:

$$P(\boldsymbol{\tau} \mid \mathcal{I}, \phi, \boldsymbol{\gamma}) = \prod_{k=1}^m P(\tau_k \mid \mathcal{I}, \phi, \gamma_k). \quad (19)$$

3.3. Model Identifiability and Estimation Consistency

Let Ω_n be the space of all permutations of $\{1,\ldots,n\}$ in which τ_k takes value, and let $\boldsymbol{\theta} = (\mathcal{I}, \phi, \boldsymbol{\gamma})$ be the vector of model parameters. The PAMA model in (19), that is, $P(\boldsymbol{\tau} \mid \boldsymbol{\theta})$, defines a family of probability distributions on Ω_n^m indexed by parameter $\boldsymbol{\theta}$ taking values in space $\Theta = \Omega_{\mathcal{I}} \times \Omega_\phi \times \Omega_{\boldsymbol{\gamma}}$, where $\Omega_{\mathcal{I}}$ is the space of all permutations of $\{1,\ldots,n_1,\mathbf{0}_{n_0}\}$, $\Omega_\phi = (0,+\infty)$ and $\Omega_{\boldsymbol{\gamma}} = [0,+\infty)^m$. We show here that the PAMA model defined in Equation (19) is identifiable and the model parameters can be estimated consistently under mild conditions.

Theorem 1. The PAMA model is identifiable, that is,

$$\begin{aligned} &\forall\, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta, \text{ if } P(\boldsymbol{\tau} \mid \boldsymbol{\theta}_1) \\ &\quad = P(\boldsymbol{\tau} \mid \boldsymbol{\theta}_2) \text{ for } \forall\, \boldsymbol{\tau} \in \Omega_n^m, \text{ then } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2. \end{aligned} \quad (20)$$

Proof. See the supplementary material. □

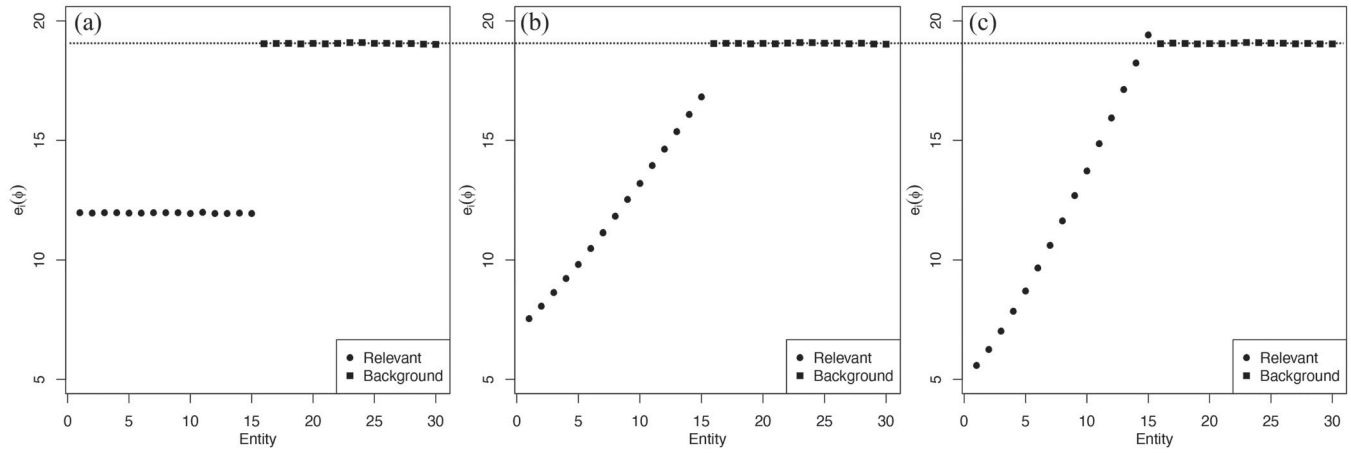

Figure 2. Average ranks of all the entities with fixed $\mathcal{I} = (1, \ldots, n_1, 0, \ldots, 0)$, n = 30, $n_1 = 15$, $m = 100{,}000$ and $F(\gamma) = U(0, 2)$. (a), (b), and (c) are the corresponding results for $\phi = 0, 0.2,$ and 0.4 respectively.

To show that parameters in the PAMA model can be estimated consistently, we will first construct a consistent estimator for the indicator vector $\mathcal{I}$ as $m \to \infty$ but with the number of ranked entities n fixed, and show later that ϕ can also be consistently estimated once $\mathcal{I}$ is given. To this end, we define $\bar{\tau}(i) = m^{-1} \sum_{k=1}^{m} \tau_k(i)$ to be the average rank of entity E_i across all m rankers, and assume that the ranker-specific quality parameters $\gamma_1, \ldots, \gamma_m$ are iid samples from a non-atomic probability measure $F(\gamma)$ defined on $[0, \infty)$ with a finite first moment (referred to as condition $\boldsymbol{C}_\gamma$ hereinafter). Then, by the strong law of large numbers we have

$$\bar{\tau}(i) = \frac{1}{m} \sum_{k=1}^{m} \tau_k(i) \to \mathbb{E}\big[\tau(i)\big] \text{ a.s. with } m \to \infty, \tag{21}$$

since $\{\tau_k(i)\}_{k=1}^{m}$ are iid random variables with expectation

$$\mathbb{E}\big[\tau(i)\big] = \mathbb{E}\Big[\mathbb{E}\big[\tau(i) \mid \gamma\big]\Big] = \int \mathbb{E}\big[\tau(i) \mid \gamma\big] dF(\gamma),$$

where $\mathbb{E}\big[\tau(i) \mid \gamma\big]$ is the conditional mean of $\tau(i)$ given the model parameters $(\mathcal{I}, \phi, \gamma)$, that is,

$$\mathbb{E}\big[\tau(i) \mid \gamma\big] = \sum_{t=1}^{n} t \cdot P\big(\tau(i) = t \mid \mathcal{I}, \phi, \gamma\big).$$

Clearly, $\mathbb{E}\big[\tau(i)\big]$ is a function of ϕ given $\mathcal{I}$ and $F(\gamma)$. We define $e_i(\phi) \triangleq \mathbb{E}\big[\tau(i)\big]$ to emphasize $\mathbb{E}\big[\tau(i)\big]$'s nature as a continuous function of ϕ. Without loss of generality, we suppose that $U_R = \{1, \ldots, n_1\}$ and $U_B = \{n_1 + 1, \ldots, n\}$, that is, $\mathcal{I} = (1, \ldots, n_1, 0, \ldots, 0)$, hereinafter. Then, the Partition structure and the Mallows model embedded in the PAMA model lead to the following facts:

$$\begin{aligned} &e_1(\phi) < \cdots < e_{n_1}(\phi) \text{ and } e_{n_1+1}(\phi) = \cdots = e_n(\phi) \\ &\quad = e_0, \ \forall\, \phi \in \Omega_\phi. \end{aligned} \tag{22}$$

Note that $e_i(\phi)$ degenerates to a constant with respect to ϕ (i.e., e_0) for all $i > n_1$ because parameter ϕ influences only the relative rankings of relevant entities in the Mallows model. The value of e_0 is completely determined by $F(\gamma)$. For the BARD model, it is easy to see that $e_1 = \cdots = e_{n_1} \le e_{n_1+1} = \cdots = e_n$.

Figure 2 shows some empirical estimates of the $e_i(\phi)$'s based on $m = 100{,}000$ independent samples drawn from PAMA models with $n = 30$, $n_1 = 15$, and $F(\gamma) = U(0, 2)$, but three different ϕ values: (a) $\phi = 0$, which corresponds to the BARD model; (b) $\phi = 0.2$; and (c) $\phi = 0.5$. One surprise is that in case (c), some relevant entities may have a larger $e_i(\phi)$ (worse ranking) than the average rank of background entities. Lemma 1 guarantees that for almost all $\phi \in \Omega_\phi$, e_0 is different from $e_i(\phi)$ for $i = 1, \ldots, n_1$. The proof of Lemma 1 can be found in the supplementary material.

Lemma 1. For the PAMA model with condition $\boldsymbol{C}_\gamma$, $\exists\ \tilde{\Omega}_\phi \subset \Omega_\phi$, s.t. $(\Omega_\phi - \tilde{\Omega}_\phi)$ contains only finite elements and

$$e_i(\phi) \neq e_0 \text{ for } i = 1, \ldots, n_1, \ \forall\, \phi \in \tilde{\Omega}_\phi. \tag{23}$$

The facts demonstrated in Equations (22) and (23) suggest the following three-step strategy to estimate $\mathcal{I}$: (a) find the subset S_0 of $n_0 = (n - n_1)$ entities from U so that the within-subset variation of the $\bar{\tau}(i)$'s is the smallest, that is,

$$\begin{aligned} &S_0 = \arg\min_{S \in U,\ |S| = n_0} \sum_{i \in S} (e_i - \bar{e}_S)^2, \ \text{ with} \\ &\bar{e}_S = n_0^{-1} \sum_{i \in S} e_i, \end{aligned} \tag{24}$$

and let $\tilde{U}_B = S_0$ be an estimate of U_B; (b) rank the entities in $U \setminus S_0$ by $\bar{\tau}(i)$ increasingly and use the obtained ranking $\tilde{\mathcal{I}}^+$ as an estimate of $\mathcal{I}^+$; (c) combine the above two steps to obtain the estimate $\tilde{\mathcal{I}}$ of $\mathcal{I}$. This can be achieved by defining $\tilde{U}_R = U \setminus \tilde{U}_B$ and $\tilde{\mathcal{I}}^+ = rank(\{\bar{\tau}(i) : i \in \tilde{U}_R\})$, and obtain $\tilde{\mathcal{I}} = (\tilde{\mathcal{I}}_1, \ldots, \tilde{\mathcal{I}}_n)$, with $\tilde{\mathcal{I}}_i = \tilde{\mathcal{I}}_i^+ \cdot \mathbb{I}(i \in \tilde{U}_R)$.

Note that $\tilde{U}_B$ is based on the mean ranks, $\{\bar{\tau}(i)\}_{i \in U}$, thus is clearly a moment estimator. Although this three-step estimation strategy is neither statistically efficient nor computationally feasible (step (a) is NP-hard), it nevertheless serves as a prototype for developing the consistency theory. Theorem 2 guarantees that $\tilde{\mathcal{I}}$ is a consistent estimator of $\mathcal{I}$ under mild conditions.

Theorem 2. For the PAMA model with condition $\boldsymbol{C}_\gamma$, for almost all $\phi \in \Omega_\phi$, the moment estimator $\tilde{\mathcal{I}}$ converges to $\mathcal{I}$ with probability 1 with m going to infinity.

Proof. Combining fact (23) in Lemma 1 with fact (21), we have for $\forall\, \phi \in \tilde{\Omega}_\phi$ that

$$e_1(\phi) < \cdots < e_{n_1}(\phi) \text{ and } e_i(\phi) \neq e_0 \text{ for } i = 1, \ldots, n_1.$$

Moreover, as fact (21) tells us that for $\forall\, \epsilon, \delta > 0, \exists\, M > 0$ s.t. for $\forall\, m > M$,

$$P\big(|\bar{\tau}(i) - e_i(\phi)| < \delta\big) \geq 1 - \epsilon,\ i = 1, \ldots, n,$$

it is straightforward to see the conclusion of the theorem. □

Theorem 2 tells us that estimating $\mathcal{I}$ is straightforward if the number of independent rankers m goes to infinity: a simple moment method ignoring the quality difference of rankers can provide us a consistent estimate of $\mathcal{I}$. In a practical problem where only a finite number of rankers are involved, however, more efficient statistical inference of the PAMA model based on Bayesian or frequentist principles becomes more attractive as effectively utilizing the quality information of different rankers is critical.

With n_0 and n_1 fixed, parameter γ_k, which governs the power-law distribution for the rank list τ_k, cannot be estimated consistently. Thus, its distribution $F(\gamma)$ cannot be determined nonparametrically even when the number of rank lists m goes to infinity. We impose a parametric form $F_\psi(\gamma)$ with ψ as the hyper-parameter and refer to the resulting hierarchical-structured model as PAMA-H, which has the following marginal likelihood of (ϕ, ψ) given $\mathcal{I}$:

$$\begin{aligned} L(\phi, \psi \mid \mathcal{I}) &= \int P(\boldsymbol{\tau} \mid \mathcal{I}, \phi, \boldsymbol{\gamma}) dF_\psi(\boldsymbol{\gamma}) \\ &= \prod_{k=1}^{m} \int P(\tau_k \mid \mathcal{I}, \phi, \gamma_k) dF_\psi(\gamma_k) = \prod_{k=1}^{m} L_k(\phi, \psi \mid \mathcal{I}). \end{aligned}$$

We show in Theorem 3 that the MLE based on the above marginal likelihood is consistent.

Theorem 3. Under the PAMA-H model, assume that (ϕ, ψ) belongs to the parameter space $\Omega_\phi \times \Omega_\psi$, and the true parameter (ϕ_0, ψ_0) is an interior point of $\Omega_\phi \times \Omega_\psi$. Let $(\hat{\phi}_\mathcal{I}, \hat{\psi}_\mathcal{I})$ be the maximizer of $L(\phi, \psi \mid \mathcal{I})$. If $F_\psi(\gamma)$ has a density function $f_\psi(\gamma)$ that is differentiable and concave with respect to ψ, then $\lim_{m\to\infty}(\hat{\phi}_\mathcal{I}, \hat{\psi}_\mathcal{I}) = (\phi_0, \psi_0)$ almost surely.

Proof. See Supplementary material. □

4. Inference with the PAMA

4.1. Maximum Likelihood Estimation

Under the PAMA model, the MLE of $\boldsymbol{\theta} = (\mathcal{I}, \phi, \boldsymbol{\gamma})$ is $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta})$, where

$$l(\boldsymbol{\theta}) = \log P(\tau_1, \tau_2, \ldots, \tau_m \mid \boldsymbol{\theta}) \tag{25}$$

is the logarithm of the likelihood function (19). Here, we adopt the *Gauss–Seidel* iterative method in Yang (2018), also known as *backfitting* or *cyclic coordinate ascent*, to implement the optimization. Starting from an initial point $\boldsymbol{\theta}^{(0)}$, the Gauss–Seidel method iteratively updates one coordinate of $\boldsymbol{\theta}$ at each step with the other coordinates held fixed at their current values. A Newton-like method is adopted to update ϕ and γ_k. Since $\mathcal{I}$ is a discrete vector, we find favorable values of $\mathcal{I}$ by swapping two neighboring entities to check whether $g(\mathcal{I} \mid \boldsymbol{\gamma}^{(s+1)}, \phi^{(s+1)})$ increases. More details of the algorithm are provided in the supplementary material.

With the MLE $\hat{\boldsymbol{\theta}} = (\hat{\mathcal{I}}, \hat{\phi}, \hat{\boldsymbol{\gamma}})$, we define $U_R(\hat{\mathcal{I}}) = \{i \in U : \hat{\mathcal{I}}_i > 0\}$ and $U_B(\hat{\mathcal{I}}) = \{i \in U : \hat{\mathcal{I}}_i = 0\}$, and generate the final aggregated ranking list $\hat{\tau}$ based on the rules below: (a) set the top-n_1 list of $\hat{\tau}$ as $\hat{\tau}_{n_1} = sort(i \in U_R(\hat{\mathcal{I}})\ by\ \hat{\mathcal{I}}_i \uparrow)$, (b) all entities in $U_B(\hat{\mathcal{I}})$ tie for positions behind. Hereinafter, we refer to this MLE-based rank aggregation procedure under PAMA model as PAMA_F.

For the PAMA-H model, a similar procedure can be applied to find the MLE of $\boldsymbol{\theta} = (\mathcal{I}, \phi, \psi)$, with the $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$ being treated as the missing data. With the MLE $\hat{\boldsymbol{\theta}} = (\hat{\mathcal{I}}, \hat{\phi}, \hat{\psi})$, we can generate the final aggregated ranking list $\hat{\tau}$ based on $\hat{\mathcal{I}}$ in the same way as in PAMA, and evaluate the quality of ranker τ_k via the mean or mode of the conditional distribution below:

$$f(\gamma_k \mid \tau_k; \hat{\mathcal{I}}, \hat{\phi}, \hat{\psi}) \propto f(\gamma_k \mid \hat{\psi}) \cdot P(\tau_k \mid \hat{\mathcal{I}}, \hat{\phi}, \gamma_k).$$

In this article, we refer to the above MLE-based rank aggregation procedure under PAMA-H model as PAMA_{HF}. The procedure is detailed in the supplementary material.

4.2. Bayesian Inference

Since the three model parameters $\mathcal{I}$, ϕ, and $\boldsymbol{\gamma}$ encode "orthogonal" information of the PAMA model, it is natural to expect that $\mathcal{I}$, ϕ and $\boldsymbol{\gamma}$ are mutually independent a priori. We thus specify their joint prior distribution as

$$\pi(\mathcal{I}, \phi, \boldsymbol{\gamma}) = \pi(\mathcal{I}) \cdot \pi(\phi) \cdot \prod_{k=1}^{m} \pi(\gamma_k).$$

Without much loss, we may restrict the range of ϕ and γ_k's to a closed interval $[0, b]$ with a large enough b. In contrast, $\mathcal{I}$ is discrete and takes value in the space $\Omega_\mathcal{I}$ of all permutations of $\{1, \ldots, n_1, \underbrace{0, \ldots, 0}_{n_0}\}$. It is convenient to specify $\pi(\mathcal{I})$, $\pi(\phi)$ and $\pi(\gamma_k)$ as uniform, that is,

$$\pi(\mathcal{I}) \sim U(\Omega_\mathcal{I}),\ \pi(\phi) \sim U[0, b],\ \pi(\gamma_k) \sim U[0, b].$$

Based on our experiences in a large range of simulation studies and real data applications, we find that it works reasonably well to set $b = 10$. In Section 3.3 we also discussed letting $\pi(\gamma_k)$ be of a parametric form, which will be further discussed later.

The posterior distribution can be expressed as

$$\begin{aligned} &f(\mathcal{I}, \phi, \boldsymbol{\gamma} | \tau_1, \tau_2, \ldots, \tau_m) \\ &\quad \propto \pi(\mathcal{I}, \phi, \boldsymbol{\gamma}) \cdot P(\tau_1, \tau_2, \ldots, \tau_m | \mathcal{I}, \phi, \boldsymbol{\gamma}) = \mathbb{I}\big(\phi \in [0, 10]\big) \times \prod_{k=1}^{m} \\ &\quad \times \Big\{ \frac{\mathbb{I}\big(\tau_k^0 \in \mathcal{A}_{U_R}(\tau_k^{0|1})\big) \times \mathbb{I}\big(\gamma_k \in [0, 10]\big)}{A^*_{\tau_k,I} \times (B^*_{\tau_k,I})^{\gamma_k} \times (C^*_{\gamma_k,n_1})^{n-n_1} \times (D^*_{\tau_k,\mathcal{I}})^{\phi\cdot\gamma_k} \times E^*_{\phi,\gamma_k}} \Big\}, \end{aligned} \tag{26}$$

with the following conditional distributions:

$$f(\mathcal{I} \mid \phi, \boldsymbol{\gamma}) \propto \prod_{k=1}^{m} \frac{\mathbb{I}\big(\tau_k^0 \in \mathcal{A}_{U_R}(\tau_k^{0|1})\big)}{A^*_{\tau_k,I} \times (B^*_{\tau_k,I})^{\gamma_k} \times (D^*_{\tau_k,\mathcal{I}})^{\phi\cdot\gamma_k}}, \tag{27}$$

$$f(\phi \mid \mathcal{I}, \boldsymbol{\gamma}) \propto \mathbb{I}\big(\phi \in [0, 10]\big) \times \prod_{k=1}^{m} \frac{1}{(D^*_{\tau_k,\mathcal{I}})^{\phi\cdot\gamma_k} \times E^*_{\phi,\gamma_k}}, \tag{28}$$

$$f(\gamma_k \mid \mathcal{I}, \phi, \boldsymbol{\gamma}_{[-k]}) \propto \frac{\mathbb{I}(\gamma_k \in [0,10])}{(B^*_{\tau_k,I})^{\gamma_k} \times (C^*_{\gamma_k,n_1})^{n-n_1} \times (D^*_{\tau_k,\mathcal{I}})^{\phi\cdot\gamma_k} \times E^*_{\phi,\gamma_k}}, \quad (29)$$

based on which posterior samples of $(\mathcal{I}, \phi, \boldsymbol{\gamma})$ can be obtained by Gibbs sampling, where $\boldsymbol{\gamma}_{[-k]} = (\gamma_1, \ldots, \gamma_{k-1}, \gamma_{k+1}, \ldots, \gamma_m)$.

Considering that conditional distributions in Equations (27)-(29) are nonstandard, we adopt the Metropolis–Hastings algorithm (Hastings 1970) to enable the conditional sampling. To be specific, we choose the proposal distributions for ϕ and γ_k as

$$q(\phi \mid \phi^{(t)}; \mathcal{I}, \boldsymbol{\gamma}) \sim \mathcal{N}(\phi^{(t)}, \sigma^2_\phi)$$
$$q(\gamma_k \mid \gamma_k^{(t)}; \mathcal{I}, \phi, \boldsymbol{\gamma}_{[-k]}) \sim \mathcal{N}(\gamma_k^{(t)}, \sigma^2_{\gamma_k}),$$

where σ^2_ϕ and $\sigma^2_{\gamma_k}$ can be tuned to optimize the mixing rate of the sampler. Since $\mathcal{I}$ is a discrete vector, we propose new values of $\mathcal{I}$ by swapping two randomly selected adjacent entities. Note that the entity whose ranking is n_1 could be swapped with any background entity. Due to the homogeneity of background entities, there is no need to swap two background entities. Therefore, the number of potential proposals in each step is $\mathcal{O}(nn_1)$. More details about MCMC sampling techniques can be found in Liu (2008).

Suppose that M posterior samples $\{(\mathcal{I}^{(t)}, \phi^{(t)}, \boldsymbol{\gamma}^{(t)})\}_{t=1}^M$ are obtained. We calculate the posterior means of different parameters as follows:

$$\bar{\mathcal{I}}_i = \frac{1}{M}\sum_{t=1}^{M}\left[\mathcal{I}_i^{(t)} \cdot I_i^{(t)} + \frac{n_1+1+n}{2}\cdot(1-I_i^{(t)})\right],\ i = 1, \ldots, n,$$
$$\bar{\phi} = \frac{1}{M}\sum_{t=1}^{M}\phi^{(t)},$$
$$\bar{\gamma}_k = \frac{1}{M}\sum_{t=1}^{M}\gamma_k^{(t)}, k = 1, \ldots, m.$$

We quantify the quality of ranker τ_k with $\bar{\gamma}_k$, and generate the final aggregated ranking list $\hat{\tau}$ based on the $\bar{\mathcal{I}}_i$'s as follows:

$$\hat{\tau} = \text{sort}(i \in U \text{ by } \bar{\mathcal{I}}_i \uparrow).$$

Hereinafter, we refer to this MCMC-based Bayesian rank aggregation procedure under the PAMA as PAMA_B.

The Bayesian inference procedure PAMA_{HB} for the PAMA-H model differs from PAMA_{B} only by replacing the prior distribution $\prod_{k=1}^m \pi(\gamma_k)$, which is uniform in $[0, b]^m$, with a hierarchical-structured prior $\pi(\psi)\prod_{k=1}^m f_\psi(\gamma_k)$. The conditional distributions needed for Gibbs sampling are almost the same as Equations (27)–(29), except an additional one

$$f(\psi \mid \mathcal{I}, \phi, \boldsymbol{\gamma}) \propto \pi(\psi) \cdot \prod_{k=1}^{m} f_\psi(\gamma_k). \quad (30)$$

We may specify $f_\psi(\gamma)$ to be an exponential distribution and let $\pi(\psi)$ be a proper conjugate prior to make Equation (30) easy to sample from. More details for PAMA_{HB} with $f_\psi(\gamma)$ specified as an exponential distribution is provided in the supplementary material.

Our simulation studies suggest that the practical performance of PAMA_B and PAMA_{HB} are very similar when n_0 and n_1 are reasonably large (see supplementary material for details). In contrast, as we will show in Section 5, the MLE-based estimates (e.g., PAMA_F) typically produce less accurate results with a shorter computational time compared to PAMA_B.

4.3. Extension to Partial Ranking Lists

The proposed PAMA can be extended to more general scenarios where partial ranking lists, instead of full ranking lists, are involved in the aggregation. Given the entity set U and a ranking list τ_S of entities in $S \subseteq U$, we say τ_S is a *full ranking list* if $S = U$, and a *partial ranking list* if $S \subset U$. Suppose τ_S is a partial ranking list and τ_U is a full ranking list of U. If the projection of τ_U on S equals to τ_S, we say τ_U is compatible with τ_S, denotes as $\tau_U \sim \tau_S$. Let $\mathcal{A}(\tau_S) = \{\tau_U : \tau_U \sim \tau_S\}$ be the set of all full lists that are compatible with τ_S. Suppose a partial list τ_k is involved in the ranking aggregation problem. The probability of τ_k can be evaluated by

$$P(\tau_k \mid \mathcal{I}, \phi, \gamma_k) = \sum_{\tau_k^* \sim \tau_k} P(\tau_k^* \mid \mathcal{I}, \phi, \gamma_k), \quad (31)$$

where $P(\tau_k^* \mid \mathcal{I}, \phi, \gamma_k)$ is the probability of a compatible full list under the PAMA model. Clearly, the probability in (31) does not have a closed-form representation due to complicated constraints between τ_k and τ_k^*, and it is very challenging to do statistical inference directly based on this quantity. Fortunately, as rank aggregation with partial lists can be treated as a missing data problem, we can resolve the problem via standard methods for missing data inference.

The Bayesian inference can be accomplished by the classic data augmentation strategy (Tanner and Wong 1987) in a similar way as described in Deng et al. (2014), which iterates between imputing the missing data conditional on the observed data given the current parameter values, and updating parameter values by sampling from the posterior distribution based on the imputed full data. To be specific, we iteratively draw from the following two conditional distributions:

$$P(\tau_1^*, \ldots, \tau_m^* \mid \tau_1, \ldots, \tau_m; \mathcal{I}, \phi, \boldsymbol{\gamma}) = \prod_{k=1}^{m} P(\tau_k^* \mid \tau_k; \mathcal{I}, \phi, \gamma_k),$$
$$f(\mathcal{I}, \phi, \boldsymbol{\gamma} \mid \tau_1^*, \ldots, \tau_m^*) \propto \pi(\mathcal{I}) \times \pi(\boldsymbol{\gamma}) \times \pi(\phi) \times \prod_{k=1}^{m} P(\tau_k^* \mid \mathcal{I}, \gamma_k, \phi).$$

To find the MLE of $\boldsymbol{\theta}$ for this more challenging scenario, we can use the Monte Carlo EM algorithm (MCEM; Wei and Tanner (1990)). Let $\tau_k^{(1)}, \ldots, \tau_k^{(M)}$ be M independent samples drawn from distribution $P(\tau_k^* \mid \tau_k, \mathcal{I}, \phi, \gamma_k)$. The E-step involves the calculation of the Q-function below:

$$Q(\mathcal{I}, \boldsymbol{\gamma}, \phi \mid \mathcal{I}^{(s)}, \boldsymbol{\gamma}^{(s)}, \phi^{(s)}) = E\left\{\sum_{k=1}^{m} \log P(\tau_k^* \mid \mathcal{I}, \boldsymbol{\gamma}, \phi) \mid \tau_k, \mathcal{I}^{(s)}, \boldsymbol{\gamma}_k^{(s)}, \phi^{(s)}\right\}$$
$$\approx \frac{1}{M}\sum_{k=1}^{m}\sum_{t=1}^{M} \log P(\tau_k^{(t)} \mid \mathcal{I}, \boldsymbol{\gamma}_k, \phi).$$

In the M-step, we use the *Gauss–Seidel* method to maximize the above Q-function in a similar way as detailed in the supplementary material.

No matter which method is used, a key step is to draw samples from

$$P(\tau_k^* \mid \tau_k; \mathcal{I}, \phi, \gamma_k) \propto P(\tau_k^* \mid \mathcal{I}, \gamma_k, \phi) \cdot \mathbb{I}\big(\tau_k^* \in \mathcal{A}(\tau_k)\big).$$

To achieve this goal, we start with τ_k^* obtained from the previous step of the data augmentation or MCEM algorithms, and conduct several iterations of the following Metropolis step with $P(\tau_k^* \mid \tau_k; \mathcal{I}, \phi, \gamma_k)$ as its target distribution: (a) construct the proposal τ_k' by randomly selecting two elements in the current full list τ_k^* and swapping them; (b) accept or reject the proposal according to the Metropolis rule, that is to accept τ_k' with probability of $\min(1, \frac{P(\tau_k' \mid \mathcal{I}, \gamma_k, \phi)}{P(\tau_k^* \mid \mathcal{I}, \gamma_k, \phi)})$. Note that the proposed list τ_k' is automatically rejected if it is incompatible with the observed partial list τ_k.

4.4. Incorporating Covariates in the Analysis

In some applications, covariate information for each ranked entity is available to assist rank aggregation. One of the earliest attempts for incorporating such information in analyzing rank data is perhaps the *hidden score model* due to Thurstone (1927), which has become a standard approach and has many extensions. Briefly, these models assume that there is an unobserved score for each entity that is related to the entity-specific covariates $X_i = (X_{i1}, \ldots, X_{ip})^T$ under a regression framework and the observed rankings are determined by these scores plus noises, that is,

$$\tau_k = \text{sort}(S_{ik} \downarrow,\ E_i \in U), \text{ where } S_{ik} = X_i^T \boldsymbol{\beta} + \varepsilon_{ik}.$$

Here, $\boldsymbol{\beta}$ is the common regression coefficient and $\varepsilon_{ik} \sim N(0, \sigma_k^2)$ is the noise term. Recent progresses along this line are reviewed by Yu (2000), Bhowmik and Ghosh (2017), and Li, Yi, and Liu (2021).

Here, we propose to incorporate covariates into the analysis in a different way. Assuming that covariate X_i provides information on the group assignment instead of the detailed ranking of entity E_i, we connect X_i and $\mathcal{I}_i$, the enhanced indicator of X_i, by a logistic regression model

$$\begin{aligned} P(\mathcal{I}_i \mid X_i) &= P(I_i \mid X_i, \boldsymbol{\psi}) \\ &= \frac{\exp\{X_i^T \boldsymbol{\psi} \cdot I_i\}}{1 + \exp\{X_i^T \boldsymbol{\psi}\}}, \quad i = 1, \ldots, n, \end{aligned} \tag{32}$$

where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_p)^T$ as the regression parameters. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be the covariate matrix, we can extend the PAMA as

$$P(\tau_1, \ldots, \tau_m, \mathcal{I} \mid \boldsymbol{X}) = P(\mathcal{I} \mid \boldsymbol{X}, \boldsymbol{\psi}) \times P(\tau_1, \ldots, \tau_m \mid \mathcal{I}, \phi, \boldsymbol{\gamma}), \tag{33}$$

where the first term

$$P(\mathcal{I} \mid \boldsymbol{X}, \boldsymbol{\psi}) = \prod_{i=1}^{n} P(I_i \mid X_i, \boldsymbol{\psi})$$

comes from the logistic regression model (32), and the second term comes from the original PAMA. In the extended model, our goal is to infer $(\mathcal{I}, \phi, \boldsymbol{\gamma}, \boldsymbol{\psi})$ based on $(\tau_1, \ldots, \tau_m; \boldsymbol{X})$. We can achieve both Bayesian inference and MLE for the extended model in a similar way as described for the Partition-Mallows Model. More details are provided in the supplementary material.

An alternative way to incorporate covariates is to replace the logistic regression model by a Naive Bayes model, which models the conditional distribution of $\boldsymbol{X} \mid \mathcal{I}$ instead of $\mathcal{I} \mid \boldsymbol{X}$, as follows:

$$f(\tau_1, \ldots, \tau_m, \boldsymbol{X} \mid \mathcal{I}) = P(\tau_1, \ldots, \tau_m, \mid \mathcal{I}, \phi, \boldsymbol{\gamma}) \times f(\boldsymbol{X} \mid \mathcal{I}), \tag{34}$$

where

$$\begin{aligned} f(\boldsymbol{X} \mid \mathcal{I}) &= \prod_{i=1}^{n} f(X_i \mid \mathcal{I}_i) = \prod_{i=1}^{n} f(X_i \mid I_i) = \prod_{i=1}^{n} \prod_{j=1}^{p} f(X_{ij} \mid I_i) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{p} \Big\{ \big[f_j(X_{ij} \mid \psi_{j0})\big]^{1-I_i} \cdot \big[f_j(X_{ij} \mid \psi_{j1})\big]^{I_i} \Big\}, \end{aligned}$$

f_j is prespecified parametric distribution for covariates X_j with parameter ψ_{j0} for entities in the background group and ψ_{j1} for entities in the relevant group. Since the performances of the two approaches are very similar, in the rest of this article, we use the logistic regression strategy to handle covariates due to its convenient form.

5. Simulation Study

5.1. Simulation Settings

We simulated data from two models: (a) the proposed PAMA, referred to as $\mathcal{S}_{PM}$, and (b) the Thurstone hidden score model, referred to as $\mathcal{S}_{HS}$. In the $\mathcal{S}_{PM}$ scenario, we specified the true indicator vector as $\mathcal{I} = (1, \ldots, n_1, 0, \ldots, 0)$, indicating that the first n_1 entities $E_1, \ldots, E_{n_1}$ belong to U_R and the rest belong to the background group U_B, and set

$$\gamma_k = \begin{cases} 0.1, & \text{if } k \le \frac{m}{2}; \\ a + (k - \frac{m}{2}) \times \delta_R, & \text{if } k > \frac{m}{2}. \end{cases}$$

Clearly, $a > 0$ and $\delta_R > 0$ control the quality difference and signal strength of the m base rankers in the $\mathcal{S}_{PM}$ scenario. We set $\phi = 0.6$ (defined in (16)), $\delta_R = \frac{2}{m}$, and a with two options: 2.5 and 1.5. For easy reference, we denoted the strong signal case with $a = 2.5$ and the weak signal case with $a = 1.5$ by $\mathcal{S}_{PM_1}$ and $\mathcal{S}_{PM_2}$, respectively.

In the $\mathcal{S}_{HS}$ scenario, we used the Thurstone model to generate the rank lists as $\tau_k = sort(i \in U \text{ by } S_{ik} \downarrow)$, where $S_{ik} \sim N(\mu_{ik}, 1)$ and

$$\mu_{ik} = \begin{cases} 0, & \text{if } k \le \frac{m}{2} \text{ or } i > n_1; \\ a^* + \frac{b^* - a^*}{m} \times k + (n_1 - i) \times \delta_E^*, & \text{otherwise.} \end{cases}$$

In this model, a^*, b^* and δ_E^* (all positive numbers) control the quality difference and signal strength of the m base rankers. We also specified two sub-cases: $\mathcal{S}_{HS_1}$, the stronger-signal case with $(a^*, b^*, \delta_E^*) = (0.5, 2.5, 0.2)$; and $\mathcal{S}_{HS_2}$, the weaker-signal case with $(a^*, b^*, \delta_E^*) = (-0.5, 1.5, 0.2)$. Table 1 shows the configuration matrix of μ_{ik} under $\mathcal{S}_{HS_1}$ when $m = 10, n = 100$ and $n_1 = 10$. In both scenarios, the first half of rankers is completely non-informative, with the other half providing increasingly strong signals.

For each of the four simulation scenarios (i.e., $\mathcal{S}_{PM_1}$, $\mathcal{S}_{PM_2}$, $\mathcal{S}_{HS_1}$, and $\mathcal{S}_{HS_2}$), we fixed the true number of relevant entities $n_1 = 10$, but allowed the number of rankers m and the total

Table 1. The configuration matrix of the μ_{ik}'s under $\mathcal{S}_{HS_1}$ with m=10, n = 100 and n_1=10.

	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
E_1	0	0	0	0	0	3.7	3.9	4.1	4.3	4.5
E_2	0	0	0	0	0	3.5	3.7	3.9	4.1	4.3
E_3	0	0	0	0	0	3.3	3.5	3.7	3.9	4.1
E_4	0	0	0	0	0	3.1	3.3	3.5	3.7	3.9
E_5	0	0	0	0	0	2.9	3.1	3.3	3.5	3.7
E_6	0	0	0	0	0	2.7	2.9	3.1	3.3	3.5
E_7	0	0	0	0	0	2.5	2.7	2.9	3.1	3.3
E_8	0	0	0	0	0	2.3	2.5	2.7	2.9	3.1
E_9	0	0	0	0	0	2.1	2.3	2.5	2.7	2.9
E_{10}	0	0	0	0	0	1.9	2.1	2.3	2.5	2.7
E_{11}	0	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
E_{100}	0	0	0	0	0	0	0	0	0	0

number of entities n to vary, resulting in a total of 16 simulation settings ($\{\text{scenarios} : \mathcal{S}_{PM_1}, \mathcal{S}_{PM_2}, \mathcal{S}_{HS_1}, \mathcal{S}_{HS_2}\} \times \{m : 10, 20\} \times \{n : 100, 300\} \times \{n_1 : 10\}$). Under each setting, we simulated 500 independent datasets to evaluate and compare performances of different rank aggregation methods.

5.2. Methods in Comparison and Performance Measures

Except for the proposed PAMA$_B$ and PAMA$_F$, we considered state-of-the-art methods in several classes, including the Markov chain-based methods MC$_1$, MC$_2$, and MC$_3$ in Lin (2010) and CEMC in Lin and Ding (2010), the partition-based method BARD in Deng et al. (2014), and the Mallows model-based methods MM and EMM in Li et al. (2020). Classic Naive methods based on summary statistics were ignored because they have been shown in the previous studies to perform suboptimally, especially in cases where base rankers are heterogeneous in quality. The Markov-chain-based methods, MM, and EMM were implemented in TopKLists, PerMallows, and ExtMallows packages in R (https://www.r-project.org/), respectively. The code of BARD was provided by its authors.

Let τ be the underlying true ranking list of all entities, $\tau_R = \{\tau(i) : E_i \in U_R\}$ be the true relative ranking of relevant entities, $\hat{\tau}$ be the aggregated ranking obtained from a rank aggregation approach, $\hat{\tau}_R = \{\hat{\tau}(i) : E_i \in U_R\}$ be the relative ranking of relevant entities after aggregation, and $\hat{\tau}_{n_1}$ be the top-n_1 list of $\hat{\tau}$. After obtaining the aggregated ranking $\hat{\tau}$ from a rank aggregation approach, we evaluated its performance by two measurements, namely the *recovery distance* κ_R and the *coverage* ρ_R, defined as below:

$$\kappa_R \triangleq d_\tau(\hat{\tau}_R, \tau_R) + n_{\hat{\tau}} \times \frac{n + n_1 + 1}{2},$$
$$\rho_R \triangleq \frac{n_1 - n_{\hat{\tau}}}{n_1},$$

where $d_\tau(\hat{\tau}_R, \tau_R)$ denotes the Kendall tau distance between $\hat{\tau}_R$ and τ_R, and $n_{\hat{\tau}}$ denotes the number of relevant entities who are classified as background entities in $\hat{\tau}$. The recovery distance κ_R considers detailed rankings of all relevant entities plus misclassification distances, while the coverage ρ_R cares only about the identification of relevant entities without considering the detailed rankings. In the setting of PAMA, $\frac{n+n_1+1}{2}$ is the average rank of a background entity. The recovery distance increases if some relevant entities are misclassified as background entities. Clearly, we expect a smaller κ_R and a larger ρ_R for a stronger aggregation approach.

5.3. Simulation Results

Table 2 summarizes the performances of the nine competing methods in the 16 different simulation settings, demonstrating the proposed PAMA$_B$ and PAMA$_F$ outperform all the other methods by a significant margin in most settings and PAMA$_B$ uniformly dominates PAMA$_F$. Figure 3 shows the quality parameter $\boldsymbol{\gamma}$ learned from the Partition-Mallows Model in various simulation scenarios with $m = 10$ and $n = 100$, confirming that the proposed methods can effectively capture the quality difference among the rankers. The results of $\boldsymbol{\gamma}$ for other combinations of (m, n) can be found in the supplementary material which demonstrates consistent performance with Figure 3.

Figure 4 (a) shows the boxplots of recovery distances and the coverages of the nine competing methods in the four simulation scenarios with $m = 10$, $n = 100$, and $n_1 = 10$. The five methods from the left outperform the other four methods by a significant gap, and the PAMA-based methods generally perform the best. Figure 4 (b) confirms that the methods based on the Partition-Mallows Model enjoy the same capability as BARD in detecting quality differences between informative and noninformative rankers. However, while both BARD and PAMA can further discern quality differences among informative rankers, EMM fails this more subtle task. Similar figures for other combinations of (m, n) are provided in the supplementary material, highlighting consistent results as in Figure 4.

5.4. Robustness to the Specification of n_1

We need to specify n_1, the number of relevant entities, when applying PAMA$_B$ or PAMA$_F$. In many practical problems, however, there may not be a strong prior information on n_1 and there may not even be clear distinctions between relevant and background entities. To examine robustness of the algorithm with respect to the specification of n_1, we design a simulation setting $\mathcal{S}_{HS_3}$ to mimic the no-clear-cut scenario and investigate how the performance of PAMA is affected by the specification of n_1 under this setting. Formally, $\mathcal{S}_{HS_3}$ assumes that $\tau_k = \text{sort}(i \in U \text{ by } S_{ik} \downarrow)$, where $S_{ik} \sim N(\mu_{ik}, 1)$, following the same data generating framework as $\mathcal{S}_{HS}$ defined in the Section 5.1, with μ_{ik} being replaced by

$$\mu_{ik} = \begin{cases} 0, & \text{if } k \le \frac{m}{2}, \\ \frac{2 \times a^* \times k/m}{1 + e^{-b^* \times (70 - i)}}, & \text{otherwise,} \end{cases}$$

where $a^* = 50$ and $b^* = 0.1$. Different from $\mathcal{S}_{HS_1}$ and $\mathcal{S}_{HS_2}$, where μ_{ik} jumps from 0 to a positive number as i ranges from background to relevant entities, in the $\mathcal{S}_{HS_3}$ scenario μ_{ik} increases smoothly as a function of i for each informative ranker k. In such cases, the concept of "relevant" entities is not well-defined.

We simulate 500 independent datasets from $\mathcal{S}_{HS_3}$ with $n = 100$ and $m = 10$. For each dataset, we try different specifications of n_1 ranging from 10 to 50 and compare PAMA to several

Table 2. Average recovery distances [coverages] of different methods based on 500 independent replicates under different simulation scenarios.

Configuration			Partition-type Models			Mallows Models		MC-based Models			
$\mathcal{S}$	n	m	PAMA$_F$	PAMA$_B$	BARD	EMM	MM	MC$_1$	MC$_2$	MC$_3$	CEMC
$\mathcal{S}_{PM_1}$	100	10	24.5	**15.2**	57.1	51.7	103.2	338.4	163.1	198.6	197.8
			[0.95]	**[0.97]**	[0.91]	[0.89]	[0.81]	[0.36]	[0.69]	[0.63]	[0.62]
	100	20	2.6	**0.3**	42.1	22.8	44.2	466.6	88.9	121.2	114.7
			[0.99]	**[1.00]**	[0.95]	[0.95]	[0.93]	[0.11]	[0.82]	[0.78]	[0.77]
	300	10	17.4	**4.0**	180.0	683.3	519.2	1268.3	997.7	1075.8	1085.7
			[0.99]	**[1.00]**	[0.89]	[0.66]	[0.55]	[0.17]	[0.34]	[0.29]	[0.28]
	300	20	7.1	**3.2**	122.3	124.4	157.1	1445.9	613.5	723.0	727.2
			[1.00]	**[1.00]**	[0.93]	[0.92]	[0.90]	[0.05]	[0.60]	[0.53]	[0.52]
$\mathcal{S}_{PM_2}$	100	10	90.0	**66.6**	115.2	108.3	152.9	404.3	285.5	307.2	313.8
			[0.82]	**[0.86]**	[0.77]	[0.77]	[0.70]	[0.24]	[0.47]	[0.43]	[0.41]
	100	20	26.9	**2.4**	81.5	59.8	91.5	468.1	217.3	245.2	249.5
			[0.94]	**[1.00]**	[0.85]	[0.87]	[0.82]	[0.11]	[0.60]	[0.55]	[0.53]
	300	10	81.1	**26.8**	468.4	609.8	472.1	1388.4	1294.7	1321.5	1328.4
			[0.95]	**[0.98]**	[0.69]	[0.68]	[0.60]	[0.09]	[0.15]	[0.13]	[0.13]
	300	20	77.2	**3.4**	313.6	267.5	337.0	1469.0	1205.9	1251.8	1258.9
			[0.95]	**[1.00]**	[0.79]	[0.82]	[0.78]	[0.04]	[0.21]	[0.18]	[0.18]
$\mathcal{S}_{HS_1}$	100	10	24.9	**20.6**	22.9	54.9	115.9	334.7	150.9	180.3	186.0
			[0.97]	[0.98]	**[0.99]**	[0.91]	[0.80]	[0.37]	[0.71]	[0.66]	[0.64]
	100	20	18.7	15.6	22.8	**8.7**	33.4	498.8	46.7	64.1	60.8
			[0.98]	[0.98]	**[1.00]**	**[1.00]**	[0.97]	[0.05]	[0.92]	[0.89]	[0.89]
	300	10	172.0	159.8	**37.9**	205.5	490.6	1098.6	627.0	752.9	769.4
			[0.89]	[0.90]	**[0.99]**	[0.87]	[0.68]	[0.28]	[0.59]	[0.50]	[0.49]
	300	20	7.4	**7.0**	22.7	11.4	114.1	1402.6	237.8	319.7	322.3
			[1.00]	**[1.00]**	**[1.00]**	**[1.00]**	[0.94]	[0.08]	[0.84]	[0.79]	[0.79]
$\mathcal{S}_{HS_2}$	100	10	92.6	74.0	**68.7**	123.7	162.3	382.4	228.2	250.2	256.6
			[0.83]	[0.86]	**[0.88]**	[0.77]	[0.70]	[0.27]	[0.56]	[0.52]	[0.50]
	100	20	24.4	20.0	22.2	**12.4**	38.3	500.3	87.5	103.5	102.9
			[0.96]	[0.97]	**[1.00]**	[0.99]	[0.95]	[0.04]	[0.83]	[0.80]	[0.80]
	300	10	319.1	463.8	**245.6**	516.9	683.5	1267.9	998.0	1076.0	1085.5
			[0.79]	[0.69]	**[0.84]**	[0.66]	[0.55]	[0.17]	[0.34]	[0.29]	[0.28]
	300	20	8.7	**8.0**	23.2	30.3	155.5	1430.7	437.6	516.2	523.3
			[1.00]	**[1.00]**	**[1.00]**	[0.99]	[0.91]	[0.06]	[0.71]	[0.66]	[0.65]

NOTE: The bold texts indicate the best performance for each row.

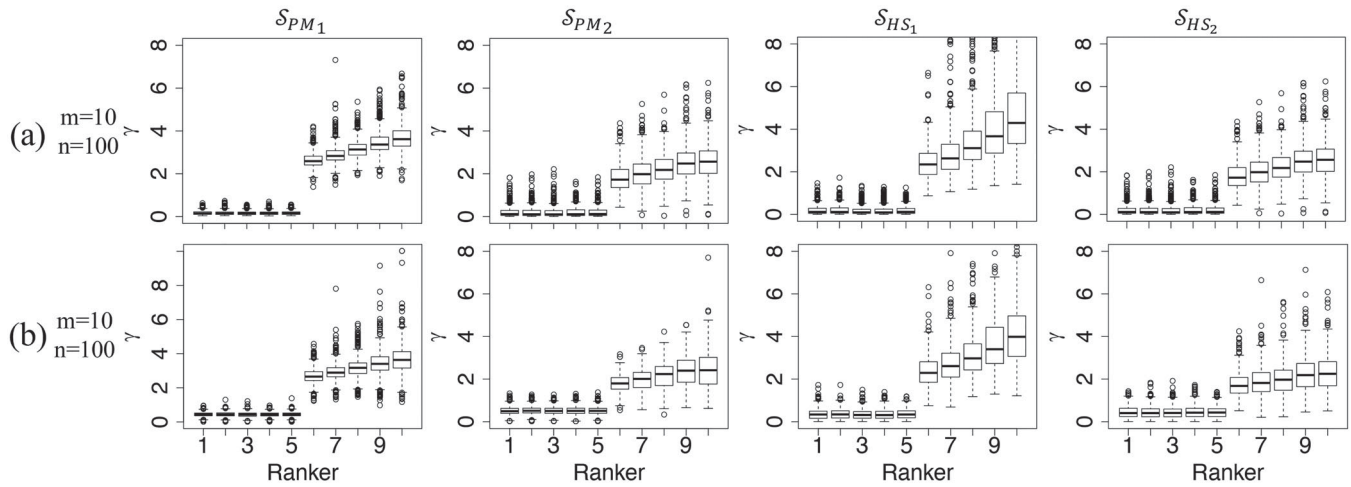

Figure 3. (a) The boxplots of $\{\bar{\gamma}_k\}$ estimated by PAMA$_B$ with $m = 10$ and $n = 100$. (b) The boxplots of $\{\hat{\gamma}_k\}$ estimated by PAMA$_F$ with $m = 10$ and $n = 100$. Each column denotes a scenario setting. The results were obtained from 500 independent replicates.

competing methods based on their performance on recovering the top-A list $[E_1 \preceq E_2 \preceq \cdots \preceq E_A]$, which is still well-defined based on the simulation design. The results summarized in Table 3 show that no matter which n_1 is specified, the partition-type models consistently outperform all the competing methods in terms of a lower recovery distance from the true top-n_1 list of items, that is, $[E_1 \preceq E_2 \preceq \cdots \preceq E_{n_1}]$. Figure 5 illustrates in details the average aggregated rankings of the top-10 entities by PAMA as n_1 increases, suggesting that PAMA is able to figure out the correct rankings of the top entities effectively. These results give us confidence that PAMA is robust to misspecification of n_1.

Noticeably, although PAMA and BARD achieve comparable coverage as shown in Table 3, PAMA dominates BARD uniformly in terms of a much smaller recovery distance in all cases, suggesting that PAMA is capable of figuring out the detailed ranking of relevant entities that is missed by BARD. In fact, since BARD relies only on $\rho_i \triangleq P(I_i = 1 \mid \tau_1, \ldots, \tau_m)$ to rank entities, in cases where the signal to distinguish the relevant and background entities is strong, many ρ_i's are very

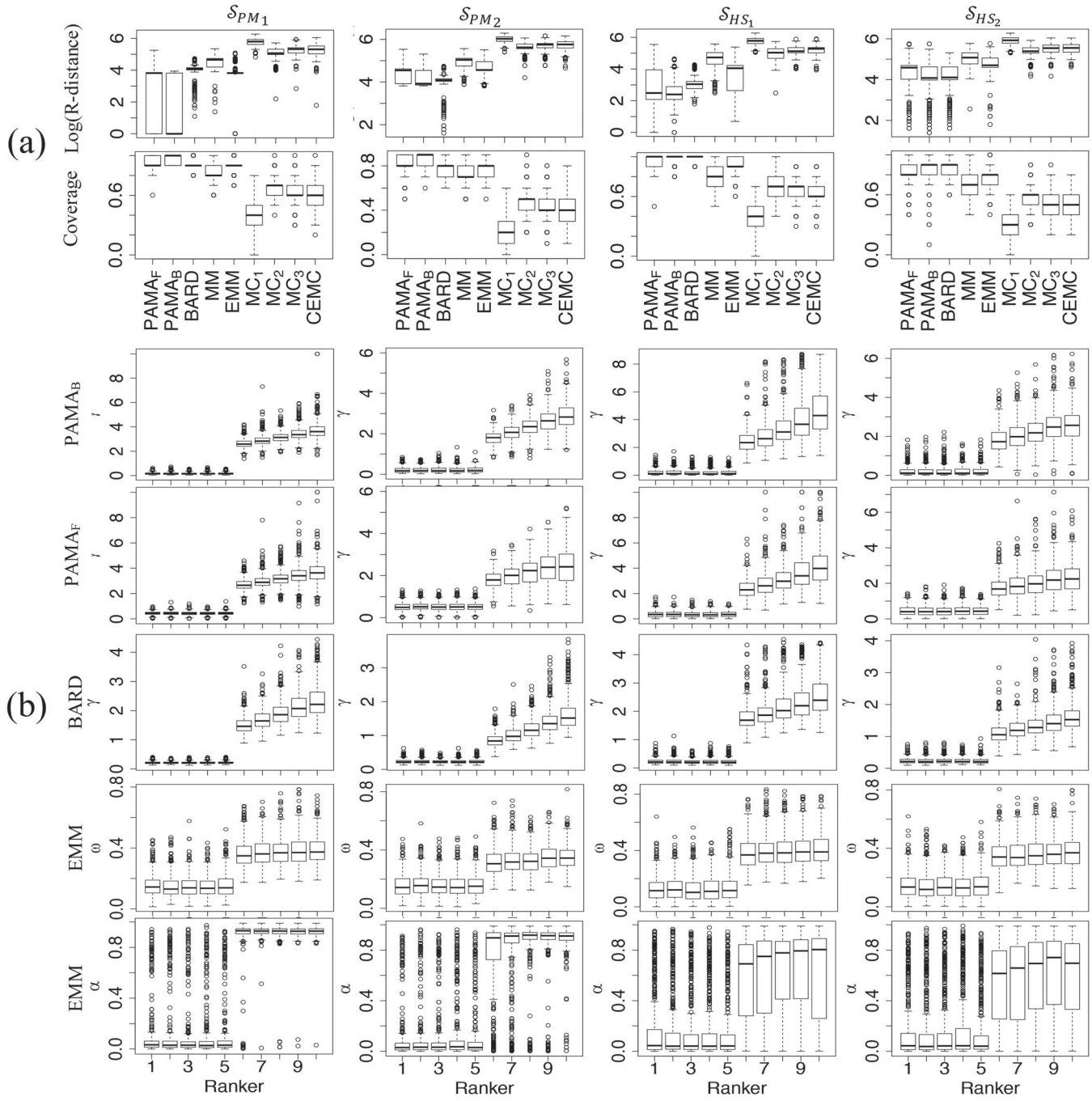

Figure 4. Boxplots of the rank aggregation results of 500 replications obtained from different methods under various scenarios with m=10, n=100, and n_1=10. (a) Recovery distances in log scale and coverage obtained from nine algorithms. (b) Quality parameters obtained by Partition-type models and EMM.

Table 3. Average recovery distances [coverages] of different methods based on 500 independent replicates under scenario $\mathcal{S}_{HS_3}$.

Configuration			Partition-type Models			Mallows Models		MC-based Models			
n	m	n_1	PAMA$_F$	PAMA$_B$	BARD	EMM	MM	MC$_1$	MC$_2$	MC$_3$	CEMC
100	10	10	44.8	**34.6**	42.6	61.5	227.7	423.8	45.6	199.1	241.3
			[0.90]	[0.93]	**[0.96]**	[0.88]	[0.58]	[0.20]	[0.92]	[0.63]	[0.54]
100	10	20	39.2	**33.9**	94.2	107.0	308.6	764.6	52.3	268.9	372.9
			[0.95]	[0.96]	**[0.99]**	[0.90]	[0.75]	[0.33]	[0.96]	[0.78]	[0.67]
100	10	30	**27.5**	29.2	207.4	126.2	360.6	1040.2	67.8	325.4	445.0
			[0.98]	**[0.98]**	**[0.98]**	[0.93]	[0.83]	[0.44]	[0.96]	[0.84]	[0.77]
100	10	40	**16.0**	17.4	363.9	131.6	408.1	1274.1	83.1	382.5	486.9
			[0.99]	**[0.99]**	[0.98]	[0.95]	[0.87]	[0.54]	[0.97]	[0.87]	[0.83]
100	10	50	**8.8**	9.3	565.3	134.6	452.2	1484.2	109.2	446.4	524.9
			[1.00]	**[1.00]**	[0.99]	[0.97]	[0.90]	[0.62]	[0.96]	[0.89]	[0.88]

NOTE: The bold texts indicate the best performance for each row.

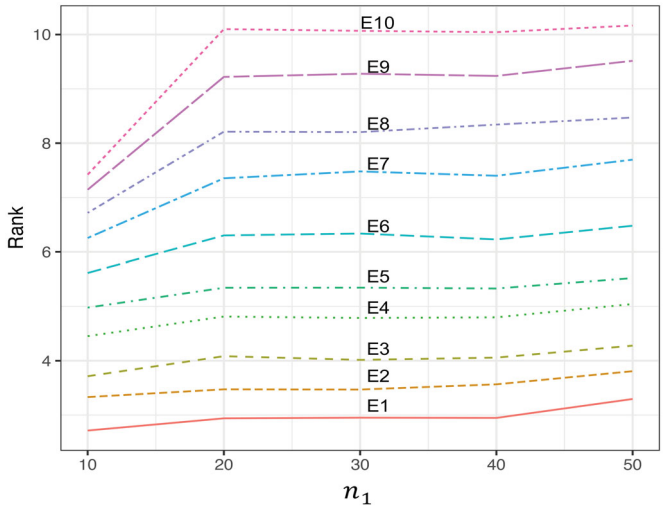

Figure 5. Average aggregated rankings of the top-10 entities by PAMA as n_1 increases from 10 to 50 for simulated datasets generated from $\mathcal{S}_{HS_3}$.

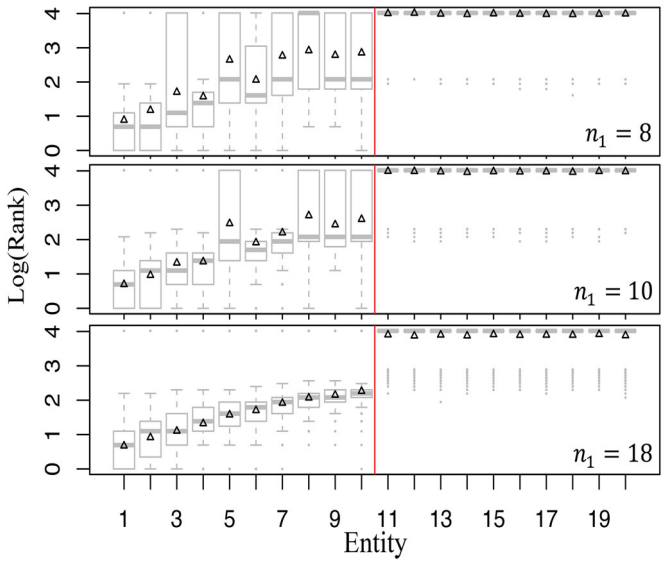

Figure 6. Boxplots of the estimated $\mathcal{I}$ from 500 replications under the setting of $\mathcal{S}_{HS_1}$ with n_1 being set as 8, 10, and 18, respectively. The true n_1 is 10. The vertical lines separate relevant entities (left) from background ones. The Y-axis shows the logarithm of the entities' ranks. The rank of a background entity is replaced by their average $\frac{100+10+1}{2}$. The triangle denotes the mean rank of each entity.

close to 1, resulting in nearly a "random" ranking among the top relevant entities. Theoretically, if all relevant entities are recognized correctly but ranked randomly, the corresponding recovery distance would increase with n_1 in an order of $\mathcal{O}({n_1}^2)$, which matches well with the increasing trend of the recovery distance of BARD shown in Table 3.

We also tested the model's performance when there is a true n_1 but it is mis-specified in our algorithm. We varied n_1 as 8, 10, and 18, respectively, for setting $\mathcal{S}_{HS_1}$ with n=100 and m=10, where the true n_1=10 (the first 10 entities). Figure 6 shows boxplots of $\mathcal{I}$ for each misspecified case. For the visualization purpose, we just show the boxplot of E_1 to E_{20}. The other entities are of the similar pattern with E_{20}. The figure shows a robust behavior of PAMA$_B$ as we vary the specifications of n_1. It also shows that the results are slightly better if we specify a n_1 that is moderately larger than its true value. The consistent results of other mis-specified cases, such as 5, 12, and 15, can be found in the supplementary material.

6. Real Data Applications

6.1. Aggregating Rankings of NBA Teams

We applied PAMA$_B$ to the NBA-team data analyzed in Deng et al. (2014), and compared it to competing methods in the literature. The NBA-team dataset contains 34 "predictive" power rankings of the 30 NBA teams in the 2011–2012 season. The 34 "predictive" rankings were obtained from 6 professional rankers (sports websites) and 28 amateur rankers (college students), and the data quality varies significantly across rankers. More details of the dataset can be found in Table 1o f Deng et al. (2014).

Figure 7 displays the results obtained by PAMA$_B$ (the partial-list version with n_1 specified as 16). Figure 7(a) shows the posterior distributions, as boxplots, of the quality parameter of each involved ranker. Figure 7(b) shows the posterior distribution of the aggregated power ranking of each NBA team. All the posterior samples that reports "0" for the rank of an entity means that the entity is a background one, for visualization purpose we replace "0" by the rank of background average rank, $\frac{n+n_1+1}{2} = \frac{30+16+1}{2} = 23.5$. The final set of 16 playoff teams are shown in blue while the champion of that season is shown in red (i.e., Heat). Figure 7(c) shows the trace plot of the log-likelihood of the PAMA model along the MCMC iteration. Comparing the results to Figure 8 of Deng et al. (2014), we observe the following facts: (i) PAMA$_B$ can successfully discover the quality difference of rankers as BARD; (ii) PAMA$_B$ can not only pick up the relevant entities effectively like BARD, but also rank the discovered relevant entities reasonably well, which cannot be achieved by BARD; and (iii) PAMA$_B$ converges quickly in this application.

We also applied other methods, including MM, EMM, and Markov-chain-based methods, to this dataset. We found that none of these methods could discern the quality difference of rankers as successfully as PAMA and BARD. Moreover, using the team ranking at the end of the regular season as the surrogate true power ranking of these NBA teams in the reason, we found that PAMA also outperformed BARD and EMM by reporting an aggregated ranking list that is the closest to the truth. Table 4 provides the detailed aggregated ranking lists inferred by BARD, EMM, and PAMA, respectively, as well as their coverage of and Kendall τ distance from the surrogate truth. Note that the Kendall τ distance is calculated for the eastern teams and western teams separately because the NBA Playoffs proceed at the eastern conference and the western conference in parallel until the NBA final, in which the two conference champions compete for the NBA champion title, making it difficult to validate the rankings between Eastern and Western teams.

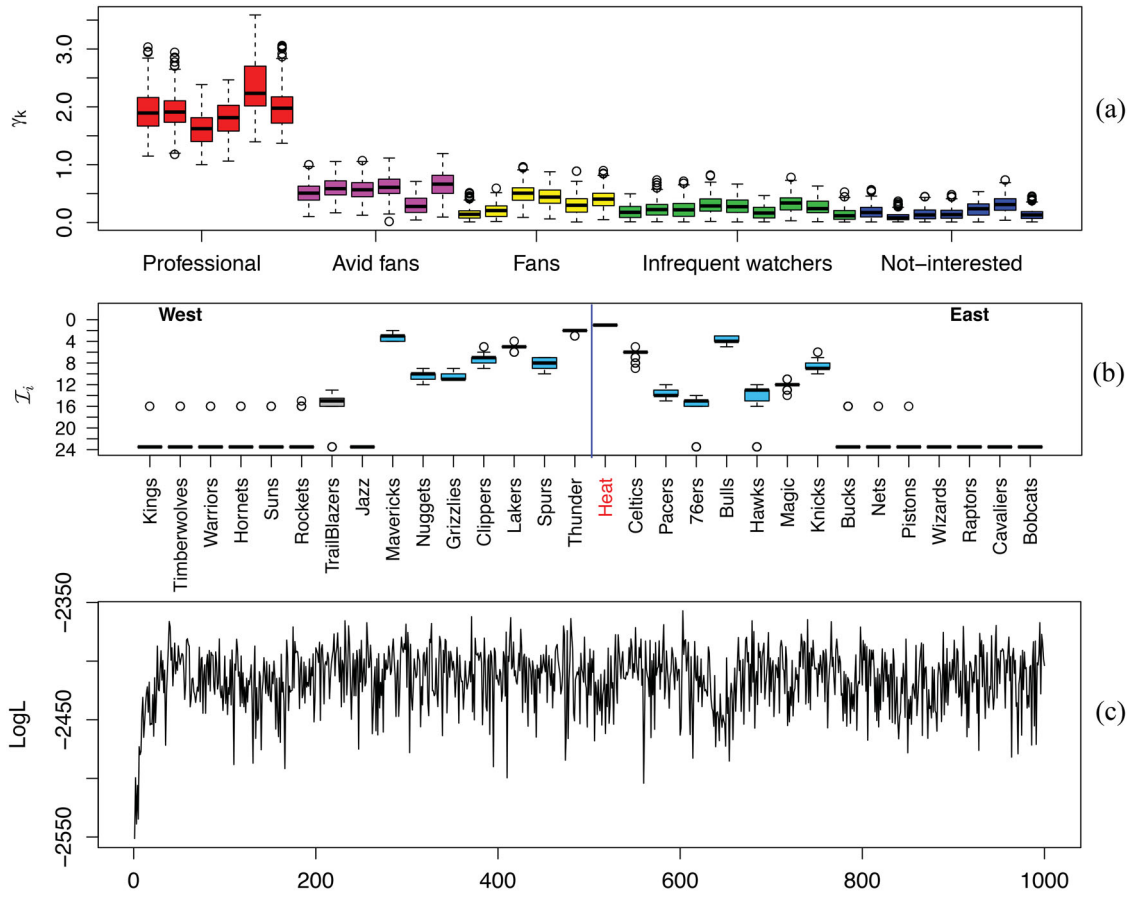

Figure 7. Results from PAMA$_B$ for the NBA-team dataset. (a) boxplots of posterior samples of γ_k. (b) barplots of $\bar{\mathcal{I}}_i$ where the vertical line divides the NBA teams in Western and Eastern Conferences. (c) the trace plot of the log-likelihood function.

Table 4. Aggregated power ranking of the NBA teams inferred by BARD, EMM, and PAMA, respectively, and the corresponding coverage of and the Kendall τ distance from the surrogate true rank based on the performances of these teams in the regular season.

Ranking	Surrogate truth		BARD		EMM		PAMA	
	Eastern	Western	Eastern	Western	Eastern	Western	Eastern	Western
1	Bulls	Spurs	*Heat*	*Thunder*	Heat	Thunder	Heat	Thunder
2	Heat	Thunder	*Bulls*	*Mavericks*	Bulls	Maverick	Bulls	Mavericksss
3	Pacers	Lakers	*Celtics*	*Clippers*	Knicks	Clippers	Celtics	Lakers
4	Celtics	Grizzlies	*Knicks*	*Lakers*	Celtics	Lakers	Knicks	Clippers
5	Hawks	Clippers	*Magic*	*Spurs*	Magic	Spurs	Magic	Spurs
6	Magic	Nuggets	*Pacers*	*Grizzlies*	Pacers	Grizzlies	Hawks	Nuggets
7	Knicks	Mavericks	76ers	*Nuggets*	76ers	Nuggets	Pacers	Grizzlies
8	76ers	Jazz	Hawks	Jazz*	Hawks*	Jazz*	76ers	Jazz*
Kendall τ	–	–	14.5	10.5	9	10	8	10
Coverage	–		$\frac{15}{16}$		$\frac{14}{16}$		$\frac{15}{16}$	

NOTE: The teams in italic indicate that they have equal posterior probabilities of being in the relevant group, and the teams with asterisk are those that were misclassified to the background group.

6.2. Aggregating Rankings of NFL Quarterback Players With the Presence of Covariates

Our next application is targeted at the NFL-player data reported by Li, Yi, and Liu (2021). The NFL-player data contains 13 predictive power rankings of 24 NFL quarterback players. The rankings were produced by 13 experts based on the performance of the 24 NFL players during the first 12 weeks in the 2014 season. The dataset also contains covariates for each player summarizing the performances of these players during the period, including the *number of games played* (G), *pass completion percentage* (Pct), the *number of passing attempts per game* (Att), *average yards per carry* (Avg), *total receiving yards* (Yds), *average passing yards per attempt* (RAvg), the *touchdown percentage* (TD), the *intercept percentage* (Int), *running attempts per game* (RAtt), *running yards per attempt* (RYds), and the *running first down percentage* (R1st). Details of the dataset can be found in (Li, Yi, and Liu 2021, table 2).

Here, we set $n_1 = 12$ in order to find which players are above average. We analyzed the NFL-player data with PAMA$_B$ (the covariate-assisted version) and the results are summarized in Figure 8: (a) the posterior boxplots of the quality parameter for all the rankers; (b) the barplot of $\bar{\mathcal{I}}_i$ for all the NFL players with the descending order; (c) the traceplot of the log-likelihood of

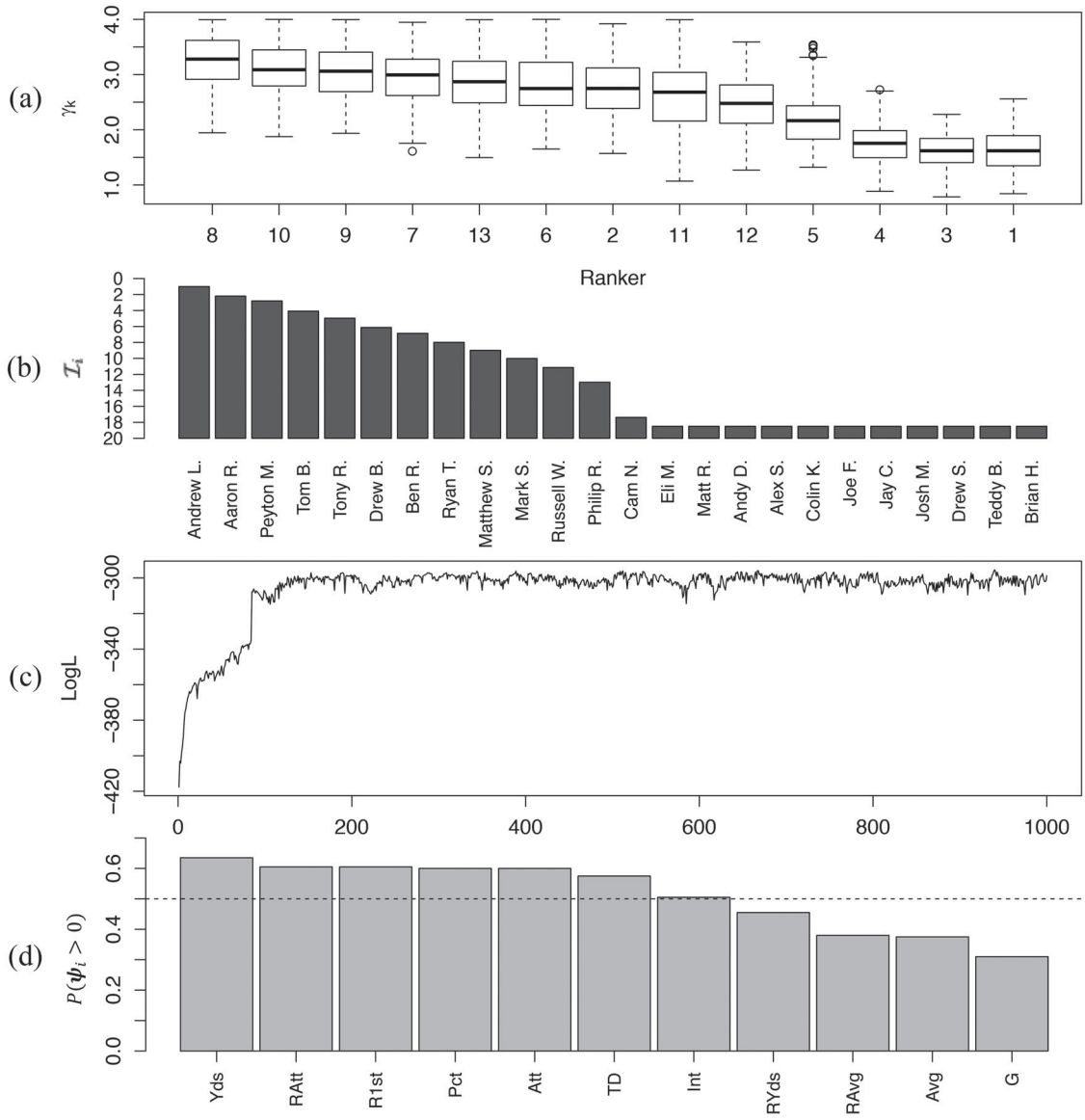

Figure 8. Key results from PAMA$_B$ for the NFL-player dataset. (a) Boxplots of posterior samples of γ_k. (b) Barplots of $\bar{\mathcal{I}}_i$. (c) Trace plot of the log-likelihood. (d) Barplots of posterior probabilities for each coefficient to be positive.

the model; and (d), the barplot of probabilities $P(\psi_j > 0)$ and the covariates are rearranged from left to right by decreasing the probability. From Figure 8(a), we observe that rankers 1, 3, 4, and 5 are generally less reliable than the other rankers. In the study of the same dataset in Li, Yi, and Liu (2021), the authors assumed that the 13 rankers fall into three quality levels, and reported that seven rankers (i.e., 2, 6, 7, 8, 9, 10, and 13) are of a higher quality than the other six (see (Li, Yi, and Liu 2021, fig. 7)). Interestingly, according to Figure 8(a), the PAMA algorithm suggested exactly the same set of high-quality rankers. In the meantime, ranker 2 is of the lowest quality among the seven high-quality rankers in both studies. From Figure 8(b), a consensus ranking list can be obtained. Our result is consistent with that of (Li, Yi, and Liu 2021, fig. 6). Figure 8(d) shows that six covariates are more probable to have positive effects.

Using the Fantasy points of the players (*https://fantasy.nfl.com/research/players*) derived at the end of the 2014 NFL season as the surrogate truth, the recovery distance and coverage of the aggregated rankings by different approaches can be calculated so as to evaluate the performances of different approaches. Note that the Fantasy points of two top NFL players Peyton Manning and Tony Romo are missing for unknown reasons, we excluded them from analysis and only report results for the top 10 positions instead of top 12. Table 5 summarizes the results, demonstrating that PAMA outperformed the other two methods.

7. Conclusion and Discussion

The proposed Partition-Mallows Model embeds the classic Mallows model into a partition modeling framework developed earlier by Deng et al. (2014), which is analogous to the well-known "spike-and-slab" mixture distribution often employed in Bayesian variable selection. Such a nontrivial "mixture" combines the strengths of both the Mallows model and BARD's

Table 5. Top players in the aggregated rankings inferred by BARD, EMM, and PAMA.

Ranking	Gold standard	BARD	EMM	PAMA
1	Aaron R.	*Andrew L.*	Andrew L.	Andrew L.
2	Andrew L.	*Aaron R.*	Aaron R.	Aaron R.
3	Ben R.	*Tom B.*	Tom B.	Tom B.
4	Drew B.	*Drew B.*	Ben R.	Drew B.
5	Russell W.	*Ben R.*	Drew B.	Ben R.
6	Matt R.	*Ryan T.*	Ryan T.	Ryan T.
7	Ryan T.	Russell W.	Russell W.	Russell W.
8	Tom B.	Philip R.*	Philip R.*	Philip R.
9	Eli M.	Eli M.*	Eli M.*	Eli M.*
10	Philip R.	Matt R.*	Matt R.*	Matt R.*
R-distance	–	35.5	32	25
Coverage	–	0.7	0.7	0.8

NOTE: The entities in italic indicate that they have equal posterior probabilities of being in the relevant group, and the players with asterisk are those that were misclassified to the background group.

partition framework, leading to a stronger rank aggregation method that can not only learn quality variation of rankers and distinguish relevant entities from background ones effectively, but also provide an accurate ranking estimate of the discovered relevant entities. Compared to other frameworks in the literature for rank aggregation with heterogeneous rankers, the PAMA enjoys more accurate results with better interpretability at the price of a moderate increase of computational burden. We also show that the Partition–Mallows framework can easily handle partial lists and and incorporate covariates in the analysis.

Throughout this work, we assume that the number of relevant entities n_1 is known. This is reasonable in many practical problems where a specific n_1 can be readily determined according to the research demands. Empirically, we found that the ranking results are insensitive to the choice of a wide range of values of n_1. If needed, we may also determine n_1 according to a model selection criterion, such as AIC or BIC.

In the PAMA model, $\pi(\tau_k^0 \mid \tau_k^{0|1})$ is assumed to be a uniform distribution. If the detailed ranking of the background entities is of interest, then we can modify the conditional distribution $\pi(\tau_k^0 \mid \tau_k^{0|1})$ to be the Mallows model or other models. A quality parameter can still be incorporated to control the interaction between relevant entities and background entities. The resulting likelihood function becomes more complicated, but is still tractable.

In practice, the assumption of independent rankers may be violated. In the literature, a few approaches have been proposed to detect and handle dependent rankers. For example, Deng et al. (2014) proposed a hypothesis-testing-based framework to detect pairs of over-correlated rankers and a hierarchical model to accommodate clusters of dependent rankers; Johnson et al. (2020) adopted an extended Dirichlet process and a similar hierarchical model to achieve simultaneous ranker clustering and rank aggregation inference. Similar ideas can be incorporated in the PAMA model as well to deal with non-independent rankers.

ORCID

Jun S. Liu http://orcid.org/0000-0002-4450-7239

Supplementary Materials

The Supplementary Materials include the numerical supports of our assumptions for the model setting, more detailed simulation results and proofs of the Theorems.

Funding

This research is supported in part by the National Natural Science Foundation of China (grant nos. 11771242 and 11931001), Beijing Academy of Artificial Intelligence (grant BAAI2019ZD0103), and the National Science Foundation of USA (grant nos. DMS-1903139 and DMS-1712714). The author Wanchuang Zhu is partially supported by the Australian Research Council (Data Analytics for Resources and Environments, grant no. IC190100031). We thank the two reviewers for their insightful comments and suggestions that helped us improve the article greatly. We also thank Miss Yuchen Wu for helpful discussions at the early stage of this work.

References

Aslam, J. A., and Montague, M. (2001), "Models for Metasearch," in *Proceedings of the 24th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, New York: ACM, pp. 276–284. [2]

Bader, M. (2011), "The Transposition Median Problem is NP-Complete," *Theoretical Computer Science*, 412, 1099–1110. [1]

Bhowmik, A., and Ghosh, J. (2017), "LETOR Methods for Unsupervised Rank Aggregation," in *Proceedings of the 26th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, pp. 1331–1340. [9]

Borda, J. C. (1781), "Mémoire sur les Élections au Scrutin," *Histoire del' Académie Royale des Sciences*. [1]

Chen, J., Long, R., Wang, X., Liu, B., and Chou, K. (2016), "dRHP-PseRA: Detecting Remote Homology Proteins Using Profile-Based Pseudo Protein Sequence and Rank Aggregation," *Scientific Reports*, 6. [1]

Chen, Y., Fan, J., Ma, C., and Wang, K. (2019), "Spectral Method and Regularized MLE are Both Optimal for Top-*K* Ranking," *Annals of Statistics*, 47, 2204. [2]

Chen, Y., and Suh, C. (2015), "Spectral MLE: Top-*K* Rank Aggregation From Pairwise Comparisons," in International Conference on Machine Learning, 371–380. [2]

Deconde, R. P., Hawley, S., Falcon, S., Clegg, N., Knudsen, B., and Etzioni, R. (2011), "Combining Results of Microarray Experiments: a Rank Aggregation Approach," *Statistical Applications in Genetics & Molecular Biology*, 5, 1544–6115. [1]

Deng, K., Han, S., Li, K. J., and Liu, J. S. (2014), "Bayesian Aggregation of Order-Based Rank Data," *Journal of the American Statistical Association*, 109, 1023–1039. [2,3,4,8,10,13,15,16]

Diaconis, P. (1988), "Group Representations in Probability and Statistics," *Lecture Notes-Monograph Series*, 11, 1–192. [1,3]

Diaconis, P., and Graham, R. L. (1977), "Spearman's Footrule as a Measure of Disarray," *Journal of the Royal Statistical Society*, Series B, 39, 262–268. [1]

Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. (2001), "Rank Aggregation Methods for the Web," in *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong: ACM, pp. 613–622. [1,2]

Fagin, R., Kumar, R., and Sivakumar, D. (2003), "Efficient Similarity Search and Classification Via Rank Aggregation," in *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, San Diego, CA: ACM, pp. 301–312. [1]

Fligner, M. A., and Verducci, J. S. (1986), "Distance Based Ranking Models," *Journal of the Royal Statistical Society*, Series B, 48, 359–369. [2,3]

Freund, Y., Iyer, R., Schapire, R. E., and Singer, Y. (2003), "An Efficient Boosting Algorithm for Combining Preferences," *Journal of Machine Learning Research*, 4, 933–969. [1]

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109. [8]

Irurozki, E., Calvo, B., and Lozano, J. A. (2014), "Permallows : An R Package for Mallows and Generalized Mallows Models," *Journal of Statistical Software*, 71, 1–30. [3]

Johnson, S. R., Henderson, D. A., and Boys, R. J. (2020), "Revealing Subgroup Structure in Ranked Data Using a Bayesian WAND," *Journal of the American Statistical Association*, 115, 1888–1901. [16]

Li, H., Xu, M., Liu, J. S., and Fan, X. (2020), "An Extended Mallows Model for Ranked Data Aggregation," *Journal of the American Statistical Association*, 115, 730–746. [2,3,10]

Li, X., Yi, D., and Liu, J. S. (2021), "Bayesian Analysis of Rank Data With Covariates and Heterogeneous Rankers," *Statistical Science*. [2,9,14,15]

Lin, S. (2010), "Space Oriented Rank-Based Data Integration," *Statistical Applications in Genetics & Molecular Biology*, 9, article 20. [1,10]

Lin, S., and Ding, J. (2010), "Integration of Ranked Lists Via Cross Entropy Monte Carlo With Applications to mRNA and microRNA Studies," *Biometrics*, 65, 9–18. [1,2,10]

Linas, B., Tadas, M., and Francesco, R. (2010), "Group Recommendations With Rank Aggregation and Collaborative Filtering," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, Barcelona, Spain: ACM, pp. 119–126. [1]

Liu, J. S. (2008), *Monte Carlo Strategies in Scientific Computing*. Springer Science & Business Media, New York: Springer-Verlag. [8]

Liu, Y., Liu, T., Qin, T., Ma, Z., and Li, H. (2007), "Supervised Rank Aggregation," in *Proceedings of the 16th International Conference on World Wide Web*, Banff Alberta, Canada: ACM, pp. 481–490. [2]

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley. [1]

Mallows, C. L. (1957), "Non-Null Ranking Models," *Biometrika*, 44, 114–130. [1,3]

Plackett, R. L. (1975), "The Analysis of Permutations," *Journal of the Royal Statistical Society*, Series C, 24, 193–202. [1]

Porello, D., and Endriss, U. (2012), "Ontology Merging as Social Choice: Judgment Aggregation Under the Open World Assumption," *Journal of Logic and Computation*, 24, 1229–1249. [1]

Rajkumar, A., and Agarwal, S. (2014), "A Statistical Convergence Perspective of Algorithms for Rank Aggregation From Pairwise Data," in *International Conference on Machine Learning*, Beijing, China, pp. 118–126. [2]

Renda, M. E., and Straccia, U. (2003), "Web Metasearch: Rank vs. Score Based Rank Aggregation Methods," in *Proceedings of the 2003 ACM Symposium on Applied Computing*, Melbourne, FL: ACM, pp. 841–846. [1]

Soufiani, H. A., Parkes, D. C., and Xia, L. (2014), "A Statistical Decision-Theoretic Framework for Social Choice," in *Advances in Neural Information Processing Systems*, eds. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Montreal, Quebec, Canada: Curran Associates, Inc., pp. 3185–3193. [1]

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540. [8]

Thurstone, L. L. (1927), "A Law of Comparative Judgment," *Psychological Review*, 34, 273–286. [1,9]

Wei, G. C. G., and Tanner, M. A. (1990), "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms," *Journal of the American Statistical Association*, 85, 699–704. [8]

Yang, K. H. (2018), Chapter 7—"Stepping Through Finite Element Analysis," in *Basic Finite Element Method as Applied to Injury Biomechanics*, ed. K.-H. Yang, Cambridge, MA: Academic Press, pp. 281–308. [7]

Young, H. P. (1988), "Condorcet's Theory of Voting," *American Political Science Review*, 82, 1231–1244. [1]

Young, H. P., and Levenglick, A. (1978), "A Consistent Extension of Condorcet's Election Principle," *SIAM Journal on Applied Mathematics*, 35, 285–300. [1]

Yu, P. L. H. (2000). "Bayesian Analysis of Order-Statistics Models for Ranking Data," *Psychometrika*, 65, 281–299. [9]