A Framework for Analyzing Rank Data in Political Science*

Yuki Atsusaka†

January 13, 2020

1 Introduction

The idea of ranked preference abounds in social science phenomena. Voters put rankings on multiple candidates in elections. People have ranked preferences on how closely they feel to different partisans. Experts consider which political institution works the best among multiple potential systems. Leaders prepare for multiple actions in the order of ranked preferences in international crises. Moreover, any binary or multinomial preference (e.g., Obama or Clinton) can be considered as a special case of ranked preferences. Along with the increasing availability of rank data, analyzing ranked preferences appears to be a promising area in political science research.

Despite the importance of ranked preferences across subfields, very little is known about how researchers can analyze and model rank data in political science. Indeed, in contrast to other types of variables – continuous, binary, count, duration, and ordinal variables – most political scientists are not familiar with how to perform simple descriptive analysis, let alone more advanced inferences on rank data. What is the empirical mean in rank data? How can we quantify variance for ranked items? When can we say that the two groups of interest have different preferences? Can we perform regression-type analyses on ranked preferences with covariates? What does hypothesis testing look like in rank data? How should we process rank data when applying various descriptive and inferential techniques?

This article answers these questions by presenting a general framework for analyzing and modeling rank data in political science.

*Any comment is appreciated.

†Graduate student, Department of Political Science, Rice University. atsusaka@rice.edu.

2 Structure of Rank Data

2.1 Concepts and Notation

Rank data arises whenever a number of units put rankings on multiple elements.[1] In the rank data literature, such units are called *judges* or *assessors* and these elements are referred to as *items*. Judges can be individuals, voters, survey respondents, leaders, groups, organizations, or countries. Items can be candidates, politically salient groups, options, public figures, political parties, or types of representation. Judges' preferences are then represented by how they "permute" available items. For example, a candidate may decide to campaign by emphasizing her experience first, passion second, and group membership the last. In this case, the candidate puts rankings of 1,2, and 3 on experience, passion, and group membership, respectively.

As this example implies, a *ranking* is considered as a projection from items to permuted numbers. Let $\mathcal{A} \in \{A_1, ..., A_n\}$ be a finite set of items (e.g., candidates) on which a judge (e.g., voter) puts ranking according to some criteria. Let $R_i \in \{1, ..., n\}$ be a rank that the judge assigns to each item A_i. Then, a full ranking of all items $\mathbf{R} = (R_1, ..., R_n)$ becomes a mapping: $\mathcal{A} \rightarrow \mathcal{P}_n$, where $\mathcal{P}_n$ is the space of n-dimensional permutations.

In the campaigning example, the candidate's ranking is expressed as $\mathcal{P}_3(\mathsf{passion}) = 2$, which means that the item $\mathsf{passion}$ has a rank of 2. A closely related concept is *ordering* and it is represented by an inverse of the ranking function $\mathcal{P}_n^{-1}$. In the example, the candidate's ordering is represented by $\mathcal{P}_3(2)^{-1} = \mathsf{passion}$, which tells that the item which has a rank of 2 is $\mathsf{passion}$. Simply put, a ranking is a function of an item which outputs a corresponding number (i.e., rank) and an ordering is a function of a rank which provides a corresponding item for each judge.

Generally, a judge is said to prefer A_i to A_j when $R_i < R_j$ and the most preferred item is represented as A_1. In our example, the candidate prefers $\mathsf{passion}$ to $\mathsf{group\ membership}$ because their corresponding ranks are 2 and 3, respectively (i.e., first choice and second choice). Here, the most preferred item is $\mathsf{experience}$ since its rank is 1. It should be noted that we usually say that $\mathsf{experience}$ has the highest rank even though its corresponding ranking takes the smallest number (1).

When judges show ranked preferences on n items, the total number of possible ranked preferences becomes $n!$. This is astonishing since the total number of ways judges can express their preferences rapidly grows as the number of items increases. Row 1 of Table 1 demonstrates this point. When voters choose one

[1] Rank data are also referred to as ranking data or ranked data.

candidate among seven, the total number of ways that voters can express their preferences is 7. When they can rank order all candidates, however, the total number becomes 5040. This property implies that rank data is often a high-dimensional data, whose analysis entails computational difficulties.

	3	4	5	6	7	8	9	10	# items
Complete rank	6	24	120	720	5040	40320	362880	3628800	
Top 3 Complete	6	24	60	120	210	336	504	720	
Top 3 Incomplete rank	16	53	126	247	428	681	1018	1451	

Table 1: **Total Number of Possible Ranked Preferences**

When there are maximum numbers of items that judges can specify, a ranking is often called as *top-k ranking*. For example, voters may be able to choose only top-3 candidates among 10 candidates in ranked-choice voting. In such case, the total number of possible rankings becomes $\binom{n}{k} \times k!$. When judges can only choose a single or subset of items to express more nuanced preferences, the resulting ranking is referred to as a *partial ranking*. In partial rankings, the total number of possible rankings becomes $2\binom{n}{k}k! + \sum_{i=1}^{k-1}\binom{n}{k-i-1}(k-i-1)!$. Rows 2 and 3 of Table 1 demonstrate the results for a top-3 ranking and top-3 partial ranking, respectively. Although the increase in the total number for the two rankings is more gradual than full rankings, the number nevertheless rapidly increases as the number of items increases.

When researchers collected rank data, it is thus critical to understand the basic structure of data. This includes examining the number of judges and items, the existence of maximum number of choices judges can make, and the presence of any partial ranking. To simplify the discussion, I assume that we observe complete rank data and it is generated by some probabilistic sampling in the analyses below.

2.2 Descriptive Analysis of Ranked Preferences

While I introduced the notion of rankings as a ranking by a single judge, in practice, what we observe is a set of rankings by multiple judges: $\mathbf{R} = (\mathbf{R}_1, ...\mathbf{R}_N)$. Here, N is used for the number of judges, whereas n is reserved for the number of items. For example, assume that you have collected data from five candidates about their campaign strategies. Transforming your data into the ranking format, your "data" can be represented by the matrix in Table 2. In the matrix, rows represent five different judges, whereas columns show three different items. In practice, however, researchers may collect their data in the ordering format shown in Table 3. Depending on the types of methods they wish to use, data format must be transformed.

	Experience	Passion	Group
Candidate A	1	2	3
Candidate B	2	1	3
Candidate C	1	3	2
Candidate D	1	2	3
Candidate E	2	3	1

Table 2: **A Hypothetical Rank Data**

	1st	2nd	3rd
Candidate A	Experience	Passion	Group
Candidate B	Passion	Experience	Group
Candidate C	Experience	Group	Passion
Candidate D	Experience	Passion	Group
Candidate E	Group	Experience	Passion

Table 3: **A Hypothetical Ordering Data**

After understanding the structure of data, researchers may wish to perform descriptive analysis to learn useful information from it.

One such information is about the popularity of each item. What is the mean ranking that judges put on a particular item? The *mean rank* of item j represents such popularity measure as a weighted average of rankings assigned to j, where the weight reflects the empirical probability that each ordered preference appears in data. The mean rank of item j becomes:

$$m_j = \sum_{i=1}^{n!} \frac{N_i}{N} \mathcal{P}_i(j), \tag{1}$$

where N is the total number of judges, N_i is the number of judges with ranked preference i, $n!$ is the total number of possible ranked preferences, and $\mathcal{P}_i(j)$ is a ranking assigned to item j in ranked preference i. Sometimes it is represented the mean rank **m** which is a k-dimensional vector having m_j is jth entry (Lee and Yu 2013, 3). Here, it must be emphasized that an index i is assigned to each ranking, not individual judge i.

Calculating the mean rank vector in our hypothetical data, I obtain **m** = (1.4, 2.2, 2.4). The result implies that the item Experience has the lowest number for the mean rank and thus is the most preferred strategy in our dataset. It also suggests that we do not observe a great difference between Passion and Group, while we see a preference gap between Experience and the other two items.

Researchers may also be interested in the popularity of an item relative to another specific item of interest. The *pairwise frequency* of item j relative to item k represents the relative frequency that item j is ranked higher than item k among all possible $\binom{n}{2}$ item pairs:

$$P_{st} = \sum_{i=1}^{n!} N_i \mathbb{1}[\mathcal{P}_i(j) > \mathcal{P}_i(k)], \tag{2}$$

where $\mathbb{1}$ is an indicator function denoting 1 if the condition inside the bracket holds. We can similarly consider the pairwise empirical probability as:

$$P_{st} = \sum_{i=1}^{n!} \frac{N_i}{N} \mathbb{1}[\mathcal{P}_i(j) > \mathcal{P}_i(k)], \tag{3}$$

Like the mean rank, the pairwise frequencies of all possible pairs are represented as $n \times n$ matrix **P** whose (s,t)-th entry corresponds to P_{st}, where the diagonal of **P** is a zero-vector.

Here, I calculate the pairwise frequency matrix **P** and obtain:

	Experience	Passion	Group
Experience	0	4	4
Passion	1	0	3
Group	1	2	0

The matrix clearly shows that Experience is preferred to the other two items. It is also clear that while the mean ranks implied that Passion and Group have similar mean ranks, candidates generally seem to prefer the former to the latter strategy.

Finally, it is sometimes useful to refer to the marginal distribution of item j being ranked s-th:

$$M_{j,s} = \sum_{i=1}^{n!} N_i \mathbb{1}[\mathcal{P}_i(j) = s] \tag{4}$$

which can be also expressed as $n \times n$ matrix **M** where (j,s)-th entry corresponds to the marginal $M_{j,s}$

In our example, the marginal distribution matrix **M** looks like the following:

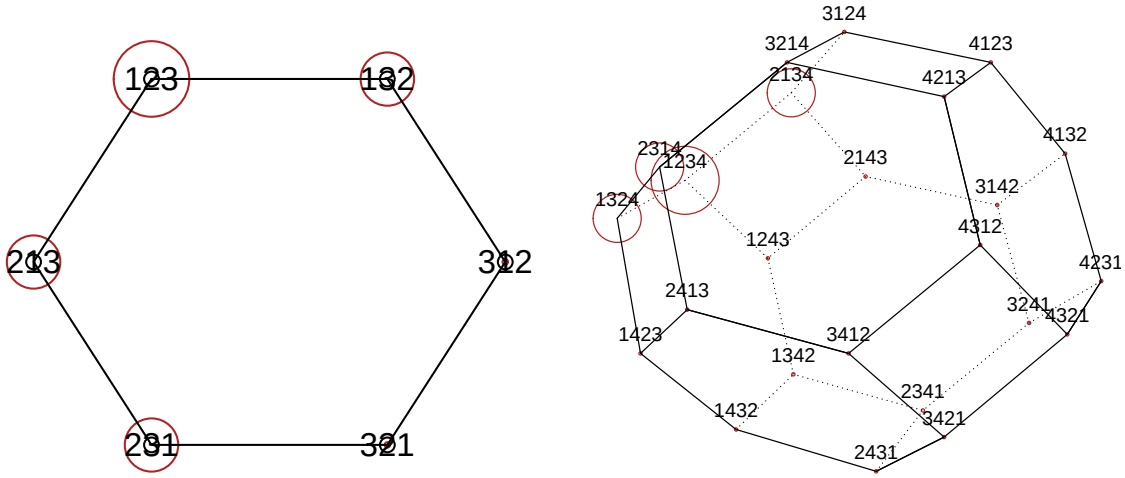
	1st	2nd	3rd
Experience	3	2	0
Passion	1	2	2
Group	1	1	3

2.3 Visualization of Rank Data

Along with summary statistics, researchers often wish to visualize their data to understand the whole picture or "distribution" of it. For more conventional types of data, it is usually done with some non-parametric

estimation such as histogram or kernel density for single variables or with scatter plot for pairs of variables. Visualizing rank data, however, is tricky in a sense that these conventional tools cannot be used. Indeed, visualizing rank data is difficult because it is, by definition, multidimensional and multivariate data, where each observation possesses values for multiple variables.

One conventional approach to visualize complete rank data employs what is called a "permutation polytope." A permutation polytope plots empirical distributions or frequencies of observed rankings (Thompson, 1993*a*,*b*). To illustrate this method, I plot two hypothetical rank data in Figure 1.



(a) Three Items (b) Four Item

Figure 1: **Permutation Polytope for Hypothetical Data**
Note: This figure presents visualization of rank data with three and four items, respectively. The three item rankings are displayed on a hexagon and the four item rankings are shown on a truncated octahedron.

On the left panel of Figure 1, I plot complete rank data with three items on a hexagon. Here, each of its vertices represents a possible ranking among all permutations of three items (i.e., 3!=6) and an open circle over each vertex (and its size) expresses the frequency of the corresponding ranking. When a particular ranking is not observed, no open circle is located on its corresponding vertex. In this example, we observe the largest circle on the vertex of ranking (1,2,3), smaller circles on the vertices of rankings (1,3,2), (2,1,3), and (2,3,1), and no circle exists on the vertices of rankings (3,1,2) and (3,2,1). Notice that this is a straightforward visualization of Table 2.

Moreover, each edge connects two vertices in a way that it represents a particular property of data. For example, edges on the top, lower left, and lower right represent rankings whose "first" choices are 1, 2,

and 3, respectively. In contrast, the remaining edges on upper left, the bottom, and upper right represent rankings whose "last" choices are 3, 1, and 2, respectively. The graph, then, suggests that 1 is the most popular and 3 is the least popular as the first choice, whereas 3 is the most popular as the last choice. While our example only has five judges and the benefit of using permutation polytope is negligible, it becomes increasingly useful as the number of judges increases to graphically understand the empirical distribution of possible rankings.

On the right panel of Figure 1, I also plot complete rank data with four items on a truncated octahedron. Now, representing unique rankings on its vertices, the three dimensional graph is drawn in the two dimension. Like before, open circles are put over the vertices and their size represents the empirical frequencies of their corresponding rankings.

More generally, a ranking of t items is represented as a coordinate in the $\mathbb{R}^{t-1}$th dimensional Euclidean space. Thus, rankings of three items are shown in the two dimension, while rankings of four items are located in the three dimension. The limitation of permutation polytope is that it cannot apply to rankings with more than five items. For such rankings, other approaches such as multidimensional scaling and unfolding as well as multidimensional preference analysis can be employed (Alvo and Yu, 2014, 11-19).

3 Hypothesis Testing

So far, we have considered a variety of descriptive tools to find similarities and dissimilarities in rank data. Beyond examining sample values, however, researchers may wish to perform a hypothesis testing to make statistical inference about properties of data in the spirit of the difference in means test. In this section, I describe two primary types of hypothesis testing: uniformity hypothesis testing and heterogeneous rankings testing. In the first test, we ask whether there is *any* pattern in a given rank data. In the second test, we ponder if two groups of interest possess different ranked preferences.

3.1 Uniformity Hypothesis Testing

A major class of hypothesis testing in rank data is called uniformity test. A rank data is said to be *uniform* when all possible rankings have the same probability of occurring. More precisely, each ranking has a probability of $\frac{1}{n!}$ under uniformity. When rank data is uniform, we usually say that there is no pattern in ranked preferences (e.g., voters, on average, are satisfied with any candidate). In uniformity test, we

establish a null hypothesis that the given rank data is uniform and alternative hypothesis that the observed rankings are not uniform.

There are several ways to test the null hypothesis. The most intuitive approach is to use the proportions of observed rankings to create test statistics as the expected frequencies of each rankings becomes $N \times \frac{1}{n!} = \frac{N}{n!}$. When $n!$ is small or moderate compared to N, the standard chi-square goodness-of-fit test can be employed to test the null hypothesis. Here, the test statistic is the sum of squared differences between observed and expected frequencies divided by the expected frequencies over all rankings, or

$$T_{prop} = \sum_{r=1}^{n!} \frac{(x_r - N/n!)^2}{N/n!} = \sum_{r=1}^{n!} \frac{x_r^2}{N/n!} - n!, \tag{5}$$

where x_r is the observed frequency for ranking r. Under the null hypothesis, the test statistic is assumed to follow a χ^2 distribution with $n! - 1$ degrees of freedom. When an observed p-value is lower than a pre-specified benchmark (e.g., 0.01), the null hypothesis can be rejected at the specified level.

Alternatively, the mean rank, pairwise frequencies, and marginal frequencies can be used to construct test statistics

$$T_{mean} = \frac{12N}{n(n+1)} \sum_{j=1}^{n} \left(m_j - \frac{n+1}{2} \right)^2 \tag{6}$$

$$T_{pair} = 12N \left\{ \sum_{a>b}^{n} \left(P_{ab} - 0.5 \right)^2 - \frac{1}{n+1} \sum_{j=1}^{n} \left(m_j - \frac{n+1}{2} \right)^2 \right\} \tag{7}$$

$$T_{mar} = N(n+1) \sum_{a>b}^{n} \left(M_{ab} - \frac{1}{n} \right)^2, \tag{8}$$

respectively (Marden, 1995, Ch.2.1). Under the null hypothesis that the rankings are uniform, each statistic becomes $\frac{(n+1)}{2}$, $\frac{N}{2}$, and $\frac{N}{n}$, and the test statistics follow a χ^2 distribution with $n-1$, $\binom{n}{2}$, and $(n-1)^2$ degrees of freedom, respectively. Similar to the above, researchers can reject the null hypothesis when observed p-values for their test statistics is are lower than their specified benchmark. These approaches are often useful when $n!$ is large compared to N, where for some rankings observed frequencies may be too small (e.g., less than 5) to perform the asymptotic hypothesis testing.

(Demonstration here)

3.2 Homogeneity Hypothesis Testing

Building on the uniformity test, researchers can also perform a test that is similar to a difference in means test. That is, for given two rank data, several homogeneity tests can be applied to rank data in order to study whether there is any difference between rankings generated by judges from two groups of interest. Here, the null hypothesis is that two groups of interest have the same distribution of rankings.

Among several available tests, the most intuitive strategy is to use the chi-square test of homogeneity between the marginal frequencies of rankings for two populations of judges. In this test, the test statistic is derived from a $n^2 \times 2$ contingency table where the vector of marginal frequencies for each group are listed side to side. To illustrate this approach, a contingency table from two hypothetical marginals are depicted in Table 4.

	1st	2nd	3rd
A	30	10	10
B	20	20	0
C	20	20	0

Group 1 Marginals

+

	1st	2nd	3rd
A	40	0	20
B	40	0	20
C	40	0	20

Group 2 Marginals

→

	Group 1	Group 2	Totals
A_1	30	40	70
A_2	10	0	10
A_3	10	20	30
B_1	20	40	60
B_2	20	0	20
B_3	0	20	20
C_1	20	40	60
C_2	20	0	20
C_3	0	20	20
Totals	130	180	310

Table 4: **Contingency Table for Homogeneity Test**

The test statistic then becomes:

$$T_h = \sum_{i=1}^{2n^2} \frac{(O_i - E_i)^2}{E_i}, \tag{9}$$

where O_i and E_i are observed and expected frequencies of table cell i, respectively. Here, the expected frequency for (rc)-th cell is calculated as $E_{rc} = T_c \times T_r/T$, where T_c refers to the total number of cell counts in Column c, T_r the total number of cell counts in Row r, and T the total number of cell counts in the table. For example, the expected frequency for A_1 in Group 1 is $130 \times 70/310 = 29.35$, which turns out to be very close to the observed frequency for the cell (30).

Under the null hypothesis, the test statistic is asymptotically distributed as a chi-squared distribution

with $(n-1)^2$ degrees of freedom. Suppose that we specify $p = 0.001$ as our benchmark for testing the null hypothesis. For Table 4, $T_h = 102.7$ and the probability of observing such value under $\chi^2_{df=4}$ distribution is less than 0.001. Then, it is plausible to reject the null and conclude that Group 1 and Group 2 do not have the same ranked preferences on the three items. When $n!$ is large compared to N, however, this test for homogeneity might lead to incorrect inferences and other approaches must be employed.

For testing homogeneity for three or more populations of judges, a multivariate analysis of variance (MANOVA) type of tests are considered (Feigin and Alvo, 1986; Marden, 1995; Alvo and Yu, 2014).

4 Statistical Models for Rank Data

In this section, I discuss various statistical models for rank data. The primary purpose of using such models is to make a probabilistic inference for what is called the *shared true consensus ranking* in rank data. The shared true consensus ranking refers to a unique ranking that most judges agree on, and it is usually estimated under parametric models. This basic idea can be also extended to the estimation of multiple clusters in ranked preferences, where researchers can discover any grouping structure in their rank data. Moreover, some probabilistic models can incorporate covariates to explore the associations between covariates and specific consensus rankings.

4.1 Estimating Shared Consensus Rankings

To date, the literature on rank data has considered four types of probabilistic models for dank data. By probabilistic models, I refer to a group of models which assume particular stochastic data generating processes. The four types of models include the order statistics models, paired comparison models, multistage models, and distance-based models.

4.1.1 Order statistics models

The *order statistics models*, as known as random utility models, assume that rank data arise from people's "comparative judgement" on items according to their random utility scores. First proposed by Thurstone (1927), the order statistics model assumes that, given n items, each judge j has a vector of random variables $\mathbf{Y}_j = (Y_{j1}, ..., Y_{jn})$ which represent her random scores or utilities for all items. Here, Y_{ji} is assumed to be independently drawn from its own probability distribution F_i. It is then assumed that observed rankings for

j, $\mathrm{R}_j = (R_{j1}, ..., R_{jn}) \in \mathcal{P}_n$, arise according to the probability defined by such random score variables:

$$P(\mathbf{R}) = P(Y_{X_1} < Y_{X_2} \cdots < Y_{X_n} | \mathbf{X} = \mathbf{R}^{-1}) \tag{10}$$

An important implication from Equation (5) is that the data generating process of observed rankings (the LHS) is determined by the unobserved orders of random score variables (the RHS). Commonly, the order statistics model "parameterizes" Equation (5) by considering an additive model for each random variable, $Y_i = u_i + \epsilon_i$, $i = 1, ..., n$, where the u_i denotes the *mean score* associated with item A_i and ϵ_i represents an error term. Statistical models with this specification are often called as Thurstone order statistics models.

Finally, a further specification is applied to $F_i(y) = F(y - u_i)$. When the normal cumulative distribution function (CDF) is used for F, the resulting model is called the Thurstone model (Thurstone, 1927). When instead the Gumbel CDF is employed, the resulting model is known as the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Luce, 1959). The goal of this class of models is to estimate an unobserved vector of the mean scores $\mathbf{u} = (u_1, ..., u_n)$ from observed rankings of N judges or $\mathbf{R} = (\mathbf{R}_1, ..., \mathbf{R}_N)$ via the maximum likelihood or other estimation methods. After estimating $\mathbf{u}$, a consensus ranking can be obtained **(more explanation needed)**.

4.1.2 Paired comparison models

The *paried comparison models* assume that the observed full rankings emerge from a series of pairwise comparisons between two items by judges. One of the most popular model in this class is the Mallows-Bradley-Terry (MBT) model, where the probability that a judge prefers one item (A_i) to another (A_j) is modeled as $P(A_i \prec A_j) = \frac{\mu_i}{\mu_i + \mu_j}$, where $\mu_i > 0$, $i = 1, ..., n$ and $\sum_{i=1}^{n} \mu_i = 1$. Like the order statistics models, the paired comparison models represent some utilities for all items as μ_i and the probability of interest is determined *only* by the relative size of the utilities for two items in comparison. The ratio of two random utilities is called the Bradley-Terry form (Bradley and Terry, 1952).

The likelihood function for a single judge under the MTB model becomes:

$$P(\mathbf{R}|\boldsymbol{\mu}) = c(\boldsymbol{\mu}) \prod_{i=1}^{n-1} \mu_i^{n-R_i}, \tag{11}$$

where $c(\boldsymbol{\mu})$ is a normalizing factor that guarantees that the probability sums up to 1. The purpose of this

family of models is to estimate an unobserved vector of $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)$ from observed full rankings of N judges. For estimation, we consider a likelihood function $\prod_{j=1}^{N} P(\mathbf{R}_j|\boldsymbol{\mu})$ for the ML or other estimation methods. Other models have been available including Mallows ϕ models and Mallows ρ models (Mallows, 1957).

4.1.3 Multistage models

The *multistage models* assume that rank data arises from a series of sampling without replacement over a set of available items. Also known as the urn models, the multistage models consider that the sample probability is a function of random utilities assigned to the most preferred item and to the rest of the items in each iteration of sampling. The most dominant model in this family is called the Plackett-Luce (PL) model (Luce, 1959; Plackett, 1975). Based on the multistage sampling process, the PL model assume that the observed *orderings* $\mathbf{X} = (X_1, ..., X_n)$ is determined by the probability of A_i being the top ranked item $\frac{\mu_i}{\sum_{k=1}^{n} \mu_k}$, $i = 1, ..., n$, the probability of A_m being the top ranked item in the remaining items $\frac{\mu_m}{\sum_{k \neq m} \mu_k}$, $i \neq m$, and a subsequent probability in each iteration. Here, μ_i represents some random utility assigned to an item A_i. Since the last item is drawn with a probability of 1, we end up with having $n - 1$ iterations of sampling.

The likelihood function based on a single judge's ordering becomes:

$$P(\mathbf{X}|\boldsymbol{\mu}) = \prod_{i=1}^{n-1} \frac{\mu_i}{\sum_{j=i}^{n} \mu_j} \tag{12}$$

Like the order statistics models and paired comparison models, the multistage models also consider some random scores assigned to all items μ_i. Thus, some authors call the above three families of models as the "score-based" methods (Liu et al., 2018, 334) in contrast to what I introduce next: the distance-based methods.

4.1.4 Distance-based models

The *distance-based models*, also known as the Mallows model (MM), assume that observed full rankings are generated by the "distance" between each observed ranking and the shared true consensus ranking (Mallows, 1957; Diaconis, 1988). In this family of models, the probability of an observed ranking is said to decay as the distance between the ranking and the consensus ranking increases. Formally, the distance-based models

consider the following likelihood function for an observed ranking for a single judge:

$$P(\mathbf{R}|\alpha, \boldsymbol{\rho}) = \frac{1}{Z_n(\alpha)} \exp\left[-\frac{\alpha}{n} d(\mathbf{R}, \boldsymbol{\rho}) \right], \tag{13}$$

where $\boldsymbol{\rho} \in \mathcal{P}_n$ is the location parameter, expressing the true consensus ranking (i.e, central location of the distribution), $\alpha > 0$ is the scale parameter, representing the dispersion around the consensus ranking (just like a variance of the distribution), $Z_n(\alpha)$ is the normalizing function, and $d(\mathbf{R}, \boldsymbol{\rho})$ is the distance between the judge's ranking and the true consensus ranking.

So far, I have introduced the distance-based family of models without defining what the *distance really means*. Since the notion of distance is critical in rank data analysis, I account for the idea of distance in more depth in the later section. In a more practical term, the distance between two rankings is a metric of how far (or close) a pair of rankings is. Suppose that there exists a true consensus ranking (1,2,3) for three items. If your own ranking is (1,3,2), the distance between the two rankings is expressed as $d(\mathbf{R}, \boldsymbol{\rho}) = d((1,3,2),(1,2,3))$.

How far is your ranking from the consensus ranking? It seems that the rankings are not *too* far from each other because they both have 1 in the first position. However, one can tell that they are not identical as the second and third items have different rankings. To quantify such notion of "how far," however, it is necessary to rely on some distance functions for rank data.

4.1.5 Distance between Rankings

4.2 Discovering Clusters of Consensus Rankings

Learning about potential grouping structures in rank data is informative when researchers want to understand the level of diversity and potential cleavages in ranked preferences. In rank data analysis, it often takes the form of discovering multiple clusters of consensus rankings based on mixtures of parametric rank models. The basic idea is that for a given rank data it is possible to estimate different model parameters for multiple clusters, including different consensus rankings.

Various parametric models have been considered under the mixture models framework.

A challenge in using these mixture models is that researchers *ex ante* need to specify the number of clusters. The general advise made in the literature is that the number of clusters must be selected after estimating models with varying numbers of clusters based on within-cluster deviation measures. Specifically, it

is argued that researches should adopt the number of clusters which minimizes the within-cluster distances between the consensus ranking and observed rankings (Liu et al., 2018).

Cluster estimation can be implemented with several R packages and details are discussed in Appendix A.

4.3 Using Covariates in Rank Models

5 Empirical Illustrations

6 Advanced Topics

6.1 Incomplete Rankings

6.2 Pairwise Comparisons

6.3 Tied Rankings

6.4 Inferences on Unobserved Individual Rankings

6.5 Rank Data Over Time

6.6 Model Selection

6.7 Computational challenge

7 Concluding Remarks

This article provided a general framework for analyzing and modeling rank data with specific applications to political science research in minds.

References

Alvo, Mayer and Philip LH Yu. 2014. *Statistical methods for ranking data*. Springer.

Bradley, Ralph Allan and Milton E Terry. 1952. "Rank analysis of incomplete block designs: I. The method of paired comparisons." *Biometrika* 39(3/4):324–345.

Diaconis, Persi. 1988. "Group representations in probability and statistics." *Lecture notes-monograph series* 11:i–192.

Feigin, Paul D and Mayer Alvo. 1986. "Intergroup diversity and concordance for ranking data: An approach via metrics for permutations." *The Annals of Statistics* pp. 691–707.

Liu, Qinghua, Marta Crispino, Ida Scheel, Valeria Vitelli and Arnoldo Frigessi. 2018. "Model-Based Learning from Preference Data." *Annual Review of Statistics and Its Application* (0).

Luce, R Duncan. 1959. "On the possible psychophysical laws." *Psychological review* 66(2):81.

Mallows, Colin L. 1957. "Non-null ranking models. I." *Biometrika* 44(1/2):114–130.

Marden, John I. 1995. *Analyzing and modeling rank data*. Chapman and Hall/CRC.

Plackett, Robin L. 1975. "The analysis of permutations." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24(2):193–202.

Thompson, GL. 1993*a*. "Generalized permutation polytopes and exploratory graphical methods for ranked data." *The Annals of Statistics* 21(3):1401–1430.

Thompson, GL. 1993*b*. Graphical techniques for ranked data. In *Probability models and statistical analyses for ranking data*. Springer pp. 294–298.

Thurstone, Louis L. 1927. "A law of comparative judgment." *Psychological review* 34(4):273.

Appendices

A Software for Rank Data Analysis

In this section, I summarize currently available statistical software for rank data analysis.