



The Analysis of Ranking Data Using Score Functions and Penalized Likelihood

Mayer Alvo
University of Ottawa

Hang Xu
University of Hong Kong

Abstract

In this paper, we consider different score functions in order to summarize certain characteristics for one and two sample ranking data sets. Our approach is flexible and is based on embedding the nonparametric problem in a parametric framework. We make use of the von Mises-Fisher distribution to approximate the normalizing constant in our model. In order to gain further insight in the data, we make use of penalized likelihood to narrow down the number of items where the rankers differ. We applied our method on various real life data sets and we conclude that our methodology is consistent with the data.

Keywords: Rankings, Score function, Rao score test, Spearman, Kendall, Penalized likelihood, von Mises distribution.

1. Introduction

Ranking data occur quite frequently in practice. There are examples in sport competitions (Deng, Han, Li, and Liu 2014), in the selection of candidates in political elections (Croon 1989; Lee and Philip 2012), in the arrangement of web-pages when using search engines (Aslam and Montague 2001) and in flagging disease related gene in bioinformatics (DeConde, Hawley, Falcon, Clegg, Knudsen, and Etzioni 2006) just to name a few. In all cases it is of interest to present summaries of the data. Some of the methods include the calculation of the average ranks. Borda counts (Baba (1986)) usually sum up the rank as each item's score and present transformations such as ordering or averaging. There are also more sophisticated methods which involve building models to explain the ranking data. For example, Kemeny rankings (MalloWS 1957; Kemeny and Snell 1962) are based on fitting a distance-based model with Kendall distance and using the modal rank as a summary analogous to the mean for numerical data. Deng *et al.* (2014) suggested a Bayesian approach to aggregate the observed rankings whereas Dwork, Kumar, Naor, and Sivakumar (2001) suggested a model using a Markov chains approach to have a summary of the word association or the web-page for example.

Here we first introduce a parametric model for the one sample case which incorporates an arbitrary score function. Such score functions include both the Spearman and Kendall scores. The former focuses attention on a single item at a time whereas the latter allows for comparisons between pairs of items. In the era of big data, the number of items being ranked is usually very large. For this reason, we may apply penalized likelihood methods to focus on those

items that are particularly preferred by the rankers. The parametric model can be extended to the two sample case whereby we now compare groups of judges. Bootstrap methods can be particularly useful for constructing a confidence interval for the parameters in the models. The article is organized as follows. In Section 2 we formally introduce the parametric model and indicate how to calculate the maximum likelihood estimates (MLE). We also obtain the estimates using penalized likelihood (PMLE). In Section 3, we introduce the Spearman score function, explain its significance and then apply it to three real data sets. In Section 4 we consider the Kendall score function, explain its significance and use it on the same data sets. In Section 5 we consider the two sample problem and illustrate our results on two real life data sets. We conclude with some remarks in Section 6.

2. The Probability Models

Neyman (1937) first proposed smooth tests for testing the null hypothesis that data come from a uniform distribution on the interval $(0, 1)$. The smooth alternative density proposed by Neyman was given by

$$p(y, \theta) = \exp \left\{ \sum_{i=1}^k \theta_i h_i(y) - K(\theta) \right\}$$

where $\theta' = (\theta_1, \theta_2, \dots, \theta_k)$ is a set of unknown parameters, $K(\theta)$ is the normalizing constant and $h_i(y)$ are the Legendre orthonormal polynomials with respect to the uniform on $(0, 1)$. Motivated by this model, we can define a similar probability function for ranking data. Assume there are n judges each of whom ranks t items. Letting $\{\omega_j\}$ be the set of $t!$ possible rankings and define the probability of the π_j is given by

$$\pi_j(\theta) = \exp \{ \theta' X(\omega_j) - K(\theta) \} \frac{1}{t!}, \quad j = 1, \dots, t!$$

where $\theta = (\theta_1, \dots, \theta_k)'$ is a k -dimensional vector of parameters, $K(\theta)$ is a normalizing constant. $X(\omega_j)$ is a k -dimensional vector transformed from the original ranking ω_j . The transformation rule $X(\omega_j) = x_j$ is defined by certain scoring function. The likelihood function is obtained from the multinomial distribution and is proportional to

$$L(\theta) \sim \prod [\pi_j(\theta)]^{n_j}$$

where n_j is the number of judges choosing ω_j . Let the total sample size $n = \sum_{j=1}^{t!} n_j$. The log-likelihood function is then given by

$$l(\theta) \sim \sum_{j=1}^{t!} n_j [\theta' x_j - K(\theta)]$$

As mentioned in the introduction, when the number of items being ranked is large, the dimension $t!$ becomes very large. Penalized likelihood is a parametric method which helps to narrow down the number of items to look at. The main idea is to add a penalty term in the log likelihood function. In this paper, we consider l_2 norm penalized terms and minimize this function in order to obtain the Maximum Penalized likelihood (MPL) estimates of the vector parameter θ :

$$\Lambda(\theta, c) = -\theta' \left[\sum_{j=1}^{t!} n_j x_j \right] + nK(\theta) + \lambda \left(\sum_{i=1}^t \theta_i^2 - c \right)$$

for some prescribed values of the constant c . When t is large (say $t \geq 10$), the exact computation of the normalizing constant $K(\theta)$ involves a summation of $t!$ items. McCullagh (1993)

suggested using the normalizing constant from the von Mises-Fisher distribution. Following on this suggestion, we approximate $K(\theta)$ with

$$K(\theta) \approx \frac{1}{t!} (2\pi)^{\frac{t}{2}} I_{\frac{t}{2}-1}(\|\theta\|) \|\theta\|^{-\frac{t}{2}+1}$$

where $\|\theta\|$ is the norm of θ and $I_\nu(z)$ is the modified Bessel function of the first kind given by

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(\nu+k+1)} \left(\frac{z}{2}\right)^{2k+\nu}$$

We may now find approximately the Maximum Penalized likelihood estimation for θ after ensuring that $\|x\| = 1$ in our model. There are several ways to minimize the target function $\Lambda(\theta, c)$. In this paper, we used three methods: the BFGS Quasi-Newton method (Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970), the Trust-Region-Reflective method Coleman and Li (1994, 1996) and the Interior Point Algorithm (Byrd, Hribar, and Nocedal 1999; Byrd, Gilbert, and Nocedal 2000; Waltz, Morales, Nocedal, and Orban 2006). As well, we imposed the constraint: $\theta \geq 0$ on θ in order to compare the results when there is no constraint. Differentiating with respect to θ we have

$$\frac{\partial \Lambda(\theta, c)}{\partial \theta} = - \sum_{j=1}^{t!} n_j x_j + 2\lambda \theta + \frac{n\theta}{\|\theta\|} \times \frac{\partial K(\theta)}{\partial \|\theta\|}$$

where

$$\frac{\partial K(\theta)}{\partial \|\theta\|} = \frac{1}{t!} (2\pi)^{\frac{t}{2}} \left[\|\theta\|^{1-\frac{t}{2}} \left(I_{\frac{t}{2}}(\|\theta\|) - \frac{1-\frac{t}{2}}{\|\theta\|} I_{\frac{t}{2}-1}(\|\theta\|) \right) + \left(1 - \frac{t}{2}\right) \|\theta\|^{-\frac{t}{2}} I_{\frac{t}{2}-1}(\|\theta\|) \right]$$

The critical points of the minimization occur at saddle points, rather than at local maxima (or minima). For the Quasi-Newton method, the target function is the magnitude of the gradient which is the square root of the sum of the squares of the partial derivatives instead of the $\Lambda(\theta, c)$ to solve the problem. In the applications that follow, the results from the three different methods yield the same solutions and the algorithms implemented in MATLAB converge very fast.

Following the estimation of θ , we proceeded to apply the basic bootstrap method in order to assess the distribution of θ . The basic idea of the bootstrap is to sample n rankings with replacement from the data. Then we find the MLE of each bootstrap sample. Repeating this procedure several times say 10^4 , leads to a distribution of θ . We can draw useful inference from the distribution θ and determine whether or not the θ'_j 's are significantly different from zero. Two sided confidence intervals can also be calculated. We illustrate this on real life data.

3. Using the Spearman score function

3.1. The Spearman score function and its meaning

In this section we restrict our attention to the Spearman score function defined as:

$$X(\omega_j) = \left(\omega_j(1) - \frac{t+1}{2}, \dots, \omega_j(t) - \frac{t+1}{2} \right), j = 1, \dots, t!$$

where $\omega_j(i)$ is the rank given to item i .

Rankings	(123)	(132)	(213)	(231)	(312)	(321)
Frequencies	1	1	1	5	7	12

Table 1: Sutton data on leisure preferences

Let T_S be the $t \times t!$ matrix of possible values of X : $T_S = (X(\pi_j))$. Taking $t = 3$ as an example, let ω_j be the possible rankings:

$$\omega_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \omega_2 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \omega_3 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \omega_4 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \omega_5 = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \omega_6 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

Then the matrix T_S becomes:

$$T_S = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \end{bmatrix}$$

Let $\hat{\pi}_i = \Pr(\text{observe } \omega_j) = n_j/n$ and let $\hat{\pi}$ be the column vector of $\hat{\pi}_i$. The first part of our likelihood function $\Lambda(\theta, c)$ becomes:

$$-n\theta' T_S \hat{\pi} = -\frac{n}{\|x\|} \times \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \begin{bmatrix} \pi_5 + \pi_6 - \pi_1 - \pi_2 \\ \pi_2 + \pi_4 - \pi_3 - \pi_5 \\ \pi_1 + \pi_3 - \pi_4 - \pi_6 \end{bmatrix}$$

We can notice that for θ_1 , $\pi_5 + \pi_6 = \Pr(\text{giving rank 3 to item 1})$ and $\pi_1 + \pi_2 = \Pr(\text{giving rank 1 to item 1})$. So here θ_1 weights the difference in probability between giving the top rank and the lowest rank to item 1 ($\Pr(\text{giving rank 3 to item 1}) - \Pr(\text{giving rank 1 to item 1})$). There is a similar interpretation for θ_2 and θ_3 . In general, the components of the matrix T_S actually focus on a special characteristic of the data, namely on the difference in weighted average of the rankings to the t objects. The weights here are $i - \frac{t+1}{2}$, $i = 1, \dots, t$. The θ_i 's here represent the coefficients attributed to each item.

Similarly, we can also compute the matrix of possible scores for $t = 4$. It can be seen that the first row element for $T_S \hat{\pi}$ is $-1.5 \Pr(\text{giving rank 1 to item 1}) - 0.5 \Pr(\text{giving rank 2}) + 0.5 \Pr(\text{giving rank 3}) + 1.5 \Pr(\text{giving rank 4})$. This is the weighted average of probability giving the high rank compare with the low rank to item 1. Notice that when t is odd, the weight of the middle item is 0 which means that our comparison is symmetric and balanced.

3.2. Application to real data

Sutton data

For the first example, we consider the Sutton data ($t = 3$). C. Sutton in her 1976 thesis on leisure preferences and attitudes on retirement of the elderly for 14 white and 13 black females in the age group 70- 79 years. Each individual was asked: with which sex do you wish to spend your leisure? Each female was asked to rank the three responses: male(s), female(s) or both, assigning rank 1 for the most desired and 3 for the least desired. The first item in the ranking corresponds to “male”, the second to “female” and the third to “both”. To illustrate the approach in the one sample case, we combined the data from the two groups as in Table 1.

We applied our penalized likelihood in this situation and the results are shown in Table 2.

To better illustrate our result, we rearrange our result (unconstrained θ , $c=1$) and data as Table 3. It can be seen that θ_1 is the largest coefficient and Item 1 (Male) shows the greatest difference between the number of judges choosing rank 1 or rank 3 which means that the

Choice of c	$\theta \geq 0$	θ_1	θ_2	θ_3	$\Lambda(\theta, c)$
c=0.5	No	0.53	-0.06	-0.47	50.00
	Yes	0.71	0.00	0.00	52.88
c=1	No	0.75	-0.09	-0.66	50.36
	Yes	1.00	0.00	0.00	54.44
c=2	No	1.06	-0.12	-0.93	54.62
	Yes	1.41	0.00	0.00	60.38
c=10	No	2.36	-0.28	-2.08	159.96
	Yes	3.16	0.00	0.00	172.84
MLE	No	0.60	-0.07	-0.53	49.90
	Yes	0.61	0.00	0.00	52.79

Table 2: Maximum Penalized likelihood Estimation for Sutton's data

Items	Number of judges	Action	Difference	θ	value
Item 1 (Male):	2	choose to rank 1	-17	θ_1	0.75
	19	choose to rank 3			
Item 2 (Female):	8	choose to rank 1	2	θ_2	-0.09
	6	choose to rank 3			
Item 3 (Both):	17	choose to rank 1	15	θ_3	-0.66
	2	choose to rank 3			

Table 3: The Sutton data and the estimation of θ

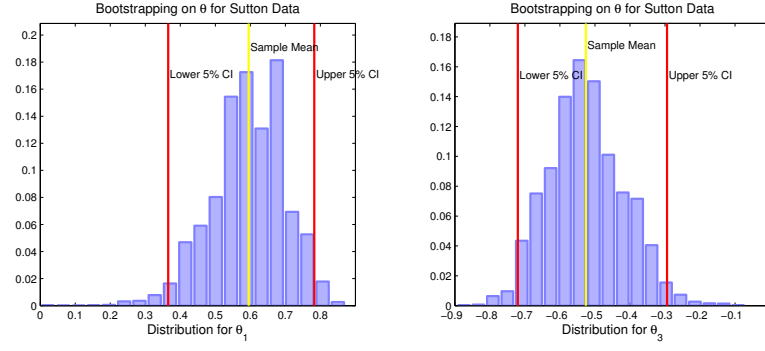
judges dislike spending leisure with male the most. For item 3 (Both), the greater value of negative θ_3 means judges prefer to spend leisure with both sex the most. θ_2 is close to zero and we deduce the judges show no strong preference on Female. This is consistent with the hypothesis that θ close to zero means randomness. To conclude, the results also show that θ_i weights the difference in probability giving the top rank and the lower rank to item i . Negative θ_i means the judges prefer item i more and positive θ_i means the judges are more likely to give a lower rank to item i .

Applying the bootstrap method on the Sutton data we plot the distribution of θ in Figure 1. The bootstrap sample size is 10^4 in this case. For $H_0: \theta_i = 0$, we can see that θ_1 and θ_3 are significantly different from 0 and θ_2 is not significantly different from zero. We can also see that the distribution of θ_1 and θ_2 is not completely bell shape. This is mainly because the sample size of the data is small. Using a traditional t-test method may be misleading in this case.

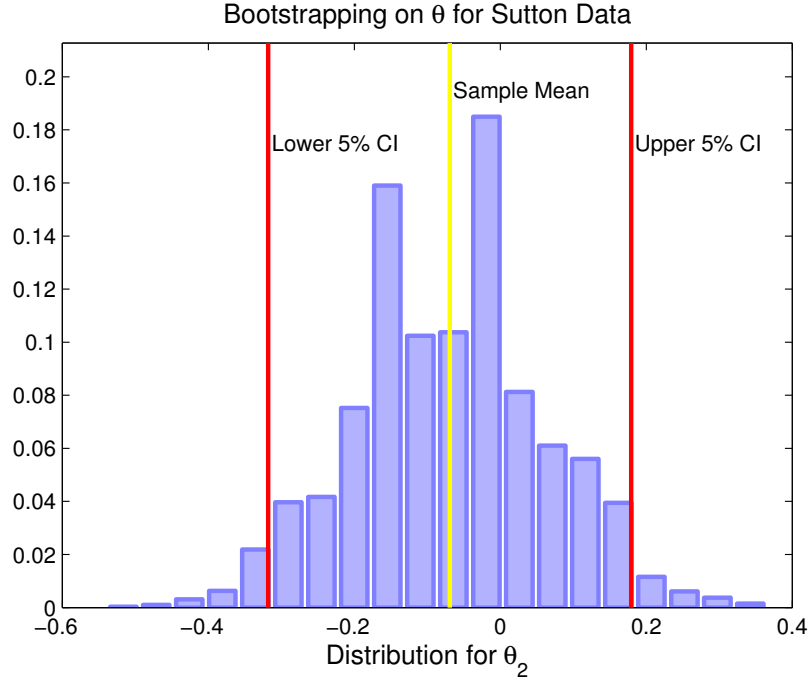
Song data

Our second example is the Song data ($t=5$) from Critchlow, Fligner, and Verducci (1991). Ninety-eight students were asked to rank 5 words, (1) score, (2) instrument, (3) solo, (4) benediction and (5) suit, according to the association with the word "song". Critchlow *et al.* (1991) reported that the average ranks for words (1) to (5) are 2.72, 2.27, 1.60, 3.71 and 4.69 respectively. However, the available data given in Critchlow *et al.* (1991) is in grouped format and the ranking of 15 students are unknown and hence discarded, resulting in 83 rankings, as shown in Table 4.

We applied the penalized likelihood in this situation and the results are shown in Table 5. We also rearranged the results (unconstrained θ , $c=1$) and data in Table 6. Note that for convenience we only show the number of judges who rank the top and the lowest but ranking 2 or 4 also is involves in determining the value of θ . From the results, we can see the value of θ successfully captures the properties of the data. θ_5 is the largest positive value and most



(a) Significantly different from zero



(b) Not significantly different from zero

Figure 1: The distribution of θ for Sutton data by bootstrap method

Rankings	Observed frequency
(32145)	19
(23145)	10
(13245)	9
(42135)	8
(12345)	7
(31245)	6
(32154)	6
(52134)	5
(21345)	4
(24135)	3
(41235)	2
(43125)	2
(52143)	2
others	0

Table 4: Song data set

Choice of c	$\theta \geq 0$?	θ_1	θ_2	θ_3	θ_4	θ_5	$A(\theta, c)$
c=0.5	No	-0.07	-0.22	-0.41	0.21	0.48	-30.15
	Yes	0.00	0.00	0.00	0.29	0.65	-17.81
c=1	No	-0.10	-0.31	-0.58	0.30	0.69	-49.60
	Yes	0.00	0.00	0.00	0.41	0.91	-32.15
c=2	No	-0.15	-0.44	-0.81	0.43	0.97	-76.45
	Yes	0.00	0.00	0.00	0.57	1.29	-51.76
c=10	No	-0.33	-0.98	-1.82	0.96	2.17	-176.13
	Yes	0.00	0.00	0.00	1.28	2.89	-120.94
MLE	No	-0.49	-1.46	-2.73	1.44	3.25	-220.60
	Yes	0.00	0.00	0.00	1.76	3.97	-141.12

Table 5: Maximum Penalized likelihood Estimation for Song's data

Items	number of judges	Action	Average ranks	θ	value
Item 1 (Score):	16	choose to rank 1	2.72	θ_1	-0.10
	7	choose to rank 5			
Item 2 (instrument):	10	choose to rank 1	2.27	θ_2	-0.31
	0	choose to rank 5			
Item 3 (solo):	55	choose to rank 1	1.60	θ_3	-0.58
	0	choose to rank 5			
Item 4 (benediction):	0	choose to rank 1	3.71	θ_4	0.30
	6	choose to rank 5			
Item 5 (suit):	0	choose to rank 1	4.69	θ_5	0.69
	70	choose to rank 5			

Table 6: The Song data and the estimation of θ

of the judges think word “suit” is not related to the word “song”. θ_3 is the largest negative value and 55 of the 83 judges think that word “solo” is the closest to the word “song”. From the results, we see that θ_i successfully weights the difference in probability giving the upper rank and the lower rank to item i .

We also applied our bootstrap method on the Song data and plotted the distribution of θ in Figure 2. The bootstrap sample size is 10^4 in this case. For $H_0: \theta_i = 0$, we can see that all the θ are significantly different from zero. As well, the distribution of θ are more bell shaped in view of the larger sample size.

Goldberg data

Our final example is due to Goldberg (1976) data ($t=10$). In the data, 143 graduates were asked to rank 10 occupations according to the degree of social prestige. These 10 occupations are: (i) Faculty member in an academic institution (Fac), (ii) Mechanical engineer (ME), (iii) Operation researcher (OR), (iv) Technician (Tech), (v) Section supervisor in a factory (Sup), (vi) Owner of a company employing more than 100 workers (Own), (vii) Factory foreman (For), (viii) Industrial engineer (IE), (ix) Manager of a production department employing more than 100 workers (Mgr) and (x) Applied scientist (Sci). The data are given in Cohen and Mallows (1980) and have been analyzed by many researchers.

Feigin and Cohen (1978) analyzed the Goldberg data and found three outliers due to the fact that the corresponding graduates wrongly presented rankings in reverse order. After reversing these 3 rankings, the average ranks received by the 10 occupations are 8.57, 4.90, 6.29, 1.90, 4.34, 8.13, 1.47, 6.27, 5.29, 7.85, with the convention that bigger rank means more prestige. Then the preference of graduates is in the order: $\text{Fac} \succ \text{Own} \succ \text{Sci} \succ \text{OR} \succ \text{IE} \succ \text{Mgr} \succ \text{ME} \succ \text{Sup} \succ \text{Tech} \succ \text{For}$.

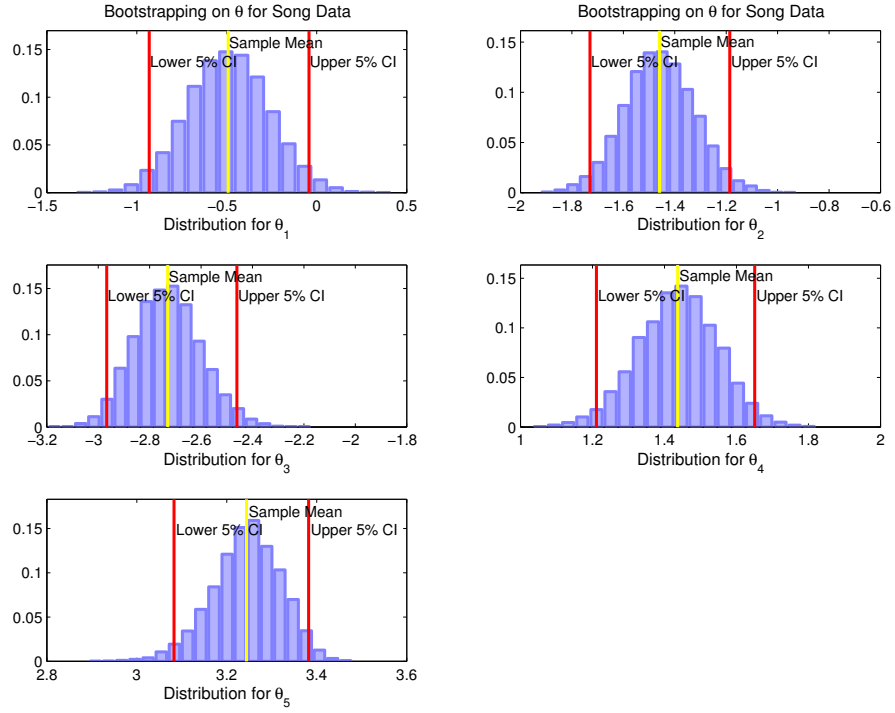


Figure 2: The distribution of θ for Song data by bootstrap method

Choice of c	$\theta \geq 0$	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	$\Lambda(\theta, c)$
c=0.5	No	0.30	-0.06	0.08	-0.35	-0.11	0.25	-0.39	0.07	-0.02	0.23	-81.89
	Yes	0.45	0.00	0.12	0.00	0.00	0.39	0.00	0.11	0.00	0.35	-53.50
c=1	No	0.42	-0.08	0.11	-0.49	-0.16	0.36	-0.55	0.11	-0.03	0.32	-115.80
	Yes	0.64	0.00	0.16	0.00	0.00	0.55	0.00	0.16	0.00	0.49	-75.66
c=2	No	0.59	-0.11	0.15	-0.69	-0.22	0.51	-0.78	0.15	-0.04	0.45	-163.77
	Yes	0.90	0.00	0.23	0.00	0.00	0.77	0.00	0.23	0.00	0.69	-107.01
c=10	No	1.32	-0.26	0.34	-1.55	-0.50	1.13	-1.73	0.33	-0.09	1.01	-366.21
	Yes	2.02	0.00	0.52	0.00	0.00	1.73	0.00	0.51	0.00	1.54	-239.27
MLE	No	8.74	-1.70	2.24	-10.25	-3.31	7.48	-11.46	2.20	-0.61	6.67	-2277.02
	Yes	13.03	0.00	3.34	0.00	0.00	11.17	0.00	3.28	0.00	9.95	-1447.59

Table 7: Maximum Penalized likelihood Estimation for Goldberg's data

We applied our penalized likelihood method and the results are shown in Table 7.

We also rearranged our results (unconstrained θ , $c=1$) and data in Table 8. For convenience we only show the number of judges who rank the top and the lowest but giving other rankings also involves in determining the value of θ . We can see that the results are consistent with the average ranks and our preference result from θ is the also consistent with the results from Feigin and Cohen (1978). For item 7 (Factory foreman), 93 of 143 graduates give rank 1 which means most of them think Factory foreman has the lowest social prestige and our θ_7 is the lowest (that bigger rank means more prestige). For $t=10$, it is almost impossible to calculate exact $K(\theta)$. Our good results of θ shows that our von Mises-Fisher distribution approximation of $K(\theta)$ actually works well and can be easily used when t is big.

The bootstrap distribution of θ is exhibited in Figure 3. The bootstrap sample size is 10^4 in this case. For $H_0: \theta_i = 0$, we see that all the θ except θ_9 are significantly different from zero. We can also see that the distribution of θ are all completely bell shaped because the sample size is large.

4. Pair Comparison using the Kendall score function

4.1. Kendall score function and its meaning

In this section, we re-consider the previous data sets through the lens of the Kendall score statistic. Specifically, for the Kendall score function, the $X(\omega_j)$ vector takes values $(t_K(\omega_j))_q$ where the q^{th} element is given by

$$(t_K(\omega_j))_q = \text{sgn}[\omega_j(m) - \omega_j(n)]$$

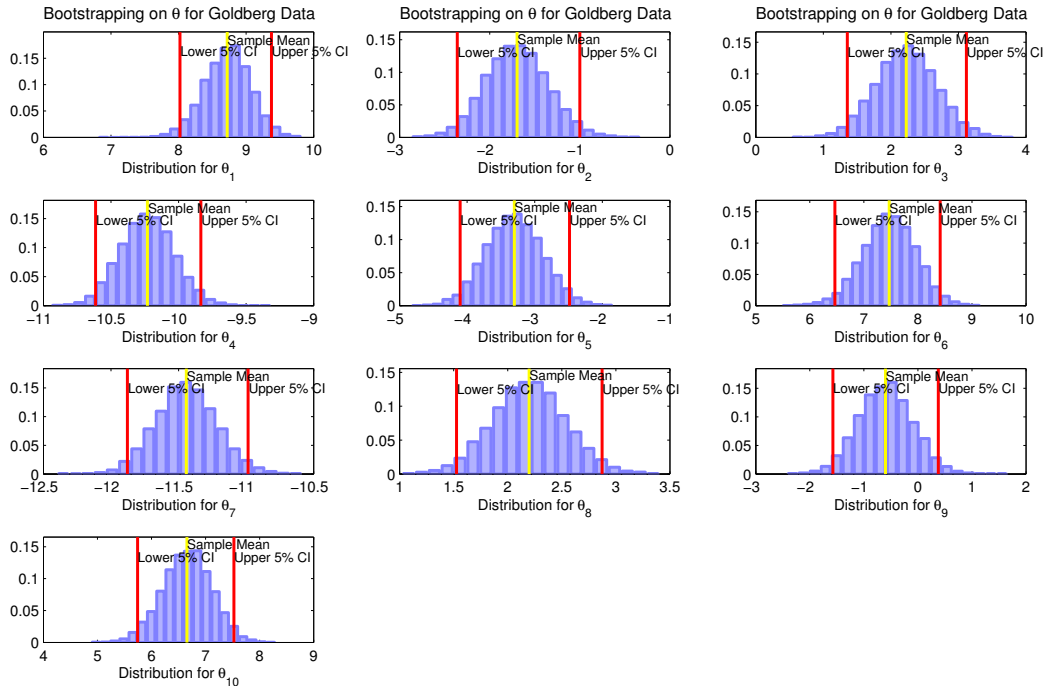
for $q = (n-1)(t - \frac{n}{2}) + (m-n)$, $1 \leq i < j \leq t$. The matrix of possible values of X becomes

$$\mathbf{T}_K = (t_K(\omega_1), \dots, t_K(\omega_{t!}))'$$

which is of dimension $\binom{t}{2} \times t!$. And θ_q from 1 to $\binom{t}{2}$ weight pair comparison between all possible pair m and n . In the case of the Kendall score function, the θ 's focus on the comparison between pairs of items ranked by the judges. As an example, consider once again the case $t = 3$. Then,

$$X(\omega) = \begin{pmatrix} \text{sgn}(\omega(2) - \omega(1)) \\ \text{sgn}(\omega(3) - \omega(1)) \\ \text{sgn}(\omega(3) - \omega(2)) \end{pmatrix}$$

Items	number of judges	Action	Average ranks	θ value
Item 1 (Fac)	0	choose to rank 1	8.57	0.42
	49	choose to rank 10		
Item 2 (ME)	0	choose to rank 1	4.90	-0.08
	0	choose to rank 10		
Item 3 (OR)	3	choose to rank 1	6.29	0.11
	1	choose to rank 10		
Item 4 (Tech)	45	choose to rank 1	1.90	-0.49
	0	choose to rank 10		
Item 5 (Sup)	0	choose to rank 1	4.34	-0.16
	0	choose to rank 10		
Item 6 (Own)	0	choose to rank 1	8.13	0.36
	54	choose to rank 10		
Item 7 (For)	93	choose to rank 1	1.47	-0.55
	0	choose to rank 10		
Item 8 (IE)	0	choose to rank 1	6.27	0.11
	1	choose to rank 10		
Item 9 (Mgr)	1	choose to rank 1	5.29	-0.03
	3	choose to rank 10		
Item 10 (Sci)	1	choose to rank 1	7.85	0.32
	35	choose to rank 10		

Table 8: The Goldberg data and the estimation of θ Figure 3: The distribution of θ for Goldberg data by bootstrap method

Pair Compare			choice of c				
item i	item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.35	-0.49	-0.70	-1.56	-0.60
1	3	θ_2	-0.56	-0.80	-1.13	-2.53	-0.97
2	3	θ_3	-0.24	-0.34	-0.48	-1.08	-0.41
$\Lambda(\theta, c)$			42.79	40.17	40.20	127.76	39.59

Table 9: MPLE using Kendall score function for Sutton data

item i	item j	number of judges	Pair comparison	θ
1	2	7	more prefer 1	-0.49
		20	more prefer 2	
1	3	3	more prefer 1	-0.80
		24	more prefer 3	
2	3	9	more prefer 2	-0.34
		18	more prefer 3	

Table 10: Pair comparison from The Sutton data and the estimation of θ

where θ_1 weights the comparison between items 1 and 2, θ_2 weights items 1, 3 and θ_3 items 2, 3. As well,

$$\mathbf{T}_K = \begin{pmatrix} 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

When $\theta_q < 0$ (for say items i and item j), it means that the judges prefer item j over item i ($\mu(i) > \mu(j)$). When θ_q is close to zero, the judges have no special preference between this pair. In the next section we apply the Kendall score function to the previous data sets.

4.2. Application to real data

Sutton data

For the first example, we consider the Sutton data ($t = 3$). We applied our penalized likelihood with Kendall score function in this situation and the results are shown in Table 9.

We rearrange the Sutton data focusing on pair comparison and our results ($c=1$) in Table 10. First, from our estimated θ , we can find that all θ'_i s are negative. This is consistent with our interpretation of θ . The judges strongly prefer Males to Both and Males to Females. They least prefer Females to Both. We can conclude that our θ' s well represent the paired comparisons among the judges.

Song data

As to the Song data ($t=5$) the results are shown in Table 11.

Since the Song data is about how much 5 words are close to the word “song”. We can summarize the results in Table 12. We note that θ_7 , θ_8 and θ_9 all have the same value 0.29. For these, all of the judges think item i is closer to item j. Once again, we conclude that θ well represents the paired preferences among judges.

Goldberg data

For the Goldberg (1976) data ($t=10$) we note that a large rank means more prestige, so our interpretation is reversed. We applied penalized likelihood with the Kendall score function in this situation and part of the results are shown in Table 13.

Pair Compare			choice of c				
item i	item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.09	-0.12	-0.17	-0.39	-0.76
1	3	θ_2	-0.15	-0.21	-0.30	-0.67	-1.30
1	4	θ_3	0.16	0.22	0.31	0.70	1.36
1	5	θ_4	0.24	0.34	0.48	1.07	2.09
2	3	θ_5	-0.16	-0.22	-0.31	-0.70	-1.36
2	4	θ_6	0.25	0.36	0.51	1.13	2.21
2	5	θ_7	0.29	0.41	0.57	1.28	2.51
3	4	θ_8	0.29	0.41	0.57	1.28	2.51
3	5	θ_9	0.29	0.41	0.57	1.28	2.51
4	5	θ_{10}	0.23	0.33	0.46	1.04	2.03
$\Lambda(\theta, c)$			-125.39	-184.29	-266.92	-602.04	-973.61

Table 11: MPLE using Kendall score function for Song data

Behavior of θ	Interpretation of θ
$\theta_1, \theta_2 < 0$	score \prec instrument, solo
$\theta_3 > 0$	score \succ benediction, suit
$\theta_5 < 0$	instrument \prec solo
$\theta_6, \theta_7 > 0$	instrument \succ benediction, suit
$\theta_8, \theta_9 > 0$	solo \succ benediction, suit
$\theta_{10} > 0$	benediction \succ suit

Table 12: Interpretation of θ in song data. If $A \succ B$, it means A is more preferred than B. In our case, it means A is closer to the word “Song”.

Pair Compare			choice of c				
item i	item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.13	-0.18	-0.25	-0.56	-4.09
1	3	θ_2	-0.09	-0.13	-0.18	-0.40	-2.89
1	4	θ_3	-0.15	-0.21	-0.29	-0.65	-4.76
1	5	θ_4	-0.12	-0.17	-0.25	-0.55	-4.02
1	6	θ_5	-0.02	-0.02	-0.03	-0.07	-0.50
1	7	θ_6	-0.15	-0.21	-0.29	-0.65	-4.76
1	8	θ_7	-0.11	-0.15	-0.21	-0.48	-3.49
1	9	θ_8	-0.12	-0.17	-0.24	-0.53	-3.89
1	10	θ_9	-0.03	-0.04	-0.05	-0.11	-0.83
2	3	θ_{10}	0.06	0.09	0.12	0.28	2.03
2	4	θ_{11}	-0.14	-0.20	-0.28	-0.63	-4.62
...
9	10	θ_{45}	0.09	0.13	0.19	0.42	3.09
$\Lambda(\theta, c)$			-490.79	-694.06	-981.55	-2194.80	-15177.07

Table 13: MPLE using Kendall score function for Goldberg data

Behavior of θ	Interpretation of θ
$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9 < 0$	Fac \succ ME, OR, <u>Tech</u> , Sup, Own, For, <u>IE</u> , Mgr, Sci
$\theta_{11}, \theta_{12}, \theta_{14} < 0$	ME \succ Tech, Sup, For
$\theta_{10}, \theta_{13}, \theta_{15}, \theta_{16}, \theta_{17} > 0$	ME \prec OR, Own, <u>IE</u> , Mgr, Sci
$\theta_{18}, \theta_{19}, \theta_{21}, \theta_{22}, \theta_{23} < 0$	OR \succ Tech, Sup, For, IE, Mgr
$\theta_{20}, \theta_{24} > 0$	OR \prec Own, Sci
$\theta_{27} < 0$	Tech \succ For
$\theta_{25}, \theta_{26}, \theta_{28}, \theta_{29}, \theta_{30} > 0$	Tech \succ Sup, <u>Own</u> , <u>IE</u> , Mgr, <u>Sci</u>
$\theta_{32} < 0$	Sup \succ For
$\theta_{31}, \theta_{33}, \theta_{34}, \theta_{35} > 0$	Sup \prec <u>Own</u> , IE, Mgr, Sci
$\theta_{36}, \theta_{37}, \theta_{38}, \theta_{39} < 0$	Own \succ For, IE, Mgr, Sci
$\theta_{40}, \theta_{41}, \theta_{42} > 0$	For \prec <u>IE</u> , <u>Mgr</u> , <u>Sci</u>
$\theta_{43} < 0$	IE \succ Mgr
$\theta_{42} > 0$	IE \prec Sci
$\theta_{45} > 0$	Mgr \prec Sci

Table 14: Interpretation of θ in Goldberg data. If $A \succ B$, it means A has more social prestige than B. The underline θ_i in left column means that its absolute value is bigger than 0.1. The underline occupation in right column means that the preference is very strong upon this comparison.

The Goldberg data compares occupations with more social prestige. We can also interpret the estimated θ to get the pair comparison results. We arrange the behavior and interpretation of θ in Table 12. In the table, the underlined θ_i in the left column means that its absolute value is larger than 0.1. The underlined occupation in the column on the right means that the preference is very strong. From the interpretation of θ , we find that the pair comparisons make sense. To conclude, the Kendall score function provides a nice way to look at the data by pair preference.

5. Extension to the two-sample ranking problem

5.1. Extension to two-sample ranking problems and its meanings

We may extend the approach to the two sample case. Let X_1, \dots, X_2 be two independent random variables under our model whose distributions are given by $p(\pi_1, \theta_1), p(\pi_2, \theta_2)$ respectively where :

$$\pi(\omega_{lj}; \theta_l) = \exp \{ \theta'_l x_j - K(\theta_l) \} \frac{1}{t!}, \quad j = 1, \dots, t!, \quad l = 1, 2$$

where $\theta_l = (\theta_{l1}, \dots, \theta_{lt})'$ represents the vector of parameters for population l and x_j is as in the one-sample case, a t -dimensional vector of scores. We shall use the Spearman scores here throughout the two-sample case. Set $\gamma = \theta_1 - \theta_2$ and write

$$\theta_l = \mu + b_l \gamma$$

for $l = 1, 2$ where

$$\mu = \frac{n_1 \theta_1 + n_2 \theta_2}{n_1 + n_2}, \quad b_1 = \frac{n_2}{n_1 + n_2}, \quad b_2 = -\frac{n_1}{n_1 + n_2}.$$

Then the density function can be written as:

$$\pi(\omega_{lj}, \theta_l) = \exp \{ \theta'_l x_j - K(\theta_l) \} \frac{1}{t!}, \quad j = 1, \dots, t!, \quad l = 1, 2$$

rankings	(123)	(132)	(213)	(231)	(312)	(321)
Frequencies for white females	0	0	1	0	7	6
Frequencies for black females	1	1	0	5	0	6

Table 15: Sutton data on leisure preferences (Two-sample problem)

Suppose that the observed vector of frequencies for the l^{th} population is

$$n_l = (n_{l1}, \dots, n_{lt!})'$$

The logarithm of the likelihood L as a function of (μ, γ) is proportional to

$$\log L(\mu, \gamma) \sim \sum_{l=1}^2 \sum_{j=1}^{t!} n_{lj} \{(\mu + b_l \gamma)' x_j - K(\theta_l)\}$$

We may now consider penalized likelihood to determine significant components of γ which most separate the populations. Hence, we consider minimizing with respect to the parameters μ and γ the function:

$$\Lambda(\mu, \gamma) = - \sum_{l=1}^2 (\mu + b_l \gamma) \sum_{j=1}^{t!} n_{lj} x_{lj} + \sum_{l=1}^2 n_l K(\mu + b_l \gamma) + \lambda \left(\sum_{i=1}^t \gamma_i^2 - c \right)$$

for some prescribed values of the constant c and λ . We may continue to use the normalizing constant from the von Mises-Fisher distribution to approximate $K(\theta)$. Differentiating we get

$$\begin{aligned} \frac{\partial \Lambda(\mu, \gamma)}{\partial \gamma} &= - \sum_{l=1}^2 b_l \sum_{j=1}^{t!} n_{lj} x_{lj} + 2\lambda \gamma + \sum_{l=1}^2 n_l b_l \times \frac{\partial K(\theta_l)}{\partial \theta_l} \\ \frac{\partial \Lambda(\mu, \gamma)}{\partial \mu} &= - \sum_{l=1}^2 \sum_{j=1}^{t!} n_{lj} x_{lj} + \sum_{l=1}^2 n_l \times \frac{\partial K(\theta_l)}{\partial \theta_l} \end{aligned}$$

where

$$\frac{\partial K(\theta)}{\partial \theta} = \frac{1}{t!} \frac{\theta}{\|\theta\|} (2\pi)^{\frac{t}{2}} \left[\|\theta\|^{1-\frac{t}{2}} \left(I_{\frac{t}{2}}(\|\theta\|) - \frac{1-\frac{t}{2}}{\|\theta\|} I_{\frac{t}{2}-1}(\|\theta\|) \right) + \left(1 - \frac{t}{2}\right) \|\theta\|^{-\frac{t}{2}} I_{\frac{t}{2}-1}(\|\theta\|) \right]$$

Here γ_i shows the difference between the two groups with respect to their preference on item i . A negative value of γ_i means that group 1 shows more preference for item i compared to population 2. A positive value of γ_i means that group 2 shows more preference on item i compared to population 1. For γ_i close to zero, there is no difference between the two groups on that item. As we shall see, this interpretation is consistent with the results in the real data applications. From the definition, we know that μ is the common part of θ_1 and θ_2 . More specifically, μ is the weight average of θ_1 and θ_2 taking into account the sample sizes of the populations.

5.2. Application to real data

Two-sample Sutton data

For the first example, we consider the Sutton data ($t = 3$) found in Table 15.

We applied penalized likelihood and the results are shown in Table 16.

We rearranged one of our estimation results ($c=1$) and the original data in Table 17. First, it is easy to see that μ is just like the θ 's in the one-sample problem. For example, μ_3 is

Choice of c	γ_1	γ_2	γ_3	μ_1	μ_2	μ_3	$A(\mu, \gamma)$
c=0.5	0.34	-0.57	0.24	0.59	-0.07	-0.52	46.88
c=1	0.48	-0.81	0.34	0.58	-0.06	-0.52	46.38
c=2	0.67	-1.15	0.48	0.57	-0.06	-0.51	46.46
c=10	1.50	-2.57	1.07	0.47	-0.04	-0.43	58.73
MLE	0.56	-0.95	0.40	0.58	-0.06	-0.52	46.30

Table 16: The Maximum Penalized likelihood Estimation for the two-sample Sutton data

Item:	No. white female	Diff.. for white	No. black female	Diff. for black	Sum	Action	γ	μ
Male	0	-13	2	-4	2	give rank 1	0.48	0.58
	13		6		19	give rank 3		
Female	8	8	0	-6	8	give rank 1	-0.81	-0.06
	0		6		6	give rank 3		
Both	6	5	11	10	17	give rank 1	0.34	-0.52
	1		1		2	give rank 3		

Table 17: The Two-Sample Sutton data and the estimation of μ, γ

the smallest value and the whole population prefers Item “Both” best. μ_3 is largest and the whole population mostly dislikes Item “Male”. This is not surprising since we know that μ is the common part of θ_1 and θ_2 . For the parameter γ , we first consider item “Female”. We see that white females prefer to spend leisure time with Females (8 assign rank 1) whereas black females do not (6 give rank 3). We find that γ_2 is negative and is largest in absolute value. There is a significant difference between the opinions with respect to item 2 “Female”. For item Male and Both, we find black females prefer them more than white females. To conclude, the results are consistent with the interpretation of μ and γ .

Game data

For the second example, we consider the Game data ($t = 6$) (Fok, Paap, and Van Dijk 2012). In this data, 91 Dutch students were asked to consider buying a new platform to play computer games. They had to rank 6 different platforms suitable to play computer games. The 6 platforms are the X-box (360), the PlayStation (2 or 3), the Gamecube (or Wii), the PlayStation Portable, the Gameboy or regular PC. In addition, we know the average number of hours that each student spends on gaming each week. We separate the students into two groups: frequent player (49 students) and seldom player (42 students). We classify a student as a frequent player if the average number of hours that he spends on gaming each week is larger than two. Otherwise he is classified as a seldom player

The game data is too large to exhibit here. Instead, we present the average ranks for the two groups in Table 18.

We applied penalized likelihood in this situation and the results are shown in Table 19.

To better illustrate the estimation results and the data, we calculate the weighted average of the observed probability of giving the high rank compared with the low rank for each

Average Rank	Xbox	PlayStation	PSPortable	GameCube	GameBoy	PC
Frequent player	2.59	2.59	3.94	4.69	5.04	2.14
Seldom player	2.86	3.17	3.64	3.88	4.24	3.21

Table 18: Average rank for Game data

Choice of c		Xbox	PlayStation	PSPortable	GameCube	GameBoy	PC	$\Lambda(\mu, \gamma)$
c=0.5	γ	0.00	-0.23	0.12	0.34	0.26	-0.50	-228.28
	μ	-2.66	-2.16	1.01	2.74	3.94	-2.88	
c=1	γ	0.09	-0.31	0.17	0.48	0.31	-0.74	-229.52
	μ	-2.67	-2.15	1.01	2.72	3.94	-2.84	
c=2	γ	0.27	-0.41	0.24	0.66	0.32	-1.08	-230.74
	μ	-2.70	-2.13	0.99	2.69	3.93	-2.78	
c=10	γ	1.24	-0.73	0.47	1.32	0.14	-2.44	-231.87
	μ	-2.84	-2.04	0.94	2.53	3.91	-2.50	
MLE	γ	0.96	-0.64	0.40	1.14	0.20	-2.06	-232.06
	μ	-2.81	-2.07	0.96	2.58	3.92	-2.58	

Table 19: The Maximum Penalized likelihood Estimation for Game data

	Weighted average of observed probability giving the high rank compare to the low rank				parameter	
Item:	Frequent players	Seldom players	Difference	Total populations	γ	μ
Xbox	-0.91	-0.64	-0.27	-0.79	0.09	-2.67
PlayStation	-0.91	-0.33	-0.57	-0.64	-0.31	-2.15
PSPortable	0.44	0.14	0.30	0.30	0.17	1.01
GameCube	1.19	0.38	0.81	0.82	0.48	2.72
GameBoy	1.54	0.74	0.80	1.17	0.31	3.94
PC	-1.36	-0.29	-1.07	-0.86	-0.74	-2.84

Table 20: Weighted average of observed probability and the estimation for the Game data

item. The weight for each rank is $i - \frac{t+1}{2}$, $i = 1, \dots, t$ (-2.5,-1.5,-0.5,0.5,1.5,2.5) here. The observations giving ranks close to the top and the bottom will get higher weight. A negative weighted average means preference in this case. We summary the results in Table 20. First, for the parameter μ , there is a strong relationship with the weighted average of the total population. For the total population, GameBoy receives the most low ranks and μ_5 is the largest, which is consistent with our interpretation of μ . Then, it can be found that the results for γ are consistent with the trend of the difference of weighted average between two samples. For example, for the item PC, frequent players exhibits the largest difference of opinion with seldom players and the γ_5 has the largest absolute value among γ . To conclude, our application on Game data also shows that our methodology is consistent with the data.

6. Conclusions

In this paper, we considered both the Spearman and Kendall score functions in the one or two sample problems as a way to summarize ranking data. A parametric model was formulated and then we applied penalized likelihood on the original parametric model to narrow down the items being ranked. We used the von Mises-Fisher distribution to approximate the normalizing constant and then determined the MLE estimates in several examples. This estimation procedure is fast and simple. In all cases, the estimation is shown to be consistent with the data. Our methodology was applied on various popular ranking data sets.

References

- Aslam JA, Montague M (2001). "Models for metasearch." In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 276–284. ACM.
- Baba Y (1986). "Graphical analysis of rank data." *Behaviormetrika*, **19**, 1–15.
- Broyden CG (1970). "The convergence of a class of double-rank minimization algorithms 1. general considerations." *IMA Journal of Applied Mathematics*, **6**(1), 76–90.
- Byrd RH, Gilbert JC, Nocedal J (2000). "A trust region method based on interior point techniques for nonlinear programming." *Mathematical Programming*, **89**(1), 149–185.
- Byrd RH, Hribar ME, Nocedal J (1999). "An interior point algorithm for large-scale nonlinear programming." *SIAM Journal on Optimization*, **9**(4), 877–900.
- Cohen A, Mallows C (1980). "Analysis of ranking data." *Technical memorandum*, AT&T Bell Laboratories, Murray Hill, N.J.
- Coleman T, Li Y (1994). "On the Convergence of Reflective Newton Methods for Large-scale Nonlinear Minimization Subject to Bounds vol. 67." *Ithaca, NY, USA: Cornell University*.
- Coleman TF, Li Y (1996). "An interior trust region approach for nonlinear minimization subject to bounds." *SIAM Journal on optimization*, **6**(2), 418–445.
- Critchlow DE, Fligner MA, Verducci JS (1991). "Probability models on rankings." *Journal of Mathematical Psychology*, **35**(3), 294–318.
- Croon MA (1989). "Latent class models for the analysis of rankings." *Advances in psychology*, **60**, 99–121.
- DeConde RP, Hawley S, Falcon S, Clegg N, Knudsen B, Etzioni R (2006). "Combining results of microarray experiments: a rank aggregation approach." *Statistical Applications in Genetics and Molecular Biology*, **5**(1).

- Deng K, Han S, Li KJ, Liu JS (2014). “Bayesian aggregation of order-based rank data.” *Journal of the American Statistical Association*, **109**(507), 1023–1039.
- Dwork C, Kumar R, Naor M, Sivakumar D (2001). “Rank aggregation methods for the web.” In *Proceedings of the 10th international conference on World Wide Web*, pp. 613–622. ACM.
- Feigin PD, Cohen A (1978). “On a model for concordance between judges.” *Journal of the Royal Statistical Society Series B*, **40**, 203–213.
- Fletcher R (1970). “A new approach to variable metric algorithms.” *The computer journal*, **13**(3), 317–322.
- Fok D, Paap R, Van Dijk B (2012). “A Rank-Ordered Logit Model With Unobserved Heterogeneity In Ranking Capabilities.” *Journal of applied econometrics*, **27**(5), 831–846.
- Goldfarb D (1970). “A family of variable-metric methods derived by variational means.” *Mathematics of computation*, **24**(109), 23–26.
- Kemeny JG, Snell JL (1962). *Mathematical models in the social sciences*, volume 9. Ginn New York.
- Lee PH, Philip L (2012). “Mixtures of weighted distance-based models for ranking data with applications in political studies.” *Computational Statistics & Data Analysis*, **56**(8), 2486–2500.
- Mallows CL (1957). “Non-null ranking models. I.” *Biometrika*, **44**, 114–130.
- McCullagh P (1993). “Models on spheres and models for permutations.” In MA Fligner, JS Verducci (eds.), *Probability Models and Statistical Analyses for Ranking Data*, pp. 278–283. Springer-Verlag.
- Neyman J (1937). “Smooth test for goodness of fit.” *Skandinavisk Aktuarietidskrift*, **20**, 149–199.
- Shanno DF (1970). “Conditioning of quasi-Newton methods for function minimization.” *Mathematics of computation*, **24**(111), 647–656.
- Waltz RA, Morales JL, Nocedal J, Orban D (2006). “An interior algorithm for nonlinear optimization that combines line search and trust region steps.” *Mathematical Programming*, **107**(3), 391–408.

Affiliation:

Mayer Alvo
 Department of Mathematics and Statistics
 University of Ottawa
 585 King Edward Ave, Ottawa, ON K1N 6N5, Canada
 E-mail: malvo@uOttawa.ca
 URL: <http://science.uottawa.ca/mathstat/en/people/alvo-mayer>