Bridging the gap: a parametric approach to a non parametric problem

Mayer Alvo

September 10, 2015

Department of Mathematics & Statistics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

Abstract

In the parametric setting in statistics, the notion of a likelihood function forms the basis for the development of tests of hypotheses and estimation of parameters. Tests in connection with the analysis of variance stem entirely from considerations of the likelihood function. On the other hand, the empirical likelihood method which is entirely data driven, presents an alternative to the parametric notion. In the present article, we define a likelihood function motivated by characteristics of the ranks of the data. Such a likelihood function can be fruitfully used in several problems in nonparametric statistics involving the use of ranks.

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

1 Introduction

Let X be a random $k-$ dimensional vector defined on the space of $t!$ permutations $\{\omega_{\mathrm{j}}\}$ of the integers $1, 2, ..., t$. Define the probability distribution of X as

$$\pi_j(\theta) = \exp\{\theta' x_j - K(\theta)\} \frac{1}{t!}, j = 1, ..., t!$$

where $\theta = (\theta_1, ..., \theta_k)'$ is a k-dimensional vector of parameters, $K(\theta)$ is a normalizing constant, and $X(\omega_j) = x_j$. When $\theta = 0$, $\pi_j(\theta) = \frac{1}{t!}$ and X has a uniform distribution. The hypothesis of uniformity can be tested using a traditional χ^2 goodness of fit approach. However, the fact that $t!$ is large even for moderate values of t diminishes the power of this test since many of the cells would be sparse. Alvo (2015) considered the Rao score test of the

$$H_0(\theta_0) : \theta = \theta_0 \ vs \ H_1(\theta_0) : \theta \neq \theta_0$$

The likelihood function for a random sample of n observations derives from the multinomial distribution and is proportional to

$$L(\theta) \quad \sim \quad \pi_1^{n_1}(\theta)\, \pi_2^{n_2}(\theta) ... \pi_{t!}^{n_{t!}}(\theta)$$

where n_j represents the observed frequency of occurrence of the ranking ω_j. The log likelihood is proportional to

$$log\, L(\theta) \quad \sim \quad n[\theta'\hat{\eta} - K(\theta)]$$

where the sample estimate of the mean $K'(\theta) = \Sigma_{j=1}^{t!} x_j \pi_j(\theta)$ is given by

$$\hat{\eta} = \left[\sum_{j=1}^{t!} x_j \hat{\pi}_n\right], \hat{\pi}_n = (n_j/n)$$

which also represents the usual sufficient statistic. Letting

$$U(\theta) = \left(\frac{\partial log\, L(\theta)}{\partial \theta_r}\right)$$

the Rao score test rejects the null hypothesis for large values of

$$S_k = [U(\theta_0)]' [I(\theta_0)]^{-1} [U(\theta_0)]$$

where

$$I(\theta) = \left(-E_\theta \frac{\partial^2 log\, L(\theta)}{\partial \theta_r \partial \theta_s}\right)$$

represents the Fisher information matrix. For large n,

$$S_k \Rightarrow_{\mathcal{L}} \chi_f^2$$

where $f = rank\ I\left(\theta_0\right)$. Alvo (2015) has shown that when $\theta_0 = 0$ and X is the t-dimensional random vector of adjusted ranks and

$$X\left(\omega_j\right) = \left(\omega_j\left(1\right) - \frac{t+1}{2}, ..., \omega_j\left(t\right) - \frac{t+1}{2}\right)', \ j = 1, ..., t!$$

$$S_k = \frac{12n}{t\left(t+1\right)} \sum_{i=0}^{t} \left(\bar{R}_i - \frac{\left(t+1\right)}{2}\right)^2$$

the usual Friedman test statistic which has asymptotically a χ^2_{t-1} distribution under the null hypothesis.

Alvo (2015) proposed penalized likelihood as a way to focus on only a small number of $\theta's$. We describe this in the next section for the Spearman score function above. The rest of the paper is organized as follows. In section 2, we apply the method to a data set using the Spearman score. In section 3 we consider the Kendall score function. In section 4, we consider the two sample problem. We conclude with some remarks in section 5.

2 Penalized likelihood estimation

Define the penalized likelihood in order to narrow down the number of items to look at. The idea is to minimize the negative likelihood function subject to a constraint on the parameters as follows:

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

for some prescribed values of the constant c. When t is large (say $t \geq 10$), the computation of exact $K(\theta)$ involves a summation of $t!$ items. Instead, we approximate this constant by following a suggestion of McCullagh (1993):

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

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

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

We proceed to find the Maximum Penalized likelihood estimation for θ after ensuring that $\|x_j\| = 1$ in our model using algorithms implemented in MATLAB that converge very fast. Following the estimation of θ, we 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 maximum likelihood estimate of each bootstrap sample. Repeating this procedure $10,000$ times leads to a distribution of θ. We

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

Table 1: Combined data on leisure preferences

c	θ_1	θ_2	θ_3	$\Lambda(\theta, c)$
0.5	0.53	-0.06	-0.47	50.00
1	0.75	-0.09	-0.66	50.36
2	1.06	-0.12	-0.93	54.62

Table 2: Penalized likelihood for the combined data

can draw useful inference from the distribution θ and in particular construct a two-sided confidence interval. We applied this to a data set with $t = 3$. In order to interpret the meaning of the coefficients, let $\hat{\pi}_i = \Pr(observing\ \omega_j) = n_j/n$ and let $\hat{\pi}$ be the column vector of $\hat{\pi}_i$. The first part of the function $\Lambda(\theta, c)$ becomes:

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

where T_S is the $t \times t!$ matrix of possible values of X. We note that for θ_1, $\pi_5 + \pi_6 = \Pr(givingrank3toitem1)$ and $\pi_1 + \pi_2 = \Pr(givingrank1toitem1)$. So here θ_1 weights the difference in probability giving the top rank and the lowest rank to item 1 $(\Pr(givingrank3toitem1) - \Pr(giving\ rank\ 1\ to\ item\ 1))$. Similarly for θ_2 and θ_3.

We can also illustrate the matrix of possible scores for $t = 4$. The first row element for $T_S\hat{\pi}$ for item 1 is

$$-1.5\Pr(giving\ rank\ 1) - 0.5\Pr(giving\ rank\ 2) + 0.5\Pr(giving\ rank\ 3) + 1.5\Pr(giving\ rank\ 4)$$

which is a weighted average of the probabilities of assigning the high ranks compared with the low ranks. The weight here is $i - \frac{t+1}{2}$, $i = 1, ..., t$. Note for t is odd, the weight of the middle item is 0 making the comparison symmetric. A similar interpretation can be made for θ_2 and θ_3

C. Sutton considered in her 1976 thesis, the 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

Item	Number of judges	Action	Difference	θ	Λ
Male	2	assign rank 1	-17	θ_1	**0.75**
	19	assign rank 3			
Female	8	assign rank 1	2	θ_2	-0.09
	6	assign rank 3			
Both	17	assign rank 1	15	θ_3	**-0.66**
	2	assign rank 3			

Table 3: The combined data re expressed

greatest difference between the number of judges choosing rank 1 or rank 3 which means that the 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.

We plot the bootstrap distribution of θ in Figure 1. For H_0: $\theta_i = 0$, we see that θ_1 and $\theta_3$are significantly different from 0 whereas θ_2 is not. We also see that the bootstrap distributions are not entirely bell shaped leading us to conclude that a traditional t-test method may not be appropriate in this case.

3 Using the Kendall score function

Suppose now that the random vector X takes values$\left(t_K\left(\mu\right)\right)_q$where the q^{th} element is given by

$$\left(t_K\left(\mu\right)\right)_q = sgn\left[\mu\left(j\right) - \mu\left(i\right)\right]$$

for $q = \left(i-1\right)\left(t - \frac{i}{2}\right) + \left(j-i\right), 1 \leq i < j \leq t$. This is the Kendall score function whose matrix of possible values of becomes the $\binom{t}{2} \times t!$ matrix

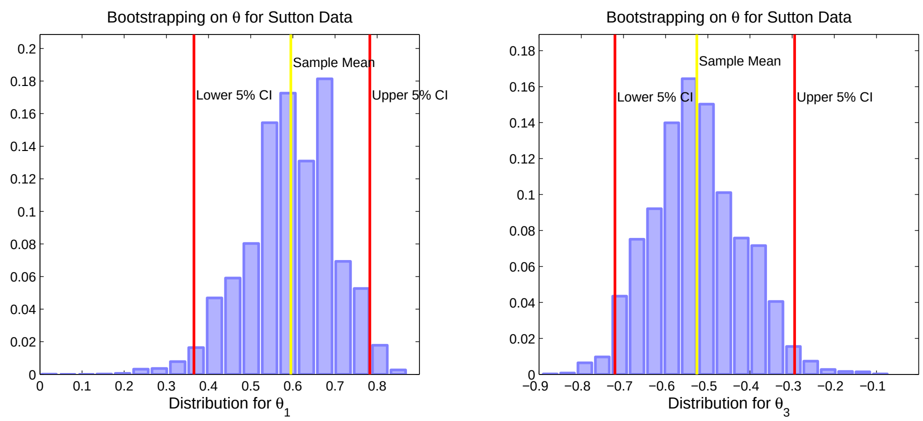
$$\mathbf{T_K} = \left(t_K\left(\mu_1\right), ..., t_K\left(\mu_{t!}\right)\right)'$$

The covariance matrix when $\theta = 0$ is given by

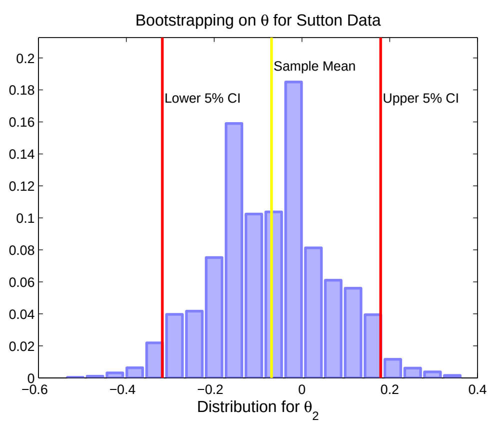
$$Cov_0\left(X\right) = \frac{1}{t!}\mathbf{T_K}\mathbf{T_K'}$$

whose entries $A\left(s, s', t, t'\right) = \frac{1}{t!}\Sigma_\nu sgn\left(\nu\left(s\right) - \nu\left(t\right)\right) sgn\left(\nu\left(s'\right) - \nu\left(t'\right)\right)$ are, by Lemma 4.1, p.58 in Alvo and Yu (2014), given by

$$A\left(s, s', t, t'\right) = \begin{cases} 0 & s \neq s', t \neq t' \\ 1 & s = s', t = t' \\ \frac{1}{3} & s = s', t \neq t' \\ -\frac{1}{3} & s = t', s' \neq t \end{cases}.$$



(a) Significantly different from zero



(b) Not significantly different from zero

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

Pair Compare			choice of c				
item i	item j	θ	c=0.5	c=1	c=2	c=10	no constraint
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 4: Penalized likelihood using the Kendall score function for the Sutton data

Moreover, the eigenvalues of $Cov_0\,(X)$ are $\frac{1}{3}, \frac{t+1}{3}$ with multiplicities $\binom{t-1}{2}, (t-1)$ respectively. The choice of $\mathbf{T_K}$ leads to the test statistic

$$n\left(\mathbf{T_K}\bar{\pi}_n\right)'\left(\mathbf{T_K}\mathbf{T_K'}\right)^{-1}\left(\mathbf{T_K}\bar{\pi}_n\right) \Rightarrow \chi^2_{\binom{t}{2}}$$

The inverse matrix which can be readily computed even for values of $t = 10$, has entries of the form

$$A\left(s, s', t, t'\right) = a\begin{cases} 0 & s \neq s', t \neq t' \\ b & s = s', t = t' \\ -1 & s = s', t \neq t' \\ 1 & s = t', s' \neq t \end{cases}$$

for constants $a = t - 1, b = \frac{3}{t+1}$. As an example, consider the case $t = 3$. Then,

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

and θ_1 weights the comparison between item 1 and item 2, θ_2 is between item 1, 3 and θ_3 is between item 2, 3. As well,

$$\mathbf{T_K}\mathbf{T_K'} = \begin{pmatrix} 6 & 2 & -2 \\ 2 & 6 & 2 \\ -2 & 2 & 6 \end{pmatrix}$$

and

$$\left(\mathbf{T_K}\mathbf{T_K'}\right)^{-\mathbf{1}} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{1}{8} & \frac{1}{4} & -\frac{1}{8} \\ \frac{1}{8} & -\frac{1}{8} & \frac{1}{4} \end{pmatrix}$$

When $\theta_q < 0$ (θ_q weights the comparison between item 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.

We consider once again the Sutton data ($t = 3$) and apply penalized likelihood. The results are shown in Table 4.

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 5: Pair comparison for The Sutton data and the estimation of θ

We rearrange the Sutton data focusing on paired comparison and the results (c=1) are displayed in Table 5. First, we note that all the $\theta_i's$ are negative. This is consistent with our interpretations. The judges show a strong preference for Males to Both and Males to Females. They least prefer Females to Both. We can conclude that the $\theta_i's$ represent well the paired preferences among the judges.

4 The two-sample ranking problem

We may extend the one sample problem to the two sample case. We shall use the Spearman scores throughout the two-sample case. Let $X_1, .X_2$ be two independent random vectors whose distributions are given by $\pi(\theta_1), \pi(\theta_2)$ respectively where

$$\pi_j(\theta_l) = \exp\{\theta_l' x_j - K(\theta_l)\} p_j, j = 1, ..., t!, l = 1, 2$$

and $\theta_l = (\theta_{l1}, ..., \theta_{lt})'$ represents the vector of parameters for population l. It follows that the covariances are for $l = 1, 2$,

$$Cov(X_l) = \mathbf{T_S} \Sigma_l \mathbf{T_S}'$$

where

$$\Sigma_l = diag(\pi_j(\theta_l)) - [\pi(\theta_l)][\pi(\theta_l)]'$$

As in Alvo (2015) let $\gamma = \theta_1 - \theta_2$ and write

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

for $l = 1, 2$ where

$$\mu = \frac{n_1\theta_1 + n_2\theta_2}{n_1 + n_2}, b_1 = \frac{n_{2,}}{n_1 + n_2}, b_2 = -\frac{n_1}{n_1 + n_2}.$$

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

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

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 6: Sutton data on leisure preferences

where $\{n_{lj}\}$ represent the observed vector of frequencies for X_l.Writing

$$\pi(\theta_l) = (\pi(\theta_{l1}), ..., \pi(\theta_{lk}))'$$

and $\hat{\pi}_l = \left(\frac{n_{lj}}{n_l}\right)$, Alvo (2015) showed that the Rao score vector evaluated under the null hypothesis $H_0 : \theta_1 = \theta_2$ is given by

$$\left(\frac{\partial \log L(\mu, \gamma)}{\partial \gamma_r}\right) = \frac{n_1 n_2}{n_1 + n_2} [\mathbf{T_S}\hat{\pi}_1 - \mathbf{T_S}\hat{\pi}_2]$$

and the test statistic is

$$n [\mathbf{T_S}\hat{\pi}_1 - \mathbf{T_S}\hat{\pi}_2]' \hat{D} [\mathbf{T_S}\hat{\pi}_1 - \mathbf{T_S}\hat{\pi}_2] \Rightarrow \chi^2_f$$

where $\hat{D}$ is the Moore-Penrose inverse of $\mathbf{T_S}\hat{\Sigma}\mathbf{T_S}'$ whenever $n_l/n \to \lambda_l > 0$ as $n \to \infty$, where $n = n_1 + n_2$.Here $\hat{\Sigma}$ is a consistent estimator of $\Sigma = \frac{\Sigma_1}{\lambda_1} + \frac{\Sigma_2}{\lambda_2}$.

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

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

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 $K(\theta)$.

Here γ_i shows the difference between the two population's preference on item i. A negative γ_i means that population 1 shows more preference on item i compared to population 2. A positive γ_i means that population 2 shows more preference on item i compare to population 1. For γ_i close to zero, there is no difference between the two populations on that item. As we shall see, this interpretation is consistent with the results in the real data applications. From the definition of μ, 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.

As an application consider the Sutton data ($t = 3$) found in Table 6.

We applied penalized likelihood in this situation and the results are shown in Table 7.

Rearranging the results for c=1 we have the original data in Table 8. First, it is seen that μ is just like the θ's in the one-sample problem. For example, μ_3 is the smallest value and the whole population prefers Item Both best. μ_3 is the largest and the whole population mostly dislikes Item Male. This is not surprising since we know that μ is the

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

Table 7: Penalized likelihood results for the Sutton data

<table>
<tr><th>Item:</th><th>#white fe- males</th><th>#black fe- males</th><th>Sum</th><th>Action</th><th>γ</th><th>μ</th></tr>
<tr><td rowspan="2">Male</td><td>0</td><td>2</td><td>2</td><td>give rank 1</td><td rowspan="2">0.48</td><td rowspan="2">0.58</td></tr>
<tr><td>13</td><td>6</td><td>19</td><td>give rank 3</td></tr>
<tr><td rowspan="2">Female</td><td>8</td><td>0</td><td>8</td><td>give rank 1</td><td rowspan="2">-0.81</td><td rowspan="2">-0.06</td></tr>
<tr><td>0</td><td>6</td><td>6</td><td>give rank 3</td></tr>
<tr><td rowspan="2">Both</td><td>6</td><td>11</td><td>17</td><td>give rank 1</td><td rowspan="2">0.34</td><td rowspan="2">-0.52</td></tr>
<tr><td>1</td><td>1</td><td>2</td><td>give rank 3</td></tr>
</table>

Table 8: The Sutton data and the estimation of μ, γ

common part of θ_1 and θ_2. For the parameter γ, we 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 in 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.

5 Conclusions

In this article, we implemented a penalized likelihood method on ranking data following the parametric reformulation introduced by Alvo (2015). We considered both the Spearman and the Kendall score functions in one and two sample ranking problems and applied the methodology on a small data set. The methodology is readily applicable to large data sets.

References

Alvo, M.: 2015, Bridging the gap: a likelihood: function approach for the analysis of ranking data, *Communications in Statistics - Theory and Methods* .

Alvo, M. and Yu, P.: 2014, *Data Analysis Based on Ranks*, Springer.

McCullagh, P.: 1993, Models on spheres and models for permutations, *in* M. A. Fligner and J. S. Verducci (eds), *Probability Models and Statistical Analyses for Ranking Data*, Springer-Verlag, pp. 278–283.