Statistica Sinica 1(1991), 301-308

THE COVARIANCE OF RANK SCORES IN ORDER-STATISTICS MODELS

R. J. Henery

University of Strathclyde

Abstract: Order-statistics models may be used to estimate rank covariances, and the choice between the various models is determined by general considerations such as maximum entropy and/or the symmetry of the ranking process. In practical cases these covariances are fairly close to the sample covariances, and may be used in preference to sample covariances when the sample sizes are small (of order 10 to 20).

Key words and phrases: Covariance matrix, logistic models, mean ranks, order-statistics models, rank preferences.

1. Introduction

To specify the general distribution of rank preferences of n objects, we would clearly require $n! - 1$ probabilities. Usually, however, we are given the rank preferences from only N judges, where N may even be smaller than $n! - 1$ and the problem is how to make reasonable inferences using a much smaller number of parameters. Linhart (1960), Koch, Freeman and Lehnen (1976), and Hollander and Sethuraman (1978) all suggest using the n means and $n(n+1)/2$ covariances in various ways. In practical cases, even this may involve too many parameters, and a popular alternative is to use the so-called order-statistics models which require only $n - 1$ parameters. Thus Pettitt (1982) uses a normal order statistics model for the permutation probabilities, but due to the computational complexity of the maximum likelihood equations he makes a further numerical approximation to the rank order probabilities. Moran (1950) also considers the normal order statistics model, in particular the trend case in which the parameters θ_i are proportional to i. Moran also gives results for the covariance of the ranks in this case, but because of the numerical labour involved does not take the method any further.

Mak (1985) indicates that the order-statistics parameters could be estimated using only the average ranks of the objects. Although it is clearly not possible to consider questions about the goodness of fit of the model if we are given only the average ranks, it is possible to give some idea of the variance of estimates of the parameters and even to compare estimates from the average ranks of different samples if we know the covariance matrix of the ranks. In this paper we present the general formula for the covariance of the ranks and base most of our remarks on the logistic order statistics model for which the covariances may be calculated without recourse to numerical integration. This theoretical covariance matrix can then be used in place of the sample covariance matrix when we are dealing with very small samples and further analysis can then proceed along the lines of Hollander and Sethuraman (1978) for example. For an instance of inference using only average ranks we use some data of Snell (1983), and to give some idea of how closely the logistic order-statistics models predict the sample covariance matrix in practical examples, we look at the data of Hollander and Sethuraman (1978). We shall conclude that such differences as appear are only apparent for very large samples and that the logistic model may be used with confidence for small samples provided that the ranking process has a suitably sequential structure.

2. Order-Statistics Models

In order-statistics models, we assume that there is an underlying random variable Z for each object which can be thought of as a measure of preference for a particular object. However the value of Z is never observed directly — only the relative order of the objects. Usually it is assumed that the distributions of the Z's differ in location only, in which case we write $Z_i = Y_i + \theta_i$, where θ_i is the location parameter and the Y's are independent identically distributed random variables.

Since the variables Z are never observed, it is not clear how best to specify their distribution. Indeed since any monotonic transformation of Z would preserve ranks, there is considerable arbitrariness in the choice of Z. Within the class of order-statistics models therefore, Henery (1986) suggests using maximum entropy principles in the search for an optimal order-statistics model for the average ranks problem, and, independently, Joe (1988) makes the same suggestion for average scores in tournaments. In particular, the principle of maximum entropy leads to the Plackett (1975) type I model for rank probabilities and by implication to an order-statistics model with an extreme-value distribution for the Z_i's. Frequently the Z's are assumed to have a normal distribution, but for most purposes the main distinguishing feature between the normal and logistic models is the symmetry or otherwise of the rank variance as a function of the mean rank.

3. Variance of Average Rank Scores

Without loss of generality, consider an order-statistics model, with the Z's now not necessarily independent nor identically distributed. To rank n objects, for example, the Z's might be chosen without replacement from the integers $1, \ldots, n$. We will write $P(ijk)$ for the probability that $Z_i < Z_j < Z_k$, and write ξ_i for the rank of Z_i. We give the highest rank to the highest Z. In the context of a race among n athletes this would correspond to giving the winner (with the lowest time) a rank of one. In other circumstances it might be more appropriate for the winner to be given the highest rank, for example in long jump competitions. Then the rank of Z_i is

$$\xi_i = 1 + \sum_{j \neq i} H(Z_i - Z_j),$$

where $H(x) = 1$ if $x > 0$, and $H(x) = 0$ if $x < 0$. Taking expectations of both sides of the above equation we find

$$E(\xi_i) = 1 + \sum_{j \neq i} P(ji). \tag{3.1}$$

Moran (1950) and Daniels (1950) consider special cases of the variance of the ranks when the variables Z_i have a normal distribution. By extending the argument of Moran (1950), we find for the moment of order m:

$$E(\xi_i^m) = 1 + \alpha_1 \sum_{j \neq i} P(ji) + \alpha_2 \sum_{j \neq k \neq i} P(jki) + \ldots + \alpha_m \sum_{j_1 \neq \ldots \neq j_m \neq i} P(j_1 \ldots j_m i). \tag{3.2}$$

Using combinatorial arguments or by special cases, the coefficients $\alpha_1, \ldots, \alpha_m$ are easily shown to be $\alpha_1 = 2^m - 1$; $\alpha_2 = 3^m - 2 \cdot 2^m + 1$; $\alpha_3 = 4^m - 3 \cdot 3^m + 3 \cdot 2^m - 1$; etc.

The covariance of ξ_i and ξ_j, $i \neq j$, can similarly be found:

$$\mathrm{Cov}(\xi_i, \xi_j) = 1 + P(ij)P(ji) - P(ij)E(\xi_i) - P(ji)E(\xi_j) + \sum_{k \neq i,j} \{2P(kij) + 2P(kji) - P(ki)P(kj)\}. \tag{3.3}$$

We shall write W for the covariance matrix of the ranks, i.e. $W_{ij} = \mathrm{Cov}(\xi_i, \xi_j)$. W has rank $n - 1$ in general, since the sum of the ranks is constrained to be $n(n+1)/2$.

Apart from the obvious difficulty that the sums in (3.3) may involve rather many terms, there is the added difficulty that, in general, the $P(ijk)$ term involves a numerical integration. Thus for a normal order statistics model $P(ijk)$

is a bivariate normal integral, which can be simplified to a univariate integral by the reduction formula of Plackett (1954). Fortunately, however, the evaluation of $P(ij)$ and $P(ijk)$ is particularly easy for the logistic model, since this is just Plackett's (1975) Type I model. Writing $F(x) = 1/(1+\exp(-x))$ for the distribution function for the logistic distribution, we find

$$P(ij) = F(\theta_j - \theta_i) = \frac{p_i}{p_i + p_j}$$

$$P(ijk) = \frac{p_i}{p_i + p_j + p_k} \cdot \frac{p_j}{p_j + p_k},$$

where $p_i = \exp(-\theta_i)$.

The mean rank $E(\xi_i)$ can then be written simply in terms of the p_i:

$$E(\xi_i) = 1 + \sum_{j \neq i} \frac{p_j}{p_j + p_i}. \tag{3.4}$$

In practice the sample mean ranks $\bar{\xi}_i$ are used in (3.4) in place of $E(\xi_i)$; the requisite p_i can be found by an iterative procedure. An appropriate starting point might be $p_i = \exp(-\theta_i^0)$ where

$$\theta_i^0 = 4n^{-1}(\bar{\xi}_i - (n+1)/2).$$

In terms of the Bradley-Terry model, we might allocate preferences p_i (for being first) to the n objects. The full ranking is then decided sequentially in the order $1, 2, \ldots, n$. Object i is ranked first with probability $p_i/(p_1 + p_2 + \ldots + p_n)$; once rank 1 has been decided, the remaining objects are independently ranked among themselves and the highest ranked of the remainder is given rank 2, and so on. We shall refer to this model, Plackett's Type I model, as the forward logistic model. The backward model, which has the same expected rank scores, is obtained by ranking the objects sequentially from last to first, with preferences for being last of $1/p_i$. Given only the mean ranks there is no way to distinguish between these two models. However their covariances are different and this may enable us to choose between the two.

When the distribution of the underlying variables Z is symmetric, it is immediately obvious that the variance-covariance matrix is unchanged if the parameters θ_i are reversed. This is most easily seen by observing that $Z_i < Z_j < \ldots < Z_k$ implies that $-Z_k < \ldots < -Z_j < -Z_i$, and, using the symmetry of Z, that this is the order statistics model with parameters $-\theta_k, \ldots, -\theta_j, -\theta_i$ and with the ranks reversed. Note that the covariance matrix of $n + 1 - \xi_i$ is the same as that for ξ_i itself. Thus the basic structure of the model is unchanged if we reverse the ranks of symmetric models, such as the normal for example. A simple way to obtain a symmetric covariance matrix, which has the benefit of avoiding numerical integration, is by averaging the covariances of the forward and backward logistic models.

4. Comparing the Mean Ranks from Two Samples

Given the mean rank vectors X_1 and X_2 from two independent samples of rank vectors, of sizes N_1 and N_2 respectively, from the same population with probabilities defined by the order statistics model with parameters θ_i, we first calculate the pooled mean rank vector

$$\bar{\xi} = (N_1 X_1 + N_2 X_2)/(N_1 + N_2),$$

then estimate the parameters θ_i using $\bar{\xi}$ in place of $E(\xi)$ in (3.4). Mak (1985) observes that there is always a unique solution to these equations (only $n - 1$ of which are independent since the mean ranks must sum to $n(n+1)/2$). The covariance matrix of the rank vector is then found from (3.2) and (3.3). We can now compare the two mean rank vectors by calculating the statistic Q:

$$Q = \frac{N_1 N_2}{N_1 + N_2}(X_1 - X_2)' \cdot W^{-1} \cdot (X_1 - X_2) \tag{4.1}$$

which is distributed asymptotically as a $\chi^2(n-1)$ random variable. Since W is not of full rank it is most convenient to take the first $(n-1)$ mean ranks for the mean vectors and the corresponding covariance of order $(n-1) \times (n-1)$. This procedure is identical to that of Hollander and Sethuraman (1978) except that (i) we use the theoretical covariance matrix W in place of the sample covariance matrix C; and (ii) we are assuming as null hypothesis that the samples come from an order statistics model with the same parameters.

5. Examples

Example 1. This is an example of inference when only the average ranks are given. Snell (1983) quotes the results of a company survey in which participants were asked to score from 1 to 20 each of seven possible aims of companies according to how important they saw the item in their company. The scores were then ranked from 1–7 for each participant, and it is required to assess whether the rankings given by the accountants differed from those of supervisors/foremen. Snell quotes only the mean ranks for the groups, which we give in Table 1 together with their variances as estimated from order-statistics models with the same pooled mean ranks.

Table 1. Mean ranks for two groups and estimated variances

	Mean ranks			Estimated variances	
Item no.	Group 1	Group 2	Pooled	Forward	Backward
1	5.333	4.583	4.958	2.789	2.279
2	3.833	5.167	4.500	2.799	2.564
3	3.083	4.000	3.542	2.441	2.801
4	5.667	3.917	4.792	2.808	2.395
5	1.167	2.667	1.917	1.106	1.961
6	2.750	2.167	2.458	1.611	2.422
7	6.167	5.500	5.833	2.363	1.495
Sample size	6	6	12		

The asymmetry of the logistic models shows up here in that the forward model gives minimum variance to item 5 which has the lowest pooled rank, whereas the backward model gives minimum variance to item 7 which has the highest pooled rank. Here the backward model fits the data as a whole since the Q-statistic for comparing the mean ranks of the two groups is $Q = 10.98$, as against the 5% critical value of 12.59 (based on 6 d.f.). We may conclude that the two groups have the same mean ranks according to the backward model. From the description of how the data were gathered it would seem that high ranks are more important and this is compatible with the backward model, with high ranks determined first. In contrast the forward model gives a Q-statistic of $Q = 12.74$, so that we should reject the null hypothesis of equal mean ranks in this case (or conclude that the forward logistic model does not fit the data).

Example 2. Hollander and Sethuraman (1978) give the full set of rank preferences for two groups for three leisure activities. The mean ranks for the two groups, and the pooled sample covariance matrix C, are given in Table 2 along with the theoretical covariance matrix W calculated from the pooled mean ranks using a forward logistic model. The Hollander-Sethuraman statistic Q_{HS} is calculated by using C in place of W in (4.1) and we find $Q_{\mathrm{HS}} = 13.8$ (based on 2 d.f.) which is just on the borderline of being significant at 0.1%. However, note that the true significance level is 0.02% for the permutation test suggested by Hollander and Sethuraman.

The main reason for calculating the theoretical covariance W in this example is to gain some additional insight into the mechanism by which rankings are made. In this case the variances predicted by the forward logistic model agree closely with the sample variances. To give some idea of how close the agreement is, let us

Table 2. Mean ranks and (common) covariances for two groups

	Mean ranks		Covariance matrix					
Option	Group 1	Group 2	Sample C			Forward Logistic W		
1	2.93	2.31	0.396	−0.259	−0.137	0.416	−0.266	−0.150
2	1.43	2.46		0.533	−0.274		0.495	−0.228
3	1.64	1.23			0.410			0.379
Sample size	13	14						

NOTE: Using (4.1), i.e. with the theoretical covariance W instead of C, the Q-statistic is $Q = 14.65$ (2 d.f.), and the conclusion is the same: reject the hypothesis of equal mean ranks at the 0.1% level.

assume that the sample variance is proportional to a χ^2 random variable with 25 d.f. This would give 95% confidence limits for the true variance of $0.6s^2 < \sigma^2 < 1.9s^2$ approximately. Clearly the elements of W are well within these implied limits and this means that a forward logistic model fits this aspect of the data. However it should be remembered that the null hypothesis of equal mean ranks (and, in the forward logistic model, equal covariance matrices) is rejected for the above example. Each group should therefore be treated separately, with its own distinct mean ranks and covariance matrix.

If the forward logistic model is correct, and the approximate equality of C and W implies that it is nearly correct, the ranking procedure would appear to be sequential in structure: choose the object which is to be given rank 1; independently of this choice choose from the remainder the object which is to be ranked 2; and so on. Vice versa if we know that the ranking process is likely to have this type of structure we may use the logistic model with some confidence even if the sample sizes are so small, as in example 1, that covariance estimates are quite unreliable.

The ranking process may have a more symmetric structure if there are many factors which influence the ranks, so that an order statistics model based on a normal score would be appropriate. The examples discussed by Koch, Freeman and Lehnen (1976) are of this type and the normal model does fit reasonably well although there are noticeable departures from the predicted variances for some alternatives, perhaps indicating that relatively few factors influence the rankings of these objects.

References

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39**, 324-345.

Daniels, H. E. (1950). Rank correlation and population models. *J. Roy. Statist. Soc. Ser. B* **12**, 171-181.

Henery, R. J. (1986). Interpretation of average ranks. *Biometrika* **73**, 224-227.

Hollander, M. and Sethuraman, J. (1978). Testing for agreement between two groups of judges. *Biometrika* **65**, 403-411.

Joe, H. (1988). Majorization, entropy and paired comparisons. *Ann. Statist.* **16**, 915-925.

Koch, G. G., Freeman, J. L. and Lehnen, R. G. (1976). A general methodology for the analysis of ranked policy preference data. *Int. Statist. Rev.* **44**, 1-28.

Linhart, H. (1960). Approximate tests for m rankings. *Biometrika* **47**, 476-480.

Mak, T. K. (1985). Interpretation and estimation in ranking theory. *Biometrika* **72**, 227-229.

Moran, P. A. P. (1950). Recent developments in ranking theory. *J. Roy. Statist. Soc. Ser. B* **12**, 153-162.

Pettitt, A. N. (1982). Parametric tests for agreement amongst groups of judges. *Biometrika* **69**, 365-375.

Plackett, R. L. (1975). The analysis of permutations. *Appl. Statist.* **24**, 193-202.

Plackett, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* **41**, 351-360.

Snell, M. C. (1983). Recent literature on testing for intergroup concordance. *Appl. Statist.* **32**, 134-140.

Department of Statistics, University of Strathclyde, Glasgow G1 1XH, Scotland.

(Received March 1989; accepted June 1990)