Biometrika (1987), **74**, 2, *pp.* 432–6
Printed in Great Britain

Ranking from unbalanced paired-comparison data

BY H. A. DAVID
Statistical Laboratory, Iowa State University, Ames, Iowa 50011, U.S.A.

SUMMARY

It is proposed that for ranking objects or players in an incomplete paired-comparison experiment or tournament with at most one comparison per pair, the score of a player, C, be the total number of (a) wins of players defeated by C minus losses of players to whom C lost, plus (b) C's wins minus C's losses. A tied match counts as half a win plus half a loss. More general tournaments can be treated similarly.

Some key words: Kendall–Wei approach; Partial balance; Preference data; Round robin tournament.

1. INTRODUCTION

Suppose that t objects $C_1, \ldots, C_t$ are compared $n \geq 1$ times in all possible $\frac{1}{2}t(t-1)$ different pairings. A good deal of attention has been given to the problem of converting the results of such a balanced paired-comparison experiment into a ranking of the objects. The simplest procedure is to rank according to the vector w of row-sum scores, or wins,

$$w = A1, \tag{1·1}$$

where 1 is the column vector of t ones and A is the matrix $((\alpha_{ij}))$ of the proportion of times C_i is preferred to C_j $(i, j = 1, \ldots, t;\ i \neq j)$. Note that $\alpha_{ij} + \alpha_{ji} = 1$, if necessary by splitting of points in case of any ties in individual comparisons. We take $\alpha_{ii} = 0$ $(i = 1, \ldots, t)$.

Row-sum scoring makes no distributional assumptions and we confine ourselves to nonparametric methods. Although for balanced experiments there is little wrong with this straightforward technique, a surprisingly large number of alternatives have been developed; see David (1971) for a review. One line of approach, the Kendall–Wei method, based on generating scores by powering the matrix A in (1·1) (Kendall, 1955), will be further discussed. Another active but very different area of research has resulted from Slater's (1961) principle of finding a ranking that minimizes the number of inconsistencies, i.e. individual paired-comparison results opposite to what would be expected from the ranking. Some methods of the Kendall–Wei type are useful primarily in providing rationales, or at least techniques, for ranking objects with equal row-sum scores or, more briefly, for breaking ties, in an overuse of the word 'tie'. If applied to all objects, many of the methods may do more than break ties by actually changing the row-sum ranking. It is important to note that questions of statistical significance play little role. Often rankings are required from quite limited data, perhaps to help in the selection of the best or the best few objects.

Abundance of methods becomes dearth when the paired-comparison experiment is not balanced, but see Cowden (1975) and Daley (1979). In this case the need for suitable methods is all the greater since the row-sum scores are clearly no longer satisfactory. The basic problems are to take into account (i) the varied calibre of the opposition encountered by each object, and (ii) possibly different numbers of comparisons involving the objects. Regarding (i), note that in scoring an object C_i the Kendall–Wei method pays attention only to the strength of the opposition defeated by C_i and not to the strength of the objects preferred to C_i. This is inappropriate in the absence of balance, a point that applies also to the use of 'fair scores' proposed, in the balanced case, by Daniels (1969) and Moon & Pullman (1970). For (ii) it must be realized that there can be no entirely satisfactory way of ranking if the number of replications of each object varies appreciably.

One special case for which good nonparametric methods of analysis exist occurs when the experiment has been run as a partially balanced incomplete block design of block size 2, with n replications. Here n must be large enough to permit an arcsine or similar transformation of the nonzero proportions α_{ij} $(i, j = 1, \ldots, t)$, for example, Bock & Jones (1968, § 7·1). This case is of limited value since for $n > 1$ one would ordinarily try to achieve greater balance by not having identical replicates. A possible general approach is through the use of Slater's (1961) principle but this is difficult to carry out except in small experiments; see, for example, Flueck & Korsh (1974). Moreover, its premise of giving equal weight to mild and gross inconsistencies is debatable (David, 1963, p. 34).

In § 2 we propose and examine a simple ranking procedure that is a generalization of the use of row-sum scores and is related to the methods of the Kendall–Wei type used in the balanced case. A detailed discussion of this relationship is postponed to § 4 since it is not needed for an understanding of our technique. A small example in § 3 illustrates the procedure.

To avoid trivialities we will assume the usual indivisibility condition for paired-comparison data (Kendall, 1955), namely that the matrix A cannot be expressed, by rearrangement of rows and columns, as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}. \qquad (1\cdot2)$$

Note that, if C_i and C_j are not compared, both α_{ij} and α_{ji} are zero. If A can be written as in (1·2), then the objects can be divided into two or more subgroups for which the indivisibility assumption is satisfied. Different subgroups are not comparable if disconnected; if connected, then all objects in one group are automatically ranked higher than all objects in another group.

2. A general ranking method

2·1. *The Kendall–Wei approach*

The problem described in § 1 can be expressed equivalently in the often more convenient language of tournaments, a balanced paired-comparison experiment with n replications corresponding to a round robin of n rounds. In chess tournaments a long-standing method of breaking ties among top-scorers, stated for $n = 1$, is to replace w_i, the ith component of w in (1·1), by the sum of the scores of the players defeated by C_i plus half the sum of the scores of players tied with C_i; that is, the ranking is based on the vector $w^{(2)}$ given by

$$w^{(2)} = Aw = A^2 1. \qquad (2\cdot1)$$

In this form n need not be 1. Since ties may not all be broken by this powering of A, the process may be continued to generate

$$w^{(d)} = A^d 1 \quad (d = 1, 2, \ldots). \qquad (2\cdot2)$$

With increasing d, rankings based on $w^{(d)}$ will tend to stabilize; see § 4. The idea of this approach (Kendall, 1955) is to give more credit to a player for defeating a high-scoring than a low-scoring opponent, but this means, in effect, that a loss to the latter is punished less than a loss to the former. Ramanujacharyulu (1964) calls $w^{(d)}$ the iterated power of order d and introduces the corresponding iterated weakness

$$l^{(d)} = (A')^d 1 \quad (d = 1, 2, \ldots). \qquad (2\cdot3)$$

In particular, $l_i^{(2)}$ is the sum of the number of losses plus half the number of ties of players who defeated C_i $(i = 1, \ldots, t)$. Actually the two authors just cited used a slightly modified procedure corresponding to setting α_{ii} equal to $\frac{1}{2}$ rather than 0. Equivalently, we can replace the matrix A by $B = A + \frac{1}{2}I$, where I is the identity matrix.

2·2. *The method and its properties*

We are now ready to propose a simple formula for scoring, and hence ranking, the t players in an incomplete or unbalanced round robin tournament. Let s, the $t\times 1$ vector of scores, be given by

$$s = w^{(2)} - l^{(2)} + w - l \tag{2·4}$$

$$= B^2 1 - (B')^2 1. \tag{2·5}$$

The idea of this approach is that s_i $(i = 1, \ldots, t)$ reflects equally the strength of players defeated by C_i and the weakness of players by whom C_i was defeated.

PROPERTY 1. *Let s^* be the vector of scores when wins and losses are interchanged. Then $s^* = -s$.*

Proof. We have that $s^* = (B')^2 1 - B^2 1 = -s$. □

Comment. The reversal of ranks implied by this property does not hold in general when (2·2) of (2·3) are used alone.

PROPERTY 2. *The scores sum to zero,* $1's = 0$.

Proof. For $1's = 1'B^2 1 - 1'(B^2)'1 = 0$. □

PROPERTY 3. *If the tournament is balanced, then s gives the same ranking as w.*

Proof. For convenience and without essential loss of generality we consider in the remainder of this section the unreplicated case, $n = 1$. We first obtain a more general result for tournaments in which each player meets r other players once $(r = 2, \ldots, t-1)$ and does not meet the remaining $t-1-r$ players. Then $l = r1 - w$, so that from (2·4) and (2·1)

$$s = Aw + A'w - rl + w - l = (A + A')w + (r+2)w - r(r+1)1. \tag{2·6}$$

Let $\Sigma^* w_j$, $\Sigma^{**} w_j$ and Σw_j denote respectively summation over players met by A_i, not met by A_i, and all t players. Then

$$s_i = (r+2)w_i + \sum^* w_j - r(r+1) \quad (i = 1, \ldots, t),$$

or, alternatively, since $\Sigma w_j = \frac{1}{2}rt$,

$$s_i = (r+1)w_i - \sum^{**} w_j + \tfrac{1}{2}rt - r(r+1).$$

In particular, if $r = t-1$ we have $s_i = t\{w_i - \frac{1}{2}(t-1)\}$ and Property 3 follows. □

Comment. The very fact that our procedure fails to break tied row-sum scores for a complete tournament makes it more attractive in the absence of balance. Property 3 does not hold if $d > 2$ is substituted for the index 2 in (2·5).

PROPERTY 4. *If tournaments T and T^* differ only in that C_i and C_k do not meet in T but $C_i \to C_k$ in T^*, then for $i = 1, \ldots, t$*

$$s_i^* = s_i + w_k + 1, \quad s_k^* = s_k - l_i - 1,$$

$$s_m^* = s_m + \alpha_{mi} - \alpha_{km} \quad (m = 1, \ldots, t;\ m \neq i, k).$$

Proof. We have $\alpha_{ik}^* = 1$, $\alpha_{ik} = 0$, $w_i^* = w_i + 1$, $l_k^* = l_k + 1$. Otherwise corresponding starred and unstarred α's, w's, and l's are equal. The results are now easily verified, e.g.

$$s_i^* - s_i = \sum_{j=1}^{t} (\alpha_{ij}^* w_j^* - \alpha_{ij} w_j) - \sum_{j=1}^{t} (\alpha_{ji}^* l_j^* - \alpha_{ji} l_j) + (w_i^* - w_i - l_i^* + l_i) = w_k + 1. \quad \square$$

Comment. Properties 1 to 3 hold whether the term $w - l$ in (2·4) is present or not, i.e. whether B or A is used as the matrix in (2·5). Here the term ensures that C_i gains a point for defeating C_k even when C_k has not won any games; without the term, C_i would be worse off than before beating such a C_k since all players who defeated C_i gain a point as a result of C_i's win.

PROPERTY 5. *If two tournaments T and T^* differ only in that C_i and C_k do not meet in T but tie in T^*, then for* $i=1,\ldots,t$

$$s_i^* = s_i + \tfrac{1}{2}(w_k - l_k), \quad s_k^* = s_k + \tfrac{1}{2}(w_i - l_i),$$

$$s_m^* = s_m + \tfrac{1}{2}(\alpha_{mi} - \alpha_{im} + \alpha_{mk} - \alpha_{km}) \quad (m = 1, \ldots, t;\ m \neq i, k).$$

Proof. This follows from Property 4, a tie being half a win plus half a loss. □

Comment. The results of Properties 4 and 5 can, of course, be aggregated to cover any situation.

PROPERTY 6. *If two tournaments T and T^* differ only in that $C_k \to C_i$ in T but $C_i \to C_k$ in T^*, then for* $i=1,\ldots,t$

$$s_i^* = s_i + w_k + l_k + 1, \quad s_k^* = s_k - w_i - l_i - 1,$$

$$s_m^* = s_m + (\alpha_{mi} + \alpha_{im}) - (\alpha_{mk} + \alpha_{km}) \quad (m = 1, \ldots, t;\ m \neq i, k).$$

Proof. This follows from Property 4. □

3. AN EXAMPLE

Table 1 shows the outcome of a small tournament in which each of $C_1, \ldots, C_7$ meets just four opponents. Also given are the vectors w, $w^{(2)}$, l, $l^{(2)}$, as well as s of (2·4) and y and s^* to be explained shortly. We note that C_2 and C_3, who are among the top row-sum scorers, tie for first place on the basis of $w^{(2)}$ or $B^2 1 = w^{(2)} + w$. This reflects the fact that in the games they won or tied, their respective opponents were equally strong as measured by w. However, C_2 lost to a weaker player than did C_3, as is reflected by $l^{(2)}$ and hence s.

The vector y is an alternative score vector obtained by utilizing the cyclic nature of the tournament schedule. If we regard this as a cyclic incomplete block design of block size 2 with observations $(1, 0)$, $(\frac{1}{2}, \frac{1}{2})$, or $(0, 1)$ in each block it is easy to estimate the players' strengths, the 'treatment effects', by the methods described in John, Wolock & David (1972). The constants $C(1, j)$ given in their Design A5 have been scaled up for comparability with s. Agreement is excellent in this example, resulting in identical rankings.

Now suppose that the game between C_3 and C_5 was not actually played. Instead of applying (2·4) from scratch, we can use Property 5, but with T and T^* reversed. This gives the vector s^*, e.g.

$$s_1 = s_1^* + \tfrac{1}{2}(\alpha_{13}^* - \alpha_{31}^* + \alpha_{15}^* - \alpha_{51}^*), \quad s_1^* = 2 - \tfrac{1}{2}(0 - 1 + 0 - 0) = 2{\cdot}5.$$

We can also suppose the tournament to be at the s^*-stage, with only the game between C_3 and C_5 to be completed. At this point C_1 and C_2 are tied for second place. After C_3 and C_5 draw, the s-stage shows that C_2 has moved ahead of C_1. Actually, whatever the outcome of C_3 versus C_5, it is easy to see that C_2 moves a point ahead of C_1. This is because $C_2 \to C_3$ but $C_1 \leftarrow C_3$. Thus we would break the tie in favour of C_2.

Table 1. *An incomplete tournament*

	1	2	3	4	5	6	7	w	$w^{(2)}$	s	y	s^*
1		$\frac{1}{2}$	0			1	1	2·5	3·25	2	1·875	2·5
2	$\frac{1}{2}$		1	1			0	2·5	5·75	3	3·25	2·5
3	1	0		1	$\frac{1}{2}$			2·5	5·75	4·5	4·625	4
4		0	0		1	1		2	3·5	0·5	0·75	0·5
5			$\frac{1}{2}$	0		1	1	2·5	3·25	1·5	1·25	1
6	0			0	0		1	1	1	−6	−6·125	−5·5
7	0	1			0	0		1	2·5	−5·5	−5·625	−5
l'	1·5	1·5	1·5	2	1·5	3	3	14	25	0	0	0
$l^{(2)\prime}$	2·25	3·75	2·25	3	2·75	5	6	25				

4. Relation to ranking methods for balanced paired comparisons

As stated in § 2·1 rankings in the balanced case based on $w^{(d)} = A^d 1$ will tend to stabilize as d increases. More precisely, it follows from Perron-Frobenius theory (Kendall, 1955; Seneta, 1973, p. 1) that, as $d \to \infty$, $\lim (A/\lambda)^d 1 = v$, where λ is the unique positive characteristic root of A with the largest absolute value and v is a vector of positive terms. Here v may be obtained, up to a constant multiplier, as a column eigenvector satisfying $Av = \lambda v$. Replacing A by $B = A + \frac{1}{2}I$ leads to the same ranking (Moon, 1968, pp. 44-6).

Kendall regards letting $d \to \infty$ as primarily of theoretical interest, writing that $d = 2$ 'is as far as one would wish to go on practical grounds, perhaps'. Many other authors have proceeded to the limit, most relevant to our proposal being the suggestion by Hasse (1961) to rank according to the vector $u = v - v^*$, where v and v^* are respectively the column eigenvectors corresponding to B and B'.

Under the indivisibility assumption of § 1, the Perron-Frobenius theory will continue to apply in the case of unbalanced paired comparisons. With u as vector of scores, Properties 1 and 2 of § 2·2 are seen still to hold, but not in general Property 3. For unbalanced data the use of u seems best restricted to the possible resolution of ties in s-scores when simpler methods fail.

Acknowledgement

This research was supported by the U.S. Army Research Office.

References

Bock, R. D. & Jones, L. V. (1968). *The Measurement and Prediction of Judgment and Choice.* San Francisco: Holden-Day.

Cowden, D. J. (1975). A method of evaluating contestants. *Am. Statistician* **29**, 82-4.

Daley, D. J. (1979). Markov chains and a pecking order problem. In *Interactive Statistics*, Ed. D. McNeil, pp. 247-54. Amsterdam: North-Holland

Daniels, H. E. (1969). Round-robin tournament scores. *Biometrika* **56**, 295-9.

David, H. A. (1963). *The Method of Paired Comparisons.* London: Griffin.

David, H. A. (1971). Ranking the players in a Round Robin tournament. *Rev. Int. Statist. Inst.* **39**, 137-47.

Flueck, J. A. & Korsch, J. F. (1974). A branch search algorithm for maximum likelihood paired comparison ranking. *Biometrika* **61**, 621-6.

Hasse, M. (1961). Über die Behandlung graphentheoretischer Probleme unter Verwendung der Matrizenrechnung. *Wiss. Zeit. Tech. Univ. Dresden* **10**, 1313-6.

John, J. A., Wolock, F. W. & David, H. A. (1972). *Cyclic Designs*, Appl. Math. Ser. 62. Washington, D.C.: National Bureau of Standards.

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics* **11**, 43-62.

Moon, J. W. (1968). *Topics on Tournaments.* New York: Holt, Rinehart & Winston.

Moon, J. W. & Pullman, N. J. (1970). On generalized tournament matrices. *SIAM Rev.* **12**, 384-99.

Ramanujacharyulu, C. (1964). Analysis of preferential experiments. *Psychometrika* **29**, 257-61.

Seneta, E. (1973). *Non-Negative Matrices.* New York: Wiley.

Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika* **48**, 303-12.

[*Received August* 1986. *Revised October* 1986]