Round-Robin Tournament Scores
Author(s): H. E. Daniels
Source: *Biometrika*, Vol. 56, No. 2 (Aug., 1969), pp. 295-299
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2334422

Biometrika (1969), **56**, 2, *p.* 295
Printed in Great Britain

Round-robin tournament scores

BY H. E. DANIELS
University of Birmingham

SUMMARY

Some methods of scoring players in a round-robin tournament are proposed, based on a concept of fair allocation of rewards to each player. The scores are related to those determined by a previously suggested principle of 'consistency'. The scoring systems lead to the π_i's of the Bradley–Terry model, and they have desirable ranking properties in the case of a general linear model. Various other properties of the scores are described.

1. INTRODUCTION

In a paired comparison experiment, preferences are expressed between all possible pairs of a set of objects, or some balanced selection of them. The problem of analysing such an experiment is basically to assign optimum scores to the objects, often with the more limited aim of ranking them in order of preference. A round-robin tournament is another example of the same situation, but in assigning the optimum scores to the players one is perhaps led to place more emphasis on questions of fairness, and on the resolution of ties which are apt to occur when simple total scores are used. Much work has been done on the various probability models, scoring systems and experimental designs which are applicable to these situations: reference may be made to the comprehensive review by David (1963).

In the present paper some methods of scoring players in a round-robin tournament are proposed which are based on a concept of fair allocation of rewards. There are a number of unresolved questions concerning these scoring systems, and the aim of the paper is to put the ideas forward without much development in depth in the hope that others will be encouraged to pursue them further.

2. MODIFIED WEI SCORES

Suppose n games are played between each pair of players $A_1, \ldots, A_k$. Score 1, $\frac{1}{2}$ or 0 according as a player wins, draws, or loses a game, and let $\mathbf{N} = \{n_{ij}\}$ be the matrix of total scores, where n_{ij} is scored by A_i playing A_j. Then $n_{ij} = n - n_{ji}$, $n_{ii} = \frac{1}{2}n$ and the marginal totals are $n_{i.}$, $n_{.j}$, where $n_{i.} = kn - n_{.i}$. The outcomes of all games are assumed independent, and n_{ij}/n estimates the probability $\pi_{ij} = 1 - \pi_{ji}$ that A_i beats A_j. Conventionally $\pi_{ii} = \frac{1}{2}$.

The total score for A_i is $n_{i.}$. T. H. Wei (Kendall, 1955) calculates new scores $\Sigma n_{ij} n_{j.}$ which redistribute the original scores to each player according to the results of his games with the other players. Here and throughout the paper the summation is with respect to j, unless indicated otherwise. The new scores are the row sums of $\mathbf{N}^2$. Repetition of the process gives the row sums of successively higher powers of $\mathbf{N}$, and the scores when rescaled ultimately settle down to the column eigenvector $\mathbf{z}$ satisfying $\Sigma n_{ij} z_j = \kappa z_i$, where κ is the largest eigenvalue of $\mathbf{N}$.

In the discussion on the paper by Williams (1967) I suggested that the limiting Wei scores could be regarded as 'consistent' in the sense that Wei's procedure reproduces them apart from the scale factor κ. But the fact that κ is not 1 detracts from the idea of consistency, and it is also not clear why on the grounds of consistency one should choose the largest rather than some other eigenvalue. I therefore proposed modifying Wei's procedure so that the original scores $n_{i.}$ are reallocated to become $\Sigma(n_{ij}/n_{.j})\,n_{j.}$. The factor $n_{ij}/n_{.j}$ is the proportion of occasions that A_i has beaten A_j out of the total number of occasions that A_j has been beaten, counting a draw as $\frac{1}{2}$. Thus if A_j is relatively easy to beat, A_i receives less credit for having beaten him, which seems a more equitable arrangement. Also, since the matrix $\{n_{ij}/n_{.j}\}$ is stochastic with column sums unity, its maximum eigenvalue is unity and the limiting scores x_i satisfy

$$\Sigma \frac{n_{ij}}{n_{.j}} x_j = x_i \quad (i = 1, \ldots, k). \tag{2·1}$$

The scores x_i reproduce themselves exactly, thus giving a unique meaning to consistency and excluding from consideration the other eigenvalues. The x_i's are estimates of population scores ξ_i satisfying

$$\Sigma \frac{\pi_{ij}}{\pi_{.j}} \xi_j = \xi_i \quad (i = 1, \ldots, k), \tag{2·2}$$

where $\pi_{.j} = \sum_i \pi_{ij}$.

Since the matrices concerned have non-negative elements and may be assumed irreducible, it is known that the elements x_i or ξ_i of the maximal eigenvectors must be all positive or all negative (Debreu & Herstein, 1953). We take them to be positive.

3. Fair scores

There is another approach to the scoring problem which turns out to be related to the one just described. Suppose A_i is 'worth' λ_i in the sense that any player who beats A_i wins λ_i from him, or $\frac{1}{2}\lambda_i$ if they draw. What is the fairest way of assigning the λ_i's? One way is to choose them so that the match as a whole is fair, i.e. so that in the long run every player wins or loses the same amount on the average. The λ_i's then satisfy the equations

$$\Sigma\pi_{ij}\lambda_j = \Sigma\pi_{ji}\lambda_i = \pi_{.i}\lambda_i \quad (i = 1, \ldots, k), \tag{3·1}$$

where the term $\pi_{ii}\lambda_i$ is added for convenience to both sides. These are in fact the same equations as (2·2) if we take $\xi_i = \pi_{.i}\lambda_i$. The score ξ_i can therefore be interpreted as the average amount A_i expects to win, which has been equated to the average amount he expects to lose to all the other players. It follows incidentally that every $\lambda_i > 0$.

The question then arises: should one use λ_i or ξ_i as the score for A_i? For the Bradley–Terry model $\pi_{ij} = \pi_i/(\pi_i+\pi_j)$ (David, 1963, Chapter 4) and, therefore, we find (3·1) is satisfied by $\lambda_i = \pi_i$, whereas ξ_i involves the π_j's for all the players. In this case λ_i is a more intrinsic property of A_i than ξ_i.

From a set of sample scores n_{ij} any $(k-1)$ of the equations

$$\Sigma n_{ij} l_j = n_{.i} l_i \tag{3·2}$$

give the same estimates l_i of λ_i. In the Bradley–Terry case they can be shown to be less efficient than the maximum-likelihood estimates, but they have the advantage of requiring no iteration.

Since there is an arbitrary scale factor in λ_i, it may be more convenient to work with $\theta_i = \log \lambda_i$ which has an arbitrary added constant. For the Bradley–Terry model, this is the parameter of the underlying logistic distribution.

4. A dual scoring system

The amount λ_j won by A_i when he beats A_j depends only on the strength of his opponent A_j. At the other extreme one might consider A_i to win an amount μ_i from every player A_j he beats, regardless of A_j's strength. The μ_i's can again be determined to make the match fair: balancing the average wins and losses as before, we get the dual equations

$$\pi_{i}.\mu_i = \Sigma\pi_{ij}\mu_i = \Sigma\pi_{ji}\mu_j. \tag{4·1}$$

Since $\eta_i = \pi_i.\mu_i$ is the maximal row eigenvector of the matrix $\{\pi_{ij}/\pi_i.\}$, every μ_i can again be taken to be positive.

Here one expects the better players to have smaller values of μ_i for the match to be fair. For the Bradley–Terry model, (4·1) is satisfied by $\mu_i = 1/\pi_i$ and we get the same θ_i as before with sign reversed. The corresponding estimating equations for μ_i are

$$\Sigma n_{ji}m_j = n_i.m_i. \tag{4·2}$$

These again involve no iteration, and $1/m_i$ has the same asymptotic efficiency as l_i in estimating π_i for the Bradley–Terry model.

Other models for π_{ij} do not in general lead to a simple relation between the λ_i's and the μ_i's and they may not even have reverse rankings. For example, if the probabilities π_{ij} are specified by the matrix

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{16} \\ \frac{3}{4} & \frac{1}{2} & \frac{3}{4} \\ \frac{15}{16} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}, \tag{4·3}$$

then $\lambda_i = (1, 6, 3)$ but $\mu_i = (51, 10, 9)$. However, in the important case where the model is linear with $\pi_{ij} = F(u_i - u_j)$, F being an absolutely continuous distribution function, λ_i and $1/\mu_i$ have the same rank order as u_i, and indeed the same is true of ξ_i and $1/\eta_i$.

The proof is as follows. From (3·1) we have

$$\Sigma\left(\frac{\pi_{ij}}{\pi_{.i}} - \frac{\pi_{rj}}{\pi_{.r}}\right)\lambda_j = \lambda_i - \lambda_r. \tag{4·4}$$

Suppose that $u_i > u_r$. Then $u_i - u_j > u_r - u_j$ so that $\pi_{ij} > \pi_{rj}$ whatever $u_j, \pi_i. > \pi_r.$ and $\pi_{.i} < \pi_{.r}$. Since every $\lambda_j > 0$ we must have $\lambda_i > \lambda_r$, and by a similar argument, $\mu_i < \mu_r$. The fact that $\xi_i > \xi_r$, $\eta_i < \eta_r$ follows from

$$\Sigma(\pi_{ij} - \pi_{rj})\lambda_j = \xi_i - \xi_r \tag{4·5}$$

and the corresponding formula for $\eta_i - \eta_r$.

The same thing need not, of course, be true of the sample scores l_i, m_i and x_i, y_i because of sampling fluctuation. Although from the ranking point of view λ_i and ξ_i (or μ_i and η_i) are equivalent for a linear model, it would be worth examining whether in the case of sample scores one is more efficient than the other in estimating the true ranking.

5. A general scoring system

Both the above scoring systems can be included in the following general scheme. Suppose that if A_i beats A_j he wins λ_j discounted by a factor μ_i, i.e. he wins $\mu_i\lambda_j$. For the match to be fair we require

$$\Sigma\pi_{ij}\mu_i\lambda_j = \Sigma\pi_{ji}\mu_j\lambda_i. \tag{5·1}$$

19-2

There is now an element of arbitrariness in the relative magnitudes of λ_i, μ_i and further conditions are needed to determine them.

When $\pi_{ij} = \pi_i/(\pi_i+\pi_j)$ the solution of (5·1) is

$$\pi_i = \lambda_i/\mu_i. \tag{5·2}$$

This suggests that λ_i/μ_i may be an appropriate score for A_i in more general cases. Actually, for the linear model $\pi_{ij} = F(u_i-u_j)$ the scores λ_i/μ_i have the same rank order as u_i. This can be deduced as before from

$$\frac{\lambda_i}{\mu_i}-\frac{\lambda_r}{\mu_r}=\frac{\Sigma\pi_{ij}\lambda_j}{\Sigma\pi_{ji}\mu_j}-\frac{\Sigma\pi_{rj}\lambda_j}{\Sigma\pi_{jr}\mu_j}. \tag{5·3}$$

On the logarithmic scale the corresponding score is $\theta_i-\phi_i$, where $\theta_i = \log\lambda_i$, $\phi_i = \log\mu_i$.

One natural restriction to impose is

$$\lambda_i\mu_i = c \quad (i = 1, \ldots, k). \tag{5·4}$$

Then (5·1) becomes

$$\Sigma\pi_{ij}\lambda_j/\lambda_i = \Sigma\pi_{ji}\lambda_i/\lambda_j \tag{5·5}$$

and the reward given to A_i on beating A_j is the ratio of their worths λ_i. Notice that a reward based on *differences* of worths does not lead to useful scores because the only solution of

$$\Sigma\pi_{ij}(\lambda_j-\lambda_i) = \Sigma\pi_{ji}(\lambda_i-\lambda_j) \tag{5·6}$$

is λ_i = constant.

6. Further properties

There are a number of other properties of the scores λ_i, μ_i which, though suggestive, are difficult to tie together, and I shall merely state them.

We can write (5·1) as

$$\frac{\Sigma\pi_{ij}\lambda_j}{\lambda_i}=\frac{\Sigma\pi_{ji}\mu_i}{\mu_i}=\kappa_i, \quad \text{say}, \tag{6·1}$$

or

$$\Sigma\pi_{ij}\lambda_j = \kappa_i\lambda_i, \quad \Sigma\pi_{ji}\mu_j = \kappa_i\mu_i, \tag{6·2}$$

where the κ_i's are determined from the further restrictions imposed on λ_i, μ_i. If $\kappa_i = \kappa$ for all i, λ_i are the Wei scores, and we observe that μ_i are the elements of the corresponding row eigenvector.

We can also generalize (2·2) in the following way. Write

$$\xi_i = \Sigma\pi_{ij}\mu_i\lambda_j = \Sigma\pi_{ji}\mu_j\lambda_i; \tag{6·3}$$

then

$$\mu_i = \xi_i/\Sigma\pi_{ij}\lambda_j, \quad \lambda_i = \xi_i/\Sigma\pi_{ji}\mu_j, \tag{6·4}$$

so that

$$\xi_i = \Sigma\left(\frac{\pi_{ij}\mu_i\xi_j}{\sum\limits_r \pi_{rj}\mu_r}\right) \tag{6·5}$$

and

$$\xi_i = \Sigma\left(\frac{\pi_{ji}\lambda_i\xi_j}{\sum\limits_r \pi_{jr}\lambda_r}\right). \tag{6·6}$$

Thus ξ_i is the column eigenvector of

$$\{\pi_{ij}\mu_i/\sum_r \pi_{rj}\mu_r\}$$

and the row eigenvector of $\{\pi_{ij}\lambda_j / \sum_r \pi_{ir}\lambda_r\}$,

corresponding to the largest eigenvalue which is 1 in both cases.

The scores λ_i, μ_i have a minimum property. If

$$G = \Sigma\Sigma\pi_{ij}\mu_i\lambda_j, \tag{6·7}$$

summed over i and j, is the total expected gain (loss) over all players, then λ_i, μ_i minimize G subject to the conditions $\lambda_i\mu_i = c_i$ $(i = 1, \ldots, k)$. The Wei scores are obtained if we minimize G subject to the single condition $\lambda_1\mu_1 + \ldots + \lambda_k\mu_k =$ constant. Another way of looking at this is in terms of $\theta_i = \log\lambda_i$, $\phi_i = \log\mu_i$, $\gamma_i = \log c_i$. The solution of

$$\frac{\partial G}{\partial\theta_i} = \frac{\partial G}{\partial\phi_i} \quad (i = 1, \ldots, k) \tag{6·8}$$

minimizes

$$G = \Sigma\Sigma\pi_{ij}\exp(\theta_i + \phi_j)$$

subject to $\theta_i + \phi_i = \gamma_i$ $(i = 1, \ldots, k)$. The minimum property follows from the convexity of G.

Finally, one can suggest yet another scoring system. Suppose that if A_i beats A_j he gains a proportion $\lambda_j/(\lambda_i + \lambda_j)$ of a fixed sum. As in (5·5) the rewards are based on the relative worths of each pair of players. Matching up as before we get

$$\Sigma\left(\frac{\pi_{ij}\lambda_j}{\lambda_i + \lambda_j}\right) = \Sigma\left(\frac{\pi_{ji}\lambda_i}{\lambda_i + \lambda_j}\right). \tag{6·9}$$

There is no further generality in considering

$$\Sigma\frac{\pi_{ij}\mu_i\lambda_j}{\mu_i\lambda_j + \mu_j\lambda_i} = \Sigma\frac{\pi_{ji}\mu_j\lambda_i}{\mu_i\lambda_j + \mu_j\lambda_i} \tag{6·10}$$

because this merely replaces λ_i by λ_i/μ_i. For the Bradley–Terry model $\pi_{ij} = \pi_i/(\pi_i + \pi_j)$, (6·9) is satisfied by $\lambda_i = \pi_i$, and in fact the sample form of (6·9) is the set of maximum likelihood estimating equations for π_i.

It is worth observing that the scoring systems proposed here apply equally well to incomplete block arrangements. For example, there might be two teams with every member of each team playing every member of the other team.

References

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

Debreu, G. & Herstein, I. N. (1953). Non-negative square matrices. *Econometrica* **21**, 597–607.

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

Williams, E. J. (1967). The analysis of association among many variates (with discussion). *J. R. Statist. Soc.* B **29**, 199–242.

[*Received November* 1968. *Revised January* 1969]