RANKING BY MAXIMUM LIKELIHOOD UNDER A MODEL FOR PAIRED COMPARISONS*

BU-389-M by September, 1971

Roger R. Davidson

University of Victoria and Cornell University

Abstract

This paper discusses a method of ranking a set of objects which are presented in pairs in a preference testing experiment. The ranking to be considered is based on the maximum likelihood estimates of the parameters of a probability model for paired comparisons. It is shown, under a weak assumption, that the maximum likelihood estimates exist and are unique. It is further noted that one can employ an iterative procedure which converges monotonically to the unique estimates. A property of the ranking in the case of a partially balanced paired comparison experiment is presented.

---

* This research was supported in part by the National Research Council of Canada under Grant No. A7166.

Paper No. BU-389-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

RANKING BY MAXIMUM LIKELIHOOD UNDER A MODEL FOR PAIRED COMPARISONS*

BU-389-M by September, 1971

Roger R. Davidson
University of Victoria and Cornell University

1. Introduction

In paired comparison experimentation, responses are obtained to all possible pairs of a set of objects. The analysis of such an experiment involves the estimation of the relative worths of the objects for the purpose of ranking them. The method of paired comparisons is commonly used in preference testing experiments. League competitions and round-robin tournaments are natural examples of where rankings are based on the performance of objects when presented in pairs. A valuable review of the method of paired comparisons has been given by David [6].

A mathematical model for paired comparisons has been presented and developed in a series of three papers: Bradley and Terry [4] and Bradley [1], [2]. This same model had been proposed earlier by Zermelo [11] in discussing chess tournaments, and was subsequently independently presented by Ford [8]. Under this model one considers a set of m objects which are presented in pairs. It is assumed the responses to the objects may be described in terms of an underlying continuum on which the "worths" of the objects can be relatively located. Let π_i denote the "worth", an index of relative preference, of the i^{th} object, $\pi_i \geq 0$, $\sum_{i=1}^{m} \pi_i = 1$. The Bradley-Terry model postulates that, if X_i and X_j are the responses to objects i and j respectively, then

* This research was supported in part by the National Research Council of Canada under Grant No. A7166.

Paper No. BU-389-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

$$P(X_i > X_j | i,j) = \pi_i/(\pi_i + \pi_j) \tag{1}$$

is the probability of an indicated preference for object i in the comparison of objects i and j. One interprets $X_i > X_j$ as meaning that the response to object i is more favorable than that to object j. A summary of various representations of the Bradley-Terry model is given in Bradley [3].

In the Bradley-Terry model no provision is made for an expression of no preference between the objects presented in a pair. A common practice is either to force a definite expression of preference, or to treat ties when they occur by ignoring, splitting or randomly allocating them. However, two extensions of the Bradley-Terry model which accommodate expressions of no preference have been proposed: Rao and Kupper [10] and Davidson [7]. The latter study gives a comparison of the two extensions.

In [7] it is assumed that $P(X_i \approx X_j | i,j)$, the probability of no indicated preference in the comparison of objects i and j, is proportional to the geometric mean of the probabilities of preference for the objects being compared, namely

$$P(X_i \approx X_j | i,j) = \nu \sqrt{P(X_i > X_j | i,j) P(X_j > X_i | i,j)} \ . \tag{2}$$

The constant of proportionality ν serves as an index of discrimination which is assumed to be independent of i and j. One interprets $X_i \approx X_j$ as meaning that the responses to objects i and j are indistinguishable. The derived model is given by

$$P(X_i > X_j | i,j) = \pi_i/(\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j}) \ ,$$

and (3)

$$P(X_i \approx X_j | i,j) = \nu\sqrt{\pi_i \pi_j}/(\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j}) .$$

The purpose of this paper is to present results concerning the maximum likelihood estimates of the parameters of the model (3). It is shown, under a weak assumption, that the maximum likelihood estimates exist and are unique. It is noted that one can employ an iterative procedure which converges monotonically to the unique estimates. A property of the ranking based on the maximum likelihood estimates in the case of a partially balanced paired comparison experiment is presented.

2. The existence and uniqueness of the maximum likelihood estimates

The parameters $\underline{\pi} = (\pi_1, \cdots, \pi_m)$ of the Bradley-Terry model can be estimated by the method of maximum likelihood. The maximum likelihood estimates $\underline{p} = (p_1, \cdots, p_m)$ can then be used to obtain a ranking of the m objects. The problem of obtaining these estimates has been discussed by Ford [8]. In the present section this discussion is extended to the maximum likelihood estimates $(\underline{p}, \hat{\nu})$ of the parameters $(\underline{\pi}, \nu)$ of the model (3).

Consider a paired comparison experiment in which m objects are presented in pairs with r_{ij} independent responses being obtained in the comparison of objects i and j. Let w_{ij}, w_{ji}, and t_{ij} be the frequencies of preference for i over j, preference for j over i, and no preference, respectively. Clearly, $r_{ij} = w_{ij} + w_{ji} + t_{ij}$. Let $\underline{W} = [w_{ij};\ i,j = 1, \cdots, m]$ and $\underline{T} = [t_{ij};\ i,j = 1, \cdots, m]$ where w_{ii} and t_{ii} are defined to be zero $i = 1, \cdots, m$.

The maximum likelihood estimates $(\underline{p}, \hat{\nu})$ are the values of $(\underline{\pi}, \nu)$ which maximize the likelihood function

$$L(\underline{\pi}, \nu;\ \underline{W}, \underline{T}) = \prod_{i<j} \frac{\pi_i^{w_{ij}} \pi_j^{w_{ji}} \ (\nu\sqrt{\pi_i \pi_j})^{t_{ij}}}{(\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j})^{r_{ij}}} \tag{4}$$

over the region $\{(\underline{\pi}, \nu) | \pi_i > 0,\ \Sigma\pi_i = 1,\ 0 < \nu < \infty\}$. The existence and uniqueness of the maximum likelihood estimates will now be demonstrated under mild assumptions on $\underline{W}$ and $\underline{T}$ by adapting the argument given by Ford [8] for the Bradley-Terry model.

As in Ford, the following assumption is made on $\underline{W}$. In every possible partition of the objects into two nonempty subsets, some object in the second set has been preferred at least once to some object in the first set. Equivalently, any two objects are linked by a chain of objects in which each object in the chain has been preferred at least once to the object which follows it: mathematically formulated this assumes that for each pair (i,j) there exists a sequence of indices $i_0, i_1, \cdots, i_n$ with $i_0 = i$ and $i_n = j$ such that $w_{i_\ell i_{\ell+1}} > 0$ for $\ell = 0, 1, \cdots, n-1$.

In addition it is assumed that at least one entry in $\underline{T}$ is positive. If this is not the case the extended model (3) need not be used, hence this assumption is nonrestrictive.

<u>Theorem 1.</u> <u>Under the assumptions on</u> $\underline{W}$ <u>and</u> $\underline{T}$, <u>the likelihood</u> $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ <u>has a maximum over</u> $R = \{(\underline{\pi},\nu) | \pi_i > 0,\ \Sigma\pi_i = 1,\ 0 < \nu < \infty\}$.

<u>Proof:</u> The likelihood is positive and continuous over the region R. The existence of a maximum in this region is established by showing that if one defines $L(\underline{\pi},\nu;\ \underline{W},\underline{T}) = 0$ for $(\underline{\pi},\nu)$ on the boundary then one obtains a uniformly continuous extension of $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ to the closed region $\{(\underline{\pi},\nu) | \pi_i \geq 0,\ \Sigma\pi_i = 1,\ 0 \leq \nu \leq \infty\}$. The boundary points are of three types and these will be considered separately.

(i) If $(\underline{\pi}^o,\nu)$ is on the boundary $\partial R = \{(\underline{\pi},\nu) | \pi_i = 0 \text{ for some } i\}$, then there is an index i such that $\pi_i^o = 0$ and an index j such that $\pi_j^o > 0$. By the assumption on $\underline{W}$ the chain linking i and j has adjacent indices k and ℓ such that $\pi_k^o = 0$, $\pi_\ell^o > 0$. Hence,

$$L(\underline{\pi},\nu;\ \underline{W},\underline{T}) \leq \left[\frac{\pi_k}{\pi_k + \pi_\ell + \nu\sqrt{\pi_k\pi_\ell}}\right]^{w_{k\ell}} \leq \left[\frac{\pi_k}{\pi_k + \pi_\ell}\right]^{w_{k\ell}}$$

which tends to zero uniformly in ν as $(\underline{\pi},\nu)$ approaches ∂R.

(ii) Consider the boundary $\partial_o R = \{(\underline{\pi},\nu) | \pi_i > 0,\ \Sigma\pi_i = 1,\ \nu = 0\}$. By the assumption on $\underline{T}$ there exists a $t_{ij} > 0$. With the use of the geometric mean-arithmetic mean inequality one obtains

$$L(\underline{\pi},\nu;\ \underline{W},\underline{T}) \le \left[\frac{\nu\sqrt{\pi_i\pi_j}}{\pi_i + \pi_j + \sqrt{\pi_i\pi_j}} \right]^{t_{ij}} \le \left[\frac{\nu}{2 + \nu} \right]^{t_{ij}}$$

which tends to zero uniformly in $\underline{\pi}$ as $(\underline{\pi},\nu)$ approaches $\partial_o R$.

(iii) Consider the boundary $\partial_\infty R = \{(\underline{\pi},\nu) | \pi_i > 0,\ \Sigma\pi_i = 1,\ \nu = \infty\}$ and let $\delta < 1/m$ be specified. For each $(\underline{\pi},\nu)$ there is an index j such that $\pi_j > \delta$. By the assumption on $\underline{W}$, the chain linking any object i to such a j has adjacent indices k and ℓ such that $\pi_\ell > \delta$ and $w_{k\ell} > 0$. With the use of the geometric mean-arithmetic mean inequality it follows that

$$L(\underline{\pi},\nu;\ \underline{W},\underline{T}) \le \left[\frac{\pi_k}{\pi_k + \pi_\ell + \nu\sqrt{\pi_k\pi_\ell}} \right]^{w_{k\ell}} \le \left[\frac{1}{\sqrt{\delta}\ (2 + \nu)} \right]^{w_{k\ell}}$$

in a neighborhood of each point of $\partial_\infty R$. Thus $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ tends to zero uniformly in $\underline{\pi}$ as $(\underline{\pi},\nu)$ approaches $\partial_\infty R$.

One now concludes that $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ achieves a maximum in the interior of a closed and bounded subset of R.

The uniqueness of the maximum likelihood estimates will be established through the use of the following lemma which can be proved as a routine exercise in differential calculus.

Lemma: *If f_{ij}, f_{ji}, and e_{ij} are positive constants then*

$$\left[\frac{\pi_i}{\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j}}\right]^{f_{ij}} \left[\frac{\pi_j}{\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j}}\right]^{f_{ji}} \left[\frac{\nu\sqrt{\pi_i\pi_j}}{\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j}}\right]^{e_{ij}}$$

has a maximum value, unique up to a constant of proportionality, at $\pi_i = cf_{ij}$ and $\pi_j = cf_{ji}$, independent of the value of ν.

Let $s_i = 2\sum_j w_{ij} + \sum_j t_{ij}$ be a "score" for object i which is based on all of its comparisons with the remaining objects. In addition, let $t = \sum_{i<j}\sum t_{ij}$ be the total number of expressions of no preference. The likelihood function (4) then becomes

$$L(\underline{\pi},\nu;\ \underline{W},\underline{T}) = \nu^t \prod_{i=1}^{m} \pi_i^{\frac{1}{2}s_i} \Big/ \prod_{i<j} (\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j})^{r_{ij}} \qquad (5)$$

and the statistic $(\underline{s},t)$, where $\underline{s} = (s_1, \cdots, s_m)$ is a sufficient statistic for $(\underline{\pi},\nu)$ of the model (3).

Maximizing $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ over R, it follows that the maximum likelihood estimates $(\underline{p},\hat{\nu})$ of $(\underline{\pi},\nu)$ are obtained as a solution to the system of equations

$$s_i/p_i - \sum_j r_{ij}(2 + \hat{\nu}\sqrt{p_j/p_i})/(p_i + p_j + \hat{\nu}\sqrt{p_ip_j}), \quad i=1,\cdots,m \qquad (6)$$

$$t/\hat{\nu} - \sum_{i<j}\sum r_{ij}\sqrt{p_ip_j}/(p_i + p_j + \hat{\nu}\sqrt{p_ip_j}) \qquad (7)$$

Theorem 2. *Under the assumptions on $\underline{W}$ and $\underline{T}$, the likelihood $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ has a unique maximum over R at $(\underline{p},\hat{\nu})$, the solution to (6) and (7) for which $\sum_{i=1}^{m} p_i = 1$.*

Proof: Let $(\underline{p},\hat{\nu})$ be any member of R which satisfies (6) and (7). Define $\underline{C} = [c_{ij};\ i,j = 1, \cdots, m]$ where $c_{ij} = r_{ij}\hat{\nu}\sqrt{p_ip_j}/(p_i + p_j + \hat{\nu}\sqrt{p_ip_j})$,

$\underline{D} = [d_{ij};\ i,j = 1, \cdots, m]$ where $d_{ij} = r_{ij}\, p_i/(p_i + p_j + \hat{\nu}\sqrt{p_i p_j})$. By letting $\underline{C}$ and $\underline{D}$ correspond to $\underline{T}$ and $\underline{W}$, one may define $a_{ij} = d_{ij} + d_{ji} + c_{ij}$, $b_i = 2\sum_j d_{ij} + \sum_j c_{ij}$, and $c = \sum_{i<j}\sum c_{ij}$ to correspond to r_{ij}, s_i, and t, respectively.

It then follows that

$$a_{ij} = \frac{r_{ij}\, p_i}{(p_i + p_j + \hat{\nu}\sqrt{p_i p_j})} + \frac{r_{ij}\, p_j}{(p_i + p_j + \hat{\nu}\sqrt{p_i p_j})} + \frac{r_{ij}\, \hat{\nu}\sqrt{p_i p_j}}{(p_i + p_j + \hat{\nu}\sqrt{p_i p_j})} = r_{ij} \tag{8}$$

and from the definitions above in conjunction with (6) and (7) that

$$b_i = 2p_i \sum_j r_{ij}/(p_i + p_j + \hat{\nu}\sqrt{p_i p_j}) + \hat{\nu} \sum_j r_{ij}\sqrt{p_i p_j}/(p_i + p_j + \hat{\nu}\sqrt{p_i p_j}) = s_i \tag{9}$$

and

$$c = \hat{\nu} \sum_{i<j}\sum r_{ij}\sqrt{p_i p_j}/(p_i + p_j + \hat{\nu}\sqrt{p_i p_j}) = t. \tag{10}$$

With the use of (8), (9), and (10) in (5), it is seen that

$$L(\underline{\pi}, \nu;\ \underline{W}, \underline{T}) = L(\underline{\pi}, \nu;\ \underline{D}, \underline{C})$$

over the region R. The lemma is now used to show that $L(\underline{\pi}, \nu;\ \underline{D}, \underline{C})$ has a unique maximum over R. Let $(\underline{p}, \hat{\nu})$ be any solution to (6) and (7). Since $p_i > 0$, $i = 1, \cdots, m$, one has $d_{ij} > 0$ for $i \neq j$. Note that $(K\underline{p}, \hat{\nu})$, where K is any positive constant, is also a solution to (6) and (7). Hence one may assign p_1 to the first object and consider the chain which links the first object and the j^{th} object. One may now proceed along the chain, at each stage maximizing the factor

$$\left[\frac{\pi_k}{\pi_k + \pi_\ell + \nu\sqrt{\pi_k \pi_\ell}}\right]^{d_{k\ell}} \left[\frac{\pi_\ell}{\pi_k + \pi_\ell + \nu\sqrt{\pi_k \pi_\ell}}\right]^{d_{\ell k}} \left[\frac{\nu\sqrt{\pi_k \pi_\ell}}{\pi_k + \pi_\ell + \nu\sqrt{\pi_k \pi_\ell}}\right]^{c_{k\ell}}$$

of $L(\underline{\pi},\nu;\ \underline{D},\underline{C})$ by assigning values through the lemma to adjacent worth parameters π_k and π_ℓ. Since $d_{k\ell}/d_{\ell k} = p_k/p_\ell$, one must assign p_ℓ to π_ℓ once p_k has been assigned to π_k. Thus a unique p_j must be assigned to π_j, $j = 2, \cdots, m$, and hence there is a unique $(\underline{p},\hat{\nu})$ in R which maximizes $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$.

The equations (6) and (7) cannot be solved explicitly when $m > 2$, and hence an iterative procedure must be employed. In Davidson [7], an iterative procedure is described for which each new iterate is shown to result in an increase in the likelihood, with the likelihood remaining unchanged if and only if (6) and (7) are satisfied. From the proof of Theorem 1, one can confine attention to a bounded subset of R which contains the sequence of iterates. Since $L(\underline{\pi},\nu;\ \underline{W},\underline{T})$ is continuous and bounded, the monotone sequence of values of the likelihood converges and the corresponding sequence of iterates converges to $(\underline{p},\hat{\nu})$, the unique maximum likelihood estimates.

3. A property of the maximum likelihood ranking

The maximum likelihood estimates $\underline{p}$ of the worth parameters $\underline{\pi}$ can be used to rank the set of m objects. This is true for both the Bradley-Terry model and for its generalization (3) to allow for expressions of no preference.

It was noted by Ford [8] that in the case of a balanced paired comparison experiment in which ties are not permitted, e.g., major league baseball, the ranking obtained from the maximum likelihood estimates of the parameters of the Bradley-Terry model is the same as that obtained from the total numbers of wins. In the event that ties are permitted a common ranking system, used for example in hockey and soccer competition, is based on the points accumulated when a team is awarded 2, 1 or 0 points for a win, tie or loss, respectively. This is precisely the ranking based on the scores $\underline{s} = (s_1, \cdots, s_m)$. For a balanced paired comparison experiment this ranking is in agreement with that obtained from $\underline{p}$ for the model (3) (cf. Davidson [7]).

Optimal properties of these ranking procedures have been established by Bühlmann and Huber [5] and Huber [9] when consideration is restricted to ranking procedures that are invariant to a relabeling of the objects. When ties are not permitted, Bühlmann and Huber [5] have shown that the ranking based on the total numbers of wins is uniformly best with respect to an "acceptable" loss function if and only if the underlying probability structure is given by the Bradley-Terry model. A loss function is said to be "acceptable" if the loss does not decrease when the ranking is made worse by interchanging two items. When ties are permitted, Huber [9] has shown that under the model (3) with ν known, the ranking based on the scores $\underline{s}$ is uniformly best with respect to an "acceptable" loss function. A key aspect of the development in [9] is that $\underline{s}$ is a sufficient statistic for the model (3) with ν known. Thus it follows that for balanced paired comparison experiments under the model (1) or the model (3) with ν known, the ranking determined by the maximum likelihood estimates $\underline{p}$ of $\underline{\pi}$ is optimal.

There are situations where a balanced paired comparison experiment is not feasible. For example, in preference testing experiments there may be groups of objects which are produced in different locales so that within group comparisons are more easily obtained. In such cases it may be possible to divide the objects into groups with a constant number of comparisons being obtained between any two objects from the same group and a constant number of comparisons between any two objects from two different groups. This system is presently employed in major league hockey (N.H.L.) and major league baseball (N.L. and A.L.) where the number of games between any two teams of the same division is greater than the number of games between two teams of different divisions. The results described in the above paragraph can be generalized, in part, to such situations.

Let the set of m objects be partitioned into G groups and let m_g be the size of group g, g = 1, $\cdots$, G. Suppose further that r_{gg} responses are obtained in the comparison of any two objects from group g, and that r_{gh} responses are obtained in the comparison of each object in group g with each object in group h. In this context the parameters for the model (3) can be designated by $(\underline{\pi},\nu)$ where $\underline{\pi} = [\pi_{ig};\ i = 1, \cdots, m_g,\ g = 1, \cdots, G]$, their maximum likelihood estimates by $(\underline{p},\nu)$ where $\underline{p} = [p_{ig};\ i = 1, \cdots, m_g,\ g = 1, \cdots, G]$, and the sufficient statistic by $(\underline{s},t)$ where $\underline{s} = [s_{ig};\ i = 1, \cdots, m_g,\ g = 1, \cdots, G]$ is the vector of scores and t is the total number of ties.

Theorem 3. For a partially balanced paired comparison experiment, the ranking of the m objects based on the maximum likelihood estimates $\underline{p}$ of $\underline{\pi}$ for the model (3) has the property that the subranking for group g based on $\underline{p}_g = (p_{1g}, \cdots, p_{m_g g})$ agrees with that based on $\underline{s}_g = (s_{1g}, \cdots, s_{m_g g})$, $g = 1, \cdots, G$.

Proof: The equations (6), which together with (7) yield the unique maximum likelihood estimates $(\underline{p},\hat{\nu})$ under the assumptions on $\underline{W}$ and $\underline{T}$, can be written

$$\frac{s_{ig}}{p_{ig}} = r_{gg} \sum_{j\neq i} \frac{2 + \hat{\nu}\sqrt{p_{jg}/p_{ig}}}{(p_{ig} + p_{jg} + \hat{\nu}\sqrt{p_{ig}p_{jg}})} + \sum_{h\neq g} r_{gh} \sum_{j=1}^{m_h} \frac{2 + \hat{\nu}\sqrt{p_{jh}/p_{ig}}}{(p_{ig} + p_{jh} + \hat{\nu}\sqrt{p_{ig}p_{jh}})}$$

for $i = 1, \cdots, m_g$, $g = 1, \cdots, G$. It then follows that

$$s_{ig} - s_{kg} = (\sqrt{p_{ig}} - \sqrt{p_{kg}})\, K_g(\underline{p},\hat{\nu})$$

where

$$K_g(\underline{p},\hat{\nu}) = \sum_{h=1}^{G} r_{gh} \sum_{j=1}^{m} \frac{2(\sqrt{p_{ig}} + \sqrt{p_{kg}})p_{jh} + \hat{\nu}\sqrt{p_{ig}p_{kg}p_{jh}} + \hat{\nu}\, p_{jh}^{3/2}}{(p_{ig} + p_{jh} + \hat{\nu}\sqrt{p_{ig}p_{jh}})(p_{kg} + p_{jh} + \hat{\nu}\sqrt{p_{kg}p_{jh}})}$$

remains positive in that $(\underline{p},\hat{\nu}) \in R$ so that each component of $\underline{p}$ is positive.

It should be noted that the ranking of all m objects based on $\underline{p}$ does not necessarily agree with that based on $\underline{s}$ unless $r_{gh} = r$ for $g, h = 1, \cdots, G$.

REFERENCES

[1] Bradley, R. A. Rank analysis of incomplete block designs II. Additional tables for the method of paired comparisons. Biometrika 41 (1954), 502-537.

[2] Bradley, R. A. Rank analysis of incomplete block designs III. Some large-sample results on estimation and power for a method of paired comparisons. Biometrika 42 (1955), 450-470.

[3] Bradley, R. A. Another interpretation of a model for paired comparisons. Psychometrika 30 (1965), 315-318.

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

[5] Bühlmann, H. and Huber, P. J. Pairwise comparison and ranking in tournaments. Ann. Math. Statist. 34 (1963), 501-510.

[6] David, H. A. The Method of Paired Comparisons. (Hafner, New York, 1963).

[7] Davidson, R. R. On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. J. Amer. Statist. Assoc. 65 (1970), 317-328.

[8] Ford, L. R., Jr. Solution of a ranking problem from binary comparisons. Amer. Math. Monthly 54 (1957), 28-33.

[9] Huber, P. J. Pairwise comparison and ranking: optimum properties of the row sum procedure. Ann. Math. Statist. 34 (1963), 511-520.

[10] Rao, P. V. and Kupper, L. L. Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. J. Amer. Statist. Assoc. 62 (1967), 194-204.

[11] Zermelo, E. Die Berechnung der Turnier-Ergebnisse als eins Maximal-problem der Wahrscheinlichketsrechnung. Math. Zeitschrift 29 (1929), 436-460.