The Annals of Statistics
1999, Vol. 27, No. 3, 1041–1060

ASYMPTOTICS WHEN THE NUMBER OF PARAMETERS TENDS TO INFINITY IN THE BRADLEY–TERRY MODEL FOR PAIRED COMPARISONS

BY GORDON SIMONS AND YI-CHING YAO

University of North Carolina and Academia Sinica

We are concerned here with establishing the consistency and asymptotic normality for the maximum likelihood estimator of a "merit vector" $(u_0, \ldots, u_t)$, representing the merits of $t+1$ teams (players, treatments, objects), under the Bradley–Terry model, as $t \to \infty$. This situation contrasts with the well-known Neyman–Scott problem under which the number of parameters grows with t (the amount of sampling), and for which the maximum likelihood estimator fails even to attain consistency. A key feature of our proof is the use of an effective approximation to the inverse of the Fisher information matrix. Specifically, under the Bradley–Terry model, when teams i and j with respective merits u_i and u_j play each other, the probability that team i prevails is assumed to be $u_i/(u_i + u_j)$. Suppose each pair of teams play each other exactly n times for some fixed n. The objective is to estimate the merits, u_i's, based on the outcomes of the $nt(t+1)/2$ games. Clearly, the model depends on the u_i's only through their ratios. Under some condition on the growth rate of the largest ratio u_i/u_j $(0 \le i, j \le t)$ as $t \to \infty$, the maximum likelihood estimator of $(u_1/u_0, \ldots, u_t/u_0)$ is shown to be consistent and asymptotically normal. Some simulation results are provided.

1. Introduction. We are concerned here with proving the consistency and asymptotic normality for the maximum likelihood estimator of a "merit vector" $(u_0, u_1, \ldots, u_t)$, with $u_i > 0$ representing the merits of $t+1$ teams (players, treatments, objects) under the Bradley–Terry model, as t goes to infinity. This situation contrasts with that of a variety of interesting classical examples, described by Neyman and Scott (1948), under which the number of parameters grows with t (the amount of sampling) and for which the maximum likelihood estimator fails even to attain consistency. A key feature of our proof is the use of a remarkably accurate approximation to the inverse of the Fisher information matrix.

The Bradley–Terry model may be described (for teams) as follows. A set of $t+1$ teams play among themselves, by pairs with independent outcomes. When teams i and j, with respective merits u_i and u_j, play each other, the probability that team i prevails is assumed to take the form

$$p_{i,j} = \frac{u_i}{u_i + u_j}, \qquad i, j = 0, \ldots, t;\ i \neq j. \tag{1.1}$$

Received June 1998; revised March 1999.

AMS 1991 *subject classifications*. Primary 60F05; secondary 62J15, 62F12, 62E20.

Key words and phrases. Central limit theorem, Bradley–Terry model, consistency, maximum likelihood estimator, Fisher information matrix.

Under this model, if team i prevails over team j a total of $a_{i,j}$ times as a result of $n_{i,j}$ pairings, then one obtains the relationships

$$a_{i,j} + a_{j,i} = n_{i,j} = n_{j,i}, \qquad i, j = 0, \dots, t;\ i \neq j,$$

and the likelihood function assumes the form

$$\prod_{\substack{i,j=0 \\ i\neq j}}^{t} p_{i,j}{}^{a_{i,j}} = \frac{\prod_{i=0}^{t} u_i{}^{a_i}}{\prod_{0\le i<j\le t} (u_i + u_j)^{n_{i,j}}},$$

where

$$a_i = \sum_{j\neq i} a_{i,j},$$

the total number of victories by team i. Observe that $(a_0, \dots, a_t)$ is a sufficient statistic.

Clearly, the model depends on the u_i's only through their ratios. Thus, for the sake of identifiability, it is commonly assumed that $\sum_{i=0}^{t} u_i = 1$. This has the effect of reducing the parameter space to t dimensions. Since we shall allow t to go to infinity, we find it more convenient to normalize the u_i's by setting $u_0 = 1$. So, in effect, we are seeking the maximum likelihood estimators $\hat{u}_1, \dots, \hat{u}_t$ (with $\hat{u}_0 = 1$) for the ratios $u_1/u_0, \dots, u_t/u_0$.

The consistency and asymptotic normality of the maximum likelihood estimators are expected when t is fixed and all of the $n_{i,j}$ tend to infinity. However, what can one expect when t is large and the $n_{i,j}$ are relatively small? For example, in the National Football League, there are 30 teams and no two teams play each other more than twice each season. While this particular case seems difficult to study, we shall restrict our attention to a relatively simple but nontrivial asymptotic setting where $n_{i,j} = n$ for all pairs (i, j) for some finite fixed n and t is allowed to go to infinity. (Indeed, many basketball conferences under the purview of the NCAA have each pair of teams play each other exactly twice.) Consequently, $a_0 + \cdots + a_t = nt(t+1)/2$.

A formal calculation shows that the maximum likelihood estimators $(\hat{u}_1, \dots, \hat{u}_t)$ satisfy the equations

$$\frac{a_i}{n} = \sum_{\substack{j=0 \\ j\neq i}}^{t} \frac{\hat{u}_i}{\hat{u}_i + \hat{u}_j}, \qquad i = 1, \dots, t\ (\hat{u}_0 = 1). \tag{1.2}$$

Also, observe that

$$E(a_i) = n \sum_{\substack{j=0 \\ j\neq i}}^{t} \frac{u_i}{u_i + u_j}, \qquad i = 1, \dots, t. \tag{1.3}$$

So the method of moments [under which each a_i is equated to $E(a_i)$] and the method of maximum likelihood both lead to the likelihood equations (1.2), and hence to the same estimators $\hat{u}_1, \dots, \hat{u}_t$.

Unfortunately, the equations in (1.2) can fail to provide a solution within the parameter space, that is, satisfying the inequalities,

$$\hat{u}_i > 0, \qquad i = 1, \dots, t. \tag{1.4}$$

But, as noted by Zermelo (1929) and (independently) by Ford (1957), a unique solution can be found to (1.2), satisfying (1.4), under what we shall call Condition A.

CONDITION A. For every partition of the teams (players, treatments, objects) into two nonempty sets, a team in the second set has beaten a team in the first at least once.

Fortunately, we can insure, in Section 2, that Condition A occurs with probability approaching one as $t \to \infty$ providing we impose some bounding on the merits. It is enough (Lemma 1) that

$$\max_{0 \le i,\, j \le t} \frac{u_i}{u_j} = o\left(\sqrt{\frac{t}{\log t}}\right) \text{ as } t \to \infty.$$

And with no more than this assumption, we are able to establish the consistency of the maximum likelihood estimators, as given in Theorem 1 below.

Let

$$\begin{aligned} M_t &:= \max_{0 \le i,\, j \le t} \frac{u_i}{u_j}, \qquad \delta_t := 8M_t \sqrt{\frac{\log(t+1)}{n\, t}} \quad \text{and} \\ \Delta u_i &:= \frac{\hat{u}_i - u_i}{u_i}, \qquad i = 0, \dots, t, \end{aligned} \tag{1.5}$$

where for definiteness when Condition A fails we set $\hat{u}_i = 1$ for $i = 1, \dots, t$. Otherwise, the $\hat{u}_i$'s ($i = 1, \dots, t$) are defined by the likelihood equations given in (1.2).

Our statement of consistency takes the following form.

THEOREM 1. *If*

$$M_t = o\left(\sqrt{\frac{t}{\log t}}\right) \quad \textit{as } t \to \infty, \tag{1.6}$$

then

$$\max_{i=0,\dots,t} |\Delta u_i| \le \max_{0 \le i,\, j \le t} |\Delta u_i - \Delta u_j| = O_p(\delta_t) = o_p(1) \quad \textit{as } t \to \infty. \tag{1.7}$$

On the other hand, deriving the asymptotic normality of the maximum likelihood estimators is much more involved and requires approximating the inverse of the Fisher information matrix. This is done in Section 3, where Theorem 2 below is proved.

Let $V_t = (v_{i,j})$ denote the covariance matrix of $(a_1, \dots, a_t)$, where

$$(1.8)\qquad v_{i,i} = \sum_{\substack{k=0\\k\neq i}}^{t} \frac{n\,u_i\,u_k}{(u_i+u_k)^2}, \qquad v_{i,j} = -\frac{n\,u_i\,u_j}{(u_i+u_j)^2}, \qquad i, j = 1, \dots, t,\ j \neq i.$$

Also, let $v_{0,0} = \sum_{i=1}^{t}((n\,u_0\,u_i)/(u_0+u_i)^2) = \sum_{i=1}^{t}((n\,u_i)/(1+u_i)^2)$. Note that V_t is the Fisher information matrix for the parameterization $(\log u_1, \dots, \log u_t)$. Furthermore, we introduce a $t \times t$ matrix $S_t = (s_{i,j})$ as a close approximation to V_t^{-1}, where

$$(1.9)\qquad s_{i,j} = \frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{0,0}}, \qquad i, j = 1, \dots, t,$$

and where $\delta_{i,j}$ is the Kronecker delta.

THEOREM 2. *If $M_t = o(t^{1/10}/(\log t)^5)$, then for each fixed $r \geq 1$, as $t \to \infty$, the vector $(\Delta u_1, \dots, \Delta u_r)'$ is asymptotically normally distributed with mean* 0 *and covariance matrix given by the upper left $r \times r$ block of S_t, defined in* (1.9).

Section 4 contains simulation results along with some discussions. The proofs of the supporting lemmas in Sections 2 and 3 are relegated to Section 5.

It should be remarked that our asymptotic setup assumes implicitly that the u_i's depend on t. But, for notational simplicity, the dependence of u_i on t is suppressed throughout the paper.

Finally, we close this section by giving a very brief history of the Bradley–Terry model. Zermelo (1929) is generally credited with being the first person to study the Bradley–Terry model, using merits to model probabilities for pairwise comparisons, as described in (1.1), and showing that the likelihood equations (1.2) uniquely determine the merit of every team when Condition A holds (apart from scaling).

Zermelo also proposed an iterative algorithm to solve equations (1.2) and established the convergence of the algorithm to the true solution (under Condition A). The model, and various parts of the theory, have been rediscovered over the intervening years: Bradley and Terry (1952), Ford (1957), Jech (1983). Moreover, Bradley (1954) proposed a statistical test of the hypothesis that the model is correct. An informative summary of these papers can be found in Stob (1984). It is also worth mentioning that a large portion of Chapter 4 of David (1988) is devoted to a discussion of the Bradley–Terry model, including maximum likelihood estimation, confidence intervals, hypothesis testing and goodness-of-fit tests of the model.

There is a vast literature on paired comparisons models and related statistical analyses. See Davidson and Farquhar (1976), which provides a list of over 350 papers on this topic. Among the many paired comparisons models, it is worth noting that the Bradley–Terry model is the only one that satisfies certain desirable properties. The interested reader is referred to Bühlmann and Huber (1963), Suppes and Zinnes [(1963), pages 49, 50], Colonius (1980) and Jech (1983).

2. Consistency. The consistency result, Theorem 1, is a simple consequence of the following three lemmas, the proofs of which are relegated to Section 5.

LEMMA 1. *If M_t satisfies* (1.6), *then P(Condition* A *is satisfied*) $\to 1$ *as* $t \to \infty$. *Thus, with probability approaching* 1 *as* $t \to \infty$, *the estimators* $\hat{u}_1, \ldots, \hat{u}_t$ *are specified uniquely by the equations given in* (1.2), *and satisfy*

$$\max_{i=0,\ldots,t} \left| \sum_{j=0}^{t} \left\{ \frac{\hat{u}_i}{\hat{u}_i + \hat{u}_j} - \frac{u_i}{u_i + u_j} \right\} \right| = \frac{1}{n} \max_{i=0,\ldots,t} |a_i - E(a_i)|.$$

LEMMA 2.

$$P\Big(\max_{i=0,\ldots,t} |a_i - E(a_i)| < \sqrt{n\, t \log(t+1)} \Big) \ge \frac{t-1}{t+1} \to 1 \quad \textit{as } t \to \infty. \tag{2.1}$$

Thus, *if* M_t *satisfies* (1.6), *then with probability approaching* 1 *as* $t \to \infty$, *the estimators* $\hat{u}_1, \ldots, \hat{u}_t$ *satisfy the inequality*

$$\max_{i=0,\ldots,t} \left| \sum_{j=0}^{t} \left\{ \frac{\hat{u}_i}{\hat{u}_i + \hat{u}_j} - \frac{u_i}{u_i + u_j} \right\} \right| < \sqrt{\frac{t \log(t+1)}{n}}. \tag{2.2}$$

LEMMA 3. *If the estimators* $\hat{u}_1, \ldots, \hat{u}_t$ *satisfy inequality* (2.2), *then*

$$\max_{0 \le i,\, j \le t} |\Delta u_i - \Delta u_j| < \frac{\delta_t}{1 - \delta_t}, \tag{2.3}$$

where δ_t *is defined in* (1.5). *Thus*, *if* M_t *satisfies* (1.6), *then*

$$P\left(\max_{0 \le i,\, j \le t} |\Delta u_i - \Delta u_j| < \frac{\delta_t}{1 - \delta_t} \right) \to 1 \quad \textit{as } t \to \infty.$$

Given these lemmas, if M_t satisfies (1.6), then $\delta_t/(1 - \delta_t) = O(\delta_t) = o(1)$ as $t \to \infty$, so that

$$\max_{i=0,\ldots,t} |\Delta u_i| \le \max_{0 \le i,\, j \le t} |\Delta u_i - \Delta u_j| = O_p(\delta_t) = o_p(1) \quad \text{as } t \to \infty,$$

since $\Delta u_0 = (1 - 1)/1 = 0$. This proves Theorem 1.

Clearly, some control on the growth of M_t is needed. For if some u_i's are very large, and/or others are very small, corresponding to a large value of M_t, the teams with relatively poor merits stand very little chance of beating those with relatively large merits, thereby making estimation very difficult. It would be of interest to see if the condition $M_t = o(\sqrt{t/\log t})$ can be relaxed.

Alternative statements of consistency are possible. For instance, it is possible to show, under alternative assumptions, that $\max_{i=1,\ldots,t} |\hat{u}_i - u_i| = o_p(1)$ as $t \to \infty$. The form appearing in Theorem 1 seems natural for the present context.

3. Central limit theorems. In this section, we establish the central limit theorem (Theorem 2) for the estimated merits $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_t$ as the number of teams $t+1 \to \infty$, with the number of repetitions n held fixed.

It is easily checked that

$$\frac{n\,t}{4M_t} \le v_{i,i} \le \frac{n\,t}{4}, \qquad i = 0, 1, \dots, t, \tag{3.1}$$

where $M_t \ge 1$ and $v_{i,i}$ are defined in (1.5) and (1.8), respectively ($t \ge 1$). It is only necessary to check, for $i, k = 0, 1, \dots, t$, $k \ne i$, that

$$4 \le \left(\frac{u_i}{u_k} + 1\right)\left(\frac{u_k}{u_i} + 1\right) \le 4\,M_t. \tag{3.2}$$

As noted before, the matrix V_t, defined in (1.8), is the Fisher information matrix for the parameterization $(\log u_1, \dots, \log u_t)$. With it, one can immediately describe a classical central limit theorem for the maximum likelihood estimator of $(\log u_1, \dots, \log u_t)$.

PROPOSITION 1. *As $n \to \infty$, with $t \ge 1$ fixed,* $(\log \hat{u}_1 - \log u_1, \dots, \log \hat{u}_t - \log u_t)$ *is asymptotically normally distributed with mean zero and covariance matrix ${V_t}^{-1}$.*

This is equivalent to the central limit theorem described by Bradley (1955). Moreover [since $\log y - \log x = (y-x)/x + O((y-x)^2)$ as $y \to x > 0$], it is also equivalent to the following proposition.

PROPOSITION 2. *As $n \to \infty$, with $t \ge 1$ fixed,* $(\Delta u_1, \dots, \Delta u_t)$ *is asymptotically normally distributed with mean zero and covariance matrix ${V_t}^{-1}$.*

It follows that both t-dimensional random vectors,

$${V_t}^{1/2}(\log \hat{u}_1 - \log u_1, \dots, \log \hat{u}_t - \log u_t)' \quad \text{and} \quad {V_t}^{1/2}(\Delta u_1, \dots, \Delta u_t)' \tag{3.3}$$

converge, as $n \to \infty$, to a normal limit with independent $N(0,1)$ elements, where ${V_t}^{1/2}$ denotes the symmetric positive definite square root of V_t.

It seems reasonable to ask whether something similar happens if we reverse the roles of n and t, holding n fixed and letting t go to infinity. What kind of a central limit theorem, if any, might one expect? Two ideas come to mind, one linked to ${V_t}^{-1}$, as in Propositions 1 and 2, the other linked to ${V_t}^{1/2}$, as in (3.3) and the remark following. Thus we might hope to be able to conclude, under appropriate assumptions, with $r \ge 1$ and n held fixed and as $t \to \infty$, that:

1. $(\Delta u_1, \dots, \Delta u_r)$ is asymptotically normally distributed with mean zero and covariance matrix the upper left $r \times r$ block of ${V_t}^{-1}$, or
2. The first r elements of the vectors in (3.3) converge, as $t \to \infty$, to a normal limit with independent $N(0,1)$ elements.

We have no idea whether a theorem of the second type is possible, and we don't have a *direct* proof of a theorem of the first type. Instead, we shall prove a theorem (Theorem 2) in the spirit of the first type but with ${V_t}^{-1}$ replaced by a close symmetric positive definite approximation, $S_t = (s_{i,j})$, defined in (1.9). The quality of the approximation improves as t grows, as described in Lemma 4 below. An implication of the first type (involving ${V_t}^{-1}$, Theorem 2a below) will be inferred from Theorem 2.

There are several advantages to working with S_t instead of ${V_t}^{-1}$. The most obvious advantage is that S_t has an explicit form, which plays a vital role in our method of proof. (This is especially clear in the proof of Lemma 7, where cancellations among various remainder terms are possible only because of the specific form of S_t.) Second, it simplifies the description of the covariance matrices associated with the limiting normal distributions. Closely related, it permits more explicit descriptions of statistical procedures based on our central limit theorem. Examples of these appear at the end of this section.

A precise statement of the quality of the approximation S_t, for V_t^{-1}, has been established by Simons and Yao (1998) elsewhere. As it applies here, it takes the following form.

LEMMA 4. *If*

$$W_t := {V_t}^{-1} - S_t, \tag{3.4}$$

then, for M_t *defined in* (1.5),

$$\| W_t \| \le \left(\frac{(M_t+1)^4}{4n{M_t}^2} + \frac{(M_t+1)^6}{16n{M_t}^3} \right) \times \frac{1}{t^2} \le \frac{4{M_t}^2(M_t+1)}{n\,t^2}, \tag{3.5}$$

where here (*and elsewhere*) $\| A \|$ *denotes* $\max_{i,j} |a_{i,j}|$ *for a general matrix* $A = (a_{i,j})$.

The proof of Lemma 4, as well as those of Lemmas 5–7, is given in Section 5.

The underpinning of any central limit theorem for the estimators $\hat{u}_1, \dots, \hat{u}_t$ is a central limit theorem for the a_i's.

PROPOSITION 3. *If*

$$M_t = o(t) \quad \text{as } t \to \infty, \tag{3.6}$$

then as $t \to \infty$:

(i) *The components of* $(a_0 - E(a_0), \dots, a_r - E(a_r))$ *are asymptotically independent and normally distributed with variances* $v_{0,0}, \dots, v_{r,r}$, *respectively, for each fixed integer* $r \ge 1$, *and*

(ii) *More generally*,

$$\sum_{i=0}^{t} c_i \frac{a_i - E(a_i)}{\sqrt{v_{i,i}}}$$

is asymptotically normally distributed with mean zero and variance $\sum_{i=0}^{\infty} c_i^2$ *whenever* $c_0, c_1, \dots$ *are fixed constants and the latter sum is finite.*

This is a bit surprising because the statistics $a_0, a_1, \dots, a_t$ sum to $n\,t(t+1)/2$, and, hence, are linearly *dependent*. But, in view of (3.1), assumption (3.6) guarantees that the variances $v_{0,0}, \dots, v_{r,r}$ diverge as $t \to \infty$, thereby eliminating the influence, on the asymptotic limit, of the outcomes of games between teams i and j, $0 \le i < j \le r$, a finite number of games. Thus, for the purpose of proving part (i), one may replace the statistics $a_0, a_1, \dots, a_r$, respectively, by the *independent* random variables

$$\tilde{a}_i := a_{i,r+1} + a_{i,r+2} + \cdots + a_{i,t}, \quad i = 0, 1, \dots, r \ (t > r).$$

Then part (i) can easily be established by referring to [Loève's (1963), page 277] "bounded case," or by verifying the Lindeberg condition.

Part (ii) shows that it is possible to include the entire minimal sufficient statistic in a single central limit theorem statement. It follows from part (i) and the fact that

$$\lim_{r\to\infty} \limsup_{t\to\infty} \operatorname{Var}\left(\sum_{k=r+1}^{t} c_i \frac{a_i - E(a_i)}{\sqrt{v_{i,i}}} \right) = 0 \tag{3.7}$$

[cf. Theorem 4.2 of Billingsley (1968)]. To prove (3.7), it suffices to show that the eigenvalues of the covariance matrix of $(a_i - E(a_i))/\sqrt{v_{i,i}}$, $i = r+1, \dots, t$ are bounded by 2 (for all $r < t$). A simple, but elegant, proof of this (communicated to us by Ancel Mewborn of the University of North Carolina), based on the Perron–Frobenius theory, is described in Lemma 5 below. We shall only use part (i) in what follows.

LEMMA 5. *If A is a symmetric positive definite matrix with diagonal elements equal to* 1, *and with negative off-diagonal elements, then its largest eigenvalue is less than* 2.

Let $\mathbf{a} := (a_1, a_2, \dots, a_t)'$, and observe that $(S_t(\mathbf{a} - E(\mathbf{a})))_i$, the ith element of $S_t(\mathbf{a} - E(\mathbf{a}))$, is equal to $(a_i - E(a_i))/v_{i,i} - (a_0 - E(a_0))/v_{0,0}$ [because the elements of $\mathbf{a} - E(\mathbf{a})$ add to $-(a_0 - E(a_0))$], which has asymptotic variance $1/v_{i,i} + 1/v_{0,0}$ [when (3.6) holds]. Likewise, the asymptotic covariance of the ith and jth elements is $1/v_{0,0}$ when $i \ne j$. Therefore, Proposition 3 leads directly to the next proposition.

PROPOSITION 4. *If M_t satisfies* (3.6), *then, as $t \to \infty$, the first r rows of $S_t(\mathbf{a} - E(\mathbf{a}))$ are asymptotically normally distributed with covariance matrix given by the upper left $r \times r$ block of S_t, defined in* (1.9) $(r \ge 1)$.

The reader is reminded that the asymptotic variances $v_{0,0}, \dots, v_{r,r}$, appearing in part (i) of Proposition 3, depend on t, and, consequently, the asymptotic covariance matrix described in this proposition also depends on t.

The importance of Proposition 4 is that it yields a central limit theorem for $\Delta u_1, \Delta u_2, \ldots$ under an additional assumption on M_t. Specifically, what is needed is a strong enough assumption that

$$\Delta u_i = (S_t(\mathbf{a} - E(\mathbf{a})))_i(1 + o_p(1)) \quad \text{as } t \to \infty,\ i = 1, 2, \ldots, r, \tag{3.8}$$

or, what implies (3.8),

$$\Delta u_i = (S_t(\mathbf{a} - E(\mathbf{a})))_i + o_p(t^{-1/2}) \quad \text{as } t \to \infty,\ i = 1, 2, \ldots, r. \tag{3.9}$$

That (3.9) implies (3.8) is a simple consequence of the second inequality in (3.1) and the content of Proposition 4.

A strong enough assumption on M_t to guarantee (3.9) [and hence (3.8)] is $M_t = o(t^{1/10}/(\log t)^5)$, as described in Theorem 2.

The task of proving Theorem 2 is to establish (3.9). This will follow directly from

$$(W_t(\mathbf{a} - E(\mathbf{a})))_i = o_p(t^{-1/2}) \quad \text{as } t \to \infty, \tag{3.10}$$

where W_t is defined in (3.4), and by

$$\Delta u_i = ({V_t}^{-1}(\mathbf{a} - E(\mathbf{a})))_i + o_p(t^{-1/2}) \quad \text{as } t \to \infty. \tag{3.11}$$

Proofs of (3.10) and (3.11) are given, respectively, in Lemmas 6 and 7 below.

LEMMA 6. *If R_t denotes the covariance matrix of $W_t\mathbf{a}$, then*

$$\| R_t \| \le \frac{4(3 + M_t){M_t}^2}{n\, t^2}. \tag{3.12}$$

Consequently, (3.10) *holds if* $M_t = o(t^{1/3})$ *as* $t \to \infty$.

LEMMA 7. *If* $M_t = o(t^{1/10}/(\log t)^5)$ *as* $t \to \infty$, *then* (3.11) *holds.*

Of course, Theorem 2 has an analog along the lines of Proposition 1. That is, Theorem 2 remains true if we replace the vector $(\Delta u_1, \ldots, \Delta u_r)'$ by the vector $(\log \hat{u}_1 - \log u_1, \ldots, \log \hat{u}_r - \log u_r)'$. In addition, the following are consequences of Theorem 2.

THEOREM 2a. *If* $M_t = o(t^{1/10}/(\log t)^5)$, *then, as* $t \to \infty$, *the vector* $(\Delta u_1, \ldots, \Delta u_r)'$ *is asymptotically normally distributed with mean* 0 *and covariance matrix given by the upper left* $r \times r$ *block of* ${V_t}^{-1}$ ($V_t = (v_{i,j})$ *defined in* (1.8); $r \ge 1$).

THEOREM 2b. *If* $M_t = o(t^{1/10}/(\log t)^5)$, *then, as* $t \to \infty$, *the vector* $(\log \hat{u}_1, \ldots, \log \hat{u}_r)'$ *is asymptotically normally distributed with mean* $(\log u_1, \ldots,$

$\log u_r)'$ *and covariance matrix given by the upper left* $r \times r$ *block of* S_t, *defined in* (1.9) $(r \geq 1)$.

Theorem 2a follows from Theorem 2 and Lemma 4. Theorem 2b is equivalent to Theorem 2. (See the remark following the statement of Proposition 1.)

We conclude this section with several remarks.

REMARK 1. While it seems reasonable to suspect the growth rate $t^{1/10}/(\log t)^5$ in Theorem 2 is far from the best possible, it is not clear to us how one can significantly improve the rate. It may be tempting to conjecture that a condition similar to (1.6) [which is much less restrictive than $t^{1/10}/(\log t)^5$] is enough for the asymptotic normality to hold. In principle, one could carry out a simulation study to test whether condition (1.6) appears to be enough. However, such a simulation study would have to be very carefully designed since the behavior of the maximum likelihood estimators depends not only on M_t, but also on the configuration of all of the merits. We have not made such an attempt. Only a very limited simulation study is done and presented in the next section.

REMARK 2. We have only considered the case of $n_{i,j} = n$ for all pairs (i, j) for some finite fixed n. What can be said when the $n_{i,j}$ are not the same? For example, some of the games might be cancelled due to poor weather conditions. If the number of cancelled games is bounded, then this will have little effect on the maximum likelihood estimators when t is large, so that the consistency and asymptotic normality still hold. A more interesting case is when the $n_{i,j}$ are quite different from one another. Then the Fisher information matrix will take a more complicated form, and hence a new approximation to its inverse is needed in place of S_t. The result in Simons and Yao (1998) appears to be applicable when the $n_{i,j}$ are bounded between two fixed numbers. We plan to investigate this and other related situations in the future.

REMARK 3. By Theorems 2b, an approximate $100(1-\alpha)\%$ confidence interval for $\log(u_i/u_j)$ is $\log(\hat{u}_i/\hat{u}_j) \pm Z_{1-\alpha/2}\sqrt{1/\hat{v}_{i,i} + 1/\hat{v}_{j,j}}$, where $\hat{v}_{i,i}$ and $\hat{v}_{j,j}$ are the natural estimates of $v_{i,i}$ and $v_{j,j}$, found by replacing all $u_1, \ldots, u_t$ [appearing in their definitions in (1.8)] by their corresponding estimates, and where Z_β refers to the 100β percentile point of the standard normal distribution. Similarly, to test whether $u_i = u_j$ at level α, the hypothesis can be rejected if $|\log(\hat{u}_i/\hat{u}_j)| > Z_{1-\alpha/2}\sqrt{1/\hat{v}_{i,i} + 1/\hat{v}_{j,j}}$. Here, as suggested by a referee, instead of using $\hat{v}_{i,i}$ and $\hat{v}_{j,j}$, one may use the estimates of $v_{i,i}$ and $v_{j,j}$ under the null hypothesis of $u_i = u_j$. The same comment applies to the next remark.

REMARK 4. More generally, Theorems 1 and 2, can be used to produce multiple and simultaneous (Scheffé-type) confidence intervals for contrasts, and for testing the equality of more than two merits. For example, to test whether $u_1 = u_2 = u_3 = u_4$ at level α, the hypothesis can be rejected if

$$\left(\log\frac{\hat{u}_1}{\hat{u}_2},\ \log\frac{\hat{u}_2}{\hat{u}_3},\ \log\frac{\hat{u}_3}{\hat{u}_4}\right)\begin{pmatrix} \dfrac{1}{\hat{v}_{1,1}}+\dfrac{1}{\hat{v}_{2,2}} & \dfrac{-1}{\hat{v}_{2,2}} & 0 \\ \dfrac{-1}{\hat{v}_{2,2}} & \dfrac{1}{\hat{v}_{2,2}}+\dfrac{1}{\hat{v}_{3,3}} & \dfrac{-1}{\hat{v}_{3,3}} \\ 0 & \dfrac{-1}{\hat{v}_{3,3}} & \dfrac{1}{\hat{v}_{3,3}}+\dfrac{1}{\hat{v}_{4,4}} \end{pmatrix}^{-1}$$

$$\times\begin{pmatrix} \log\hat{u}_1/\hat{u}_2 \\ \log\hat{u}_2/\hat{u}_3 \\ \log\hat{u}_3/\hat{u}_4 \end{pmatrix} > \chi^2_{3,\,1-\alpha},$$

where $\chi^2_{3,\,1-\alpha}$ refers to the upper 100α percental point for the chi-square distribution with 3 degrees of freedom. A more succinct description of this rejection condition can be given as

$$\sum_{i=1}^{4}\hat{v}_{i,\,i}(\log\hat{u}_i)^2 - (\hat{v}_{1,\,1}+\cdots+\hat{v}_{4,\,4})^{-1}\left(\sum_{i=1}^{4}\hat{v}_{i,\,i}\log\hat{u}_i\right)^2 > \chi^2_{3,\,1-\alpha}.$$

While Theorem 2b enables one to construct confidence intervals for linear combinations of a finite number of log-merits, we do not know how to deal with the case of linear combinations involving all merits.

REMARK 5. As suggested by a referee, an interesting related problem is to impose a linear model on the merit parameters u_i: $\mathbf{u} = X\boldsymbol{\beta}$, and to study the large sample behavior of the maximum likelihood estimator of $\boldsymbol{\beta}$. Here is a challenging question: under what conditions on X and on the dimension of $\boldsymbol{\beta}$ (which grows with t) can one obtain consistency and asymptotic normality results? While we have no answer to this question, the interested reader is referred to Portnoy (1984, 1985, 1988) for some problems of a similar type, where the number of parameters tends to infinity.

4. Numerical studies. The question remains: how accurate are statistical applications based on Theorem 2? To assess this, we shall focus attention on 95% confidence intervals for $\log(u_j/u_i)$, as described in Remark 3, for particular pairs i and j, when the log-merits $\log u_k$, $k = 0, 1, \ldots, t$, assume the linear form $c\,k$ for a fixed value of $c \geq 0$.

As a practical matter, for any application of Theorem 2, it is essential that Condition A hold with high probability [so that, with high probability, the likelihood equations (1.2) properly define maximum likelihood estimates $\hat{u}_k$, $k = 1, \ldots, t$]. This probability, which depends on the size of t and the value of c, will be denoted by $p(t, c)$.

We have evaluated the function $p(t, c)$ by simulation for a variety of (t, c) values and found that $p(t, c)$ appears to increase in t and decrease in c. Moreover, while, by Lemma 1, $p(t, 0)$ tends to 1 as $t \to \infty$, it appears, based on our limited simulation results with $t \leq 200$, that for fixed $c > 0$, $p(t, c)$ converges to something less than 1 as $t \to \infty$. On the other hand, it is clear for fixed t that $p(t, c)$ goes to 0 as $c \to \infty$ since the strongest team beats everyone else with probability approaching 1.

Some sample calculations follow:

$p(t, 0) = 0.979$ and 0.99997 for $t = 10$ and 20;
$p(t, 0.05) = 0.972$ and 0.99948 for $t = 10$ and 20;
$p(t, 0.1) = 0.951$, 0.9945, 0.99863, 0.99867, 0.99873 for $t = 10$, 20, 50, 100, 200, respectively;
$p(t, 0.2) = 0.847$, 0.913, 0.92184, 0.92253, 0.92228 for $t = 10$, 20, 50, 100, 200, respectively.

Each number was generated from 100,000 simulations, yielding an approximate standard error of $0.003\sqrt{1 - p(t, c)}$. Thus the standard errors are roughly equal to 0.0009, 0.0003, 0.00009 for $p(c, t)$-values of about 0.9, 0.99, 0.999, respectively, and are large enough to account for the apparent lack of monotonicity in the last two entries. These calculations suggest that it is probably acceptable to evaluate the accuracy of 95% confidence intervals, based on Theorem 2, when $c \leq 0.1$ and $t \geq 20$. In Table 1, we consider the three values $t = 20$, 50, 100 for the six cases $c = 0.02$, 0.04, 0.06, 0.08, 0.10, 0.12, with the latter case (exceeding 0.1) thrown in for the sake of illustration.

Double entries in certain cells of Table 1 indicate that one or more failures of Condition A occurred during the simulation. The first entry, without parentheses, is the probability that Condition A and coverage of the indicated parameter both occur. The second entry, enclosed by parentheses, is the probability that Condition A fails *or* coverage occurs (the first probability plus the probability that Condition A fails). Consequently, the "true" coverage probability lies between these two numbers (assuming an accurate simulation) regardless of how one sets up the confidence interval when Condition A fails.

When $c = 0$, all merits are equal to 1, so that the choice of the index pair (i, j) is immaterial. The corresponding coverage probabilities are 95.71 (95.71), 94.05, 94.33, 94.86 for $t = 20$, 50, 100, 200, respectively.

These coverage probabilities, and most of those in the table, look quite respectable. Surprisingly, all are close to the targeted 95% when $t = 20$, even

TABLE 1
Coverage probabilities

t	(i,j)	$c = 0.02$	$c = 0.04$	$c = 0.06$	$c = 0.08$	$c = 0.1$	$c = 0.12$
20	**(0, 1)**	95.82 (95.82)	95.65 (95.67)	95.18 (95.28)	95.50 (95.74)	95.85 (96.44)	95.82 (96.98)
20	**(0, 20)**	95.80 (95.80)	95.42 (95.44)	94.97 (95.07)	94.82 (95.05)	94.62 (95.21)	94.25 (95.40)
20	**(9, 10)**	95.66 (95.67)	95.56 (95.58)	95.12 (95.22)	94.26 (94.49)	94.34 (94.93)	94.08 (95.24)
50	**(0, 1)**	95.08	95.39	95.34 (95.34)	95.48 (95.51)	95.86 (96.01)	95.83 (96.35)
50	**(0, 50)**	95.02	94.88	94.71 (94.71)	94.11 (94.13)	93.41 (93.56)	92.40 (92.92)
50	**(24, 25)**	95.32	95.00	94.97 (94.98)	94.84 (94.87)	94.86 (95.01)	94.12 (94.63)
100	**(0, 1)**	95.07	95.24	95.45	95.58 (95.60)	95.84 (95.98)	95.71 (96.26)
100	**(0, 100)**	94.97	94.66	93.80	92.36 (92.38)	89.95 (90.09)	86.03 (86.58)
100	**(49, 50)**	94.26	94.69	94.41	94.63 (94.65)	94.46 (94.60)	94.18 (94.72)

when c is as large as 0.12. Of course, it could be countered that the individual merits are close together when $t = 20$, with the ratio of largest to smallest merit equal to $e^{2.4} \doteq 11$. Even so, these coverage probabilities are very encouraging, for t as small as 20.

However, it will be observed that for each fixed $c > 0$, the quality of the approximation deteriorates with increasing t, thereby appearing to contradict the central limit theorem. But this is not so. For the ratio of largest to smallest merit grows exponentially fast with t when c is fixed. For instance, the ratios for $c = 0.1$ are $e^2 \doteq 7.4$, $e^5 \doteq 148$ and $e^{10} \doteq 22,000$ for $t = 20$, 50 and 100, respectively. Since Theorem 2 does not even accommodate a linear growth rate in t, an exponential growth rate is definitely excessive.

More meaningful comparisons can be made by letting c depend on t, with the product ct fixed. For instance, a one-to-one comparison between corresponding entries for $(c, t) = (0.1, 20)$ and $(c, t) = (0.04, 50)$, with $ct = 2$, shows modest but noticeable improvements.

Not surprisingly, the worst coverage probabilities occur when $(i, j) = (0, t)$ with $t = 50$ and 100, where two entries fall slightly below 90%. As a more extreme example, the corresponding coverage probability is only about 54% when $(c, t) = (0.1, 200)$.

In order to gain additional insight from the numbers in Table 1, we have added a second table. Two empirically determined numbers are provided for each cell in Table 2. The first (leftmost) examines those cases when Condition A occurs and the 95% interval estimate fails to cover the true parameter $[\log(u_j/u_i)]$; we record the percentage of time this failure occurs with the true parameter falling to the *left* of the confidence interval. The smallest such percentage, 44.6%, occurs when $t = 20$; it is only slightly smaller than (the ideal) 50%. Most of these percentages exceed 50%, a few by a significant amount when $(i, j) = (0, t)$.

The second (rightmost) numbers in Table 2, within parentheses, record the corresponding biases of the maximum likelihood estimators for the parameters in question divided by the estimated standard deviation. These are then multiplied by 100 and rounded to the nearest whole number. Thus the entries describe the "normalized bias" measured in units of 1/100. When computing these normalized biases, we have chosen to ignore failures of Condition A. Thus, the entries should properly be viewed as indicative of the size of the normalized bias *conditional on Condition* A *holding*. We could conceive of no satisfactory way of including the (usually relatively few) failures of Condition A in our computations.

Each entry in Table 2 is based on 100,000 simulations with the exception of the parenthetical entries for $t = 100$, which are based on 50,000.

Observe that none of the normalized biases is negative. It seems that the maximum likelihood estimator systematically gives rise, on average, to overestimation of $\log(u_j/u_i)$ whenever $u_j > u_i$, sometimes a slight overestimation, and sometimes a substantial overestimation. In this regard, the many examples of near unbiasedness, seem fairly remarkable, with every entry in Table 2 for $(i, j) \neq (0, t)$ assuming one of the four values $0-3$. In sharp contrast, most

TABLE 2
Percent of noncoverage probability in left tail

t	(i,j)	$c=0.02$	$c=0.04$	$c=0.06$	$c=0.08$	$c=0.1$	$c=0.12$
20	**(0,1)**	51.7 (0)	53.3 (1)	46.7 (1)	44.6 (2)	49.2 (2)	51.4 (1)
20	**(0,20)**	52.4 (7)	58.9 (13)	61.0 (19)	60.5 (24)	62.4 (29)	64.1 (35)
20	**(9,10)**	54.4 (0)	56.0 (1)	51.9 (1)	46.0 (0)	49.2 (1)	52.1 (2)
50	**(0,1)**	53.4 (0)	50.7 (0)	49.9 (1)	48.9 (1)	50.1 (1)	47.4 (2)
50	**(0,50)**	53.7 (10)	61.7 (19)	67.3 (28)	72.1 (36)	78.3 (44)	83.7 (52)
50	**(24,25)**	45.6 (0)	54.2 (1)	54.5 (1)	46.0 (1)	56.1 (1)	46.6 (1)
100	**(0,1)**	49.9 (0)	49.5 (0)	50.1 (0)	49.4 (1)	49.4 (1)	47.5 (3)
100	**(0,100)**	57.8 (13)	68.2 (26)	77.6 (38)	85.9 (51)	90.8 (64)	95.0 (80)
100	**(49,50)**	49.8 (0)	54.6 (0)	49.0 (0)	50.0 (0)	50.6 (0)	51.8 (2)

of the normalized biases for $(i, j) = (0, t)$ are quite large, with a normalized bias of 80/100 arising when $(c, t) = (0.12, 100)$.

From Tables 1 and 2, it appears, for the configurations of the merits considered in our simulation study, that the central limit theorem described in Theorem 2 performs very well (for our present need) for most values of t whenever u_i and u_j are relatively close, that is, whenever the parameter $\log(u_j/u_i)$ is small (even when M_t, appearing in Theorem 2, is large), and it performs poorly whenever they are far apart. However, in general, the actual size of the parameter $\log(u_j/u_i)$ being estimated is not the only main factor influencing the quality of the normal approximation. The accuracy of the approximation seems to depend on the configuration of all the merits in some complicated way. Our Theorem 2 is just a first attempt in this regard.

A relatively simple pattern in the sizes of the normalized bias is discernible in Table 2: for fixed t and for $(i, j) = (0, t)$ $(t = 20, 50, 100)$, they grow nearly linearly in c, at least for smaller values of c. Beyond this, numerical evidence reveals, for each fixed (c, t), a roughly linear growth in the sequence of normalized biases of the maximum likelihood estimates of $\log u_i$, $i = 1, 2, \dots, t$. Of course, such patterns depend on the present configurations of the merits under investigation.

5. Proofs of lemmas.

PROOF OF LEMMA 1. Let P_t denote the probability that Condition A fails, depending on the merits $u_1, \dots, u_t$ $(t = 1, 2, \dots)$. Since $M_t := \max_{0\le i,\, j\le t} u_i/u_j \ge 1$, it follows from (1.1) that

$$\max_{0\le i,\, j\le t} p_{i,j} = \max_{0\le i,\, j\le t} \frac{u_i}{u_i+u_j} \le \frac{M_t}{M_t+1} = \frac{1}{1+1/M_t} \le \left(\frac{1}{2}\right)^{1/M_t}.$$

So the probability that a particular nonempty subset S of the $t+1$ teams loses no game to a team not in S is bounded above by

$$\left(\tfrac{1}{2}\right)^{n\,|S|\,(t+1-|S|)/M_t},$$

where $|S|$ denotes the cardinality of S. Consequently, Condition A fails to occur with probability

$$P_t \leq \sum_{r=1}^{t} \binom{t+1}{r} \left(\tfrac{1}{2}\right)^{n\,r\,(t+1-r)/M_t}.$$

The task is to show that this sum converges to zero as $t \to \infty$ when M_t satisfies (1.6). Observe that the latter summands are symmetric about $(t+1)/2$, so that

$$\begin{aligned}\frac{P_t}{2} &\leq \sum_{r=1}^{\lceil t/2\rceil} \binom{t+1}{r}\left(\frac{1}{2}\right)^{n\,r\,(t+1-r)/M_t}\\ &\leq \sum_{r=1}^{\lceil t/2\rceil} \binom{t+1}{r}\left(\frac{1}{2}\right)^{n\,r\,(t+1)/2M_t}\\ &\leq \left(1+\left[\frac{1}{2}\right]^{n\,(t+1)/2M_t}\right)^{t+1} - 1.\end{aligned}$$

This does go to zero as $t \to \infty$ when M_t satisfies assumption (1.6), since (as is easily checked) $(1/2)^{n\,(t+1)/2M_t} = o(1/t)$ as $t \to \infty$. □

PROOF OF LEMMA 2. Since $a_i := \sum_{j=0,\, j\neq i}^{t} a_{i,\,j}$ is a sum of $n\,t$ independent Bernoulli random variables, Hoeffding's (1963) inequality yields

$$P(|a_i - E(a_i)| \geq x) \leq 2\exp(-2\,x^2/n\,t), \qquad x > 0\ (i = 0, \ldots, t).$$

Consequently,

$$\begin{aligned}&P\Big(\max_{i=0,\ldots,t} |a_i - E(a_i)| \geq \sqrt{n\,t\,\log(t+1)}\Big)\\ &\qquad \leq (t+1)\cdot 2\exp(-2\,n\,t\,(\log(t+1))/nt) \leq \frac{2}{t+1},\end{aligned}$$

which is equivalent to (2.1). □

PROOF OF LEMMA 3. Let $a, b \in \{0, \ldots, t\}$ be such that

$$\alpha := \max_{j=0,\ldots,t} \frac{\hat{u}_j}{u_j} = \frac{\hat{u}_a}{u_a} \geq \frac{\hat{u}_0}{u_0} = 1 \quad \text{and} \quad \beta := \min_{j=0,\ldots,t} \frac{\hat{u}_j}{u_j} = \frac{\hat{u}_b}{u_b} \leq \frac{\hat{u}_0}{u_0} = 1$$

and observe that, for $j = 0, \ldots, t$,

$$\begin{aligned}\frac{\hat{u}_a}{\hat{u}_a + \hat{u}_j} - \frac{u_a}{u_a + u_j} &= \frac{u_a\,u_j\,(\hat{u}_a/u_a - \hat{u}_j/u_j)}{(\hat{u}_a + \hat{u}_j)(u_a + u_j)}\\ &= \frac{\hat{u}_a/u_a - \hat{u}_j/u_j}{((\hat{u}_a/u_a) + (u_j/u_a)(\hat{u}_j/u_j))((u_a/u_j) + 1)}\end{aligned}$$

$$\geq \frac{\alpha - \hat{u}_j/u_j}{\alpha(1+u_j/u_a)(u_a/u_j+1)}$$

$$\geq \frac{(\alpha - \hat{u}_j/u_j)\, M_t}{\alpha\,(M_t+1)^2},$$

in view of the definition of M_t in (1.5), and since $[(1+1/x)(x+1)]^{-1} = x/(x+1)^2$ is decreasing on the range of M_t ($x \geq 1$). Consequently,

$$\sum_{j=0}^{t}\left(\frac{\hat{u}_a}{\hat{u}_a+\hat{u}_j} - \frac{u_a}{u_a+u_j}\right) \geq \frac{M_t}{(M_t+1)^2}\sum_{j=0}^{t}\left(\frac{\alpha - \hat{u}_j/u_j}{\alpha}\right). \tag{5.1}$$

Likewise, we have

$$\frac{u_b}{u_b+u_j} - \frac{\hat{u}_b}{\hat{u}_b+\hat{u}_j} = \frac{u_b\, u_j\,(\hat{u}_j/u_j - \hat{u}_b/u_b)}{(\hat{u}_b+\hat{u}_j)(u_b+u_j)}$$

$$= \frac{\hat{u}_j/u_j - \hat{u}_b/u_b}{((\hat{u}_b/u_b)+(u_j/u_b)(\hat{u}_j/u_j))(u_b/u_j+1)}$$

$$\geq \frac{\hat{u}_j/u_j - \beta}{\alpha(1+u_j/u_b)(u_b/u_j+1)}$$

$$\geq \frac{(\hat{u}_j/u_j - \beta)\, M_t}{\alpha\,(M_t+1)^2}$$

and, consequently,

$$\sum_{j=0}^{t}\left(\frac{u_b}{u_b+u_j} - \frac{\hat{u}_b}{\hat{u}_b+\hat{u}_j}\right) \geq \frac{M_t}{(M_t+1)^2}\sum_{j=0}^{t}\left(\frac{\hat{u}_j/u_j - \beta}{\alpha}\right). \tag{5.2}$$

Combining (2.2), (5.1) and (5.2) yields

$$2\sqrt{\frac{t\,\log(t+1)}{n}} > 2\max_{i=0,\dots,t}\left|\sum_{j=0}^{t}\left\{\frac{\hat{u}_i}{\hat{u}_i+\hat{u}_j} - \frac{u_i}{u_i+u_j}\right\}\right|$$

$$\geq \frac{M_t}{(M_t+1)^2}\sum_{j=0}^{t}\left(\frac{\alpha}{\alpha} - \frac{\beta}{\alpha}\right)$$

$$\geq \frac{t\,M_t}{(M_t+1)^2}\left(1 - \frac{\beta}{\alpha}\right).$$

Thus (since $M_t \geq 1$),

$$1 - \frac{\beta}{\alpha} < 2\,M_t\,\frac{(M_t+1)^2}{M_t^{\,2}}\sqrt{\frac{\log(t+1)}{n\,t}} \leq 8\,M_t\sqrt{\frac{\log(t+1)}{n\,t}} = \delta_t$$

and, consequently,

$$\max_{0\leq i,\, j\leq t}|\Delta u_i - \Delta u_j| = \max_{0\leq i,\, j\leq t}\left|\frac{\hat{u}_i}{u_i} - \frac{\hat{u}_j}{u_j}\right| = (\alpha - \beta) \leq \frac{\alpha-\beta}{\beta} < \frac{\delta_t}{1-\delta_t},$$

as required. □

PROOF OF LEMMA 4. Simons and Yao (1998) provide a bound for $\| W_t \|$ of the form $C(m, M)/t^2$, where $C(m, M) = (1+M/m)\, M/m^2$, and where, for the present context, m and M must satisfy the bounding inequalities

$$m \le \frac{n\, u_i\, u_j}{(u_i + u_j)^2} \le M, \qquad i = 1, \dots, t;\; j = 0, 1, \dots, t;\; j \ne i.$$

We want to express these in terms of M_t, defined in (1.5), independently of any particular choice of $u_1, \dots, u_t$ satisfying (1.5). Thus one is led to set $M = n/4$ and $m = nM_t/(M_t + 1)^2$, from which one obtains the coefficient $C(m, M)$ of $1/t^2$ appearing to the right of the first inequality in (3.5). The second inequality is just algebra together with the fact that $M_t \ge 1$. □

PROOF OF LEMMA 5. Clearly $B = I - A$ (where I is the identity matrix of the same dimension as A) has only nonnegative elements. Observe that λ is an eigenvalue of A if and only if $1-\lambda$ is an eigenvalue of B. The Perron–Frobenius theory asserts that B has a positively valued eigenvalue μ which is at least as large as the modulus of any other of its eigenvalues. Necessarily, $\mu < 1$ since the eigenvalues of A are strictly positive. Consequently, the smallest eigenvalue of B is greater than -1, and it follows that the largest eigenvalue of A is less than $1 - (-1) = 2$. □

PROOF OF LEMMA 6. Since the covariance matrix of $\mathbf{a}$ is V_t, we find that

$$R_t = W_t\, V_t\, W_t = ({V_t}^{-1} - S_t)\, V_t\, ({V_t}^{-1} - S_t) = (S_t\, V_t\, S_t - S_t) + ({V_t}^{-1} - S_t).$$

By direct calculation, the (i, j)th element of $S_t\, V_t\, S_t - S_t$ is

$$\begin{aligned} &\frac{n\, u_i}{(1 + u_i)^2\, v_{i,i}\, v_{0,0}} + \frac{n\, u_j}{(1 + u_j)^2\, v_{j,j}\, v_{0,0}} \\ &\quad - \frac{n(1 - \delta_{i,j})\, u_i\, u_j}{(u_i + u_j)^2\, v_{i,i}\, v_{j,j}}, \qquad i, j = 1, \dots, t. \end{aligned} \tag{5.3}$$

By using the bounds in (3.1) and observing that $0 < u_i u_j/(u_i + u_j)^2 \le 1/4$, it is easy to bound each of the terms in (5.3) by $4{M_t}^2/nt^2$. Hence,

$$\| S_t\, V_t\, S_t - S_t \| \le 8{M_t}^2/nt^2.$$

Moreover, as described in Lemma 4,

$$\| {V_t}^{-1} - S_t \| \le \frac{4(1 + M_t){M_t}^2}{n\, t^2}.$$

This establishes (3.12). □

PROOF OF LEMMA 7. Let A_t be the event that Condition A is satisfied, so that, from (1.2) and (1.3),

$$a_i - E(a_i) = n \sum_{j=0,\, j\ne i}^{t} \left\{ \frac{\hat u_i}{\hat u_i + \hat u_j} - \frac{u_i}{u_i + u_j} \right\}, \qquad i = 1, \dots, t \tag{5.4}$$

and let B_t be the event that [cf. (1.7) and (2.3)]

$$\max_{i=0,\dots,t} |\Delta u_i| \le \max_{0\le i,\, j\le t} |\Delta u_i - \Delta u_j| < \frac{\delta_t}{1-\delta_t}, \tag{5.5}$$

where δ_t is defined, in terms of M_t, in (1.5). It follows from Lemmas 1–3 that $P(A_t \cap B_t) \to 1$ as $t \to \infty$ provided M_t satisfies (1.6), which is less stringent than what we are assuming here.

Derivations in what follows are on the event $A_t \cap B_t$. By simple algebra,

$$n\left\{\frac{\hat u_i}{\hat u_i + \hat u_j} - \frac{u_i}{u_i + u_j}\right\} = \frac{n u_i u_j(\Delta u_i - \Delta u_j)}{(u_i + u_j)^2} \times \frac{u_i + u_j}{\hat u_i + \hat u_j} = \alpha_{i,\,j} + \beta_{i,\,j},$$

where

$$\alpha_{i,\,j} = \frac{n u_i u_j(\Delta u_i - \Delta u_j)}{(u_i + u_j)^2}, \qquad \beta_{i,\,j} = \alpha_{i,\,j} \times \gamma_{i,\,j},$$

and

$$\gamma_{i,\,j} = \frac{u_i + u_j}{\hat u_i + \hat u_j} - 1 = \frac{-(u_i\Delta u_i + u_j\Delta u_j)/(u_i + u_j)}{1 + (u_i\Delta u_i + u_j\Delta u_j)/(u_i + u_j)},$$

for $j = 0, 1, \dots, t$ and $i = 1, \dots, t$. Inserting this into the summation on the right side of (5.4) produces a Taylor series expansion of the functions (of the variables $\hat u_1, \dots, \hat u_t$) on the right side of (5.4) about the point $(\hat u_1, \dots, \hat u_t) = (u_1, \dots, u_t)$ up to linear terms α_i with nonstandard but convenient remainder terms β_i,

$$a_i - E(a_i) = \alpha_i + \beta_i, \qquad i = 1, \dots, t, \tag{5.6}$$

where

$$\alpha_i = \sum_{j=0,\, j\ne i}^{t} \alpha_{i,\,j} \quad \text{and} \quad \beta_i = \sum_{j=0,\, j\ne i}^{t} \beta_{i,\,j}.$$

Observe that $|\Delta u_i - \Delta u_j| \le \delta_t/1 - \delta_t$ [cf. (5.5)], and, moreover, $|\gamma_{i,\,j}| \le \delta_t/(1-2\delta_t)$ since the function $g(x) = x/(1+x)$ is nondecreasing and $|(u_i\Delta u_i + u_j\Delta u_j)/(u_i + u_j)| \le \delta_t/(1-\delta_t)$ [cf. (5.5)]. Consequently,

$$\begin{aligned}|\beta_{i,\,j}| &= \frac{n u_i u_j}{(u_i + u_j)^2} \times |\Delta u_i - \Delta u_j| \times |\gamma_{i,\,j}| \\ &\le \frac{n u_i u_j}{(u_i + u_j)^2} \times \frac{{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)}, \qquad i = 1, \dots, t;\ j = 0, \dots, t,\end{aligned}$$

and, hence [see (1.8)], for $i = 1, \dots, t$,

$$\begin{aligned}|\beta_i| \le \sum_{j=0,\, j\ne i}^{t} |\beta_{i,\,j}| &\le \sum_{j=0,\, j\ne i}^{t} \frac{n u_i u_j}{(u_i + u_j)^2} \times \frac{{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)} \\ &= \frac{v_{i,\,i}\, {\delta_t}^2}{(1-\delta_t)(1-2\delta_t)}.\end{aligned} \tag{5.7}$$

Further, $\beta_{i,j} + \beta_{j,i} = 0$ for $i, j \geq 1$, so that

$$\sum_{i=1}^{t} \beta_i = \sum_{i=1}^{t} \sum_{j=0,\, j\neq i}^{t} \beta_{i,j} = \sum_{i=1}^{t} \beta_{i,0},$$

and, consequently [see (1.8)],

$$\left|\sum_{i=1}^{t} \beta_i\right| \leq \sum_{i=1}^{t} |\beta_{i,0}| \leq \sum_{i=1}^{t} \frac{n u_i}{(1+u_i)^2} \times \frac{{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)} = \frac{v_{0,0}\,{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)}. \tag{5.8}$$

Now express (5.6) in matrix notation,

$$\mathbf{a} - E(\mathbf{a}) = \boldsymbol{\alpha} + \boldsymbol{\beta},$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_t)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_t)'$. By direct calculation, one finds that $\boldsymbol{\alpha}$ can be expressed as $V_t \boldsymbol{\Delta}\mathbf{u}$, where $\boldsymbol{\Delta}\mathbf{u} = (\Delta u_1, \ldots, \Delta u_t)'$. Consequently,

$$\boldsymbol{\Delta}\mathbf{u} = {V_t}^{-1}(\mathbf{a} - E(\mathbf{a})) - {V_t}^{-1}\boldsymbol{\beta} = {V_t}^{-1}(\mathbf{a} - E(\mathbf{a})) - (W_t + S_t)\boldsymbol{\beta}.$$

So it remains to show $(W_t\,\boldsymbol{\beta})_i = o_p(t^{-1/2})$ and $(S_t\,\boldsymbol{\beta})_i = o_p(t^{-1/2})$ as $t \to \infty$ for each fixed i.

In view of (3.1), (3.5) and (5.7), we see that

$$|(W_t\,\boldsymbol{\beta})_i| \leq \frac{4(1+M_t){M_t}^2}{n\,t^2} \times \frac{{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)} \times \sum_{j=1}^{t} v_{j,j}$$
$$\leq (1+M_t){M_t}^2 \times \frac{{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)} = O\left(\frac{{M_t}^5 \log t}{t}\right),$$

which equals $o_p(t^{-1/2})$ when $M_t = o(t^{1/10}/(\log t)^{1/5})$ as $t \to \infty$.

In view of (1.9), (5.7) and (5.8),

$$|(S_t\,\boldsymbol{\beta})_i| \leq \frac{1}{v_{i,i}}|\beta_i| + \frac{1}{v_{0,0}}\left|\sum_{j=1}^{t} \beta_j\right| \leq \frac{2\,{\delta_t}^2}{(1-\delta_t)(1-2\delta_t)} = O(\frac{{M_t}^2 \log t}{t}),$$

which equals $o_p(t^{-1/2})$ when $M_t = o(t^{1/4}/(\log t)^{1/2})$ as $t \to \infty$. □

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BRADLEY, R. A. (1954). Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. *Biometrics* **10** 375–390.

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

BRADLEY, R. A. and TERRY, M. E. (1952). The rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika* **39** 324–345.

BÜHLMANN, H. and HUBER P. J. (1963). Pairwise comparisons and ranking in tournaments. *Ann. Math. Statist.* **34** 501–510.

COLONIUS, H. (1980). Representation and uniqueness of the Bradley–Terry–Luce model for pair comparisons. *British J. Math. Statist. Psych.* **33** 99–103.

DAVID, H. A. (1988). *The Method of Paired Comparisons*, 2nd ed. Griffin, London.

DAVIDSON, R. R. and FARQUHAR, P. H. (1976). A bibliography on the method of paired comparisons. *Biometrics* **32** 241–252.

FORD, L. R., JR. (1957). Solution of the ranking problem from binary comparisons. *Amer. Math. Monthly* **64** 28–33.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

JECH, T. (1983). The ranking of incomplete tournaments: a mathematician's guide to popular sports. *Amer. Math. Monthly* **90** 246–266.

LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, New York.

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large I. Consistency. *Ann. Statist.* **12** 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366.

SIMONS, G. and YAO, Y.-C. (1998). Approximating the inverse of a symmetric positive definite matrix. *Linear Algebra Appl.* **281** 97–103.

STOB, M. (1984). A supplement to "Mathematician's Guide to Popular Sports." *Amer. Math. Monthly* **91** 277–282.

SUPPES, P. and ZINNES, J. L. (1963). Basic measurement theory. In *Handbook of Mathematical Psychology* (R. D. Luce, R. R. Bush and E. Galanter, eds.) **1** 1–76. Wiley, New York.

ZERMELO, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Zeit.* **29** 436–460.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260
E-MAIL: simons@stat.unc.edu

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI
TAIWAN
E-MAIL: yao@stat.sinica.edu.tw