Testing equality of players in a round-robin tournament

F. Thomas Bruss, Université Libre de Bruxelles, Belgium
Thomas S. Ferguson, University of California at Los Angeles, USA

This paper is dedicated to Professor Joseph M. Gani (1924-2016) in honor of all he has achieved for the development of Applied Probability and in commemoration of this Journal which he founded in 1975 and which ends with this issue.

Abstract: In a round-robin tournament with n players, each player plays every other player once, resulting in $\binom{n}{2}$ games. Let X_{ij} denote the score by which player i beats player j, with $X_{ji} = -X_{ij}$ for all $i \neq j$. If we take $X_{ii} = 0$ for all i, then $S_i = \sum_{j=1}^{n} X_{ij}$ denotes the total score of Player i, for $i = 1, 2, \ldots, n$. To test the hypothesis, H_0, that the players are equally skillful, in the sense that the X_{ij} for $i < j$ are i.i.d. with mean 0 and common variance, we suggest rejecting H_0 if $V_n = \sum_1^n S_i^2$ is too large. It is shown that an associated statistic, W_n, a generalization of the circular triads statistic of Kendall and Babington Smith, is easier to work with and more stable. We establish the asymptotic normality of V_n and W_n under general conditions. As an illustration, the results are applied to data obtained on the Greek Soccer League 2016-2017.

Keywords Skill vs. Luck, Paired Comparisons, Circular Triads, Stein's bound on approximate normality, Greek Soccer league

Math. subj. classification 2010 primary 62G10, secondary 91A06

1. Introduction

This paper deals with the question "skill or luck" in team games, and more specifically in *round-robin tournaments.*

In a round-robin tournament with n players, each player plays every other player once, resulting in $\binom{n}{2}$ games. We may of course, interpret a *player* as a team of players all (within the same team) having exactly the same objective to win, and this is why we only use the term *player.* We also mention that a tournament where each player faces each other exactly twice is called a *double* round-robin tournament. This often occurs in team sports where there is a noticable home field advantage, ond in games like chess in which the player that moves first has an advantage. We confine here our interest to simple round-robin tournaments. However in the last Section, we treat a double round-robin tournament, namely of the Greek Soccer League 2016-2017, but we treat it as a single round-robin tournament by combining the two games.

Let X_{ij} denote the score by which player (or team) i beats player j, with $X_{ji} = -X_{ij}$ for all $i \neq j$. We take $X_{ii} = 0$ for all i, so that

$$S_i = \sum_{j=1}^{n} X_{ij} \tag{1}$$

denotes the total score of Player i, for $i = 1, 2, \ldots, n$. We wish to test the null hypothesis that the players are equally skilful in the sense that the X_{ij} for $i < j$ are i.i.d. with mean zero.

Note that $\sum_i S_i = 0$, so that the S_i will be spread out about 0. We tend to think of the players with larger S_i as the more skillful, even though the null hypothesis may be true. Therefore, we should first test this hypothesis against alternatives in which some individuals are generally more skillful than others. When some players are better than others, we expect the scores will be more spread out than under the null hypothesis. To test the null hypothesis that the players are equally skillful, we may use some measure of spread of the distribution of the S_i. One simple such measure is the sum of squares,

$$V_n = \sum_{i=1}^{n} S_i^2. \tag{2}$$

We reject the null hypothesis if V_n is too large. There is a large literature on the problem of ranking the players based on the Bradley-Terry model in paired comparisons. For exampie, see the paper of Caron and Doucet (2012). See also Iida (2009).

In the case of win-lose outcomes, where the score for a game is $+1$ for the winner and -1 for the loser, this problem becomes the one introduced by Kendall and Babington Smith (1940) in their treatment of paired comparisons. There in the testing problem, the statistic, V_n, is seen to be equivalent to the circular triads statistic, d_n, which counts the number of triads of players, i, j and k, in which i beats j, j beats k and k beats i. This is discussed in detail in Section 4.

In Section 2, we introduce a statistic, W_n, related to V_n in much the same way that d_n is related to V_n in the win-lose outcome case. We see that V_n and W_n are asymptotically equivalent in the testing problem, and why W_n should be preferred to V_n. In Section 3, we show the asymptotically normality of V_n and W_n under the null hypothesis.

In Section 5, as an example, we apply our results to data of the Greek Soccer League 2016-2017, where 16 teams face each other twice. For comparison, we also test the null hypothesis using a randomization test based on W_n. If one looks at these data, one feels that one needs no mathematics to decide that the null-hypothesis should be rejected. And so do our tests. This is only for illustration, and our choice is for two reasons:

First, the size of the table of data from the Greek Soccer League is large enough to give an idea of the variance of the data, and not too big to fit nicely in a single table.

Second, and in particular, we know that Joe Gani had a close relationship with Greece. His grandparents were from western Greece. His paternal grandfather came from Ioannina and his mother's family from the Greek island of Corfu. His attachment to Greece was very visible at the Conference in honour of his retirement, co-organised by faculty members from UC Santa Barbara and several Greek universities, and held in Athens in 1995. See the special volume i.h. of J.M. Gani (Heyde et al, editors (1996) containing a wide range of subjects presented at the conference, to which we also had the honour to contribute.

2. Towards a more suitable test

If we use V_n as the test statistic, then we will need to approximate the variance of V_n. This will require that the fourth moment of the X_{ij} is finite. Therefore, we take our null hypothesis, that the players are equally skillful, to be

$$H_0\text{: The } X_{ij} \text{ for } i < j \text{ are i.i.d. with mean 0 and finite fourth moment.} \tag{3}$$

The common variance is denoted by σ^2 and will be estimated from the data. Under this hypothesis, all the S_i have mean zero and variance $(n-1)\sigma^2$.

We present another way to test this hypothesis. We look at triplets of players and see if one of the players dominates the other two, while another player is dominated. If among players i, j and k, player i is the strongest one and player k is the weakest, then X_{ij}, X_{ik} and X_{jk} would have positive expectation with the expectation X_{ik} largest of the three. As a measure of skill vs. luck for the triplet $\{ijk\}$, we may use the quantity

$$Y_{ijk} = X_{ij}X_{ik} - X_{ij}X_{jk} + X_{ik}X_{jk}. \tag{4}$$

For example in a win/lose tournament where each X_{ij} is $+1$ or -1, Y_{ijk} will be $+1$ if one of the players beats the other two, and -3 in the non-transitive case where, say, i beats j, j beats k and k beats i. Under the null hypothesis, $\mathrm{E}(Y_{ijk}) = 0$ for all $\{ijk\}$.

Note that Y_{ijk} is independent of the order of the subscripts, i.e. $Y_{ijk} = Y_{jik} = Y_{ikj}$ etc. By this symmetry, $\mathrm{E}(Y_{ijk})$ will be positive if any of the three players is the dominant one. The sum of these quantities over all triplets

$$W_n = \sum\sum\sum\nolimits_{i<j<k} Y_{ijk}. \tag{5}$$

may be used to test the null hypothesis of equality of skill of the players involved in the game. We reject the null hypothesis if W_n is too large.

It is interesting to see that these two tests are asymptotically equivalent for large n. We see this using the following lemma.

Lemma 1. $V_n = \sum\sum\nolimits_{i\neq j} X_{ij}^2 + 2W_n = 2\sum\sum\nolimits_{i<j} X_{ij}^2 + 2W_n$.

Proof. First note that

$$S_i^2 = \sum_j \sum_k X_{ij}X_{ik} = \sum_j X_{ij}^2 + 2\sum\sum\nolimits_{j<k} X_{ij}X_{ik}. \tag{6}$$

Therefore,

$$\begin{aligned} V_n &= \sum_{i=1}^{n} S_i^2 = \sum_i \sum_j X_{ij}^2 + 2\sum_{i=1}^{n} \sum \sum\nolimits_{j<k} X_{ij}X_{ik} \\ &= 2\sum \sum\nolimits_{i<j} X_{ij}^2 + 2\sum \sum \sum\nolimits_{i<j<k} [X_{ij}X_{ik} - X_{ij}X_{jk} + X_{ik}X_{jk}] \\ &= 2\sum \sum\nolimits_{i<j} X_{ij}^2 + 2W_n. \ \blacksquare \end{aligned} \tag{7}$$

Under the assumption that the X_{ij} for $i < j$ are i.i.d. with finite fourth moment, the term $\sum \sum_{i<j} X_{ij}^2$ is asymptotically normal with variance of order n^2. The term W_n, as we shall see in Theorem 1, is also asymptotically normal, but with variance of order n^3. Thus in the testing problem, the term W_n dominates $\sum \sum_{i<j} X_{ij}^2$, implying that the tests based on the statistics V_n and W_n are asymptotically equivalent.

We now collect several results which will be used in the proof of asympotic normality of W_n. We shall speak of the triplets $\{ijk\}$ as triangles and use the notation Y_t for Y_{ijk} of (4) when t is the triangle $\{ijk\}$, with vertices i, j, k and edges ij, ik and jk.

Lemma 2. *Assume the X_{ij} are independent for $i < j$ and that $\mathrm{E}(X_{ij}) = 0$.*
(a) If triangles s and t do not have a common edge, then Y_s and Y_t are independent.
(b) If $\{ij\}$ is an edge of triangle t then $\mathrm{E}(Y_t|X_{ij}) = 0$ a.s.
(c) If triangle t has at most one of its edges in common with a collection $\{s : s \in S\}$ of triangles, then $\mathrm{E}(Y_t U) = 0$ for any measurable function U of $\{Y_s : s \in S\}$ with finite expectation.
(d) If triangles s and t are distinct, then $\mathrm{E}(Y_s Y_t) = 0$.

Proof. (a) If s and t do not have a common edge, then Y_s and Y_t are constructed of independent X's and so are independent.
(b) If $t = \{ijk\}$, then

$$E(Y_t|X_{ij}) = X_{ij}\mathrm{E}(X_{ik}) - X_{ij}\mathrm{E}(X_{jk}) + \mathrm{E}(X_{ik}X_{jk}) = 0. \tag{8}$$

(c) If t has no edge in common with $\{s : s \in S\}$, then Y_t is independent of $\{Y_s : s \in S\}$ and hence U, so $\mathrm{E}(Y_s U) = 0$. If t has edge $\{ij\}$ in common with $\{s : s \in S\}$, then, conditionally given X_{ij}, Y_t is independent of U, so

$$\mathrm{E}(Y_t U) = \mathrm{E}(\mathrm{E}(Y_t U|X_{ij})) = \mathrm{E}(\mathrm{E}(Y_t|X_{ij})\mathrm{E}(U|X_{ij})) = 0 \tag{9}$$

from (b). Part (d) follows immediately from (c), since distinct triangles have at most one edge in common. $\blacksquare$

Part (d) says that the Y_t are pairwise uncorrelated.

Lemma 3. *Assume that the X_{ij} are independent for $i < j$ with means zero and variances σ^2. Then,*
(a) $\mathrm{E}(W_n) = 0$.
(b) $\mathrm{Var}(Y_t) = 3\sigma^4$.
(c) $\mathrm{Var}(W_n) = \binom{n}{3} 3\sigma^4$.

Proof. (a) Since $\mathrm{E}(Y_t) = 0$ for all t, $\mathrm{E}(W_n) = 0$.
(b) In $\mathrm{Var}(Y_t) = \mathrm{E}(X_{ij}X_{ik} - X_{ij}X_{jk} + X_{ik}X_{jk})^2$, all cross product terms disappear and we are left with $\mathrm{E}(X_{ij}X_{ik})^2 + \mathrm{E}(X_{ij}X_{jk})^2 + \mathrm{E}(X_{ik}X_{jk})^2 = 3\sigma^4$.
(c) From Lemma 2(d), $\mathrm{Cov}(Y_s, Y_t) = 0$ for distinct s and t, so $\mathrm{Var}(W_n) = \mathrm{Var}(\sum_t Y_t) = \sum_t \mathrm{Var}(Y_t) = \binom{n}{3} 3\sigma^4$. ∎

Lemma 4. *Assume that the X_{ij} for $i < j$ are independent with means zero, variances σ^2, and fourth moments $\mu_4 = \mathrm{E}(X_{ij}^4)$. Then,*
(a) W_n *and* $\sum_{i<j} X_{ij}^2$ *are uncorrelated.*
(b) $\mathrm{E}(V_n) = n(n-1)\sigma^2$.
(c) $\mathrm{Var}(V_n) = 2n(n-1)[\mu_4 + (n-3)\sigma^4]$.

Proof. (a) $\mathrm{Cov}(W_n, \sum_{i<j} X_{ij}^2) = \sum_t \sum_{i<j} \mathrm{Cov}(Y_t, X_{ij}^2)$. If $\{ij\}$ is not an edge of t, then Y_t and X_{ij} are independent and so have covariance zero. If $\{ij\}$ is an edge of t, say $t = \{ijk\}$, then $\mathrm{Cov}(Y_t, X_{ij}^2) = \mathrm{Cov}(X_{ij}(X_{ik} - X_{jk}), X_{ij}^2) = 0$ since $\mathrm{E}(X_{ik} - X_{jk}) = 0$.
(b) $\mathrm{E}(V_n) = \sum_1^n \mathrm{E}(S_i^2) = n\mathrm{E}(S_1^2)$, so (b) follows from

$$\mathrm{E}(S_1^2) = \mathrm{E}\left(\sum_{j=2}^n X_{1j} \sum_{k=2}^n X_{1k}\right) = \mathrm{E}\left(\sum_{j=2}^n X_{1j}^2\right) = (n-1)\sigma^2.$$

(c) Since from part (a) W_n and $\sum_{i<j} X_{ij}^2$ are uncorrelated, we have from Lemma 1,

$$\begin{aligned}\mathrm{Var}(V_n) &= 4\mathrm{Var}(\sum_{i<j} X_{ij}^2) + 4\mathrm{Var}(W_n) = 4[\sum_{i<j} \mathrm{Var}(X_{ij}^2) + \mathrm{Var}(W_n)] \\ &= 4[\binom{n}{2}(\mu_4 - \sigma^4) + \binom{n}{3} 3\sigma^4] = 4\binom{n}{2}[\mu_4 + (n-3)\sigma^4]. \quad ∎\end{aligned}$$

As a test of the null hypothesis, V_n is easier to understand than W_n, but W_n has two advantages. First, to use V_n, estimates of its mean and variance are needed, and to estimate its variance, an estimate of μ_4 is needed. The mean of W_n is zero and its variance does not involve μ_4. Secondly, from Lemma 1, V_n is obtained from $2W_n$ by adding a random quantity which is uncorrelated with W_n as seen from Lemma 4(a). Since the distribution of this random quantity seems to have little to do with the null hypothesis — divided by $\binom{n}{2}$, it is just an estimate of σ^2 — we conclude that W_n is more sensitive to deviations from the null hypothesis.

3. Asymptotic Normality of W_n.

We now show the approximate normality of W_n (and hence V_n) for large n.

Theorem 1. *Under the assumption that the X_{ij} for $i < j$ are i.i.d. with mean zero and finite fourth moment,*

$$\frac{W_n}{\sqrt{\mathrm{Var}(W_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{10}$$

The method of proof uses the following theorem of Stein (1986), pg. 110, on the approximation of the distribution of a sum of possibly dependent nonidentically distributed random variables by a normal distribution.

Stein's Theorem. *Let $\mathcal{T}$ be a finite index set, and for $t \in \mathcal{T}$, let Z_t be a random variable with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}(Z_t^4) < \infty$. Let $W = \sum_{t\in\mathcal{T}} Z_t$. Suppose that for every $t \in \mathcal{T}$, there is a set $S_t \subset \mathcal{T}$ such that*

$$\mathrm{E}[\sum_{t\in\mathcal{T}} Z_t \sum_{s\in S_t} Z_s] = 1. \tag{11}$$

Then

$$\begin{aligned}\sup_w |\mathrm{P}(W \le w) - \Phi(w)| \le 2\sqrt{\mathrm{E}[\sum_{t\in\mathcal{T}}\sum_{s\in S_t}(Z_tZ_s - \sigma_{ts})]^2} \\ + \sqrt{\frac{\pi}{2}}\mathrm{E}\left(\sum_{t\in\mathcal{T}}\left|\mathrm{E}(Z_t|\{Z_s\}_{s\notin S_t})\right|\right) \\ + 2^{3/4}\pi^{-1/4}\sqrt{\mathrm{E}\Big(\sum_{t\in\mathcal{T}}|Z_t|\big(\sum_{s\in S_t} Z_s\big)^2\Big)}\end{aligned} \tag{12}$$

where $\sigma_{ts} = \mathrm{E}(Z_tZ_s)$ and $\Phi(w)$ represents the standard normal distribution function.

In the application of this theorem to our problem, $\mathcal{T}$ is taken to be the set of all triangles, $\mathcal{T} = \{t = \{ijk\} : 1 \le i < j < k \le n\}$. For $t = \{ijk\} \in \mathcal{T}$, let

$$Z_t = Y_t/\delta_n = [X_{ij}X_{ik} - X_{ij}X_{jk} + X_{ik}X_{jk}]/\delta_n, \tag{13}$$

where δ_n is a normalizing constant chosen to satisfy (11). We have $\mathrm{E}(Z_t) = 0$. For all $t \in \mathcal{T}$, S_t is taken to be the set of all triangles that have an edge in common with t (including t itself). The Z_t are pairwise uncorrelated from Lemma 2(d), so from (11)

$$1 = \mathrm{E}[\sum_{t\in\mathcal{T}} Z_t \sum_{s\in S_t} Z_s] = \mathrm{E}[\sum_{t\in\mathcal{T}} Z_t^2] = \binom{n}{3}3\sigma^4/\delta_n^2. \tag{14}$$

where the last equality in (14) follows from Lemma 3(c). Thus,

$$\delta_n = \sigma^2\sqrt{n(n-1)(n-2)/2}. \tag{15}$$

The following lemma will be useful.

Lemma 5. *(a) For distinct triangles r, s and t, $\mathrm{E}(Z_r Z_s Z_t) = 0$ except when r, s and t are three of the faces of some tetrahedron as in Figure 1(a).*
(b) For distinct r, s, t and u, $\mathrm{E}\{Z_r Z_s Z_t Z_u\} = 0$ unless r, s, t and u form the four faces of a tetrahedron or four triangles arranged as in Figure 1(b).

Figure 1.

Proof. (a) If t has at most one edge in common with r and s, then $\mathrm{E}(Z_t Z_r Z_s) = 0$ from Lemma 2(c). Therefore, to be dependent, each of r, s and t have two edges in common with the other two. This only happens if r, s and t form three faces of a tetrahedron.

(b) If t has at most one edge in common with $\{r, s, u\}$, then again from Lemma 2(c) with $U = Z_r Z_s Z_u$, we have $\mathrm{E}(Z_r Z_s Z_t Z_u) = 0$. Otherwise, each of r, s, t and u have at least two edges in common with the others. This can only happen if they form the four faces of a tetrahedron or four triangles arranged as in Figure 1(b). ■

Proof of Theorem 1. We now apply Stein's Theorem to conclude asymptotic normality of W_n by showing that each of the three terms on the right side of (12) tend to zero as n goes to infinity. This will imply asymptotic normality of W_n.

The sets S_t have been chosen so that Z_t is independent of $\{Z_s\}_{s \notin S_t}$. Therefore, $\mathrm{E}(Z_t | \{Z_s\}_{s \notin S_t}) = \mathrm{E}(Z_t) = 0$ for all $t \in \mathcal{T}$. Thus the second term on the right of (12) is equal to zero.

The expectation in the third term on the right of (12) may be written

$$\mathrm{E}\left(\sum_{t \in \mathcal{T}} |Z_t| \Big(\sum_{s \in S_t} Z_s \Big)^2 \right) = \binom{n}{3} \mathrm{E}\left(|Z_t| \Big(\sum_{s \in S_t} Z_s \Big)^2 \right) \tag{16}$$

for an arbitrary $t \in \mathcal{T}$. Each Z_t is of the order of $n^{-3/2}$, but each sum over $s \in S_t$ is of order n and so, without cancellation, (16) would be of order $n^{1/2}$. If we let S_t' denote the set S_t with the point t removed, $S_t' = S_t \backslash \{t\}$, we may expand (16) as follows

$$\begin{aligned} \mathrm{E}\Big(\sum_{t \in \mathcal{T}} |Z_t| \big(\sum_{s \in S_t} Z_s \big)^2 \Big) &= \binom{n}{3} \mathrm{E}\Big(|Z_t| \big(Z_t + \sum_{s \in S_t'} Z_s \big)^2 \Big) \\ &= \binom{n}{3} \Big(\mathrm{E}(|Z_t|^3) + 2\mathrm{E}(|Z_t| Z_t \sum_{s \in S_t'} Z_s) + \mathrm{E}\Big(|Z_t| \sum_{r \in S_t'} \sum_{s \in S_t'} Z_r Z_s \Big) \Big) \end{aligned} \tag{17}$$

The term $\binom{n}{3}\mathrm{E}|Z_t|^3$ is of order $n^{-3/2}$ and so is asymptotically negligible. The central term is equal to zero from Lemma 2(c). The last term may be written

$$\binom{n}{3}\mathrm{E}\Big(|Z_t|\sum_{s\in S_t'} Z_s^2\Big) + \binom{n}{3}\mathrm{E}\Big(|Z_t|\sum_{r\in S_t'}\sum_{s\in S_t', r\neq s} Z_r Z_s\Big) \tag{18}$$

Since there are $3(n-3)$ elements in S_t', the first term is of order $n^{-1/2}$. By Lemma 5(a), the expectation $\mathrm{E}(|Z_t|Z_rZ_s)$ is zero unless r, s and t are three of the faces of a tetrahedron. Since t is fixed, there are only $n-3$ choices for the fourth vertex of the tetrahedron, so the double sum reduces to a single sum of order n. Thus, the second term in (18) is also of order $n^{-1/2}$. This shows the third term on the right of (12) converges to zero as $n \to \infty$.

Now we treat the first term on the right of (12). Since $\sigma_{st} = 0$ for $s \neq t$, the expectation in this term may be written

$$\begin{aligned}
\mathrm{E}\Big(\sum_{t\in\mathcal{T}}\sum_{s\in S_t}(Z_tZ_s - \sigma_{ts})\Big)^2 &= \mathrm{E}\Big(\sum_{t\in\mathcal{T}}\big[(Z_t^2-\sigma_{tt}) + \sum_{s\in S_t'} Z_tZ_s\big]\Big)^2 \\
&= \mathrm{E}\Big(\sum_{t\in\mathcal{T}}\big[(Z_t^2-\sigma_{tt}) + \sum_{s\in S_t'} Z_tZ_s\big]\sum_{u\in\mathcal{T}}\big[(Z_u^2-\sigma_{uu}) + \sum_{r\in S_u'} Z_uZ_r\big]\Big) \\
&= \mathrm{E}\Big(\sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{T}}(Z_t^2-\sigma_{tt})(Z_u^2-\sigma_{uu})\Big) + 2\mathrm{E}\Big(\sum_{t\in\mathcal{T}}(Z_t^2-\sigma_{tt})\sum_{u\in\mathcal{T}}\sum_{r\in S_u'} Z_uZ_r\Big) \\
&\quad + \mathrm{E}\Big(\sum_{t\in\mathcal{T}}\sum_{s\in S_t'}\sum_{u\in\mathcal{T}}\sum_{r\in S_u'} Z_tZ_sZ_uZ_r\Big)
\end{aligned} \tag{19}$$

If $u \notin S_t$, then Z_t and Z_u are independent, so $\mathrm{E}((Z_t^2-\sigma_{tt})(Z_u^2-\sigma_{uu})) = 0$. Thus the inside sum in the first expectation on the right of (19) reduces to a sum over $u \in S_t$. There are order n terms in this sum and order n^3 terms in the sum over t, but each term is of order n^{-6}, so the first term on the right of (19) is of order n^{-2} and so is negligible. The expectation in the second term vanishes if $u = t$ or if $r = t$, so we may take the summation assuming t, u and r are distinct. By Lemma 5(a), the expectation is zero unless t, u and r are three faces of a tetrahedron. As in the analysis of the last term of (18), the inside two summations reduce to a single summation of order n. So this term is of order n^{-2} and is negligible.

Now consider the last term of (19). In summing when $t = u$, the quadruple sum reduces to

$$\sum_{t\in\mathcal{T}}\sum_{s\in S_t'}\sum_{r\in S_t'} \mathrm{E}(Z_t^2Z_sZ_r). \tag{20}$$

This expectation is zero unless $s = r$, or unless t, s and r are faces of some tetrahedron. The total of both of these possibilities is of order n, so the summation is of order n^{-2} and may be neglected. Thus we may assume $t \neq u$ in the quadruple summation in (19). The same argument follows if $u = s$: $\mathrm{E}(Z_tZ_s^2Z_r)$ is not zero only if $r = t$ or if t, s and r

are faces of a tetrahedron. And similarly for summing when $t = r$ and when $r = s$. All such terms are of order n^{-2}. We may now assume that r, s, t and u are distinct in the quadruple summation. By Lemma 5(b), the expectation is zero unless r, s, t and u are a tetrahedron or they form an arrangement as in Figure 1(b). With t fixed, the number of tetrahedrons is of order n and the number of arrangements as in Figure 1(b) is of order n^2 (corresponding to the choice of the other two vertices). Since there are $\binom{n}{3}$ terms in the choice of t, there are order n^5 non-zero terms in the quadruple sum. Thus the quadruple sum is of order n^{-1}. This shows that the first term on the right of (12) goes to zero as $n \to \infty$, completing the proof. ■

It is nice to see that this proof goes through under the weaker assumptions that the X_{ij} are independent with zero means and uniformly bounded fourth moments, and with $\mathrm{Var}(W_n)/n^3$ bounded away from zero. Also, and in particular, it is not required that the variances of the X_{ij} be equal.

4. Specialization to Win/Lose Outcomes.

Suppose n players play a round-robin tournament with win or lose outcomes. Then there are $\binom{n}{2}$ games and

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ beats } j \\ -1 & \text{if } j \text{ beats } i \end{cases} \tag{21}$$

for all $i \neq j$ with $X_{ii} = 0$ for all i. It is assumed that the X_{ij} are independent for all i and j with $i < j$. The null hypothesis that no skill is involved becomes $P(X_{ij} = 1) = 1/2$ for all $i \neq j$. Under this hypothesis, $(S_i + n - 1)/2$ has a binomial distribution $\mathcal{B}(n-1, 1/2)$ for all i. We may use V_n or W_n to test this hypothesis.

This problem goes back to Kendall and Babington Smith (1940) in their treatment of paired comparisons. They suggest using the number of circular triads as the test statistic. A circular triad is a triangle, ijk, in which i beats j, j beats k and k beats i, or conversely j beats i, i beats k and k beats j. In terms of the X_{ij}, the indicator of this event is

$$U_{ijk} = \begin{cases} 1 & \text{if } X_{ij} = X_{jk} = X_{ki} = 1 \text{ or } X_{ij} = X_{jk} = X_{ki} = -1 \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

The number of circular triads is then

$$d_n = \sum\sum\sum_{i<j<k} U_{ijk}. \tag{23}$$

We reject the null hypothesis if d_n is too small.

This test is equivalent (not just asymptotically) to the test that rejects the null hypothesis if V_n or W_n is too large. To see this, note that for the Y_{ijk} of (4), we have $Y_{ijk} = +1$ if $U_{ijk} = 0$ and $Y_{ijk} = -3$ if $U_{ijk} = 1$. Thus,

$$Y_{ijk} = 1 - 4U_{ijk},$$

and from (5),

$$W_n = \sum \sum \sum_{i<j<k} (1 - 4U_{ijk}) = \binom{n}{3} - 4d_n,$$

and from Lemma 1,

$$V_n = 2\binom{n}{2} + 2W_n = 2\binom{n}{2} + 2\binom{n}{3} - 8d_n = \frac{n(n-1)^2}{3} - 8d_n.$$

In Kendall and Babington Smith (1940), it is shown that under the null hypothesis, the distribution of d_n has the the following properties.

Lemma 6. *(a) For $n > 2$ odd, the support of d_n is $\{0, 1, \ldots, n(n^2-1)/24\}$.*
(b) For $n > 2$ even, the support of d_n is $\{0, 1, \ldots, n(n^2-4)/24\}$.

Lemma 7. *(a)* $\mathrm{E}(d_n) = \binom{n}{3}\frac{1}{4}$.
(b) $\mathrm{Var}(d_n) = \binom{n}{3}\frac{3}{16}$.

In Moran (1947) it is shown that d_n is asymptotically normal. This also follows from Theorem 1. This means that, asymptotically, d_n acts as the sum of $\binom{n}{3}$ independent Bernoulli variables with success probability 1/4. Specifically,

$$\sqrt{\frac{16}{3}\binom{n}{3}}\left(\frac{d_n}{\binom{n}{3}} - \frac{1}{4}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

This may not be surprising since d_n is the sum of $\binom{n}{3}$ identically distributed Bernoulli variables. However, lack of independence of the summands skews the distribution of d_n considerably for smaller values of n. Kendall and Babington Smith provide tables of the exact distribution of d_n for $n \leq 7$. These tables were extended to $n \leq 10$ by Alway (1962). These tables show a significant negative skewness. Based on the third and fourth moments of U_n, Kendall and Babington Smith suggest using a chi-square distribution as an approximation to the distribution of d_n for small n.

A detailed analysis of the accuracy of the chi-square approximation to the distribution of d_n was undertaken by Knezek, Wallace, and Dunn-Rankin (1998). In addition, they extend the distribution tables for U_n for n up to 15. (Getting the exact distribution for $n = 15$ is quite a feat, since there are $2^{105} \approx 4.E31$ outcomes to consider.) They conclude that the chi-square approximation is quite good for these values and should continue to be good for larger n. The exact description of the chi-square distribution used for the approximation may be found in this paper and in Kendall and Gibbons (1990), pp 186-187.

To get an idea of the accuracy of the normal approximation to the distribution of d_n, consider the case of $n = 15$ and the one-sided tests at the 5% and 1% levels if significance.

d_{15} has mean $\binom{15}{3}\frac{1}{4} = 113.75$ and variance 85.3125. The 5% cutoff point is then $113.75 - 1.645 \cdot \sqrt{85.3125} = 98.56$. From the tables of Knezek et al., the true $P(d_{15} \leq 98$ is .0634. Both cutoff points 96 and 97 have probability of rejection closer to .05. The 1% is $113.75 - 2,33 \cdot \sqrt{85.3125} = 92.23$. The true $P(d_{15} \leq 92)$ is .022, while the true $P(d_{15} \leq 88)$ is .010088. We see that the normal approximation can be misleading even for n as large as 15.

5. An Example.

We now illustrate our appooach to testing the null hypothesis by V_n and W_n. The Greek Soccer League consists of 16 teams from around Greece. The tournament schedule calls for each team to play every other team twice during the season, once at home and once as a visitor. This results in $15 \cdot 16 = 240$ games.

The original data consists of the final scores of all the games of the 2016-2017 season. This allows some leeway in choosing the payoff for each game. For instance, if a team wins by score of 3 goals to 1, we might consider the difference, +2, to be the score awarded to that team. However, there are various reasons to consider just win, lose or tie as the outcome of each game, the main reason being that this is the true objective of each team, the actual margin of victory or defeat being essentially unimportant. So we consider each game to have three outcomes for a team, a win, a loss or a tie.

An important consideration here is home-team advantage. Of the 240 games played, 66 of them ended in a tie. Of the remaining 174 games, 116 were won by the home team and only 58 were won by the visiting team. This translates to a home-team advantage of $116/174 = 67\%$ that a decisive game will be won by the home team. There is no need to perform a significance test to see that the home-team advantage plays a very significant role in the outcome. However, since each team plays every other team once at home and once as visitor, we may combine the two games played. Thus, we take as the score, X_{ij}, of team i over team j, to be the number of wins of i over j minus the number of wins of j over i. The scores, X_{ij}, then take as possible values the numbers $-2, -1, 0, 1, 2$, and of course we have $X_{ij} = -X_{ji}$.

In Figure 2, we present the names of the sixteen teams followed by the Table of the X_{ij}. The entry, X_{ij}, is the score of the team of row i against the team of column j. In the final column of the table we list the total scores, $S(i)$ of the teams, which is just the sum of the numbers in each row.

For the null hypothesis that the teams are equally matched, it is assumed that the X_{ij} for $i < j$ are independent, identically and symmetrically distributed about zero. That the distribution be symmetric about zero, although natural in this context, is not really needed; it may be replaced by the assumption that the mean be zero.

It is easy to compute the statistic V_n. It is just the sum of the squares of the numbers in the last column of the table, namely, $V_n = 1268$. The easiest way to compute the statistic, W_n, is to use Lemma 1. We find $\sum\sum_{i<j} X_{ij}^2 = 192$, so $W_n = (V_n/2) - 192 = 634 - 192 = 442$.

1. AEK
2. Asteras Tripolis
3. Atromitos
4. Iraklis
5. AOK Kerkira
6. Larissa
7. Levadiakos
8. Olimpiakos
9. Panathinaikos
10. Panetolikos
11. Panionios
12. PAOK
13. PAS Gianena
14. Platanias
15. Veria
16. Xanthi

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	sum
1	*	0	1	0	1	2	2	0	−1	−1	0	0	0	2	2	1	9
2	0	*	−2	0	−2	1	1	−1	−2	1	−1	−2	1	0	−1	−1	−8
3	−1	2	*	2	−1	1	1	−2	−2	−2	−2	0	−1	0	2	1	−2
4	0	0	−2	*	1	0	0	−2	−1	0	−2	−1	2	−1	1	−2	−7
5	−1	2	1	−1	*	1	0	−1	−1	−1	0	−2	−2	−2	0	1	−6
6	−2	−1	−1	0	−1	*	1	0	−1	0	0	−2	−1	−1	1	0	−8
7	−2	−1	−1	0	0	−1	*	−1	−1	0	0	−2	1	−2	0	0	−10
8	0	1	2	2	1	0	1	*	0	2	0	0	2	1	2	2	16
9	1	2	2	1	1	1	1	0	*	1	1	0	1	0	1	−2	11
10	1	−1	2	0	1	0	0	−2	−1	*	−2	−2	−1	−1	1	−2	−7
11	0	1	2	2	0	0	0	0	−1	2	*	0	−1	0	0	2	7
12	0	2	0	1	2	2	2	0	0	2	0	*	0	2	1	0	14
13	0	−1	1	−2	2	1	−1	−2	−1	1	1	0	*	0	0	−1	−2
14	−2	0	0	1	2	1	2	−1	0	1	0	−2	0	*	1	−2	1
15	−2	1	−2	−1	0	−1	0	−2	−1	−1	0	−1	0	−1	*	−2	13
16	−1	1	−1	2	−1	0	0	−2	2	2	−2	0	1	2	2	*	5

Figure 2. Table of the Scores X_{ij}

To perform the large sample test based on W_n as implied by Theorem 1, it is necessary to obtain an estimate of the variance, σ^2. Since $\mathrm{E}X_{ij} = 0$, this is easily done with the estimate

$$\hat{\sigma}^2 = \sum_{i<j} X_{ij}^2 \Big/ \binom{n}{2} = 192/120 = 1.6$$

From Lemma 3, we may then estimate of the standard deviation of W_n by $\hat{\sigma}^2 \sqrt{\binom{16}{3} 3} = 1.6 \cdot 40.99 = 65.58$. This leads to a z-score of $442/65.58 = 6.74$, clear evidence that the teams are not evenly matched.

We may also test the null hypothesis using the statistic V_n and Lemma 4. For this, we estimate the mean of V_n by

$$\hat{\mu} = 16 \cdot 15 \cdot \hat{\sigma}^2 = 384$$

and the fourth moment of X by

$$\hat{\mu}_4 = \sum_{i<j} X_{ij}^4 \Big/ \binom{n}{2} = 624/120 = 5.2.$$

Then Lemma 4 leads to an estimate of the variance of V_n as

$$2 \cdot 16 \cdot 15 \cdot (5.2 + 13 \cdot 1.6^2) = 18470.4$$

whose square root is about 135.9. The z-score is $(1268 - 384)/135.9 = 6.50$, a slightly weaker indication of the falsity of the null hypothesis. But the need to estimate both the mean of the statistic and the fourth moment of X from the data makes this test less reliable.

A *randomization test* based on the statistic W_n may also be used to test the null hypothesis. This is a nonparametric test that avoids computing the estimate of the variance and also avoids wondering if n is large enough for approximate normality to be reasonable, To impliment the randomization test for the Greek Soccer League data, we take the $\binom{16}{2} = 120$ data values X_{ij} for $i < j$, rearrange them in random order, and randomly change the signs of these values with probability 1/2 each. Then the data is completed by assigning $X_{ji} = -X_{ij}$ for $j > i$. From this, one randomized value of W_n is computed. This is done many times to obtain an estimate of the null distribution of W_n, to which the observed value of W_n may be compared. Since $\sum\sum_{i<j} X_{ij}^2$ stays constant in ths computation, we see from Lemma 1 that the randomization test using V_n is equivalent to the one using W_n.

This computation was carried out 10,000 times. We estimate the 5% cutoff point for the statistic to be about 118 and the 1% cutoff point to be about 179. Since the observed statistic was $W_n = 464$, we clearly reject the null hypothesis of equally matched teams. This may be compared to the similar cutoff points using the normal approximation to the distribution of W_n. Using $\hat{\sigma} = 65.58$, the 5% cutoff point is 108 (compared to 118 above) and the 1% cutoff point is 153 (compared to 179). This indicates that the normal approximation rejects the null hypothesis when true somewhat more often than advertised at the 1% and 5% levels.

For help with the Greek Soccer League data, we would like to thank Costis Melolidakis of the Mathematics Department at the University of Athens, and Nicolas Christou of the Statistics Department at UCLA.

References.

G. G. Alway (1962) "The distribution of the number of circular triads in paired comparisons", *Biometrika,* **49**, 265-269.

F. Caron and A. Doucet (2012) "Efficient Bayesian Inference for Generalized BradleyTerry Models", *Journal of Computational and Graphical Statistics,* **21**, 174-196.

C.C. Heyde, Yu V. Prohorov, R. Pyke, and S.T. Rachev (Editors), "Athens Conference on Applied Probability and Time Series", Vol. I, in honour of J. M. Gani, Lect. Notes in Statist. **114**, Springer (1996).

Youichi Iida (2009) "The number of circular triads in a pairwise comparison matrix and a consistency test in the AHP", *J. Operations Research Society of Japan,* **52**, 174-185.

M. G. Kendall and B. Babington Smith (1940) "On the Method of Paired Comparisons", *Biometrika,* **31**, 324-345.

M. G. Kendall, and J. D. Gibbons (1990) *Rank correlation methods*, (5th ed.; Chapter 11) London: Edward Arnold.

Gerald Knezek, Susan Wallace, and Peter Dunn-Rankin (1998) "Accuracy of Kendall's chi-square approximation to circular triad distributions" *Psychometrika,* **63**, 2334.

P. A. P. Moran (1947) "On the method of paired comparisons", *Biometrika,* **34**, 363-365.

Charles Stein (1986) *Approximate Computation of Expectations*, IMS Lecture Notes-Monograph Series, vol. **7**.