Tournaments and Paired Comparisons

Author(s): H. A. David

Source: *Biometrika*, Jun., 1959, Vol. 46, No. 1/2 (Jun., 1959), pp. 139-149

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2332816

TOURNAMENTS AND PAIRED COMPARISONS*

BY H. A. DAVID

Virginia Polytechnic Institute and University of Melbourne

1. INTRODUCTION AND SUMMARY

In an experiment involving paired comparisons t 'stimuli' such as flavours or colours are presented in pairs to a panel of n judges. The basic experimental unit is the comparison of two stimuli by a single judge who must state which one he prefers. Such an individual paired comparison may be regarded as a game between two contestants—the stimuli—while the experiment as a whole is analogous to a tournament of t players. If all the possible $\frac{1}{2}t(t-1)$ comparisons are made by each of the n judges, this corresponds to a Round Robin tournament repeated n times. It will be convenient in this paper to use the language of tournaments rather than that of paired comparisons.

The analogy between a paired comparison experiment and a Round Robin tournament, first pointed out by Kendall (1955), suggests the investigation of another well-known way of running competitions, the Knock-out tournament. It is perhaps surprising that such tournaments do not appear to have been studied at all in the statistical literature, although this may be due to their inherent lack of balance†. However, while this objection is valid if a ranking of all players is desired, it loses much of its force when the aim is to pick the best player. In § 2 some properties of a simple (i.e. unrepeated) Knock-out tournament are investigated; in particular, an expression is found for the probability with which the best player wins such a tournament, various assumptions about the strengths of the players being made.

Round Robin tournaments (in effect) have received considerable attention both from statisticians and psychometricians. One problem is the enumeration of the possible outcomes, expressed as scores, of a tournament of a given size and the determination of the frequencies of these outcomes. The matter has been investigated by Bradley & Terry (1952) and Bradley (1954) and is further considered in § 3 and Table 1. Bradley & Terry have also proposed a model to represent the strengths π_i $(i = 1, 2, \ldots, t; \Sigma\pi_i = 1)$ of the players which allows the π_i to be estimated from the players' scores. A different model had been suggested earlier by Thurstone (1927). Tests of the over-all equality of scores may be based on these methods or on combinatorial arguments leading to a test criterion involving the sum of squares of the scores (Durbin, 1951). For a simple Round Robin tournament the latter approach is equivalent to the 'circular triad test' of Kendall & Babington Smith (1940).

However, these approaches do not provide an answer to such questions as what constitutes a significant difference in the scores of two specified players. To deal with problems of this type, the joint probability distribution of the scores of any $s(\leqslant t)$ players in a *simple* Round Robin tournament is found in § 3·1 on the hypothesis, H_0, that the players are all of equal strength. This result is used to obtain tests of significance based on what are essentially upper percentage points of (i) the score of a specified player, (ii) the difference in the scores of two specified players, and (iii) the top score in the tournament.

Some aspects of repeated tournaments are discussed in § 4.

* Research supported, in part, by the Office of Ordnance Research, U.S. Army.

† [See, however, a paper by Rita J. Maurice (*Biometrika* (1958), **45**, 581–6) published since this paper was written. ED.].

2. Knock-out tournaments

We suppose that the number t of players has been reduced, by preliminary play-offs or otherwise, to a power of 2, say $t = 2^p$. The tournament will then be completed in p rounds. If all contestants are of equal strength, each has probability $t^{-1} = 2^{-p}$ of winning the tournament; or, more briefly, each has *winning probability* 2^{-p}. Likewise, if a particular player beats any one opponent with probability π, he will win the tournament with probability π^p.

In more complicated cases the winning probability of a specified player will evidently depend on the initial tournament draw, which we take to be strictly random. Let us call the t contestants $C_1, C_2, \ldots, C_t$ and write

$$C_i \xrightarrow[\pi]{} C_j \quad (i,j = 1, 2, \ldots, t;\ i \neq j),$$

to express that C_i defeats C_j with probability π. Let $P_i^{(r)}$ $(r = 1, 2, \ldots, p)$ denote the probability that C_i reaches the rth round of the tournament *and* wins his game in this round, so that $P_i^{(p)}$ is his probability of winning the tournament. Finally, let $M_{ij}^{(r)}$ be the probability that C_i meets C_j in the rth round and M_{ij} be the probability of their meeting in the course of the tournament. Clearly

$$M_{ij} = \sum_{r=1}^{p} M_{ij}^{(r)}. \tag{2·1}$$

Suppose now that

$$\begin{aligned} &C_1 \xrightarrow[\pi_1]{} C_3, C_4, \ldots, C_t,\\ &C_2 \xrightarrow[\pi_2]{} C_3, C_4, \ldots, C_t,\\ &C_1 \xrightarrow[\pi_{12}]{} C_2. \end{aligned}$$

Then C_1's probability of winning the tournament is given by

$$P_1^{(p)} = M_{12}\pi_{12}\pi_1^{p-1} + (1 - M_{12})\,\pi_1^p. \tag{2·2}$$

If the draw is known, M_{12} and hence $P_1^{(p)}$ can be evaluated. Instead of considering this, we shall obtain expressions for $\bar{M}_{12}$ and $\bar{P}_1^{(p)}$, the values of M_{12} and $P_1^{(p)}$ *before* the draw. These prior probabilities will still be linked by (2·2). We have

$$\bar{M}_{12}^{(1)} = 1/(t-1).$$

If C_1 and C_2 are to meet in the second round, in which there will be $\frac{1}{2}t$ contestants, they must both be successful in the first round, so that

$$\begin{aligned} \bar{M}_{12}^{(2)} &= (1 - \bar{M}_{12}^{(1)})\,\pi_1\pi_2/(\tfrac{1}{2}t - 1)\\ &= 2\pi_1\pi_2/(t-1). \end{aligned}$$

Continuing the argument, we find

$$\bar{M}_{12}^{(r)} = (2\pi_1\pi_2)^{r-1}/(t-1),$$

and hence from (2·1)

$$\bar{M}_{12} = \sum_{r=1}^{p} (2\pi_1\pi_2)^{r-1}/(t-1).$$

It is interesting to note that the $\bar{M}_{ij}^{(r)}$ depend on π_1 and π_2 only through the product $\pi_1\pi_2$. We see also that the probability of a final between C_1 and C_2 is

$$\bar{M}_{12}^{(p)} = (2\pi_1\pi_2)^{p-1}/(t-1).$$

If for the moment we imagine C_1 and C_2 to be 'seeded' players placed in opposite halves of an otherwise random draw, this probability would become

$$(\pi_1\pi_2)^{p-1} \doteqdot 2\bar{M}_{12}^{(p)}.$$

We turn now to the general case where C_i defeats C_j with probability π_{ij}, the only restriction on the π_{ij} being $\pi_{ji} = 1 - \pi_{ij}$ (due to the absence of ties). It should be noted that the possibility

$$\pi_{ij} > \tfrac{1}{2}, \quad \pi_{jk} > \tfrac{1}{2}, \quad \pi_{ki} > \tfrac{1}{2}$$

is not excluded (compare Kendall & Babington Smith, 1940). This situation can occur when the scale of playing strength is not linear, in which case we define the best player as the one with the largest expected score. The models of Thurstone and Bradley & Terry assume a linear scale corresponding to the choice of suitable expressions for the π_{ij}. As we are not concerned with estimation, they need no separate treatment here.

To illustrate the general case take $t = 8$ and suppose (without loss of generality) that $C_1, \ldots, C_8$ have been drawn in that order. Then we have, for example,

$$P_1^{(1)} = \pi_{12}, P_2^{(1)} = \pi_{21}, P_3^{(1)} = \pi_{34}, \ldots, \tag{2·3}$$

$$P_1^{(2)} = P_1^{(1)}(\pi_{13}P_3^{(1)} + \pi_{14}P_4^{(1)}) = \pi_{12}(\pi_{13}\pi_{34} + \pi_{14}\pi_{43}), \tag{2·4}$$

$$P_1^{(3)} = P_1^{(2)}(\pi_{15}P_5^{(2)} + \pi_{16}P_6^{(2)} + \pi_{17}P_7^{(2)} + \pi_{18}P_8^{(2)}). \tag{2·5}$$

The corresponding probabilities before the draw can be obtained by averaging over all possible draws. Let $P_1^{(2)}(i,j,k)$ denote the joint probability that C_1 beats C_i in the first round and the winner of $C_j v. C_k$ in the second round; then $P_1^{(2)}(2,3,4)$ is the expression of (2·4). Likewise, we may (for $t = 8$) write $P_1^{(3)}$ of (2·5) more fully as $P_1^{(3)}(2,3,4)$. It follows that

$$\bar{P}_1^{(1)} = \frac{1}{7}\sum_{i=2}^{8}\pi_{1i}, \tag{2·6}$$

$$\left.\begin{aligned}\bar{P}_1^{(2)} &= \tfrac{1}{105}\sum_{\substack{i\\1\neq i\neq j\neq k}}\sum_{j}\sum_{<k} P_1^{(2)}(i,j,k),\\ \bar{P}_1^{(3)} &= \tfrac{1}{105}\sum_{\substack{i\\1\neq i\neq j\neq k}}\sum_{j}\sum_{<k} P_1^{(3)}(i,j,k).\end{aligned}\right\} \tag{2·7}$$

For $t > 8$ the number of terms on the right-hand side may become formidable. Thus $\bar{P}_1^{(r)}$ is the mean of

$$(t-1)\binom{t-2}{2}\binom{t-4}{4}\cdots\binom{t-2^{r-1}}{2^{r-1}} \text{ terms.}$$

Example. In a tournament of 8 players, let C_i's 'strength' in his encounter with C_j be represented by

$$x_{ij} = \xi_i + v_{ij},$$

where ξ_i is C_i's expected strength and the error variate v_{ij} is $N(0, \frac{1}{2}\sigma^2)$, the $56\,v_{ij}$'s being mutually independent (compare Mosteller, 1951).

Then

$$\pi_{ij} = \Pr(\xi_j + v_{ji} < \xi_i + v_{ij}) = \int_{-\infty}^{(\xi_i-\xi_j)/\sigma}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}u^2}\,du.$$

We now take ξ_i to be the expected value of the ith largest unit normal deviate in samples of 8, so that the players are arranged in decreasing order of strength with C_1 the strongest. The ξ_i have been tabulated by Teichroew (1956). For $\sigma = 1$ the following array of π_{ij}'s results:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	—	0·7161	0·8291	0·8982	0·9425	0·9710	0·9886	0·9978
C_2	0·2839	—	0·6478	0·7579	0·8425	0·9074	0·9558	0·9886
C_3	0·1709	0·3522	—	0·6256	0·7341	0·8278	0·9074	0·9710
C_4	0·1018	0·2421	0·3744	—	0·6198	0·7341	0·8425	0·9425
C_5	0·0575	0·1575	0·2659	0·3802	—	0·6256	0·7579	0·8982
C_6	0·0290	0·0926	0·1722	0·2659	0·3744	—	0·6478	0·8291
C_7	0·0114	0·0442	0·0926	0·1575	0·2421	0·3522	—	0·7161
C_8	0·0022	0·0114	0·0290	0·0575	0·1018	0·1709	0·2839	—

It will be seen that due to the symmetry of the normal distribution

$$\pi_{ij} = \pi_{9-j, 9-i}.$$

Applying (2·6) and (2·7) we obtain

$$\bar{P}_i^{(1)} = 0{\cdot}906,\quad 0{\cdot}769,\quad 0{\cdot}656,\quad 0{\cdot}551,\quad 0{\cdot}449,\quad 0{\cdot}344,\quad 0{\cdot}231,\quad 0{\cdot}094;$$
$$\bar{P}_i^{(2)} = 0{\cdot}781,\quad 0{\cdot}494,\quad 0{\cdot}319,\quad 0{\cdot}200,\quad 0{\cdot}117,\quad 0{\cdot}060,\quad 0{\cdot}024,\quad 0{\cdot}004.$$

It is interesting to compare these probabilities before the draw with those in a draw in which considerable seeding has been used, the first round games being $C_1\,v.\,C_8, C_5\,v.\,C_4$, $C_3\,v.\,C_6$, $C_7\,v.\,C_2$. From (2·3) to (2·5) we find

$$P_i^{(1)} = 0{\cdot}998,\quad 0{\cdot}956,\quad 0{\cdot}828,\quad 0{\cdot}620,\quad 0{\cdot}380,\quad 0{\cdot}172,\quad 0{\cdot}044,\quad 0{\cdot}002;$$
$$P_i^{(2)} = 0{\cdot}913,\quad 0{\cdot}662,\quad 0{\cdot}312,\quad 0{\cdot}064,\quad 0{\cdot}023,\quad 0{\cdot}020,\quad 0{\cdot}006,\quad 0{\cdot}002;$$
$$P_i^{(3)} = 0{\cdot}692,\quad 0{\cdot}216,\quad 0{\cdot}066,\quad 0{\cdot}019,\quad 0{\cdot}005,\quad 0{\cdot}001,\quad 0{\cdot}000,\quad 0{\cdot}000.$$

3. Round Robin tournaments

The results of a Round Robin tournament can be presented fully in the familiar two-way table of 1's and 0's, with the scores (i.e. number of wins) of each player listed in a column of totals. As a summary of the scores, we may write $(a_1 a_2 \ldots a_t)$, meaning thereby that C_1 has won a_1 games, C_2 has won a_2 games, and so on; clearly $\Sigma a_i = \frac{1}{2}t(t-1)$. If the a's differ greatly, a real difference between at least some of the players is indicated. Thus a Round Robin tournament, unlike a Knock-out tournament—or a ranking—provides the possibility of a test of the null hypothesis, H_0, that all players are of equal strength even when there has been no replication. From this point of view the identity of the players is of no consequence, and we may replace $(a_1 a_2 \ldots a_t)$ by the partition $[x_1^{r_1} x_2^{r_2} \ldots x_m^{r_m}]$, where the x's are simply a re-arrangement of the a's in decreasing order of magnitude, and r_1 is the number of occurrences of the largest score x_1, etc.* It follows that

$$\sum_{u=1}^{m} r_u = t, \quad \sum_{u=1}^{m} r_u x_u = \tfrac{1}{2}t(t-1).$$

One test of H_0 is the 'circular triad test' due to Kendall & Babington Smith (1940); another has been given by Bradley & Terry (1952) and is based on their model referred to earlier. In both cases the test-criterion is a symmetric function of the scores, and its null distribution is easily deduced from a knowledge of the different permissible partitions of $\frac{1}{2}t(t-1)$ together with their frequencies of occurrence. Such basic tables have been constructed by Bradley & Terry (1952), and Bradley (1954) for $t \leqslant 5$. These authors also show how results for several complete repetitions may be built up from those for $n = 1$ and their tables cover many such cases. In Table 1 we have extended their results for $n = 1$ to $t \leqslant 8$, using a score of 1 for a win and 0 for a loss, rather than 1 for a win and 2 for a loss.

A possible approach to the enumeration problem involved is to consider the generating function

$$G(t) = \prod_{i<j} (b_i + b_j), \tag{3·1}$$

a product of $\frac{1}{2}t(t-1)$ terms corresponding to the games played. The expansion of $G(t)$ contains $2^{\frac{1}{2}t(t-1)}$ terms, the exponents in each of which correspond to a possible set of scores. For example,

$$G(3) = (b_1+b_2)(b_1+b_3)(b_2+b_3) = \Sigma b_1{}^2 b_2 + 2b_1 b_2 b_3, \tag{3·2}$$

* We depart slightly from standard notation in showing zeros explicitly.

Table 1. *Partitions of scores and their frequencies in a Round Robin tournament of t players*

$t = 3$

[210]	6
1^{3}	2
Total	8

$t = 4$

[3210]	24
31^{3}	8
$2^{3}0$	8
$2^{2}1^{2}$	24
Total	64

$t = 5$

[43210]	120
431^{3}	40
$42^{3}0$	40
$42^{2}1^{2}$	120
$3^{3}10$	40
$3^{2}2^{2}0$	120
$3^{2}21^{2}$	240
$32^{3}1$	280
2^{5}	24
Total	1,024

$t = 6$

[543210]	720
5431^{3}	240
$542^{3}0$	240
$542^{2}1^{2}$	720
$53^{3}10$	240
$53^{2}2^{2}0$	720
$53^{2}21^{2}$	1,440
$532^{3}1$	1,680
52^{5}	144
$4^{3}210$	240
$4^{3}1^{3}$	80
$4^{2}3^{2}10$	720
$4^{2}32^{2}0$	1,440
$4^{2}321^{2}$	2,880
$4^{2}2^{3}1$	1,680
$43^{3}20$	1,680
$43^{3}1^{2}$	1,680
$43^{2}2^{2}1$	8,640
432^{4}	2,400

$t = 6$ (*cont.*)

[$3^{3}2^{3}$]	2,640
$3^{4}21$	2,400
$3^{5}0$	144
Total	32,768

$t = 7$

[6543210]	5,040
65431^{3}	1,680
$6542^{3}0$	1,680
$6542^{2}1^{2}$	5,040
$653^{3}10$	1,680
$653^{2}2^{2}0$	5,040
$653^{2}21^{2}$	10,080
$6532^{3}1$	11,760
652^{5}	1,008
$64^{3}210$	1,680
$64^{3}1^{3}$	560
$64^{2}3^{2}10$	5,040
$64^{2}32^{2}0$	10,080
$64^{2}321^{2}$	20,160
$64^{2}2^{3}1$	11,760
$643^{3}20$	11,760
$643^{3}1^{2}$	11,760
$643^{2}2^{2}1$	60,480
6432^{4}	16,800
$63^{5}0$	1,008
$63^{4}21$	16,800
$63^{3}2^{3}$	18,480
$5^{3}3210$	1,680
$5^{3}31^{3}$	560
$5^{3}2^{3}0$	560
$5^{3}2^{2}1^{2}$	1,680
$5^{2}4^{2}210$	5,040
$5^{2}4^{2}1^{3}$	1,680
$5^{2}43^{2}10$	10,080
$5^{2}432^{2}0$	20,160
$5^{2}4321^{2}$	40,320
$5^{2}42^{3}1$	23,520
$5^{2}3^{3}20$	11,760
$5^{2}3^{3}1^{2}$	11,760
$5^{2}3^{2}2^{2}1^{2}$	60,480
$5^{2}32^{4}$	16,800
$54^{3}310$	11,760
$54^{3}2^{2}0$	11,760

$t = 7$ (*cont.*)

[$54^{3}21^{2}$]	23,520
$54^{2}3^{2}20$	60,480
$54^{2}3^{2}1^{2}$	60,480
$54^{2}32^{2}1$	206,640
$54^{2}2^{4}$	28,560
$543^{4}0$	16,800
$543^{3}21$	223,440
$543^{2}2^{3}$	183,120
$53^{5}1$	17,808
$53^{4}2^{2}$	72,240
$4^{5}10$	1,008
$4^{4}320$	16,800
$4^{4}31^{2}$	16,800
$4^{4}2^{2}1$	28,560
$4^{3}3^{3}0$	18,480
$4^{3}3^{2}21$	183,120
$4^{3}32^{3}$	99,680
$4^{2}3^{4}1$	72,240
$4^{2}3^{3}2^{2}$	233,520
$43^{5}2$	72,240
3^{7}	2,640
Total	2,097,152

$t = 8$

[76543210]	40,320
765431^{3}	13,440
$76542^{3}0$	13,440
$76542^{2}1^{2}$	40,320
$7653^{3}10$	13,440
$7653^{2}2^{2}0$	40,320
$7653^{2}21^{2}$	80,640
$76532^{3}1$	94,080
7652^{5}	8,064
$764^{3}210$	13,440
$764^{3}1^{3}$	4,480
$764^{2}3^{2}10$	40,320
$764^{2}32^{2}0$	80,640
$764^{2}321^{2}$	161,280
$764^{2}2^{3}1$	94,080
$7643^{3}20$	94,080
$7643^{3}1^{2}$	94,080
$7643^{2}2^{2}1$	483,840
76432^{4}	134,400
$763^{5}0$	8,064

$t = 8$ (*cont.*)

[$763^{4}21$]	134,400
$763^{3}2^{3}$	147,840
$75^{3}3210$	13,440
$75^{3}31^{3}$	4,480
$75^{3}2^{3}0$	4,480
$75^{3}2^{2}1^{2}$	13,440
$75^{2}4^{2}210$	40,320
$75^{2}4^{2}1^{3}$	13,440
$75^{2}43^{2}10$	80,640
$75^{2}432^{2}0$	161,280
$75^{2}4321^{2}$	322,560
$75^{2}42^{3}1$	188,160
$75^{2}3^{3}20$	94,080
$75^{2}3^{3}1^{2}$	94,080
$75^{2}3^{2}2^{2}1$	483,840
$75^{2}32^{4}$	134,400
$754^{3}310$	94,080
$754^{3}2^{2}0$	94,080
$754^{3}21^{2}$	188,160
$754^{2}3^{2}20$	483,840
$754^{2}3^{2}1^{2}$	483,840
$754^{2}32^{2}1$	1,653,120
$754^{2}2^{4}$	228,480
$7543^{4}0$	134,400
$7543^{3}21$	1,787,520
$7543^{2}2^{3}$	1,464,960
$753^{5}1$	142,464
$753^{4}2^{2}$	577,920
$74^{5}10$	8,064
$74^{4}320$	134,400
$74^{4}31^{2}$	134,400
$74^{4}2^{2}1$	228,480
$74^{3}3^{3}0$	147,840
$74^{3}3^{2}21$	1,464,960
$74^{3}32^{3}$	797,440
$74^{2}3^{4}1$	577,920
$74^{2}3^{3}2^{2}$	1,868,160
$743^{5}2$	577,920
73^{7}	21,120
$6^{3}43210$	13,440
$6^{3}431^{3}$	4,480
$6^{3}42^{3}0$	4,480
$6^{3}42^{2}1^{2}$	13,440
$6^{3}3^{3}10$	4,480
$6^{3}3^{2}2^{2}0$	13,440

Table 1 (*cont.*)

$t = 8$ (*cont.*)		$t = 8$ (*cont.*)		$t = 8$ (*cont.*)		$t = 8$ (*cont.*)	
$[6^3 3^2 21^2]$	26,880	$[6^2 43^2 2^3]$	1,464,960	$[6543^4 1]$	3,790,080	$[5^3 43^2 21]$	7,889,280
$6^3 32^3 1$	31,360	$6^2 3^5 1$	142,464	$6543^3 2^2$	12,230,400	$5^3 432^3$	4,282,880
$6^3 2^5$	2,688	$6^2 3^4 2^2$	577,920	$653^5 2$	1,876,224	$5^3 3^4 1$	1,025,920
$6^2 5^2 3210$	40,320	$65^3 4210$	94,080			$5^3 3^3 2^2$	3,306,240
$6^2 5^2 31^3$	13,440			$64^5 20$	142,464	$5^2 4^4 20$	577,920
		$65^3 41^3$	31,360	$64^5 1^2$	142,464	$5^2 4^4 1^2$	577,920
$6^2 5^2 2^3 0$	13,440	$65^3 3^2 10$	94,080	$64^4 3^2 0$	577,920	$5^2 4^3 3^2 0$	1,868,160
$6^2 5^2 2^2 1^2$	40,320	$65^3 32^2 0$	188,160	$64^4 321$	3,790,080		
$6^2 54^2 210$	80,640	$65^3 321^2$	376,320	$64^4 2^3$	1,025,920	$5^2 4^3 321$	12,230,400
$6^2 54^2 1^3$	26,880	$65^3 2^3 1$	219,520			$5^2 4^3 2^3$	3,306,240
$6^2 543^2 10$	161,280			$64^3 3^3 1$	3,924,480	$5^2 4^2 3^3 1$	9,461,760
		$65^2 4^2 310$	483,840	$64^3 3^2 2^2$	9,461,760	$5^2 4^2 3^2 2^2$	22,780,800
$6^2 5432^2 0$	322,560	$65^2 4^2 2^2 0$	483,840	$64^2 3^4 2$	7,203,840	$5^2 43^4 2$	11,518,080
$6^2 54321^2$	645,120	$65^2 4^2 21^2$	967,680	643^6	725,760		
$6^2 542^3 1$	376,320	$65^2 43^2 20$	1,653,120	$5^5 210$	8,064	$5^2 3^6$	577,920
$6^2 53^3 20$	188,160	$65^2 43^2 1^2$	1,653,120			$54^5 30$	577,920
$6^2 53^3 1^2$	188,160			$5^5 1^3$	2,688	$54^5 21$	1,876,224
		$65^2 432^2 1$	5,644,800	$5^4 4310$	134,400	$54^4 3^2 1$	7,203,840
$6^2 53^2 2^2 1$	967,680	$65^2 42^4$	779,520	$5^4 42^2 0$	134,400	$54^4 32^2$	11,518,080
$6^2 532^4$	268,800	$65^2 3^4 0$	228,480	$5^4 421^2$	268,800		
$6^2 4^3 310$	94,080	$65^2 3^3 21$	3,037,440	$5^4 3^2 20$	228,480	$54^3 3^3 2$	23,157,120
$6^2 4^3 2^2 0$	94,080	$65^2 3^2 2^3$	2,486,400			$54^2 3^5$	5,201,280
$6^2 4^3 21^2$	188,160			$5^4 3^2 1^2$	228,480	$4^7 0$	21,120
		$654^4 10$	134,400	$5^4 32^2 1$	779,520	$4^6 31$	725,760
$6^2 4^2 3^2 20$	483,840	$654^3 320$	1,787,520	$5^4 2^4$	107,520	$4^6 2^2$	577,920
$6^2 4^2 3^2 1^2$	483,840	$654^3 31^2$	1,787,520	$5^3 4^3 10$	147,840		
$6^2 4^2 32^2 1$	1,653,120	$654^3 2^2 1$	3,037,440	$5^3 4^2 320$	1,464,960	$4^5 3^2 2$	5,201,280
$6^2 4^2 2^4$	228,480	$654^2 3^3 0$	1,464,960			$4^4 3^4$	3,230,080
$6^2 43^4 0$	134,400			$5^3 4^2 31^2$	1,464,960		
		$654^2 3^2 21$	14,515,200	$5^3 4^2 2^2 1$	2,486,400	Total	268,435,456
$6^2 43^3 21$	1,787,520	$654^2 32^3$	7,889,280	$5^3 43^3 0$	797,440		

the sum being over 6 terms. This identity may be interpreted as follows. In a tournament of 3 players there are 6 outcomes of type [210] where one player wins both his games, a second wins one game and the third wins none; and 2 outcomes of type $[1^3]$ where each player wins one game. $G(t)$ is symmetric in the a's and may be expanded as in (3·2) as a sum of monomial symmetric functions. Table 1 has been constructed essentially by noting that

$$G(t+1) = G(t) \prod_{i=1}^{t} (b_{t+1} + b_i) = G(t)[b_{t+1}^t + (1)\, b_{t+1}^{t-1} + (1^2)\, b_{t+1}^{t-2} + \ldots + (1^t)],$$

where $$(1^s) = \Sigma b_1 b_2 \ldots b_s, \quad (s = 1, 2, \ldots, t).$$

Three generalizations of (3·1) are perhaps worth pointing out, although they do not necessarily provide an easy solution to the problems involved.

(*a*) If C_i and C_j play n_{ij} games instead of merely one, the appropriate generating function is

$$\prod_{i<j} (b_i + b_j)^{n_{ij}}.$$

(*b*) If $C_i \xrightarrow[\pi_{ij}]{} C_j$ for all $i \neq j$ then $G(t)$ can be replaced by

$$\prod_{i<j} (b_i \pi_{ij} + b_j \pi_{ji})$$

to generate the *probabilities* of the various partitions.

(*c*) Suppose that instead of being two-point (1 and 0) the scoring scale is $(r+1)$-point $\{1, (r-1)/r, \ldots, (1/r), 0\}$, with the total score in any one game still unity. Then

$$\prod_{i<j} (b_i + b_i^{(r-1)/r} b_j^{1/r} + b_i^{(r-2)/r} b_j^{2/r} + \ldots + b_j)$$

generates all possible outcomes. This includes the case where games may end in either a win, a loss or a tie ($r = 2$).

3·1. *Significance tests*

As already mentioned, several methods are available for testing the null hypothesis H_0 that all t players in a Round Robin tournament are of equal strength. These tests are exact only for small values of t, but approximations for large t have also been given. However, they are overall tests and do not provide an answer to the more detailed questions we may wish to ask. We consider the following questions:

(i) What score must a *specified* player reach to be regarded as significantly better than average?

(ii) What constitutes a significant difference in scores between two specified players?

(iii) On the basis only of the scores obtained, what constitutes a significantly high score?

In order to deal with these problems the joint distribution of scores, on H_0, of $s\,(\leqslant t)$ players is required. C_i's score clearly has the binomial distribution

$$f(a_i) = \binom{t-1}{a_i} \frac{1}{2^{t-1}}.$$

To find the joint distribution of a_i and a_j, we note that these scores arise if C_i defeats C_j and wins $a_i - 1$ of his other $t-2$ games, while C_j wins a_j; and similarly if C_j defeats C_i. Thus

$$f(a_i, a_j) = \frac{1}{2^{2t-3}} \left[\binom{t-2}{a_i - 1} \binom{t-2}{a_j} + \binom{t-2}{a_i} \binom{t-2}{a_j - 1} \right]. \tag{3·3}$$

The argument may be extended to give the joint distribution of s players $C_i, C_j, \ldots, C_k$ by a consideration of the $2^{\frac{1}{2}s(s-1)}$ outcomes of the $\frac{1}{2}s(s-1)$ games between these s players. If their scores in this sub-tournament are $a'_i, a'_j, \ldots, a'_k$, their final scores will be $a_i, a_j, \ldots, a_k$ with probability

$$f(a_i, a_j, \ldots, a_k \mid a'_i, a'_j, \ldots, a'_k) = \frac{1}{2^{s(t-s)}} \binom{t-s}{a_i - a'_i} \binom{t-s}{a_j - a'_j} \cdots \binom{t-s}{a_k - a'_k},$$

provided we use the convention

$$\binom{n}{r} = 0 \quad \text{for } r < 0 \quad \text{and} \quad r > n. \tag{3·4}$$

It follows that

$$f(a_i, a_j, \ldots, a_k) = \frac{1}{2^{\frac{1}{2}s(2t-s-1)}} \Sigma \prod_{l=i,j,\ldots,k} \binom{t-s}{a_l - a'_l}, \tag{3·5}$$

where the sum is, in effect, over all outcomes of the sub-tournament compatible with the final scores.

We turn now to questions (i)–(iii) above.

(*i*) *Test of the score of a specified player*

Let C_1 be the player in whose performance we are interested. Since, on H_0, a_1 is a binomial variate with parameters $t-1$, $\frac{1}{2}$, it is easy to test H_0 against the alternative that C_1 is above average strength (one-sided binomial test) or different from the average strength of the other players (two-sided binomial test). For example, at the 5 % level of significance the following are the smallest values of a_1 which are significantly higher than the expected value $\frac{1}{2}(t-1)$:

t	a_1
$\leqslant 5$	no significantly high score possible
6–8	$t-1$
9–11	$t-2$
> 11	smallest integer $> \frac{1}{2}[t+1\cdot 645\sqrt{(t-1)}]$.

The large-sample result follows from the normal approximation to the binomial, with continuity correction.

Next, suppose that the players are not of equal strength, but that it is desired to test whether C_1's average probability of success $\pi_{1.}$ equals $\frac{1}{2}$. We have

$$\mathscr{E}(a_1) = (t-1)\,\pi_{1.},$$
$$\operatorname{var} a_1 = \sum_{j=2}^{t} \pi_{1j}(1-\pi_{1j}) \leqslant \tfrac{1}{4}(t-1).$$

Since the variance of a_1 is a maximum under H_0, we may conclude that the above test will be conservative if H_0 is not true. In fact, as

$$\operatorname{var} a_1 \leqslant (t-1)\,\pi_{1.}(1-\pi_{1.})$$

we can also test the hypothesis that $\pi_{1.}$ has a specified value π_0 by referring a_1 to the binomial distribution with parameters $t-1$, π_0.

(ii) *Test of the equality of scores of two specified players*

We shall find the probability distribution $p(d)$ of $d = a_1 - a_2$, assuming H_0 to hold. From (3·3)

$$p(d) = \sum_{a_2} f(a_2+d, a_2) = \frac{1}{2^{2t-3}}\left[\Sigma\binom{t-2}{a_2+d-1}\binom{t-2}{a_2} + \Sigma\binom{t-2}{a_2+d}\binom{t-2}{a_2-1}\right],$$

where in view of the convention (3·4), the summations may be taken over all non-negative integers a_2. Thus

$$p(d) = \frac{1}{2^{2t-3}}\left[\binom{2t-4}{t-d-1} + \binom{2t-4}{t-d-3}\right], \quad d = -(t-1), \quad -(t-2)\ldots, \quad (t-1),$$

is a symmetrical distribution whose variance is $\frac{1}{2}t$. It can be shown that the characteristic function of d is

$$\phi(u) = (\cos \tfrac{1}{2}u)^{2t-4} \cos u,$$

and hence that $d/\sqrt{(\frac{1}{2}t)}$ tends to a unit normal distribution. Numerical work indicates this tendency to be rapid.

At the 5 % level of significance, the smallest significant values of $|d|$ are as follows:

t	$\lvert d\rvert$
$\leqslant 4$	no significant values of $\lvert d\rvert$ possible
5–6	4
7–10	5
> 10	smallest integer $> 1\cdot 96\sqrt{(\frac{1}{2}t)} + 0\cdot 5$.

While this approach tests equality of the two scores under H_0 it will also provide a conservative test of the null hypothesis

$$H_0'\colon \quad \pi_{1.} = \pi_{2.} \quad \text{against the alternative } \pi_{1.} \neq \pi_{2.};$$

for, in the general case, we find

$$\operatorname{var} d = 4\pi_{12}(1-\pi_{12}) + \sum_{j=3}^{t} [\pi_{1j}(1-\pi_{1j}) + \pi_{2j}(1-\pi_{2j})],$$

which is a maximum under H_0.

(iii) *Test of the top score*

The test developed in this section is essentially different from those in (i) and (ii) in that it is not based on the scores of *specified* players, but on the score x_1 of the winning player (or players). We obtain, on H_0, the probability of large values of x_1 from which upper percentage points may be constructed. If in a particular tournament the appropriate upper percentage point is attained or exceeded by one or more players, H_0 is rejected and, as in the analogous problem of testing for 'outliers', these players may be regarded as 'superior'.

It is easy to see that for $t \leqslant 8$ the probability distribution of x_1 could be built up from Table 1. We shall employ a different approach which is useful also for larger values of t. Since x_1 is the largest of the scores $a_1, a_2, \ldots, a_t$, the method of inclusion and exclusion gives (see, for example, David, 1956)

$$\Pr(x_1 \geqslant t-r) = \sum_{s=1}^{t} (-)^{s-1} \binom{t}{s} \Pr(a_1 \geqslant t-r, a_2 \geqslant t-r, \ldots, a_s \geqslant t-r), \tag{3·6}$$

where r is any integer from 1 to t. For small values of r, the later terms in the sum will be zero, since only a few players can have a large score. We illustrate the method on the evaluation of the typical term

$$\Pr(a_1 \geqslant t-3, a_2 \geqslant t-3, a_3 \geqslant t-3, a_4 \geqslant t-3).$$

Consider the sub-tournament played by C_1, C_2, C_3 and C_4. From Table 1 we see that the possible outcomes are

$$[3210], \quad [31^3], \quad [2^30], \quad [2^21^2]$$

with respective frequencies

$$24, \quad 8, \quad 8, \quad 24.$$

It is, however, clear that only $[31^3]$ and $[2^21^2]$ can result in final scores all greater than $t-4$. The possible outcomes (ignoring the order of C_1 to C_4), and the frequencies with which they can arise from $[31^3]$ and $[2^21^2]$, are as follows:

				$[31^3]$	$[2^21^2]$
$t-1$,	$t-3$,	$t-3$,	$t-3$	1	0
$t-2$,	$t-2$,	$t-3$,	$t-3$	0	1
$t-2$,	$t-3$,	$t-3$,	$t-3$	$t-4$	$2(t-4)$
$t-3$,	$t-3$,	$t-3$,	$t-3$	$\frac{1}{2}(t-4)(t-5)$	$(t-4)^2$

This gives a *total* frequency of $28t^2 - 172t + 272$, which can be converted into the required probability on division by $2^{2(2t-5)}$, the factor of equation (3·5). In this way we obtain from (3·6)

$$\Pr(x_1 = t-1) = \frac{t}{2^{t-1}},$$

$$\Pr(x_1 \geqslant t-2) = \frac{t^2}{2^{t-1}} - \frac{\binom{t}{2}(t-1)}{2^{2t-4}} + \frac{\binom{t}{3}}{2^{3t-7}},$$

$$\Pr(x_1 \geqslant t-3) = \frac{t}{2^t}(t^2 - t + 2) - \frac{\binom{t}{2}}{2^{2t-3}}(t^3 - 4t^2 + 7t - 4)$$
$$+ \frac{\binom{t}{3}}{2^{3t-6}}(5t^3 - 33t^2 + 78t - 64) - \frac{\binom{t}{4}}{2^{4t-12}}(7t^2 - 43t + 68) + \frac{3\binom{t}{5}}{2^{5t-18}}.$$

These three probabilities are as follows for $t \leqslant 20$:

t	$\Pr(x_1 \geqslant t-1)$	$\Pr(x_1 \geqslant t-2)$	$\Pr(x_1 \geqslant t-3)$
3	0·7500	1·0000	1·0000
4	0·5000	1·0000	1·0000
5	0·3125	0·9766	1·0000
6	0·1875	0·8418	1·0000
7	0·1094	0·6447	0·9987
8	0·0625	0·4526	0·9637
9	0·0352	0·2989	0·8536
10	0·0195	0·1891	0·6868
11	0·0107	0·1161	0·5093
12	0·0059	0·0696	0·3548
13	0·0032	0·0410	0·2360
14	0·0017	0·0238	0·1517
15	0·0009	0·0137	0·0950
16	0·0005	0·0078	0·0584
17	0·0003	0·0044	0·0353
18	0·0001	0·0025	0·0211
19	0·0001	0·0014	0·0125
20	0·0000	0·0008	0·0073

We see from this table that an unbroken sequence of wins is not significant, at the 5 % level, until $t = 9$ in the case of a one-sided test and $t = 10$ for a two-sided test. For a one-sided test a score of $t-2$ is significantly high for $t \geqslant 13$, and a score of $t-3$ for $t \geqslant 17$.

4. Discussion

Some properties of Knock-out and Round Robin tournaments have been presented. We now make a few remarks about their respective merits and drawbacks.

The attraction of Knock-out tournaments is, at least in part, that a large number of players can be handled quickly. With $t = 2^p$ competitors a decision is reached after p rounds and $t-1$ games as against $t-1$ rounds and $\frac{1}{2}t(t-1)$ games required for a Round Robin tournament if no play-offs are necessary. A natural objection to Knock-out tournaments is that they are too unreliable. This is often true in the simple cases discussed in § 2, but in view of the relative brevity of such tournaments it may well be practicable to repeat them. The simplest way of doing this is to start with a fresh random draw for each repetition and to declare as winner the player with the largest total number of wins, possibly after a play-off. Alternatively, the initial draws of each repetition could be arranged to satisfy certain symmetry requirements. We do not recommend either of these approaches if the aim of the tournament is to rank all players in order of merit. For this purpose a simple Round Robin tournament will be preferable to $\frac{1}{2}t$ repetitions of a Knock-out tournament, although both require the same (minimum) number of games. However, the situation will be very different when we are primarily concerned with picking the best player, since the repeated Knock-out tournament will tend to include a larger number of encounters between the stronger players. Some work along these lines is in progress. There are also possibilities, for both types of tournaments, of some sequential procedure whereby the number of contestants is progressively reduced.

In this paper attention has been concentrated on tournaments without repetition of any kind. When comparing a number of treatments in a paired comparison experiment, we will often require some repetition, possibly by different judges. If repeated Round Robin tournaments are impracticably large, the incomplete block type designs put forward by Kendall (1955) and Bose (1956) are helpful. However, to answer the type of question considered in § 3, generalizations of tests there given as well as new tests are required. It is hoped to deal with some of these issues in a future publication.

I should like to thank Dr R. A. Bradley for some helpful discussions and Dr I. R. Savage for a valuable comment. My thanks are also due to Miss Betty Laby and Miss G. E. Jacobs of the University of Melbourne and to Mr A. Romano of the Virginia Polytechnic Institute for assistance with the computations.

REFERENCES

BOSE, R. C. (1956). Paired comparison designs for testing concordance between judges. *Biometrika*, **43**, 113–21.

BRADLEY, R. A. (1954). Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502–37.

BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324–45.

DAVID, H. A. (1956). On the application to statistics of an elementary theorem in probability. *Biometrika*, **43**, 85–91.

DURBIN, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psych.* (Statistics Section), **4**, 85–90.

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

KENDALL, M. G. & BABINGTON SMITH, B. (1940). On the method of paired comparisons. *Biometrika*, **31**, 324–45.

MOSTELLER, F. (1951). Remarks on the method of paired comparisons. I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, **16**, 3–9.

TEICHROEW, D. (1956). Tables of expected values of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Statist.* **27**, 410–26.

THURSTONE, L. L. (1927). Psychophysical analysis. *Amer. J. Psychol.* **38**, 368–89.