SOME EQUIVALENCE CLASSES IN PAIRED COMPARISONS[1]

By Joseph B. Kadane

Stanford University

Introduction. In a paired comparison experiment n judges give a preference in some or all of the $\binom{t}{2}$ pairs of t items. Frequently the purpose of the experiment is to test the null hypothesis that every preference is equally likely against a vaguely defined alternative of consistency. Our purpose is to study several of the tests used, from the point of view of a natural equivalence relation which arises in graph theory. In the first section we introduce graph theory notation, the equivalence relation, and some results on partial and strict orderings on the equivalence classes. The succeeding section applies these notions to Kendall and Babington Smith's statistic in detail (hereafter simply referred to as Kendall's statistic), and mentions applications in the Bradley-Terry model, and the strong-stochastic ordering model.

1. Notions from graph theory. We define a paired comparison experiment, for these purposes, to consist of

(i) a set X of t items, which are the items being compared by the judges, and

(ii) n ordered relations R_k $(k = 1, \cdots, n)$, subsets of $X \times X$, which are the preferences of the n judges. Thus $(x_i, x_j) \varepsilon R_k$ is interpreted to mean that item x_i is preferred to item x_j by the kth judge. We require that these n ordered relations be

(a) *anti-symmetric* $[(x, y) \varepsilon R_k \Rightarrow (y, x) \not\varepsilon R_k]$, thus each pair is judged at most once.

(b) *anti-reflexive* $[(x, x) \not\varepsilon R_k]$. No item is to be thought of as being preferred to itself.

A *path* K in $(R_1, \cdots, R_n) = R$ from y_1 to y_k, denoted $(y_1, \cdots, y_k)$ is a finite collection of ordered pairs $(y_1, y_2) \varepsilon R_{i_1}, \cdots, (y_{k-1}, y_k) \varepsilon R_{i_{k-1}}$. If $y_1 = y_k$, the path is called a *circuit.* If x and y are in some circuit together or $x = y$ then x and y are said to be *equivalent* (written $x \equiv y$). It is immediate that $\equiv$ is an equivalence relation. If $(x, y) \varepsilon R$ but x and y are not equivalent, then we may say x is an *ancestor* of y, or y is a *descendant* of x.

We will now study a natural ordering on the equivalence classes of the above equivalence relation.

Theorem 1. *There is a natural partial ordering on the equivalence classes. This ordering is given by*

Received 7 October 1965; revised 13 December 1965.

[1] This research was begun under the sponsorship of the Edgar Stern Family Fund, and completed under USPHS Training Grant 5T1GM 25-07. I wish to thank Professor L. Moses for his patience in reading several versions and suggesting improvements, as well as Professor H. Rubin for his helpful comments.

$$\alpha_1 > \alpha_2 \qquad \textit{if} \qquad \exists\, a_1 \,\varepsilon\, \alpha_1 , a_2 \,\varepsilon\, \alpha_2 \ni (a_1 , \cdots , a_2)$$

is a path in R.

PROOF. First we must show that the concept is well defined. Suppose that α_1 and α_2 are two distinct equivalence classes and that $\alpha_1 > \alpha_2$ and $\alpha_2 > \alpha_1$. Then there exist items $a_1, a_1' \,\varepsilon\, \alpha_1, a_2, a_2' \,\varepsilon\, \alpha_2$ such that $(a_1, \cdots, a_2)$ and $(a_2', \cdots, a_1')$ are paths in R. Since α_1 and α_2 are equivalence classes, there exist paths $(a_2, \cdots, a_2')$ and $(a_1', \cdots, a_1)$ in R. Then $(a_1, \cdots, a_2, \cdots, a_2', \cdots, a_1', \cdots, a_1)$ is a circuit. Since equivalence classes are either identical or disjoint, $\alpha_1 = \alpha_2$, contradiction.

Second we must show that $>$ is transitive. If $\alpha_1 > \alpha_2$ and $\alpha_2 > \alpha_3$ then there exist $a_1 \,\varepsilon\, \alpha_1, a_2, a_2' \,\varepsilon\, \alpha_2, a_3 \,\varepsilon\, \alpha_3$ such that $(a_1, \cdots, a_2)$ and $(a_2', \cdots, a_3)$ are paths in R. Since there also exists a path $(a_2, \cdots, a_2')$, there is a path $(a_1, \cdots, a_2, \cdots, a_2', \cdots, a_3)$ in R, so $\alpha_1 > \alpha_3$, which proves transitivity. Q.E.D.

COROLLARY 1. *If for every two equivalence classes* α_1 *and* α_2, $\exists\, a_1 \,\varepsilon\, \alpha_1, a_2 \,\varepsilon\, \alpha_2 \ni (a_1, a_2) \,\varepsilon\, R$ *or* $(a_2, a_1) \,\varepsilon\, R$, *then the above order is strict.*

R is *complete* if each distinct pair is considered once in R, i.e., if $(x, y) \not\varepsilon R \Rightarrow (y, x) \,\varepsilon\, R$ or $x = y$.

COROLLARY 2. *If* R *is complete then the above order is strict.*

We define the *score* of an item x in the comparison R, written $\text{sc}\,(x \mid R)$, to be the number of times it is preferred to other items in R.

LEMMA 1. *If* R_i, $i = 1, \cdots, n$ *are complete, then* $\text{sc}\,(x \mid R) \geqq \text{sc}\,(y \mid R) \Rightarrow x$ *is an ancestor of* y *or* $x \equiv y$.

PROOF. We must show that there exists a path from x to y. Suppose first that $n = 1$. If $(x, y) \,\varepsilon\, R$ we are done. If not, $\exists\, z \,\varepsilon\, X$ such that $(x, z) \,\varepsilon\, R$, $(y, z) \not\varepsilon R$. Then by completeness, $(z, y) \,\varepsilon\, R$. Then (x, z, y) is a path in R.

For any n, we have

$$\begin{aligned} \text{sc}\,(x \mid R) &= \text{sc}\,(x \mid R_1, \cdots, R_n) \\ &= \textstyle\sum_{i=1}^{n} \text{sc}\,(x \mid R_i) \geqq \sum_{i=1}^{n} \text{sc}\,(y \mid R_i) = \text{sc}\,(y \mid R) \end{aligned}$$

therefore $\exists\, i$ such that $\text{sc}\,(x \mid R_i) \geqq \text{sc}\,(y \mid R_i)$. Then the above applies to complete the proof. Q.E.D.

COROLLARY 3. *If* R_i, $i = 1, \cdots, n$, *are complete, then* $\text{sc}\,(x \mid R) = \text{sc}\,(y \mid R) \Rightarrow x \equiv y$.

LEMMA 2. *If* R_i, $i = 1, \cdots, n$, *are complete and* α_1 *and* α_2 *are two distinct equivalence classes, then* $\alpha_1 > \alpha_2$ *if and only if* $(\forall\, a_1 \,\varepsilon\, \alpha_1)(\forall\, a_2 \,\varepsilon\, \alpha_2)$, $\text{sc}\,(a_1 \mid R) > sc(a_2 \mid R)$.

PROOF. Suppose $\alpha_1 > \alpha_2$ but $\exists\, a_1^* \,\varepsilon\, \alpha_1, a_2^* \,\varepsilon\, \alpha_2 \ni \text{sc}\,(a_2^* \mid R) \geqq \text{sc}\,(a_1^* \mid R)$. Then by Lemma 1, $(a_2^*, \cdots, a_1^*)$ is a path. $\alpha_1 > \alpha_2 \Rightarrow \exists\, a_1 \,\varepsilon\, \alpha_1, a_2 \,\varepsilon\, \alpha_2 \ni (a_1, \cdots, a_2)$ is a path. Then $(a_1, \cdots, a_2, \cdots, a_2^*, \cdots, a_1^*, \cdots, a_1)$ is a circuit so $\alpha_1 = \alpha_2$, contradiction. The other way is trivial. Q.E.D.

2. Kendall and Babington Smith's statistic. In this section we restrict ourselves to $n = 1$, R complete, and consider Kendall's statistic d, the number of distinct circuits of three items.

LEMMA 3. *$d = 0$ if and only if each equivalence class α has exactly one element.*

PROOF. Suppose $d = 0$, but some α has more than one element, say a and b. Suppose arbitrarily that $(a, b) \varepsilon R$. There must be a path from b to a, say $(b, \cdots, c, \cdots, a)$ since $(b, a) \not\varepsilon R$ by anti-symmetry. Then $(a, b, \cdots, c, \cdots, a)$ is a circuit with possibly more than three items. A lemma of Kendall and Babington Smith [5] assures us that it must contain at least one circuit of three items ["circular triads" in their terminology]. But then $d > 0$, contradiction.

If $d \neq 0$, then there is a circuit of three items, and so some equivalence class must contain at least those three items. Q.E.D.

Then $d = 0$ if and only if there are t equivalence classes, and hence each score from 0 to $t - 1$ is represented once and only once. Thus ties in score are associated with $d > 0$. Suppose two items, say a and b, both have score s, and we suppose without loss of generality, that $(a, b) \varepsilon R$. Now we consider $R' = [R - (a, b)] \cup (b, a)$, so that R' is the relation which results when the preference (a, b) is reversed. Then sc $(a \mid R') = s - 1$ and sc $(b \mid R') = s + 1$. In R, a and b were in the same equivalence class, by Corollary 3, but in R, a and b may or may not be in the same equivalence class.

We will call breaking ties in this fashion *restricted changes* since the only items whose preference can be reversed are those with tied score. Let K be the number of times ties are broken in this way before there are no ties left, arriving at $d = 0$.

THEOREM 2. $K = d$.

PROOF. This proof is essentially due to Kendall and Babington Smith [5], although they did not state the theorem. Suppose a and b both have score s. From a single restricted change of $(a, b) \varepsilon R$ to $(b, a) \varepsilon R'$ the only trials which might become circular or those which might cease to be so are those including both a and b. If x represents the third item, there are four possible configurations:

(1) $(a, x) \varepsilon R$, $(b, x) \varepsilon R$, say y in number.

(2) $(x, a) \varepsilon R$, $(x, b) \varepsilon R$, say z in number.

(3) $(a, x) \varepsilon R$, $(x, b) \varepsilon R$, which are $s - y - 1$ in number since sc $(a \mid R') = s - 1 =$ number in catagories (1) and (3).

(4) $(x, a) \varepsilon R$, $(b, x) \varepsilon R$, which are $s - y$ in number, since sc $(b \mid R') = s + 1 =$ number in catagories (1) and (4) $+ 1$ for $(b, a) \varepsilon R'$.

In the change from R to R', items in the fourth catagory cease to be circular and items in the third become so. The change in number of distinct circuits of three items is

$$(s - y) - (s - y - 1) = 1.$$

Thus K and d are the same except for a constant c.

$$K = d + c.$$

But $K = 0$ if and only if $d = 0 \Rightarrow c = 0$. Thus $K = d$. Q.E.D.

Then Kendall's statistic is the number of times ties must be broken by changing the preference between tied items to arrive at a strict ordering, that is, the number of restricted changes required to achieve a strict ordering.

Another statistic, Slater's i, [7], is the number of unrestricted changes required to achieve $d = 0$. Theorem 2 provides a natural comparison between Slater's i and Kendall's d. The former weights every inconsistency equally, whereas the latter weights more heavily switches of items with more disparate scores as noted in [3], p. 34. Thus Slater's is useful if we want to protect ourselves against errors of recording, where every error is equally likely.

However, in the case of a judge, who, scaling on some continuum, should be able to distinguish items "far apart" more easily than those "close together," Kendall's d is more appropriate. This is the situation, for example, in international relations where actions are scaled for the degree of violence or potential violence in them. To check the reliability of the scaling, each judge is given a small sample of items to be examined in pairs. If the judge is nearly consistent, the scaled data can be accepted as reliable, but if the judge's choices are not significantly different from those chosen by a fair coin, then the scaled data should be rejected, (see Zinnes [8]). Such a judge should be able to distinguish between a declaration of war and a signing of a peace treaty more easily than he can between two vaguely threatening military maneuvers. Failure to do so should be counted more heavily against the alternative hypothesis (of "consistent" ordering) in the first case than in the second.

Further, from Theorem 2 we immediately have

COROLLARY 4. *Slater's $i \leqq$ Kendall's d.*

The K-representation also means that the relative scores within equivalence classes are sufficient for d. This opens up the possibility of finding the distribution of d under the null hypothesis that every choice is equally likely.

From Lemma 3, we know immediately that

$$P_t(d = 0) = t!/2^{\binom{t}{2}}.$$

The only way we can have $d = 1$ is to have one element in every equivalence class except one, which must have three items of the same score s, and there must be none with scores $s - 1$ nor $s + 1$. This will be written in terms of relative scores as 0-3-0. Only in this way can breaking one tie lead to $d = 0$.

In how many ways can this happen? There are $\binom{t}{3,1,\cdots,1} = t!/3!$ ways of assigning items to the equivalence classes, $(t - 2)$ different orderings for the equivalence classes, two possible preference orderings among the three items in the equivalence class ($A > B > C > A$ and $A < B < C < A$), and a requirement that t be at least three. Then to summarize, we have

$$P_t(d = 1) = [t!/2^{\binom{t}{2}}](2/3!)(t - 2)e_3(t)$$

where

$$\begin{aligned} e_p(t) &= 1 \qquad \text{for } t \geqq p \\ &= 0 \qquad\quad t < p. \end{aligned}$$

For the case $d = 2$, there are two possibilities: one equivalence class with 0-2-2-0 or two, each 0-3-0, as the reader may verify by examining the ways by which, breaking one tie, we arrive at $d = 1$, i.e., 0-3-0. The pattern {0-2-2-0}, is called *simple* since it contains only one equivalence class with more than one element. The pattern {0-3-0, 0-3-0} is, by distinction, called compound. The corresponding formula is

$$P_t(d = 2) = [t!/2^{\binom{t}{2}}][(t - 3)e_4(t) + \tfrac{1}{18}(t - 4)(t - 5)e_6(t)].$$

Similarly we have

$$\begin{aligned} P_t(d = 3) = {} & (t!/2^{\binom{t}{2}})[2(t - 4)e_5(t) + \tfrac{1}{3}(t - 5)(t - 6)e_7(t) \\ & + \tfrac{1}{162}(t - 6)(t - 7)(t - 8)e_9(t)] \end{aligned}$$

$$\begin{aligned} P_t(d = 4) = {} & [t!/2^{\binom{t}{2}}][\tfrac{7}{3}(t - 4)e_5(t) + 4(t - 5)e_6(t) + \tfrac{7}{6}(t - 6)(t - 7)e_8(t) \\ & + \tfrac{1}{18}(t - 7)(t - 8)(t - 9)e_{10}(t) \\ & + \tfrac{1}{1944}(t - 8)(t - 9)(t - 10)(t - 11)e_{12}(t)] \end{aligned}$$

and

$$\begin{aligned} P_t(d = 5) = {} & [t!/2^{\binom{t}{2}}][\tfrac{1}{5}(t - 4)e_5(t) + \tfrac{14}{3}(t - 5)e_6(t) + 8(t - 6)e_7(t) \\ & + \tfrac{7}{9}(t - 6)(t - 7)e_8(t) + \tfrac{10}{3}(t - 7)(t - 8)e_9(t) \\ & + \tfrac{5}{18}(t - 8)(t - 9)(t - 10)e_{11}(t) \\ & + \tfrac{1}{162}(t - 9)(t - 10)(t - 11)(t - 12)e_{13}(t) \\ & + \tfrac{1}{29{,}160}(t - 10)(t - 11)(t - 12)(t - 13)(t - 14)e_{15}(t)]. \end{aligned}$$

In general, the same reasoning leads us to the formula

$$P_t(d = i) = \frac{t!}{2^{\binom{t}{2}}} \sum_{j=1}^{p_i} m_{ij} \binom{t + 1 - l_{ij}}{n_1^{ij}, n_2^{ij}, \cdots, n_k^{ij}} e_{l_{ij}}(t) \qquad i \geq 1$$

where l_{ij} is the number of items not in equivalence classes of one element in the jth pattern yielding $d = i$, n_N^{ij} is the number of equivalence classes with Nth relative score pattern, and m_{ij} is a multiplicity factor explained below.

The possible simple patterns are determined from previous simple patterns by seeing how, with one change, one can get $d = i - 1$. The possible compound patterns are found by unions of simple patterns, when the sum of the changes required for $d = 0$ is i.

For simple patterns, l_{ij} is obtained immediately, and m_{ij} is taken from the table of David [2], dividing the number he gives by $l_{ij}!$ For compound patterns, l_{ij} is the sum over the component equivalence classes, and m_{ij} is the product over the components.

For instance, let us derive carefully the formula for $P_t(d = 4)$. We begin with a knowledge of the simple patterns for $d = 1$, 2 and 3:

Score	Pattern	l_{ij}	m_{ij}
1	0-3-0	3	$\frac{1}{3}$
2	0-2-2-0	4	1
3	0-2-1-2-0	5	2

In order to find the simple patterns for $d = 4$, we must find those patterns which, by the breaking of one tie can change into 0-2-1-2-0. A tie could be broken where a zero is; if so the previous pattern was 0-2-1-1-2-0. A tie could not have been broken where either 2 is, since there is an adjacent zero. However, a tie could have been broken where the 1 is, leading to the pattern 0-1-3-1-0. The reader may verify that these are the only possibilities. Thus there are two simple patterns for $d = 4$: 0-1-3-1-0 and 0-2-1-1-2-0. Their respective lengths are 5 and 6. To discover their multiplicities we convert them into the notation of David [2] as $[32^31]$ and $[4^2321^2]$. The table gives respective values 280 and 2880, and dividing by the length factorial yields multiplicities of $\frac{7}{3}$ and 4. Thus to summarize we have

Pattern	l_{ij}	m_{ij}	Contribution to Formula
0-1-3-1-0	5	$\frac{7}{3}$	$\frac{7}{3}(t-4)e_5(t)$
0-2-1-1-2-0	6	4	$4(t-5)e_6(t)$

for the simple patterns. In addition we have compound patterns composed of the simple patterns for $d = 1, 2, 3$. In particular we have

Pattern	l_{ij}	m_{ij}	Contribution to Formula
0-3-0, 0-2-1-2-0	8	$\frac{1}{3} \times 2 = \frac{2}{3}$	$\frac{2}{3}\binom{t-7}{1,1}e_8(t)$
0-2-2-0, 0-2-2-0	8	$1 \times 1 = 1$	$1\binom{t-7}{2}e_8(t)$
0-2-2-0, 0-3-0, 0-3-0	10	$1 \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$	$\frac{1}{9}\binom{t-9}{2,1}e_{10}(t)$
0-3-0, 0-3-0, 0-3-0, 0-3-0	12	$\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{81}$	$\frac{1}{81}\binom{t-11}{4}e_{12}(t)$

The sum of the contributions of these six patterns, two simple and four compound, gives the formula for $P_t(d = 4)$ above.

Thus it is possible, in principle, to extend these formulae indefinitely.

The equivalence classes discussed here in the context of Kendall's statistic also occur in the study of other paired comparison statistics. For example, Ford's criterion [4] quoted in David [3], for the convergence of estimates of the Bradley-Terry model [1] reduces to the existence of only one equivalence class. The same considerations apply to all linear models (see Noether [6], David [3],) and to the strong-stochastic ordering model.

REFERENCES

[1] Bradley, R., and Terry, M. (1952). The rank analysis of incomplete block designs I the method of paired comparisons. *Biometrika* **39** 324–345.

[2] David, H. A. (1959). Tournaments and paired comparisons. *Biometrika* **46** 139–149.

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

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

[5] KENDALL, M. G. and BABINGTON SMITH, B. (1940). On the method of paired comparisons. *Biometrika* **31** 324–345.

[6] NOETHER, G. E. (1960). Remarks about a paired comparison model. *Psychometrika* **25** 357–367.

[7] SLATER, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika* **48** 303–312.

[8] ZINNES, D. (1963). Expression and perception of hostility in interstate relations. Unpublished dissertation, Department of Political Science, Stanford University.