The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 2002–2015

A TREATMENT OF TIES IN PAIRED COMPARISONS[1]

By Jagbir Singh[2] and W. A. Thompson, Jr.

The Florida State University

1. Introduction. This paper extends the results of Thompson and Remage (1964) and Remage and Thompson (1966) to cover the treatment of ties in obtaining maximum likelihood paired comparison rankings.

An easily understood, nonmathematical description of this problem is the following: Given the win-loss-tie records of the various major college football teams in the nation, but ignoring the scores of the games, under minimal assumptions determine the national rankings of the teams. In particular, since all teams do not play one another, determine how pairs of teams may be ranked indirectly and which pairs cannot be ranked at all.

More important scientific motivations for the research arise in psychology and other social sciences. For various sound experimental reasons the statistician will frequently encounter paired comparison data. The members of a set X of m objects are compared two at a time by a "subject" who states his preferences and indifferences. In this way two basic comparisons $x_i \rightarrow x_j$ and $x_k — x_l$, read "x_i preferred to x_j" and "x_k ties x_l", are established among the objects of X. If these objects stand in some order then as a minimum we expect that the ordering relation will be transitive and asymmetric. Such a relation is called a preference. In general, the basic comparisons will not immediately yield a preference relation; some pairs of objects will not have been compared directly and an indirect method of comparison will have to be found. Indirect comparison is the topic of Section 2. But indirect comparisons will usually be self-contradictory and some basic comparisons will have to be deleted or changed in order to obtain a preference. The effect on ranking of changing the orientation of lines is studied in Section 3.

In Section 4, to find a criterion for altering basic comparisons, we advance the following stochastic model for paired comparison data with ties. The m objects of X are independently (in the probability sense) compared in pairs. x_i and x_j are compared on $n_{ij} \geqq 0$ independent trials; each trial having three possible outcomes denoted respectively by $x_i \rightarrow x_j$, $x_j \rightarrow x_i$, and $x_i — x_j$. Let $\mathcal{I}$ denote the set of subscript pairs (ij) such that $1 \leqq i < j \leqq m$ and $n_{ij} > 0$ and let n be the number of pairs in $\mathcal{I}$. We have $n \leqq \binom{m}{2}$ with equality holding in the important case where every pair of objects is compared at least once. We consider these basic paired comparisons to constitute a sample and we introduce the population parameters $\pi_{ij} = P(x_i \rightarrow x_j)$ and $\gamma_{ij} = P(x_i — x_j)$. The probability

Received 21 September 1967.

[1] Research supported in part under contracts NONR-988(08) and NSF GP-6359. Reproduction is permitted for any purpose of the United States Government.

[2] Now at Ohio State University.

that $x_i \to x_j$ occurs s_{ij} times and x_i—x_j occurs t_{ij} times ($s_{ij} + s_{ji} + t_{ij} = n_{ij}$) for each and every pair of items is given by

$$(1) \qquad \prod_{\mathcal{I}} n_{ij}!(s_{ij}!\, s_{ji}!\, t_{ij}!)^{-1}\pi_{ij}^{s_{ij}}\pi_{ji}^{s_{ji}}\gamma_{ij}^{t_{ij}}.$$

We may as well restrict our attention to those parameters with subscripts in $\mathcal{I}$; since $\pi_{ij} + \pi_{ji} + \gamma_{ij} = 1$ and $\gamma_{ij} = \gamma_{ji}$, there are exactly $2n$ such functionally independent parameters. Thus the parameter space Ω consists of all π_{ij} and γ_{ij} with subscripts in $\mathcal{I}$; it is a subset of $2n$ dimensional space. We denote by π a typical point of Ω. In Ω the maximum likelihood estimate of π is $\hat{\pi}$, the point with coordinates $\hat{\pi}_{ij} = s_{ij}/n_{ij}$ and $\hat{\gamma}_{ij} = t_{ij}/n_{ij}$. From (1) we have that the log-likelihood function is a constant plus $L(\pi)$, where

$$(2) \qquad L(\pi) = \sum_{\mathcal{I}} n_{ij}(\hat{\pi}_{ij} \log \pi_{ij} + \hat{\pi}_{ji} \log \pi_{ji} + \hat{\gamma}_{ij} \log \gamma_{ij}).$$

For $T(\pi)$, a relation in X to be defined in Section 4, let ω be that portion of the parameter space Ω where $T(\pi)$ is a preference relation. Given paired comparison data on X we propose to estimate an order among the objects of X by maximizing (2) with respect to π over the region ω. Writing $\tilde{\pi}$ for the resulting restricted maximum likelihood estimate of π, then the estimated order is the one determined by $T(\tilde{\pi})$. The principle result of Section 4 (Theorem 12) is a necessary condition for a ranking to be optimal.

Section 5 is a discussion of alternative approaches for treating paired comparison data with ties.

2. Use of indirect comparisons. Mathematically, paired comparisons are relations. We will use the more precise mathematical language. Relations are defined between pairs of elements of the set $X = \{x_1, x_2, \cdots, x_m\}$, they are subsets of $X \times X = \{(x_i, x_j)\}$; if x stands in relation R to y then we write xRy or $(x, y) \varepsilon R$. An introduction to relation theory may be found, for example in Cogan et al. (1958). The identity relation $\equiv$ is fundamental; $x \equiv y$ means that x and y are the same elements. Definitions of various other relations and their properties appear in the following two tables.

We now study various relations and their associations to one another. Our

TABLE 1

Properties which a relation R may have

Name of property	Property (for any x, y and z in X)
reflexive	$(x, x) \varepsilon R$
antireflexive	$(x, x) \not\varepsilon R$
symmetric	$(x, y) \varepsilon R$ implies $(y, x) \varepsilon R$
antisymmetric	$(x, y) \varepsilon R$ and $(y, x) \varepsilon R$ implies $x \equiv y$
asymmetric	$(x, y) \varepsilon R$ implies $(y, x) \not\varepsilon R$
transitive	$(x, y) \varepsilon R$, $(y, z) \varepsilon R$ implies $(x, z) \varepsilon R$
complete	either $(x, y) \varepsilon R$, $(y, x) \varepsilon R$ or $x \equiv y$

TABLE 2

Common relations

Relation	Defining properties			
Equivalence	reflexive	symmetric	transitive	
Preference		asymmetric	transitive	
Weak order	reflexive		transitive	
Partial order	reflexive	antisymmetric	transitive	
Simple order	reflexive	antisymmetric	transitive	complete

goal in doing this is ultimately to determine when a set of basic paired comparisons will yield weak, partial, and simple orders as well as preference relations.

A weak order W determines (i) an equivalence relation E defined by

$$(x, y) \varepsilon E \text{ iff } (x, y) \varepsilon W \quad \text{and} \quad (y, x) \varepsilon W$$

and (ii) a preference relation Q defined by

$$(x, y) \varepsilon Q \text{ iff } (x, y) \varepsilon W \quad \text{but} \quad (y, x) \not\varepsilon W.$$

Note that a simple order is a partial order. Also a partial order is a weak order whose determined equivalence relation is the identity relation: distinct elements cannot belong to the same equivalence class.

An equivalence relation E is *consistent* with any other relation R if $(x, y) \varepsilon E$ and $(y, z) \varepsilon R$ imply $(x, z) \varepsilon R$, while $(x, y) \varepsilon E$ and $(z, y) \varepsilon R$ imply $(z, x) \varepsilon R$. It is easy to prove that the equivalence relation determined by a weak order is consistent with the determined preference relation and also with the determining weak order. On the other hand we have:

THEOREM 1. *If E is an equivalence relation consistent with a transitive relation T, then W defined by*

$$(x, y) \varepsilon W \quad \text{iff } either \quad (x, y) \varepsilon E \quad or \quad (x, y) \varepsilon T$$

is a weak order. E and T are the equivalence and preference relations determined by W iff *T is antireflexive.*

This is an extension of a result proved in Cogan.

Given a relation R, then a path K in R from y_1 to y_l, denoted by $[y_1, y_2, \cdots, y_l]$, is a collection of pairs $(y_1, y_2) \varepsilon R$, $(y_2, y_3) \varepsilon R$, $\cdots$, $(y_{l-1}, y_l) \varepsilon R$. If $y_1, y_2, \cdots, y_l$ are distinct then K is called an *elementary path*. A *loop* is a path whose first and last points coincide. As a generalization of completeness we say that R is *semicomplete* if for distinct x and y there is a path either from x to y or from y to x.

An ordering of the elements of X will ordinarily be denoted by P or $(p_1, p_2, \cdots, p_m)$. A relation R on X is said to determine a *partial rank ordering* (pro) P if $(p_j, p_i) \not\varepsilon R$ whenever $j > i$. In the sequel we determine when a pro exists and when it is unique. Proof of the following result can be found in Thompson and Remage.

THEOREM 2. (i) *A relation R on X determines at least one* pro iff *it is loop free (has no loops).* (ii) *A relation R on X determines a unique* pro iff *it is loop free and semicomplete.*

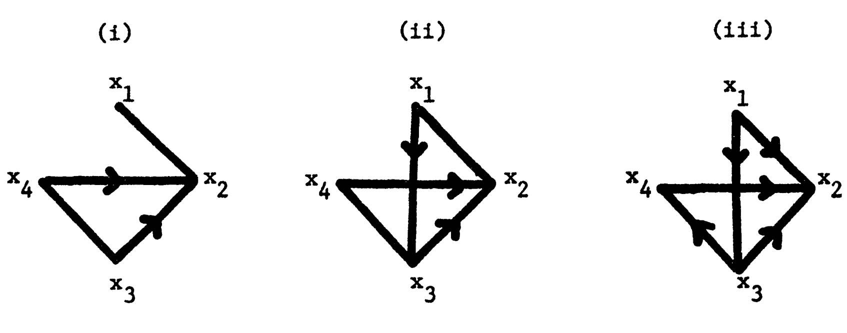
THEOREM 3. *A transitive relation T is asymmetric (and hence a preference)* iff (i) *T is loop free or* (ii) *T is antireflexive.*

PROOF. To prove that asymmetry implies (i), let $(x_1, x_2) \varepsilon T$, $(x_2, x_3) \varepsilon T$, $\cdots$, $(x_{k-1}, x_k) \varepsilon T$, $(x_k, x_1) \varepsilon T$ be a loop. By transitivity of T, $(x_1, x_k) \varepsilon T$ and T is not asymmetric. That (i) implies (ii) is a trivial special case. Finally assume (ii) but that asymmetry does not hold, i.e. $(x, y) \varepsilon T$ and $(y, x) \varepsilon T$. By transitivity $(x, x) \varepsilon T$, denying the assumption (ii).

A *bigraph* consists of a set $X = \{x_1, x_2, \cdots, x_m\}$ of points and two disjoint relations C and D, called undirected and directed lines respectively. Letting x and y be arbitrary elements of X then C and D will have the properties (i) $(x, x) \not\varepsilon C \cup D$, (ii) if $(x, y) \varepsilon D$ then $(y, x) \not\varepsilon D$, but (iii) $(y, x) \varepsilon C$ whenever $(x, y) \varepsilon C$. Geometrically, we represent a directed line $(x, y) \varepsilon D$ by $x \rightarrow y$ and an undirected line $(u, v) \varepsilon C$ by $u—v$.

We will speak of the bigraph $[X, C \cup D]$, or simply $C \cup D$ when the set X is understood. Bigraphs having no undirected lines are digraphs. Digraphs have been studied from the present point of view by Thompson and Remage (1964) and in general by many authors including Harary and Moser (1966). Our reason for studying bigraphs is that they are a very convenient way of representing paired comparisons geometrically. The objects are represented by points, the preferences and indifferences by directed and undirected lines. We think of C and D as basic or direct comparisons which are usually incomplete and inconsistent and we wish to define indirect comparisons in terms of C and D.

A loop (path) in $C \cup D$ which is not entirely in C and hence includes at least one directed line is a *circuit* (*directed path*). Note that the bigraph $C \cup D$ is loop free iff it is circuit free and $C = \varnothing$, the null set. The following figure illustrates a few concepts.



The bigraph (i) is circuit free but is not loop free since $[x_3, x_4, x_3]$ and $[x_1, x_2, x_1]$ are two paths in C with end points the same. The bigraph (ii) has circuits $[x_1, x_3, x_2, x_1]$ and $[x_3, x_4, x_2, x_1, x_3]$ whereas the bigraph (iii) is loop free.

Given a bigraph $C \cup D$ of direct comparisons, how may we establish order among the objects of X? We begin by defining various indirect relations in terms of $C \cup D$. Let x and y be arbitrary elements in X and define an equivalence relation E, a weak order W and a transitive relation T in the following way:

$(x, y) \varepsilon E$ iff $n \equiv y$ or they are in the same loop of $C \cup D$,

(3) $(x, y) \varepsilon W$ iff $x \equiv y$ or there is a path from x to y in $C \cup D$,

$(x, y) \varepsilon T$ iff there is a directed path from x to y in $C \cup D$.

If D determines several pros then T uses the indirect comparisons to decide among them; a pro determined by T is a pro determined by D but the converse is not true. Note that E is the equivalence determined by W that E and T are consistent, and also that $(x, y) \varepsilon W$ iff either $(x, y) \varepsilon E$ or $(x, y) \varepsilon T$. However, T is not necessarily the preference relation determined by W. In fact, T may not be a preference relation. We have:

THEOREM 4. *The following conditions are equivalent.*

(i) *$C \cup D$ is circuit free.*

(ii) *T is a preference relation (necessarily determined by the weak order W).*

(iii) *T is loop free.*

(iv) *T determines at least one* pro *on X.*

PROOF. It is easy to see that (i) implies (iii). Conversely, assuming (iii), then T is antireflexive from Theorem 3. If there is a circuit in $C \cup D$, then it will contradict the antireflexive property of T. Hence, (i) and (iii) are equivalent. Since E and T are consistent and $(x, y) \varepsilon W$ iff $(x, y) \varepsilon E$ or $(x, y) \varepsilon T$, it follows directly from Theorems 1 and 3 that (ii) and (iii) are equivalent. Finally (iii) and (iv) are equivalent from Theorem 2.

COROLLARY. *T determines a* pro *$(p_1, p_2, \cdots, p_m)$ on X such that $(p_i, p_{i+1}) \varepsilon W$ for $i = 1, 2, \cdots, m - 1$* iff *$C \cup D$ is semicomplete and circuit free.*

PROOF. If T determines a pro then $C \cup D$ is circuit free. If in addition $(p_i, p_{i+1}) \varepsilon W$ for $i = 1, 2, \cdots, m - 1$ then since W is transitive, $C \cup D$ is semicomplete. On the other hand if $C \cup D$ is circuit free then T determines a pro. We only need to show that $(p_i, p_{i+1}) \varepsilon$ W, but if $C \cup D$ is semicomplete then either $(p_i, p_{i+1}) \varepsilon W$ or $(p_{i+1}, p_i) \varepsilon T$. The latter possibility denies that $(p_1, \cdots, p_m)$ is a pro.

Theorem 4 says that if indirect comparisons between objects are to yield some asymmetric and transitive relation then $C \cup D$ must be circuit free; certain pairs of objects may however not be comparable. The corollary says that if in addition all pairs are to be comparable, either by equivalence or preference, then $C \cup D$ must be semicomplete.

Note that W need not be a partial order on X even if $C \cup D$ is circuit free since the equivalence relation E determined by W is not necessarily the identity relation. We have

LEMMA. *W is a partial order* iff *$C = \varnothing$ and D is circuit free.*

The proof consists of observing that the equivalence relation E determined by W is the identity relation iff $C \cup D$ is loop free.

In part, the next theorem treats the question of when it is possible to determine a unique pro on X.

THEOREM 5. *The following conditions are equivalent:*

(i) $C = \emptyset$, *D is circuit free and semicomplete.*

(ii) *W is a simple order.*

(iii) *T is loop free and complete.*

(iv) *T determines a unique* pro *on X.*

PROOF. From the lemma, $C = \emptyset$ while D is circuit free iff W is a partial order. But D is semicomplete iff W is complete. Thus (i) and (ii) are equivalent. From Theorem 2, T determines a unique pro iff T is loop free and semicomplete. Therefore, (iii) and (iv) are equivalent. Finally we see directly that (i) implies (iii). On the other hand, when T is complete then $C \cup D$ is semicomplete and from Theorem 4, $C \cup D$ is circuit free when T is loop free. It remains to be seen that $C = \emptyset$. Assume $(x_i, x_j) \varepsilon C$. Since T is complete there is a directed path K either from x_i to x_j or from x_j to x_i in $C \cup D$. In either case K combined with $(x_i, x_j) \varepsilon C$ forms a circuit in $C \cup D$. This contradicts that T is loop free and the theorem is proved.

The next result gives the canonical form of a unique pro; its proof is from the previous corollary.

COROLLARY. *T determines a unique* pro $(p_1, p_2, \cdots, p_m)$ *on X* iff $C = \emptyset$, $(p_i, p_{i+1}) \varepsilon D$ *for* $i = 1, 2, \cdots, m - 1$, *and* $(p_i, p_j) \varepsilon D$ for $i > j$.

3. Changing the orientation of lines. In general a bigraph will not be circuit free and it will be necessary to change the orientation of some of the lines in order to obtain a pro. By "changing the orientation" we mean changing the direction of a line, assigning a direction to an undirected line, or changing a directed to an undirected line. $C_1 \cup D_1$ is a subbigraph of $C \cup D$ if $C_1 \subset C$ and $D_1 \subset D$. A subbigraph $C_1 \cup D_1$ is maximal circuit free (mcf) iff it is circuit free but not properly contained in any other circuit free subbigraph of $C \cup D$. In the next section, for a particular context, we will show that rankings determined by mcf subbigraphs constitute an optimal class.

Let $C_1 \cup D_1$ be a mcf subbigraph of $C \cup D$. Define a set of directed lines F_1 such that $(x_i, x_j) \varepsilon F_1$ iff $(x_i, x_j) \varepsilon C/C_1$ and there exists a directed path from x_i to x_j in $C_1 \cup D_1$. For $(x_i, x_j) \varepsilon C/C_1$ there is a directed path from either x_i to x_j or x_j to x_i but not both, otherwise $C_1 \cup D_1$ would not be mcf. It follows that F_1 is asymmetric and $F_1 = \emptyset$ iff $C/C_1 = \emptyset$. Next we may observe that for $(x_i, x_j) \varepsilon D/D_1$, there is either a directed path from x_j to x_i in $C_1 \cup D_1$ or a path joining x_i and x_j in C_1 but not both. Define another set of directed lines F_2 such that $(x_j, x_i) \varepsilon F_2$ iff $(x_i, x_j) \varepsilon D/D_1$ and there is a directed path from x_j to x_i in $C_1 \cup D_1$. Finally define a set of undirected lines F_3 such that $(x_i, x_j) \varepsilon F_3$ iff $(x_i, x_j) \varepsilon D/D_1$ and there is a path joining x_i and x_j in C_1. Note that F_2 is asymmetric whereas F_3 is symmetric. Also $F_2 \cup F_3 = \emptyset$ iff $D/D_1 = \emptyset$. $F_1 \cup F_2 \cup F_3$ is the set of lines obtained by changing the orientation of the lines in $(C/C_1) \cup (D/D_1)$.

Our next theorem says that $C_1 \cup D_1$ and $C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3$ determine the same preference relation. That is, it is immaterial whether we delete lines or change their orientation. We will write, for example, $T(C_1 \cup D_1)$ to mean that the relation T of (3) is defined by $C_1 \cup D_1$.

THEOREM 6. *Let* $C_1 \cup D_1$ *be a* mcf *subbigraph of* $C \cup D$. *Then* $T(C_1 \cup D_1)$ *determines the* pro P iff *it is a* pro *determined by* $T(C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3)$.

PROOF. The "if" part of the theorem is obvious since $C_1 \cup D_1$ is a subbigraph of $C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3$. Conversely, it will suffice to show that there is a directed path from x to y in $C_1 \cup D_1$ whenever there is one in $C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3$. Consider a directed path K from x to y in $C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3$. If (x_i, x_j) is a line of K belonging to F_1, then there exists a directed path from x_i to x_j in $C_1 \cup D_1$; replace the line (x_i, x_j) by the corresponding directed path. Similarly for each line of K belonging to F_2 or F_3 replace the line by the corresponding path in $C_1 \cup D_1$. In this way we have a directed path from x to y in $C_1 \cup D_1$ corresponding to K.

From Theorem 4 we have the

COROLLARY. *If* $C_1 \cup D_1$ *is a* mcf *subbigraph of* $C \cup D$, *then* $C_1 \cup D_1 \cup F_1 \cup F_2 \cup F_3$ *is circuit free.*

The following theorem says that there is a possibility of determining a unique pro even if the original bigraph $C \cup D$ is not loop free, that is, there are some tie relations.

THEOREM 7. *Let* $C_1 \cup D_1$ *be a* mcf *subbigraph of a complete bigraph* $C \cup D$. *Then* $T(C_1 \cup D_1)$ *determines a unique* pro iff $C_1 = \emptyset$.

PROOF. From Theorem 5, $T(C_1 \cup D_1)$ determines a unique pro only if $C_1 \cup D_1$ is loop free which implies that $C_1 = \emptyset$. To prove the converse observe that if $C_1 = \emptyset$ then loop free is the same as circuit free. It remains to show that D_1 is semicomplete. Consider any two elements x and y in X. Either $(x, y) \in C \cup D$ or $(y, x) \in C \cup D$ since $C \cup D$ is complete. Now if there is no path between x and y in D_1, then $D_1 \cup \{(x, y)\}$ if $(x, y) \in C \cup D$ or $D_1 \cup \{(y, x)\}$ if $(y, x) \in D$ is also circuit free. This is a contradiction to the maximality of D_1.

The significance of Theorem 7 is that given a complete bigraph $C \cup D$ if it is desired to find a unique pro, then we must look only for mcf subgraphs D_1 of $C \cup D$. But given a bigraph $C \cup D$, there may not exist such a subgraph and in that event it is not possible to determine a unique pro.

The next question is: when there are mcf subgraphs D_1 of $C \cup D$, how may we enumerate them? Theorem 8 is a result in this direction, we give it without proof. If $[p_1, p_2, \cdots, p_m]$ is an elementary path in $C \cup D$ such that $(p_i, p_{i+1}) \in D$ for $i = 1, 2, \cdots, m - 1$, then we will call it a *Hamiltonian* (H) *path* in $C \cup D$.

THEOREM 8. *Given* $C \cup D$ *and an* H*-path* P *in it, then* $D_1 = \{(p_i, p_j) \in D : i < j\}$ *is the unique* mcf *subgraph of* $C \cup D$ *such that* $T(D_1)$ *determines* P *as a unique* pro.

The following is essentially a result due to Thompson and Remage; it can be proved from Theorems 7 and 8.

THEOREM 9. *If* $C \cup D$ *is a complete bigraph, then there is a one-to-one correspondence, given by Theorem* 8, *between the* mcf *subgraphs* D_1 *of* $C \cup D$ *and the* H*-paths in* $C \cup D$.

4. Estimating a maximum likelihood preference relation. The previous section gives ad hoc procedures for determining ordering relations on X when a subject's preferences and indifferences can be given as a single bigraph. But frequentlv a pair of objects will be compared more than once and, in any case, it is desirable to attempt to derive ordering relations from a general theory of statistical inference.

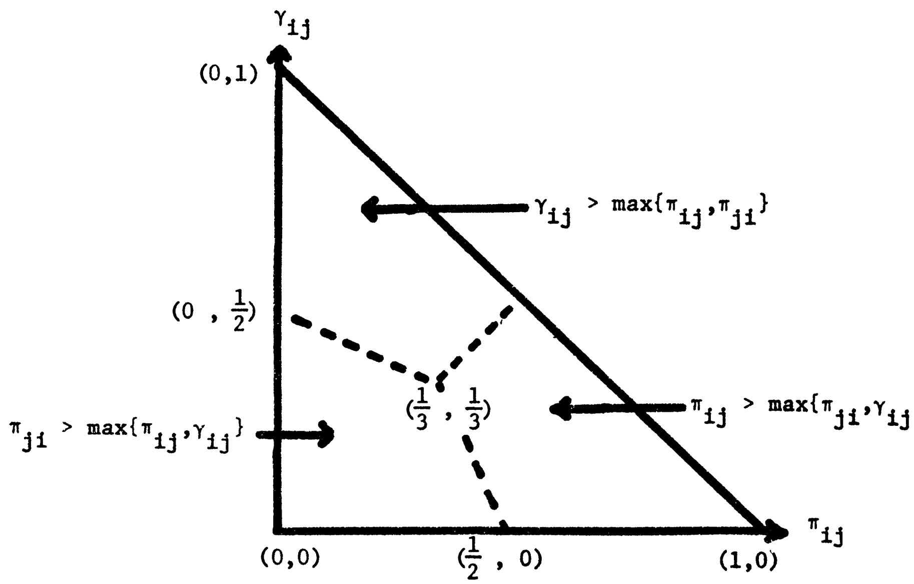
We are now ready to define the relation $T(\pi)$ announced in the introduction.

DEFINITION. $(x_i, x_j) \varepsilon D(\pi)$ iff $\pi_{ij} > \max(\pi_{ji}, \gamma_{ij})$. $(x_i, x_j) \varepsilon C(\pi)$ iff $\gamma_{ij} > \max(\pi_{ij}, \pi_{ji})$. $(x_i, x_j) \varepsilon T(\pi)$ iff there is a directed path from x_i to x_j in $C(\pi) \cup D(\pi)$.

If $\pi_{ij} = \pi_{ji} \geqq \frac{1}{3}$ or $\pi_{ij} = \gamma_{ij} > \frac{1}{3}$ or $\pi_{ji} = \gamma_{ij} > \frac{1}{3}$ (this region is indicated by the dotted lines in Figure 1) then no line in $C \cup D(\pi)$ is defined between x_i and x_j. This definition satisfies all the specifications of a bigraph for which results were developed in Section 2. In the future for brevity we will write $C \cup D(\pi)$ for the bigraph $[X, C(\pi) \cup D(\pi)]$. Note that this definition is not an assumption about the behavior of the subject; it is a statement of that aspect of his behavior which we wish to study.

Now to review. We wish to estimate an order among the objects of X by maximizing the likelihood (2) with respect to π over that portion of the sample space where $T(\pi)$ is a preference relation. According to Theorem 4 this is the region where $C \cup D(\pi)$ is circuit free. Writing $\tilde{\pi}$ for the resulting restricted maximum likelihood estimate of π, then the estimated order is the one determined by $T(\tilde{\pi})$. As noted earlier $\tilde{\pi}$ may not be unique and we may obtain several rank orders. Let μ be the subset of ω where (2) assumes its maximum.

Figure 1



The following lemma will be needed.

LEMMA. *Let* $f(t_1, t_2) = a_1 \log t_1 + a_2 \log t_2 + a_3 \log t_3$ *and* $S = \{(t_1, t_2): t_3 > \max(t_1, t_2)\}$ *where* $t_1 + t_2 + t_3 = 1$, $t_i \geqq 0$, $0 \leqq a_1 \leqq a_2 \leqq a_3$, *and* $a_1 + a_2 + a_3 = 1$. *The unique maximum of f over the complementary region* $\bar{S}$ *occurs at* $(t_1, t_2) = (a_1, \frac{1}{2}(a_2 + a_3))$.

PROOF. $\nabla^2(-f)$ is positive semidefinite, hence f is a concave function with absolute maximum at (a_1, a_2). The maximum of f over $\bar{S}$ occurs on one of the bounding faces: $B_1 = \{0 \leqq t_1 \leqq t_2 = t_3\}$ or $B_2 = \{0 \leqq t_2 \leqq t_1 = t_3\}$. The maximum on B_2 is $f(\frac{1}{3}, \frac{1}{3})$; on B_1 the maximum is $f(a_1, \frac{1}{2}(a_2 + a_3))$. But $f(a_1, \frac{1}{2}(a_2 + a_3)) \geqq f(\frac{1}{3}, \frac{1}{3})$ equality holding when $a_1 = \frac{1}{2}(a_2 + a_3) = \frac{1}{3}$.

THEOREM 10. *For any* $(ij) \varepsilon \mathscr{I}$ *and any* $\tilde{\pi} \varepsilon \mu$ *either* $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) = (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ *or* $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij})$ *is obtained by pooling the two largest members of the triple* $\hat{\pi}_{ij}, \hat{\gamma}_{ij}, \hat{\pi}_{ji}$, i.e.,

$$
\begin{aligned}
(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) &= (\tfrac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij}), \tfrac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij})) \text{ if } \hat{\pi}_{ij} \geqq \hat{\pi}_{ji} \text{ and } \hat{\gamma}_{ij} \geqq \hat{\pi}_{ji} \\
&= (\tfrac{1}{2}(\hat{\pi}_{ij} + \hat{\pi}_{ji}), \hat{\gamma}_{ij}) \text{ if } \hat{\pi}_{ij} \geqq \hat{\gamma}_{ij} \text{ and } \hat{\pi}_{ji} \geqq \hat{\gamma}_{ij} \\
&= (\hat{\pi}_{ij}, \tfrac{1}{2}(\hat{\pi}_{ji} + \hat{\gamma}_{ij})) \text{ if } \hat{\pi}_{ji} \geqq \hat{\pi}_{ji} \text{ and } \hat{\gamma}_{ij} \geqq \hat{\pi}_{ji}.
\end{aligned}
$$

PROOF. We prove the case where $\hat{\pi}_{ij} \geqq \hat{\gamma}_{ij} \geqq \hat{\pi}_{ji}$; the other five cases are analogous. Assume $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) \neq (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$. Define $\pi(t_1, t_2)$ to be obtained from $\tilde{\pi}$ by substituting t_1 for $\tilde{\pi}_{ij}$ and t_2 for $\tilde{\gamma}_{ij}$ while holding the other components of $\tilde{\pi}$ fixed.

Let $\pi' = \pi(\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ (we retain this definition of π' throughout) and $\bar{\pi} = \pi(\frac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij}), \frac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij}))$. $\tilde{\pi}$ and $\bar{\pi}$ are circuit free but π' is not since $L(\pi') > L(\tilde{\pi})$. Any circuit in $C \cup D(\pi')$ must involve (x_i, x_j), therefore $\hat{\pi}_{ij} > \max(\hat{\pi}_{ji}, \hat{\gamma}_{ij})$ but $\tilde{\pi}_{ij} \leqq \max(\tilde{\pi}_{ji}, \tilde{\gamma}_{ij})$. Now from the Lemma we have

$$
\hat{\pi}_{ij} \log \tfrac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij}) + \hat{\gamma}_{ij} \log \tfrac{1}{2}(\hat{\pi}_{ij} + \hat{\gamma}_{ij}) + \hat{\pi}_{ji} \log \hat{\pi}_{ji} \geqq \hat{\pi}_{ij} \log \tilde{\pi}_{ij} + \hat{\gamma}_{ij} \log \tilde{\gamma}_{ij} + \hat{\pi}_{ji} \log \tilde{\pi}_{ji}.
$$

Since $\bar{\pi}$ differs from $\tilde{\pi}$ only in the components corresponding to (ij) it follows that $L(\bar{\pi}) \geqq L(\tilde{\pi})$. But since $\tilde{\pi}$ is circuit free, $\tilde{\pi} = \bar{\pi}$ which proves the theorem.

Observe from the theorem that if $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) \neq (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ then there is no line between x_i and x_j in $C \cup D(\tilde{\pi})$; hence, $C \cup D(\tilde{\pi})$ is a circuit free subbigraph of $C \cup D(\hat{\pi})$. Let $(\bar{\pi}_{ij}, \bar{\gamma}_{ij})$ denote the pooled alternative of the theorem and let $\beta = \{\pi: (\pi_{ij}, \gamma_{ij}) = (\hat{\pi}_{ij}, \hat{\gamma}_{ij}) \text{ or } (\bar{\pi}_{ij}, \bar{\gamma}_{ij}) \text{ for } ij \varepsilon \mathscr{I}\}$. β depends on the particular collection of paired comparisons which have been observed and contains only a finite number of distinct points, many of which will not be circuit free. Note that in view of the previous theorems, $\mu \subset \beta$. The following is obvious:

COROLLARY. $\max_\omega L(\pi) = \max_{\omega \cap \beta} L(\pi)$.

Following Thompson and Remage it is interesting to observe that the determination of $\tilde{\pi}$ preserves an important information theory interpretation which adds to its intuitive appeal. From information theory we define the uncertainty

U_{ij} of a single comparison of x_i and x_j to be

$$U_{ij} = -(\pi_{ij} \log \pi_{ij} + \pi_{ji} \log \pi_{ji} + \gamma_{ij} \log \gamma_{ij}).$$

The uncertainty of all n_{ij} independent comparisons of x_i and x_j is $n_{ij}U_{ij}$ and the uncertainty of all comparisons is $\sum_{\mathcal{G}} n_{ij}U_{ij} = U(\pi)$, say. Without any loss of generality we follow the information theory convention of taking 2 as base for the logarithms.

Lemma. *If* $\pi \varepsilon B$, *then* $L(\pi) = -U(\pi)$.

Proof. Consider any $\pi \varepsilon \beta$. The Lemma will be proved if for all $(ij) \varepsilon \mathcal{G}$, (π_{ij}, γ_{ij}) is such that

$$-U_{ij} = \hat{\pi}_{ij} \log \pi_{ij} + \hat{\pi}_{ji} \log \pi_{ji} + \hat{\gamma}_{ij} \log \gamma_{ij}. \tag{4}$$

We know that $(\pi_{ij}, \gamma_{ij}) = (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ or $(\bar{\pi}_{ij}, \bar{\gamma}_{ij})$. In either case we obtain (4) after some algebraic simplification.

Theorem 11. *Maximizing the likelihood over* ω *is equivalent to minimizing the uncertainty over the set* $\omega \cap \beta$.

Proof. From the previous lemma and corollary

$$\max_\omega L(\pi) = \max_{\omega\cap\beta} L(\pi) = -\min_{\omega\cap\beta} U(\pi).$$

If $m = 3$, we can determine $\tilde{\pi}$ without much difficulty. However, when $m > 3$, it will be worthwhile to have a systematic method for determining $\tilde{\pi}$ and μ in general. The following result helps by establishing a connection with the material of Section 2.

Theorem 12. $C \cup D(\tilde{\pi})$ *is a* mcf *subbigraph of* $C \cup D(\hat{\pi})$.

Proof. We have observed that $C \cup D(\tilde{\pi})$ is a circuit free subbigraph of $C \cup D(\hat{\pi})$. If $C \cup D(\tilde{\pi})$ is not mcf, then there must exist a line l, say between x_i and x_j, in $C \cup D(\hat{\pi})$ but not in $C \cup D(\tilde{\pi})$ such that $l \cup D \cup D(\tilde{\pi}) = C \cup D(\pi')$ remains circuit free (π' is as previously defined). Also $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) \neq (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ since l is not in $C \cup D(\tilde{\pi})$. Therefore $L(\pi') > L(\tilde{\pi})$. This contradiction to the definition of $\tilde{\pi}$ proves the theorem.

Define $\eta = \{\pi \varepsilon \beta : C \cup D(\pi)$ is a mcf subbigraph of $C \cup D(\hat{\pi})\}$. η is called the estimation set.

Corollary. $$\mu \subset \eta.$$

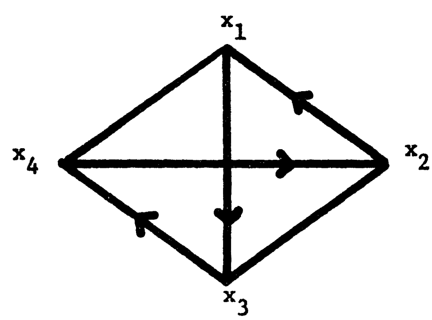
The number of points in η equals the number of distinct mcf subbigraphs of $C \cup D(\hat{\pi})$. One may, therefore, enumerate all the points in η by enumerating all the mcf subbigraphs of $C \cup D(\hat{\pi})$. The points in η which minimize $U(\pi)$ are the ones in μ. Before considering a numerical example we present one further result. Define $\xi = \{\pi \varepsilon \beta : D(\pi)$ is a mcf subbigraph of $C \cup D(\hat{\pi})\}$.

Theorem 13. *If* $C \cup D(\hat{\pi})$ *is complete, then* $T(\tilde{\pi})$ *determines a unique* pro iff $\xi \cap \mu \neq \varnothing$.

Proof. If $T(\tilde{\pi})$ determines a unique pro, then from Theorem 5 $C \cup D(\tilde{\pi}) = D(\tilde{\pi})$. Also from Theorem 12 $\tilde{\pi} \varepsilon \xi$, but $\tilde{\pi} \varepsilon \mu$ by definition so $\tilde{\pi} \varepsilon \xi \cap \mu$. On the other hand if $\xi \cap \mu \neq \varnothing$, then there exists $\tilde{\pi}$ such that $D(\tilde{\pi})$ is a mcf subbigraph of $C \cup D(\hat{\pi})$. Since $C \cup D(\hat{\pi})$ is complete, it follows from Theorem 7 that $T(\hat{\pi})$ determines a unique pro. This completes the proof.

Note that $\xi \cap \mu$ may be a null set for two reasons. First, $\xi = \emptyset$, that is, there exists no mcf subbigraph $D(\pi)$ of $C \cup D(\hat{\pi})$ and one cannot determine a unique maximum likelihood pro on X. Second, $\xi \neq \emptyset$ but $\xi \cap \mu = \emptyset$, this means $L(\hat{\pi}) > L(\pi)$ for all $\pi \in \xi$. In this case, for the sake of determining a unique pro, we may wish to select only those points $\pi \in \xi$ for which $L(\pi)$ is maximum or equivalently $U(\pi)$ is minimum. However, the unique pro determined this way would not be maximum likelihood. In short, to determine a unique pro, it is necessary that $\xi \neq \emptyset$. The determined pro will be maximum likelihood whenever $\xi \cap \mu \neq \emptyset$.

We now consider a numerical example. The data used is hypothetical. There are four treatments and each possible pair has been compared 6 times. The data is presented in Table 3 using the notation defined in Section 1.
The resulting bigraph is not circuit free:



All possible mcf subgraphs are enumerated in Figure 2.

We see from the uncertainty column of Table 4 that μ contains only one point π_1 but, from Figure 2, ξ consists of four points π_2, π_3, π_5 and π_6. Obviously $\xi \cap \mu = \emptyset$. In agreement with Theorem 13, we find that $T(\hat{\pi})$ cannot determine a unique maximum likelihood pro. However, if we restrict ourselves to ξ, then points π_2 and π_3 yield minimum uncertainty. The preference relations $T(\pi_2)$ and $T(\pi_3)$ respectively determine the orders (x_1, x_3, x_4, x_2) and (x_2, x_1, x_3, x_4).

TABLE 3

Hypothetical paired comparison data, $m = 4$, $n_{ij} = 6$ for all i and j

$(ij) \in I$	s_{ij}	t_{ij}	s_{ji}	$(\hat{\pi}_{ij}, \hat{\gamma}_{ij})$
(12)	2	1	3	$(\frac{1}{3}, \frac{1}{6})$
(13)	4	1	1	$(\frac{2}{3}, \frac{1}{6})$
(14)	0	4	2	$(0, \frac{2}{3})$
(23)	1	3	2	$(\frac{1}{6}, \frac{1}{2})$
(24)	1	2	3	$(\frac{1}{6}, \frac{1}{3})$
(34)	4	0	2	$(\frac{2}{3}, 0)$

Figure 2

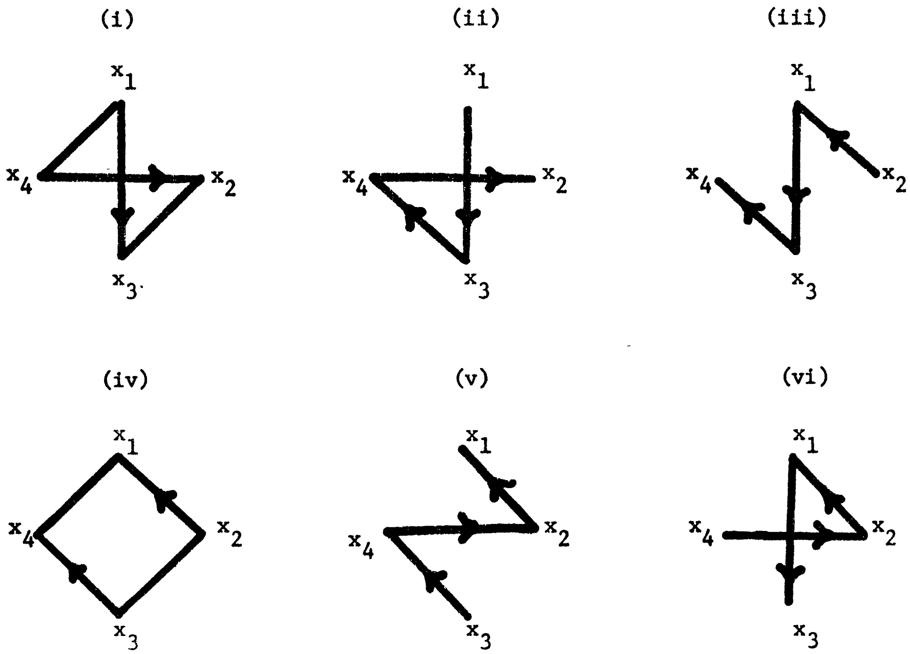


TABLE 4

Points in the estimation set

mcf subgraphs and the corresponding π	$\pi \varepsilon \eta$						Uncertainty $\cdot(\log_2 10)\cdot 6$
	(π_{12}, γ_{12})	(π_{13}, γ_{13})	(π_{14}, γ_{14})	(π_{23}, γ_{23})	(π_{24}, γ_{24})	(π_{34}, γ_{34})	
(i), π_1	$(\frac{5}{12}, \frac{1}{6})$	$(\frac{2}{3}, \frac{1}{6})$	$(0, \frac{2}{3})$	$(\frac{1}{6}, \frac{1}{2})$	$(\frac{1}{6}, \frac{1}{3})$	$(\frac{1}{2}, 0)$	2.279224
(ii), π_2	$(\frac{5}{12}, \frac{1}{6})$	$(\frac{2}{3}, \frac{1}{6})$	$(0, \frac{1}{2})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{1}{6}, \frac{1}{3})$	$(\frac{2}{3}, 0)$	2.286507
(iii), π_3	$(\frac{1}{3}, \frac{1}{6})$	$(\frac{2}{3}, \frac{1}{6})$	$(0, \frac{1}{2})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{2}{3}, 0)$	2.286507
(iv), π_4	$(\frac{1}{3}, \frac{1}{6})$	$(\frac{5}{12}, \frac{1}{6})$	$(0, \frac{2}{3})$	$(\frac{1}{6}, \frac{1}{2})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{2}{3}, 0)$	2.324382
(v), π_5	$(\frac{1}{3}, \frac{1}{6})$	$(\frac{5}{12}, \frac{1}{6})$	$(0, \frac{1}{2})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{1}{6}, \frac{1}{3})$	$(\frac{2}{3}, 0)$	2.348982
(vi), π_6	$(\frac{1}{3}, \frac{1}{6})$	$(\frac{2}{3}, \frac{1}{6})$	$(0, \frac{1}{2})$	$(\frac{1}{6}, \frac{5}{12})$	$(\frac{1}{6}, \frac{1}{3})$	$(\frac{1}{2}, 0)$	2.303824

5. Alternative approaches. The following alternative model was suggested by a referee. It is appealing and has interest here because it reflects a difference in philosophy. Suppose that

$$\gamma_{ij} = (\pi_{ij}\pi_{ji})^{\frac{1}{2}}. \tag{5}$$

Then the likelihood function is

$$\prod_{\mathcal{S}} n_{ij}!(s_{ij}!\, x_{ji}!\, t_{ij}!)^{-1}\pi_{ij}^{s_{ij}+\frac{1}{2}t_{ij}}\pi_{ji}^{s_{ji}+\frac{1}{2}t_{ij}},$$

so that, in effect, (5) makes a tie equivalent to half a preference in each direction. Now the whole machinery of Thompson and Remage (1964) and Remage and Thompson (1966) can be used to solve the maximization problem.

Equation (5) is an appealing assumption about the subject's behavior but it is an assumption and he may or may not behave this way. The assumption amounts to regarding a tie as a case where the subject had a slight preference one way or another but random error obscured the direction. In some cases we should regard ties in this way but in others we should regard them as actual ties. The present paper concerns the case when ties should be taken literally.

Two other approaches for treating actual ties were considered, both of them use a concept of stochastic bigraphs different from that of Section 4.

DEFINITION. $(x_i, x_j) \varepsilon D(\pi)$ iff $\pi_{ij} > \pi_{ji}$; and $(x_i, x_j) \varepsilon C(\pi)$ iff $\pi_{ij} = \pi_{ji}$. $(x_i, x_j) \varepsilon T$ iff there is a directed path from x_i to x_j in $C \cup D(\pi)$.

The first alternative approach, like Section 4, maximizes the likelihood (2) subject to the constraint that $T(\pi)$ be a preference, but with the new concept of stochastic bigraph. We obtain

THEOREM. *For any* $(ij) \varepsilon \mathcal{I}$, $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) \neq (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ *only if* $\hat{\pi}_{ij} = \hat{\pi}_{ji}$. *In that case* $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) = (\frac{1}{2}(\hat{\pi}_{ij} + \hat{\pi}_{ji}), \hat{\pi}_{ij})$.

The theorem says: to determine $\tilde{\pi}$, directed lines are made undirected and no undirected lines are made directed. Consequently we can never estimate a unique pro (recall Theorem 5) unless, of course, $C \cup D(\hat{\pi})$ is loop free to begin with. Often, in fact, the estimated preference relation $T(\tilde{\pi})$ would be an empty set with no statements of preference being made. For example,

THEOREM. *If* $C \cup D(\hat{\pi})$ *has an elementary circuit* $[y_1, y_2, \cdots, y_m, y_1]$, *then* $T(\tilde{\pi}) = \varnothing$.

The second alternative approach leads to the relation

(6) $(x, y) \varepsilon R(\pi)$ iff there is a path in $D(\pi)$ from x to y.

A *weak stochastic ranking* (wsr) is defined to be a pro determined by $D(\pi)$. That is $(p_1, \cdots, p_m)$ is a wsr iff $\pi_{p_i p_j} \geqq \pi_{p_j p_i}$ whenever $i < j$. This concept is a straightforward generalization of a familiar definition of wsr when $\pi_{ij} + \pi_{ji} = 1$. From Theorem 2 we have that $D(\pi)$ determines a wsr iff $D(\pi)$ is circuit free. Thus this approach ignores the ties which we wish to treat.

We may prove that $R(\pi)$ of (6) is a preference relation iff $D(\pi)$ determines a wsr. Hence maximizing the likelihood subject to the constraint that $R(\pi)$ be a preference is the same as obtaining the maximum likelihood wsr. We find

THEOREM. *For all* $(ij) \varepsilon \mathcal{I}$ *either* $(\tilde{\pi}_{ij}, \tilde{\gamma}_{ij}) = (\hat{\pi}_{ij}, \hat{\gamma}_{ij})$ *or* $(\frac{1}{2}(\hat{\pi}_{ij} + \hat{\pi}_{ji}), \hat{\gamma}_{ij})$.

When no ties are observed the Theorem reduces to that of Thompson and Remage. Also when $n_{ij} = 1$ or 0 this second approach yields Slater's (1961) nearest adjoining order.

REFERENCES

COGAN, E. J., KEMENY, J. G., NORMAN, R. Z., SNELL, J. L. and THOMPSON, G. L. (1958). *Modern Mathematical Methods and Models*. Cushing-Malloy, Ann Arbor.

HARARY, F. and MOSER, L. (1966). The theory of round robin tournaments. *Amer. Math. Monthly* **73** 231–245.

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

REMAGE, RUSSELL, JR. and THOMPSON, W. A., JR. (1966). Maximum likelihood paired comparison rankings. *Biometrika* **53** 143–149.

THOMPSON, W. A., JR. and REMAGE, R., JR. (1964). Rankings from paired comparisons. *Ann. Math. Statist.* **35** 739–747.