The Annals of Statistics
1988, Vol. 16, No. 2, 915–925

MAJORIZATION, ENTROPY AND PAIRED COMPARISONS[1]

BY HARRY JOE

University of British Columbia

Constrained majorization orderings and entropy functions are used to study the class of probability matrices associated with paired comparisons. Majorization orderings are also defined to handle the cases of order effects and/or ties. Results are obtained for maximal and minimal probability matrices with respect to the majorization ordering; these are related to transitivity conditions. The Bradley–Terry and Thurstone–Mosteller models are shown to be maximum entropy models. New models based on maximum entropy are obtained for the cases of order effects and ties; these models are compared with the Davidson and Beaver, the Rao and Kupper and the Davidson models. Applications to professional baseball and hockey are given.

1. Introduction. In paired comparisons, n players (teams) or choices (options) compete or are compared with each other in pairs. Let $P = (p_{ij})_{i \neq j}$ be the $n \times n$ probability matrix, where p_{ij} is the probability or proportion of times that player i beats player j or choice i is preferred to choice j. Note that $p_{ij} + p_{ji} = 1$ for all $i \neq j$ and p_{ii} is undefined. Let $p_i = \sum_{j \neq i} w_{ij} p_{ij}$ be the probability or proportion of times that i wins or is preferred over one of the other players or choices, where $w_{ij} > 0$, $\sum_{j \neq i} w_{ij} = 1$, $i = 1, \dots, n$. w_{ij} is a known weight variable representing the relative frequency that i is compared with j relative to other choices. If only the p_i's and w_{ij}'s are known, we consider the class $\mathscr{P} = \mathscr{P}(p_1, \dots, p_n)$ of possible probability matrices and define an ordering on $\mathscr{P}$ in order to obtain qualitative comparisons among different probability matrices in $\mathscr{P}$.

This problem is motivated by a professional sport such as baseball where a pair of teams play against each other a certain number of times in a season. At the end of the season, one can find out from newspapers the relative standings of the teams. Unless one is following all of the games in a season, one does not have information on how teams do against each other. In this case, the principle of maximum entropy [see, for example, Good (1963)] might be used to get rough estimates.

A majorization ordering will be defined on $\mathscr{P}$ in the balanced case where all the w_{ij} are equal to $(n-1)^{-1}$ and it will be interpreted as an ordering of transitivity. The probability matrices at the two extremes of the ordering and the relationship between the majorization ordering and forms of stochastic transitivity are a major focus of this paper. A probability matrix at the lower end of the ordering will have "large entropy" and will satisfy the transitivity condition $p_{ij} \geq 0.5$, $p_{jk} \geq 0.5 \Rightarrow p_{ik} \geq \max[p_{ij}, p_{jk}]$ and a probability matrix at

Received April 1987; revised August 1987.

[1]Research supported by the Natural Sciences and Engineering Research Council of Canada through Grant A-8698.

AMS 1980 *subject classifications*. Primary 62J15, 06A10.

Key words and phrases. Majorization, entropy, paired comparison, Bradley–Terry model.

the upper end of the ordering will have intransitivities. In this paper, we use entropy as a measure of closeness to uniformity for a set of numbers, with larger entropy meaning closer to uniformity. The majorization ordering is another example of a constrained vector majorization ordering. Vector majorization is the subject of Marshall and Olkin (1979); a constrained majorization ordering for two-way contingency tables is studied in Joe (1985) and a constrained majorization ordering for distributions of k-tuples is studied in Joe (1988). A good introduction to the method of paired comparisons is David (1963) and a bibliography on paired comparisons is in Davidson and Farquhar (1976).

In Section 2, results are obtained on maximal and minimal probability matrices for the majorization ordering. The Bradley–Terry (1952), uniform and Thurstone–Mosteller models [see David (1963), Chapter 4] are obtained from considering certain minimal probability matrices. In Section 3, generalizations are mentioned. These include:

1. An extension of the majorization ordering to the case where $w_{ij} = w_{ji}$ but otherwise w_{ij}'s are arbitrary.
2. An extension to allow for within-pair order effects (or the effect of playing at home versus away).
3. An extension to allow for ties.

From these generalizations, new paired comparison models are obtained. These models are compared with the Davidson and Beaver (1977), the Rao and Kupper (1967) and the Davidson (1970) models. As in Good (1963), these models are considered as "null hypotheses" that may be worth testing. Section 4 consists of applications to professional baseball and hockey.

2. Majorization and transitivity. In this section, we consider the constrained majorization ordering on $\mathscr{P}(p_1, \dots, p_n)$ with $w_{ij} = (n-1)^{-1}$ for all i, j. Our terminology for paired comparisons will be in terms of n teams ($n \geq 3$) who compete against each other in pairs. Let $x_i = (n-1)p_i$, $i = 1, \dots, n$. Then $\sum_{i=1}^n x_i = n(n-1)/2$. In addition, by generalizing Landau (1953), a necessary condition for there to exist a probability matrix in $\mathscr{P}$ is that $\sum_{k=1}^j x_{i_k} \geq j(j-1)/2$, for a j-tuple $(i_1, \dots, i_j)$, where $j \geq 2$. We assume that this condition holds throughout this section.

DEFINITION. Let $y = (y_1, \dots, y_m)$ and $z = (z_1, \dots, z_m)$ be vectors in $\mathscr{R}^m$. Then y majorizes z (written $y \succ z$) if

$$\sum_{i=1}^k y_{[i]} \geq \sum_{i=1}^k z_{[i]}, \qquad k = 1, \dots, m-1, \quad \text{and} \quad \sum_{i=1}^m y_i = \sum_{i=1}^m z_i,$$

where $y_{[1]} \geq \cdots \geq y_{[m]}$ and $z_{[1]} \geq \cdots \geq z_{[m]}$.

DEFINITION. Let P, $Q \in \mathscr{P}(p_1, \dots, p_n)$. Then P majorizes Q (written $P \succ Q$) if $P^* \succ Q^*$, where P^* and Q^* are vectors consisting of the $n(n-1)$ elements of P and Q, respectively, in a nonincreasing order. $P \in \mathscr{P}$ is minimal (maximal) if for any other $Q \in \mathscr{P}$ such that $P \succ Q$ ($P \prec Q$), then $P^* = Q^*$.

EXAMPLE. Let $(x_1, x_2, x_3) = (1.3, 1.1, 0.6)$. Then

$$\begin{bmatrix} - & 0.3 & 1 \\ 0.7 & - & 0.4 \\ 0 & 0.6 & - \end{bmatrix} \succ \begin{bmatrix} - & 0.6 & 0.7 \\ 0.4 & - & 0.7 \\ 0.3 & 0.3 & - \end{bmatrix}.$$

The majorization ordering allows a comparison of the probability matrices in $\mathscr{P}$ so it is of interest to study the minimal and maximal probability matrices. Relationships with transitivity conditions and paired comparison models are obtained and the majorization ordering is interpreted as an ordering of transitivity.

DEFINITION. A probability matrix P is weak (stochastic) transitive if

$$p_{ij} \geq 0.5 \quad \text{and} \quad p_{jk} \geq 0.5 \Rightarrow p_{ik} \geq 0.5.$$

P is strong (stochastic) transitive if

$$p_{ij} \geq 0.5 \quad \text{and} \quad p_{jk} \geq 0.5 \Rightarrow p_{ik} \geq \max[p_{ij}, p_{jk}].$$

If a probability matrix is weak or strong transitive, then there is the linear ordering of strength among the n teams with team i ahead of team j if $p_{ij} > 0.5$. Only for strong transitivity is this ordering the same as the ordering of the p_i or x_i. For example, with $(x_1, x_2, x_3) = (1.2, 1.1, 0.7)$,

$$P = \begin{bmatrix} - & 0.45 & 0.75 \\ 0.55 & - & 0.55 \\ 0.25 & 0.45 & - \end{bmatrix}$$

is weak but not strong transitive with the ordering of strength being team 2, team 1, team 3.

THEOREM 2.1. *If $P \in \mathscr{P}(p_1, \dots, p_n)$ is strong transitive, then the ordering induced by the transitivity condition is the same as the ordering of the p_i.*

PROOF. If $p_i > p_j$, then strong transitivity implies $p_{ik} \geq p_{jk}$ for all $k \neq i, j$. □

LEMMA 2.2. *A necessary condition for P to be minimal is that $p_{ij} \geq 0.5$ whenever $p_i \geq p_j$.*

PROOF. Suppose P does not satisfy the condition. Then there exists i, j such that $x_i \geq x_j$ and $p_{ij} < 0.5 < p_{ji}$. Hence there is a k such that $p_{ik} > p_{jk}$. Let $a = 0.5 \min[p_{ji} - p_{ij}, p_{ik} - p_{jk}]$. Replace $p_{ij}, p_{ji}, p_{ik}, p_{ki}, p_{jk}, p_{kj}$, respectively, by $p_{ij} + a, p_{ji} - a, p_{ik} - a, p_{ki} + a, p_{jk} + a, p_{kj} - a$ to get a new matrix P'. Then $P \succ P'$. □

THEOREM 2.3. *A necessary condition for P to be minimal is that P is strong transitive.*

PROOF. Suppose P is not strong transitive. Then by Theorem 2.1, there exists i, j, k such that $x_i \geq x_j$ and $p_{ik} < p_{jk}$. If $p_{ij} = 0.5$, then $n \geq 4$ and there exists l such that $p_{il} > p_{jl}$. In this case, replace $p_{ik}, p_{jk}, p_{il}, p_{jl}$, respectively, by $p_{ik} + a, p_{jk} - a, p_{il} - a, p_{jl} + a$, where $a = 0.5\min[p_{jk} - p_{ik}, p_{il} - p_{jl}]$ and make the corresponding changes for $p_{ki}, p_{kj}, p_{li}, p_{lj}$. The new matrix is smaller with respect to the majorization ordering. If $p_{ij} > 0.5$, then replace $p_{ij}, p_{ji}, p_{ik}, p_{ki}, p_{jk}, p_{kj}$, respectively, by $p_{ij} - a, p_{ji} + a, p_{ik} + a, p_{ki} - a, p_{jk} - a, p_{kj} + a$, where $a = 0.5\min[p_{ij} - p_{ji}, p_{jk} - p_{ik}]$. The new matrix is again smaller with respect to the majorization ordering. □

REMARK. For $n \geq 4$, strong transitivity is not a sufficient condition for minimality. An example with $n = 4$ and $(x_1, x_2, x_3, x_4) = (2.45, 1.75, 1.25, 0.55)$ is

$$\begin{bmatrix} - & 0.69 & 0.86 & 0.9 \\ 0.31 & - & 0.65 & 0.79 \\ 0.14 & 0.35 & - & 0.76 \\ 0.1 & 0.21 & 0.24 & - \end{bmatrix} \succ \begin{bmatrix} - & 0.7 & 0.85 & 0.9 \\ 0.3 & - & 0.65 & 0.8 \\ 0.15 & 0.35 & - & 0.75 \\ 0.1 & 0.2 & 0.25 & - \end{bmatrix}.$$

Theorem 2.4 shows that strong transitivity is a sufficient condition for minimality when $n = 3$.

THEOREM 2.4 (Minimality and maximality for $n = 3$). *Suppose $n = 3$. If P is strong transitive, then P is minimal. Suppose without loss of generality that $x_1 \geq x_2 \geq x_3$. If $x_2 \geq 1$, then*

$$\begin{bmatrix} - & 2 - x_2 & 1 - x_3 \\ x_2 - 1 & - & 1 \\ x_3 & 0 & - \end{bmatrix}$$

is maximal and if $x_2 \leq 1$, then

$$\begin{bmatrix} - & 1 & x_1 - 1 \\ 0 & - & x_2 \\ 2 - x_1 & 1 - x_2 & - \end{bmatrix}$$

is maximal.

PROOF. The general form of a matrix in $\mathscr{P}$ is

$$\begin{bmatrix} - & y & x_1 - y \\ 1 - y & - & x_2 - 1 + y \\ 1 - x_1 + y & 2 - x_2 - y & - \end{bmatrix},$$

where $x_1 - 1 \leq y \leq \min[1, 2 - x_2]$. The maximality result follows from: (a) if $x_2 \geq 1$, then $(x_3, 0) \succ (1 - x_1 + y, 2 - x_2 - y), (1, 1 - x_3) \succ (x_1 - y, x_2 - 1 + y)$ and $(2 - x_2, x_2 - 1) \succ (y, 1 - y)$; (b) if $x_2 \leq 1$, then $(1, x_1 - 1) \succ (y, x_1 - y)$, $(2 - x_1, 0) \succ (1 - x_1 + y, 1 - y)$ and $(x_2, 1 - x_2) \succ (x_2 - 1 + y, 2 - x_2 - y)$.

For the minimality, in analogy to Theorem 2 of Joe (1985), it suffices to show that if $P, P' \in \mathscr{P}$ are both strong transitive and $P \succ P'$, then $P = P'$. Let a_m be the difference between the largest m numbers in P and the largest m numbers in P'. $P - P'$ has the form

$$\begin{bmatrix} - & -d & d \\ d & - & -d \\ -d & d & - \end{bmatrix}.$$

Supposing $x_1 \geq x_2 \geq x_3$, $a_1 = d$ and $a_3 = -d$. If $P \succ P'$, then $d = 0$ and hence $P = P'$. □

THEOREM 2.5. *If $n = 4$, then a necessary condition for P to be maximal is for P to be an extreme point of the convex set $\mathscr{P}$.*

PROOF. By considering many cases, it can be shown that if i, j, k, l are distinct, then

$$\begin{bmatrix} p_{ij} & p_{il} \\ p_{kj} & p_{kl} \end{bmatrix}\begin{bmatrix} p_{ji} & p_{li} \\ p_{jk} & p_{lk} \end{bmatrix}$$

cannot be maximal unless at least one of the terms is a 1 or 0. Also, by Theorem 2.4, every 3×3 submatrix of P obtained from deleting a row and its corresponding column must contain a 1 and a 0 to be maximal. By enumeration, if $n = 4$, then P can be maximal only if it has at least three 1's and three 0's or if it is an extreme point in convex set $\mathscr{P}$. □

REMARK. In general there will be a maximal probability matrix among the extreme points of $\mathscr{P}$, but for $n \geq 5$ there can be maximal probability matrices which are not extreme points. An example of a maximal probability matrix which is not an extreme point is with $(x_1, \ldots, x_5) = (3.58, 1.78, 1.6, 2.52, 0.52)$:

$$\begin{bmatrix} - & 1 & 1 & 0.89 & 0.69 \\ 0 & - & 0.99 & 0 & 0.79 \\ 0 & 0.01 & - & 0.59 & 1 \\ 0.11 & 1 & 0.41 & - & 1 \\ 0.31 & 0.21 & 0 & 0 & - \end{bmatrix}$$

is a convex combination of the extreme points

$$\begin{bmatrix} - & 1 & 1 & 0.88 & 0.7 \\ 0 & - & 1 & 0 & 0.78 \\ 0 & 0 & - & 0.6 & 1 \\ 0.12 & 1 & 0.4 & - & 1 \\ 0.3 & 0.22 & 0 & 0 & - \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} - & 1 & 1 & 1 & 0.58 \\ 0 & - & 0.88 & 0 & 0.9 \\ 0 & 0.12 & - & 0.48 & 1 \\ 0 & 1 & 0.52 & - & 1 \\ 0.42 & 0.1 & 0 & 0 & - \end{bmatrix};$$

the latter extreme point is not maximal.

The next theorem implies that a maximal probability matrix cannot be strong transitive unless there is one matrix in $\mathscr{P}$, such as with $(x_1, x_2, x_3) = (2, 1, 0)$. From Theorem 2.4, it is easy to obtain a maximal probability matrix that is weak transitive, for example, with $(x_1, x_2, x_3) = (1.4, 1.3, 0.3)$. Often for a maximal probability matrix there will be cycles of intransitivities.

THEOREM 2.6. *If a probability matrix P is strictly strong transitive, that is, $p_{ik} > \max[p_{ij}, p_{jk}]$ for some (i, j, k) with $p_i \ge p_j \ge p_k$, then P is not maximal.*

PROOF. This follows by comparison with Theorem 2.4 with row sum vector $(x_1', x_2', x_3') = (p_{ij} + p_{ik}, p_{ji} + p_{jk}, p_{ki} + p_{kj})$. □

THEOREM 2.7 (Sufficient condition for minimality or maximality). *Let $\Psi(P)$ be strictly Schur convex over $P \in \mathscr{P}$, that is, $\Psi(P) > \Psi(Q)$ if $P \succ Q$ and $P^* \neq Q^*$. Then P is minimal (maximal) if it minimizes (maximizes) Ψ over $\mathscr{P}$.*

PROOF. This follows essentially by the definition of strict Schur convexity. □

Special cases of strictly Schur convex functions are $\Psi(P) = \sum\sum_{i \neq j} \psi(p_{ij})$, where ψ is strictly convex. In this case, Ψ is strictly convex over $\mathscr{P}$ and the minimum of Ψ will be the root of the derivative equations of the Lagrangian if this root is nonnegative. With Lagrangian multipliers λ_k, the solution without the nonnegativity constraint on the p_{ij}'s satisfies

$$\psi'(p_{ij}) - \psi'(1 - p_{ij}) - \lambda_i + \lambda_j = 0, \qquad i \neq j. \tag{2.1}$$

By taking $\psi'(u) = 0.5F^{-1}(u)$ or $\psi(u) \doteq \int_{0.5}^{u} F^{-1}(v)\,dv$ for a cumulative distribution function F which is symmetric about 0, the linear model $F^{-1}(p_{ij}) = \lambda_i - \lambda_j$ results, so that the linear model is a "maximum entropy" model with respect to the majorization ordering. Taking F to be logistic, uniform and normal lead, respectively, to the well known Bradley–Terry, uniform and Thurstone–Mosteller models. The Bradley–Terry model, which has the form $p_{ij} = \theta_i/(\theta_i + \theta_j)$, where $\theta_i = \exp[\lambda_i]$, also results from taking $\psi(u) = u \log u$, the negative of Shannon entropy. For the Bradley–Terry model but not for the other linear models, the maximum entropy estimates from (2.1) given p_i correspond to the maximum likelihood estimates. This is because, for the Bradley–Terry model, the minimum sufficient statistics are the winning proportions. Bradley (1976) summarizes several ways of looking at the Bradley–Terry model; the maximum entropy view appears to be new for the paired comparisons literature.

3. Generalizations.

3.1. *Extended majorization.* In this section, we consider an extended majorization ordering associated with the constraints $\sum_{j \neq i} w_{ij} p_{ij} = p_i$, $i = 1, \dots, n$, where $w_{ij} > 0$ and $\sum_{j \neq i} w_{ij} = 1$, $i = 1, \dots, n$. We also assume the symmetry condition $w_{ij} = w_{ji}$. This extension would be applicable for modeling professional sports leagues in more recent decades because pairs of teams do not play against each other the same number of times but teams play the same number of games over the season.

The extended majorization is a special case of the majorization in Joe (1987) for functions on a measure space. Here the measure space consists of the finite set $\{(i, j): 1 \le i \le n, 1 \le j \le n, i \neq j\}$ and the measure which puts mass w_{ij} on (i, j).

DEFINITION. Let P, Q be in $\mathscr{P}(p_1, \dots, p_n)$ (as defined in Section 1). Then P majorizes Q (written $P \succ Q$) if $\sum\sum_{i \neq j} w_{ij}\psi(p_{ij}) \geq \sum\sum_{i \neq j} w_{ij}\psi(q_{ij})$ for all convex continuous real-valued functions ψ on $[0,1]$. Minimality and maximality are defined as in Section 2.

The theorems in Section 2 do not easily generalize, but Theorem 2.7 becomes

THEOREM 3.1. *Let ψ be a strictly convex real-valued function on* $[0,1]$. *If $P \in \mathscr{P}$ minimizes (maximizes) $\sum\sum_{i \neq j} w_{ij}\psi(p_{ij})$ over $\mathscr{P}$, then P is minimal (maximal).*

Without the nonnegativity constraint on the p_{ij}'s, the solution of the Lagrangian again satisfies (2.1), so that the linear models are minimal with respect to the extended majorization ordering.

3.2. *Order effects.* In this section, we consider a generalization which allows for within-pair order effects [compare Beaver and Gokhale (1975) and Davidson and Beaver (1977)]. This takes into account the situation where a choice may be preferred with two different probabilities depending on whether it is presented first or second within a pair. In the context of sports, for each match one team may be considered "at home" and the other "away." We will use the superscripts H and A on p_{ij} to denote the two probabilities.

Suppose we know $\sum_{j \neq i} p_{ij}^H = (n-1)p_i^H = x_i^H$ and $\sum_{j \neq i} p_{ij}^A = (n-1)p_i^A = x_i^A$, $i = 1, \dots, n$. Note that $\sum x_i^H + \sum x_i^A = n(n-1)$. Denote by $\mathscr{P} = \mathscr{P}(p_1^H, \dots, p_n^H, p_1^A, \dots, p_n^A)$ the class of matrix pairs

$$P = \left\{P^A = \left(p_{ij}^A\right)_{i \neq j}, P^H = \left(p_{ij}^H\right)_{i \neq j}\right\}$$

that satisfy $p_{ij}^H + p_{ji}^A = 1$, $i \neq j$, $\sum_{j \neq j} p_{ij}^H = x_i^H$, $\sum_{j \neq i} p_{ij}^A = x_i^A$, $p_{ij}^H \geq 0$, $p_{ij}^A \geq 0$. The definition of majorization over $\mathscr{P}$ is similar to Section 2.

Theorem 2.7 generalizes here. (2.1) becomes

$$\psi'\left(p_{ij}^H\right) - \psi'\left(1 - p_{ij}^H\right) - \lambda_i^H + \lambda_j^A \quad \text{and} \quad \psi'\left(p_{ij}^A\right) - \psi'\left(1 - p_{ij}^A\right) - \lambda_i^A + \lambda_j^H, \qquad i \neq j.$$

For $\psi(u) = u \log u$, the solution of these equations leads to

$$p_{ij}^H = \theta_i^H/\left(\theta_i^H + \theta_j^A\right), \qquad p_{ij}^A = \theta_i^A/\left(\theta_i^A + \theta_j^H\right), \qquad i \neq j, \tag{3.2.1}$$

for certain nonnegative numbers θ_i^H, θ_i^A, where $\theta_i^H = \exp(\lambda_i^H)$ and $\theta_i^A = \exp(\lambda_i^A)$. The Davidson and Beaver model (1977) overlaps with this model. The Davidson and Beaver model has the form

$$p_{ij}^H = \theta_i/\left(\theta_i + \gamma_{ij}\theta_j\right), \qquad p_{ij}^A = \gamma_{ij}\theta_i/\left(\theta_j + \gamma_{ij}\theta_i\right), \qquad i \neq j, \tag{3.2.2}$$

where $\gamma_{ij} = \gamma_{ji}$. If $\theta_i^2 = \theta_i^H\theta_i^A$, $i = 1, \dots, n$, and $\gamma_{ij} = \theta_i\theta_j/\theta_i^H\theta_j^H$, $i \neq j$, then (3.2.1) and (3.2.2) are equivalent. A useful way to compare the models is through the likelihoods and minimum sufficient statistics given data.

The maximum likelihood estimates of the parameters in (3.2.1) correspond to the maximum entropy estimates when the number of times a pair of teams compete against each other at home and away is a constant r. One can still consider maximum likelihood estimation for the model (3.2.1) if the number of times that team i competes against j at home is r_{ij}, $i \neq j$, where the r_{ij} need not all be the same. From the likelihood equation in the parameters for the model (3.2.1), it can be seen that the minimum sufficient statistics are m_i^H, m_i^A, $i = 1, \dots, n$, where m_i^H and m_i^A are, respectively, the total number of wins at home and away for team i against the other teams. For the model (3.2.2), the minimum sufficient statistics are $m_i = m_i^H + m_i^A$, $i = 1, \dots, n$, and $m_{ij}^A + m_{ji}^A$, $i < j$, where m_{ij}^A, $i \neq j$, is the number of times that team i wins away over team j. If all the γ_{ij} are equal to a constant γ in (3.2.2), then the minimum sufficient statistics are m_i, $i = 1, \dots, n$, and $m^A = \sum_{i \neq j} m_{ij}^A$, and this is a special case of (3.2.1) with $\theta_i^H = \theta_i$, $\theta_i^A = \gamma\theta_i$, $i = 1, \dots, n$. Furthermore, in this case the model is a maximum entropy model with $\psi(u) = u \log u$ if the constraints are $\sum_j (p_{ij}^H + p_{ij}^A) = x_i^H + x_i^A = x_i$, $i = 1, \dots, n$, and $\sum_{i \neq j} p_{ij}^A = y$.

3.3. *Ties.* In this section, we consider a generalization which allows for ties between two choices or teams [compare Rao and Kupper (1967) and Davidson (1970)]. Let $p_{ij}^W, p_{ij}^T, p_{ij}^L$ be, respectively, the probabilities that i wins over, ties, loses to j, with $p_{ij}^W = p_{ji}^L$, $p_{ij}^T = p_{ji}^T$, $p_{ij}^L = p_{ji}^W$ and $p_{ij}^W + p_{ij}^T + p_{ij}^L = 1$. Suppose that the probabilities p_i^W, p_i^T, p_i^L of a win, tie and loss for team i are known when an opponent is chosen at random from one of the other teams. This leads to the constraints $\sum_{j \neq i} p_{ij}^W = (n-1)p_i^W = x_i^W$, $\sum_{j \neq i} p_{ij}^T = (n-1)p_i^T = x_i^T$ and $\sum_{j \neq i} p_{ij}^L = (n-1)p_i^L = x_i^L$, where $\sum_i (x_i^W + x_i^T + x_i^L) = n(n-1)/2$.

A majorization ordering can be defined in a similar way to Section 2. Theorem 2.7 extends directly. For $\psi(u) = u \log u$, the maximum entropy model has the form

$$\begin{aligned} p_{ij}^W &= \theta_i^W\theta_j^L / \left(\theta_i^W\theta_j^L + \theta_i^L\theta_j^W + \theta_i^T\theta_j^T\right), \\ p_{ij}^L &= \theta_i^L\theta_j^W / \left(\theta_i^W\theta_j^L + \theta_i^L\theta_j^W + \theta_i^T\theta_j^T\right), \\ p_{ij}^T &= \theta_i^T\theta_j^T / \left(\theta_i^W\theta_j^L + \theta_i^L\theta_j^W + \theta_i^T\theta_j^T\right). \end{aligned} \tag{3.3.1}$$

The Davidson model has the form

$$\begin{aligned} p_{ij}^W &= \theta_i / \left[\theta_i + \theta_j + \nu\left(\theta_i\theta_j\right)^{1/2}\right], \\ p_{ij}^L &= \theta_j / \left[\theta_i + \theta_j + \nu\left(\theta_i\theta_j\right)^{1/2}\right], \\ p_{ij}^T &= \nu\left(\theta_i\theta_j\right)^{1/2} / \left[\theta_i + \theta_j + \nu\left(\theta_i\theta_j\right)^{1/2}\right]. \end{aligned} \tag{3.3.2}$$

(3.3.2) is a special case of (3.3.1) with $\theta_i^W = \theta_i$, $\theta_i^L = 1$ and $\theta_i^T = (\nu\theta_i)^{1/2}$, $i = 1, \dots, n$. The Rao and Kupper model has the form

$$\begin{aligned} p_{ij}^W &= \theta_i/(\theta_i + \gamma\theta_j), \qquad p_{ij}^L = \theta_j/(\theta_j + \gamma\theta_i), \\ p_{ij}^T &= \theta_i\theta_j(\gamma^2 - 1)/(\theta_i + \gamma\theta_j)(\theta_j + \gamma\theta_i). \end{aligned} \tag{3.3.3}$$

The three models will be compared through likelihoods and minimum sufficient statistics given data.

The maximum likelihood estimates of the parameters in (3.3.1) correspond to the maximum entropy estimates when the number of times a pair of teams compete against each other is a constant r. As in Section 3.2, one can still consider maximum likelihood estimation for the model (3.3.1) if the number of times that teams i and j compete is r_{ij}, $i \neq j$, where the r_{ij} need not all be the same. For the model (3.3.1), the minimum sufficient statistics are m_i^W, m_i^T, $i = 1, \dots, n$, where m_i^W and m_i^T are, respectively, the total number of wins and ties for team i with other teams. For the model (3.3.2), the minimum sufficient statistics are $s_i = 2m_i^W + m_i^T$, $i = 1, \dots, n$, and $t = \sum_{i<j} m_{ij}^T$, where m_{ij}^T is the number of ties between teams i and j. The model (3.3.2) arises as a maximum entropy model with $\psi(u) = u \log u$ if the constraints are $\sum_{j \neq i}(2p_{ij}^W + p_{ij}^T) = 2x_i^W + x_i^T = y_i$, $i = 1, \dots, n$, and $\sum_{i \neq j} p_{ij}^T = z$. For the model (3.3.3), the minimum sufficient statistics are $m_i^W + m_i^T$, $i = 1, \dots, n$, and $m_{ij}^W + m_{ij}^T$, $i \neq j$, where m_{ij}^W is the number of wins for team i over team j. (3.3.3) is not a maximum entropy model.

4. Applications. In this section, we consider examples that in part motivated this work. We use professional baseball and hockey data from several decades ago because then there were fewer teams and each pair played the same number of times.

For professional baseball (NL and AL), the yearly record for each pair of teams (at home, away and combined) can be found in the annual *Official Baseball Guide*. The "probability" matrices are at the lower end of the majorization ordering, but not close to minimal; they are not strong transitive but quite close to being weak transitive. For illustration of the estimates based on the Bradley–Terry model, (3.2.1) and (3.2.2) with $\gamma_{ij} \equiv \gamma$, we use the 1964 NL winning proportions. For baseball in 1964, each pair of teams played $r = 9$ times at home and away.

In Table 1, the $\hat{\theta}_i$ have been scaled to have a sum of 1 (BT stands for Bradley–Terry and DB stands for Davidson–Beaver) and the $\hat{\theta}_i^H, \hat{\theta}_i^A$ have been scaled to have a sum of 2; for the model (3.2.2), $\hat{\gamma} = 0.893$. For these data, the $\hat{\theta}_i$

TABLE 1

Team	%	% home	% away	$\hat{\theta}$ BT	$\hat{\theta}^H$	$\hat{\theta}^A$	$\hat{\theta}$ DB
St. Louis	0.574	0.593	0.556	0.127	0.129	0.123	0.127
Cincinnati	0.568	0.580	0.556	0.124	0.122	0.124	0.124
Philadelphia	0.568	0.568	0.568	0.124	0.115	0.132	0.124
San Francisco	0.556	0.543	0.568	0.118	0.104	0.133	0.118
Milwaukee	0.543	0.556	0.531	0.113	0.111	0.114	0.113
Los Angeles	0.494	0.506	0.481	0.094	0.092	0.095	0.094
Pittsburgh	0.494	0.519	0.469	0.094	0.098	0.090	0.094
Chicago	0.469	0.494	0.444	0.086	0.089	0.082	0.086
Houston	0.407	0.506	0.309	0.068	0.100	0.045	0.068
New York	0.327	0.407	0.247	0.050	0.068	0.035	0.050

for the Bradley–Terry and Davidson–Beaver models are the same to three decimal places. For the Bradley–Terry model, the estimates based on the combined proportion of wins form a strong transitive probability matrix; the average of the estimates of the winning proportions at home and away from (3.2.1) or (3.2.2) need not form a strong transitive probability matrix.

Majorization provides a qualitative way of looking at the data, but the adequacies of fit of the different models can be studied through log likelihood ratios [Bradley (1954)]. For the combined home and away data, the log likelihood ratios statistic for the Bradley-Terry model is 33.76 with 36 degrees of freedom. For (3.2.1) and (3.2.2) with $\gamma_{ij} \equiv \gamma$, the log likelihood ratio statistics are, respectively, 67.57 with 71 degrees of freedom and 76.33 with 80 degrees of freedom. The Bradley–Terry model fits the combined data adequately, and both (3.2.1) and (3.2.2) fit the at home and away data adequately. (3.2.2) with $\gamma_{ij} \equiv \gamma$ is an adequate simpler model [the log likelihood ratio statistic for (3.2.2) relative to (3.2.1) is 8.76 with 9 degrees of freedom]. (3.2.1) would be expected to be a better model than (3.2.2) with $\gamma_{ij} \equiv \gamma$ if the winning proportions at home and away are ordered quite differently among the teams.

We now go to professional hockey (NHL). The yearly standings and pairwise records can be found in the annual *National Hockey League Guide*. For illustration of the results in Section 3.3, we use data from the 1952–1953 season consisting of the number of games won, tied and lost by each of the 6 teams. Each pair of teams played 14 times. The estimates from models (3.3.1) and (3.3.2) are given in Table 2.

For model (3.3.1), the $\hat{\theta}_i^W, \hat{\theta}_i^T, \hat{\theta}_i^L$ have been standardized to have a sum of 3, and for (3.3.2), the $\hat{\theta}_i$ have been standardized to have a sum of 1 and $\hat{\nu} = 0.590$. From these values, a team is estimated to have a winning record against another team which is lower in the standings in Table 2. A comparison with the actual record for each pair revealed that the estimates from (3.3.1) (when rounded to the nearest multiple of 1/14) can be considered only as a first order approximation; there are discrepancies due to a couple of intransitivities relative to the standings in Table 2. In other words, the actual proportions $P = (p_{ij}^W, p_{ij}^T, p_{ij}^L)_{i<j}$ here are at the lower end of the majorization ordering, but not minimal.

The adequacies of fit of the different models can be studied through log likelihood ratios as in the preceding example. The log likelihood ratio statistics for (3.3.1) and (3.3.2) are, respectively, 23.63 with 28 degrees of freedom and 33.06

TABLE 2

Team	W	T	L	$\hat{\theta}^W$	$\hat{\theta}^T$	$\hat{\theta}^L$	$\hat{\theta}$
Detroit	36	18	16	0.228	0.166	0.114	0.256
Montreal	28	19	23	0.181	0.175	0.153	0.181
Boston	28	13	29	0.189	0.108	0.195	0.158
Chicago	27	15	28	0.181	0.129	0.186	0.158
Toronto	27	13	30	0.184	0.108	0.200	0.150
New York	17	16	37	0.122	0.144	0.236	0.098

with 39 degrees of freedom. Both (3.3.1) and (3.3.2) fit the data adequately and (3.3.2) is an adequate simpler model [the log likelihood ratio statistics for (3.3.2) relative to (3.3.1) is 9.43 with 11 degrees of freedom]. (3.3.1) would be expected to be a better model than (3.3.2) if the number of ties per team is more variable.

Acknowledgments. I would like to thank Professor R. A. Bradley for comments on an earlier draft of this paper and the referees for their suggestions for improvement of this paper.

REFERENCES

BEAVER, R. J. and GOKHALE, D. V. (1975). A model to incorporate within-pair order effects in paired comparisons. *Comm. Statist.* **4** 923–939.

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

BRADLEY, R. A. (1976). Science, statistics, and paired comparisons. *Biometrics* **32** 213–232.

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

DAVID, H. A. (1963). *The Method of Paired Comparisons.* Griffin, London.

DAVIDSON, R. R. (1970). On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. *J. Amer. Statist. Assoc.* **65** 317–328.

DAVIDSON, R. R. and BEAVER, R. J. (1977). On extending the Bradley–Terry model to incorporate within-pair order effects. *Biometrics* **33** 693–702.

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

GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.

JOE, H. (1985). An ordering of dependence for contingency tables. *Linear Algebra Appl.* **70** 89–103.

JOE, H. (1987). Majorization, randomness and dependence for multivariate distributions. *Ann. Probab.* **15** 1217–1225.

JOE, H. (1988). An ordering of dependence for distributions of k-tuples, with applications to lotto games. *Canad. J. Statist.* To appear.

LANDAU, H. G. (1953). Dominance relations in the structure of animal societies. III. The condition for score structure. *Bull. Math. Biophys.* **15** 143–148.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

NATIONAL HOCKEY LEAGUE (1950–). *National Hockey League Guide.* Montreal.

RAO, P. V. and KUPPER, L. L. (1967). Ties in paired-comparison experiments. *J. Amer. Statist. Assoc.* **62** 194–204.

SPORTING NEWS (1960–). *Official Baseball Guide.* St. Louis, Mo.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
2021 WEST MALL
VANCOUVER, BRITISH COLUMBIA V6T 1W5
CANADA