On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments

ROGER R. DAVIDSON*

This study is concerned with the extension of the Bradley-Terry model for paired comparisons to situations which allow an expression of no preference. A new model is developed and its performance compared with a model proposed by Rao and Kupper. The maximum likelihood estimates of the parameters are found using an iterative procedure which, under a weak assumption, converges monotonically to the solution of the likelihood equations. It is noted that for a balanced paired comparison experiment the ranking obtained from the maximum likelihood estimates agrees with that obtained from a scoring system which allots two points for a win, one for a tie and zero for a loss. The likelihood ratio test of the hypothesis of equal preferences is shown to have the same asymptotic efficiency as that for the Rao-Kupper model. Two examples are presented, one of which introduces a set of data for an unbalanced paired comparison experiment. Initial applications of the test of goodness of fit suggest that the proposed model yields a reasonable representation of actual experimentation.

1. INTRODUCTION

A mathematical model for paired comparisons has been presented and developed in a series of three basic papers: Bradley and Terry [7] and Bradley [3 and 4]. Since that time the Bradley-Terry model has received considerable attention in the literature.[1] The relationship of the Bradley-Terry model to the logistic distribution was noted by Bradley [5] and subsequently examined by W. A. Thompson (see [1]). Luce [16] introduced a "choice axiom" and related it, through a derived ratio scale, to the Bradley-Terry model. The notion of "intrinsic worth" models was considered by Brunk [8] who represented the Bradley-Terry model in this way. The association of the Bradley-Terry model with a class of distributions proposed by Lehmann [15] is given by Bradley [1]. Bradley [2] developed and applied a procedure for testing the appropriateness of the model. A valuable review of the method of paired comparisons is given by David [9].

Most models for paired comparisons do not allow for an expression of no preference between the treatments being compared. The usual practice is either to force a definite expression of preference, or to treat tries when they occur by ignoring, splitting or randomly allocating them. There are, however, two models which make provision for tied observations: a modification of the

[1] Researchers in the field of paired comparisons have recently become aware of a paper by E. Zermelo [19] in which the model attributed to Bradley and Terry was introduced and developed. Attention was drawn to the work of Zermelo by Professors J. Moon and L. Moser, University of Alberta, Canada.

Thurstone-Mosteller model due to Glenn and David [12] and a generalization of the Bradley-Terry model due to Rao and Kupper [17]. Both of these models postulate that a tie will be declared when the absolute difference between two responses lies below a certain threshold; in the former the difference distribution is an "angular" distribution, which replaces the normal distribution of the Thurstone-Mosteller model, while in the latter it is the logistic distribution of the Bradley-Terry model.

In the present study, a new extension of the Bradley-Terry model to allow for tied observations is proposed and developed. A detailed comparison of this model with that due to Rao and Kupper is made. The asymptotic properties of the Rao-Kupper model have been extended to unbalanced paired comparison experiments in order to permit such a comparison under more general conditions.

2. THE MATHEMATICAL MODELS

In paired comparisons one considers a set of t treatments which are presented in pairs. It is assumed that the responses to the treatments may be described in terms of an underlying continuum on which the "worths" of the treatments can be relatively located. Let π_i denote the "worth," an index of relative preference, of the ith treatment,

$$\pi_i \geq 0, \qquad \sum_{i=1}^{t} \pi_i = 1.$$

The Bradley-Terry model postulates that, if X_i and X_j are the responses to treatments i and j respectively, then

$$P(X_i > X_j) = \pi_i/(\pi_i + \pi_j) \tag{2.1}$$

in the comparison of treatments i and j. One interprets $X_i > X_j$ as indicative of preference for treatment i over treatment j.

It is noted by Bradley [5] that in replacing the normal density of the Thurstone-Mosteller model by the logistic (squared hyperbolic secant) density, one obtains the Bradley-Terry model. Specifically, if the difference $Z_{ij} = X_i - X_j$ between the responses is assumed to have a logistic distribution with location parameter $(\ln\pi_i - \ln\pi_j)$ and with distribution function

$$P(Z_{ij} \leq z) = 1/\{1 + \exp[-(z - \ln \pi_i + \ln \pi_j)]\}, \quad -\infty < z < \infty, \tag{2.2}$$

for each treatment pair (i, j), then (2.1) follows by setting $z = 0$, and subtracting from unity (cf. [1]).

With the use of a threshold parameter, $\eta = \ln\theta$, in conjunction with (2.2), Rao and Kupper [17] obtain

$$p^*(i \mid i, j) = P(Z_{ij} > \eta) = \frac{\pi_i}{\pi_i + \theta\pi_j}$$

$$p^*(j \mid i, j) = P(Z_{ij} < -\eta) = \frac{\pi_j}{\theta\pi_i + \pi_j} \tag{2.3}$$

$$p^*(0 \mid i, j) = P(\mid Z_{ij} \mid < \eta) = \frac{(\theta^2 - 1)\pi_i\pi_j}{(\pi_i + \theta\pi_j)(\theta\pi_i + \pi_j)}$$

for $i \neq j$, i, $j = 1, \cdots, t$. The quantities $p^*(i \mid i, j)$, $p^*(j \mid i, j)$ and $p^*(0 \mid i, j)$ represent the Rao-Kupper probabilities of preference for i, preference for j, and no preference respectively, when the treatment pair (i, j) is presented.

In his book [16], Luce introduces his "choice axiom." A direct consequence of this axiom is contained in Lemma 3 of Luce which in the present context states the following: if $p(\ell \mid i, j)$ denotes the probability that treatment ℓ is preferred when the treatment pair (i, j) is offered, $\ell = i, j$, and if π_i denotes the probability that treatment i is preferred when the entire set of t treatments is offered, $i = 1, \cdots, t$, then under the choice axiom

$$p(i \mid i, j)/p(j \mid i, j) = \pi_i/\pi_j \tag{2.4}$$

for all (i, j) provided $p(i \mid i, j) \neq 0, 1$ (in Luce's terminology this is the assumption that all choices are imperfect). If ties are not permitted then $p(i \mid i, j) + p(j \mid i, j) = 1$ and (2.4) immediately yields the Bradley-Terry model.

The extension of the Bradley-Terry model to be developed in this study will be assumed to satisfy the ratio scale given by (2.4). The determination of a model satisfying (2.4) requires an additional condition relating the probability, $p(0 \mid i, j)$, of no preference between i and j to the probabilities $p(i \mid i, j)$ and $p(j \mid i, j)$. One such condition is that the probability of no preference be proportional to the geometric mean of the probabilities of preference for the treatments being compared, namely

$$p(0 \mid i, j) = \nu \sqrt{p(i \mid i, j)p(j \mid i, j)} \tag{2.5}$$

where $\nu \geq 0$ is a constant of proportionality which does not depend on i and j. The parameter ν, or more appropriately $1/\nu$, can be thought of as an index of discrimination which is peculiar to each specific problem. Assumption (2.5) makes the probability of no preference dependent on the extent to which the pair of treatments being compared are distinguishable. Use of the geometric mean is suggested by the fact that, under the Bradley-Terry model, the merits of the t treatments can be represented by the values $\ln \pi_1, \cdots, \ln \pi_t$ on a linear scale.

Through the use of (2.4), (2.5) and the constraint $p(i \mid i, j) + p(j \mid i, j) + p(0 \mid i, j) = 1$, the following model is obtained:

$$
\begin{aligned}
p(\ell \mid i, j) &= \frac{\pi_\ell}{\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j}}, \qquad \ell = i, j \\[2ex]
p(0 \mid i, j) &= \frac{\nu\sqrt{\pi_i\pi_j}}{\pi_i + \pi_j + \nu\sqrt{\pi_i\pi_j}}.
\end{aligned}
\tag{2.6}
$$

An immediate difference between the models (2.3) and (2.6) as extensions of (2.1) is that the additional terms in the denominators of $p^*(i \mid i, j)$ and $p^*(j \mid i, j)$ depend on π_j and π_i respectively, while those in the denominators of $p(\ell \mid i, j)$, $\ell = i, j$, depend on their geometric mean $\sqrt{\pi_i\pi_j}$, and hence are the same. This results in a number of simplifications as the properties of the model (2.6) are developed in subsequent sections. It is noted that the Bradley-Terry model is obtained from (2.3) when $\theta = 1$ and from (2.6) when $\nu = 0$.

3. MAXIMUM LIKELIHOOD ESTIMATION

In paired comparison experimentation the set of t treatments are presented in pairs with r_{ij} independent responses being obtained in the comparison of the treatments i and j, the total number of comparisons being $N = \sum \sum_{i<j} r_{ij}$. Let w_{ij}, w_{ji} and t_{ij} be the frequencies of preference for i over j, preference for j over i, and no preference respectively. Clearly $r_{ij} = w_{ij} + w_{ji} + t_{ij}$. The total number of wins (preferences) and ties for treatment i are given by $w_i = \sum_j w_{ij}$ and $t_i = \sum_j t_{ij}$ respectively, where r_{ii}, w_{ii} and t_{ii} are defined to be zero $i = 1, \cdots, t$.

The method of maximum likelihood can be used to obtain estimates $(p, \hat{v})$ of the parameters (π, v), where $p = (p_1, \cdots, p_t)$ and $\pi = (\pi_1, \cdots, \pi_t)$. Setting $s_i = 2w_i + t_i$, $i = 1, \cdots, t$, and $T = \sum \sum_{i<j} t_{ij}$, the logarithm of the likelihood function becomes

$$\ln L(\pi, v) = \frac{1}{2} \sum_{i=1}^{t} s_i \ln \pi_i + T \ln v - \sum \sum_{i<j} r_{ij} \ln(\pi_i + \pi_j + v\sqrt{\pi_i \pi_j}). \quad (3.1)$$

Maximizing $\ln L(\pi, v)$, subject to the restraint

$$\sum_{i=1}^{t} \pi_i = 1,$$

it follows that the maximum likelihood estimates $(p, \hat{v})$ of (π, v) are obtained as a solution to the system of equations

$$s_i/p_i - g_i(p, \hat{v}) = 0, \qquad i = 1, \cdots, t \quad (3.2)$$

$$T/\hat{v} - h(p, \hat{v}) = 0, \quad (3.3)$$

where

$$g_i(p, \hat{v}) = \sum_j r_{ij}(2 + \hat{v}\sqrt{p_j/p_i})/(p_i + p_j + \hat{v}\sqrt{p_i p_j}), \quad (3.4)$$

$$h(p, \hat{v}) = \sum \sum_{i<j} r_{ij}\sqrt{p_i p_j}/(p_i + p_j + \hat{v}\sqrt{p_i p_j}). \quad (3.5)$$

The existence and uniqueness of the $(p, \hat{v})$ which maximizes $L(\pi, v)$ over the region $\{\pi_i > 0, \sum \pi_i = 1; 0 < v < \infty\}$ can be demonstrated, under a weak restriction on the matrix $W = [w_{ij}; i, j = 1, \cdots, t]$, by adapting the argument given by Ford [11] for the Bradley-Terry model. As in Ford, the following assumption is made on W.

Assumption. In every possible partition of the objects into two non-empty subsets, some object in the second set has been preferred at (3.6) least once to some object in the first set.

In addition, it is assumed that $T > 0$, for if this is not the case the extended model need not be used.

It is clear that $L(\pi, v)$ is positive and continuous over the region $\{\pi_i > 0, \sum \pi_i = 1, 0 < v < \infty\}$. The existence of a maximum in this region is established by noting that if one defines $L(\pi, v) = 0$ for (π, v) on the boundary, then one obtains a uniformly continuous extension of $L(\pi, v)$ to the closed region $\{\pi_i \geq 0,$

$\sum \pi_i = 1, 0 \leq \nu \leq \infty$ }. The uniqueness of the $(\boldsymbol{\pi}, \nu)$ which yields the maximum is established by a straightforward extension of the argument of Ford [11].

In the special case $t = 2$ Equations 3.2 and 3.3 can be solved explicitly to give

$$p_i = w_i/(w_1 + w_2), \qquad i = 1, 2, \quad \text{and} \quad \hat{\nu} = T/\sqrt{w_1 w_2}.$$

However when $t > 2$, the solution to the equations cannot be found directly and thus one must employ an iterative procedure.

The following is the proposed iterative scheme for obtaining a solution to the likelihood equations, (3.2) and (3.3). The iterations are indexed by M, $M = 1, 2, \cdots$, since one revised value of each p_i and $\hat{\nu}$ is obtained for each value of M; for clarity of explanation, successive values of $\boldsymbol{p}$ are indexed by n, related to M below, with only one component of $\boldsymbol{p}$ being revised for each value of n. Details for the Mth iteration follow in two parts:

1. A new estimate $\boldsymbol{p}$ of $\boldsymbol{\pi}$ is generated cyclically through change of one element of $\boldsymbol{p}$ at a time. The $(n+1)$th stage value $\boldsymbol{p}^{(n+1)}$ is obtained from the nth stage value $\boldsymbol{p}^{(n)}$ through replacement only of the element $p_i^{(n)}$ for which $n = (M-1)t+i-1$, $n = (M-1)t, \cdots, Mt-1$. Then

$$p_i^{(n+1)} = s_i/g_i(\boldsymbol{p}^{(n)}, \hat{\nu}^{(M-1)}). \tag{3.7}$$

2. A new estimate $\hat{\nu}^{(M)}$ of ν is obtained from (3.3) as follows:

$$\hat{\nu}^{(M)} = T/h(\boldsymbol{p}^{(Mt)}, \hat{\nu}^{(M-1)}). \tag{3.8}$$

As initial estimates one may use $p_i^{(0)} = 1/t, i = 1, \cdots, t$ and $\nu^{(0)} = 2T/(N-T)$, the latter being the solution to (3.3) under equal preferences. The speed with which the iterative solution can be performed on a computer does not necessitate the use of more sophisticated initial estimates.

Under the partitioning assumption (3.6), the iterative scheme converges monotonically to the unique solution to (3.2) and (3.3). Again this is established by adapting the proof given by Ford [11]. The main point in the proof is that the likelihood $L(\boldsymbol{\pi}, \nu)$ is increased at every step of the iterative scheme if and only if the corresponding parameter value is changed. This will now be demonstrated.

Let $\partial \ln L/\partial \pi_i |_n$ denote the value of $\partial \ln L/\partial \pi_i$ when $(\boldsymbol{\pi}, \nu)$ is replaced by $(\boldsymbol{p}^{(n)}, \hat{\nu}^{(M-1)})$. It then follows from (3.2) and (3.7) that

$$\partial \ln L/\partial \pi_i \big|_n = g_i(\boldsymbol{p}^{(n)}, \hat{\nu}^{(M-1)}) \Delta p_i/p_i^{(n)}$$

where $\Delta p_i = \{p_i^{(n+1)} - p_i^{(n)}\}$ so that Δp_i has the same sign as $\partial \ln L/\partial \pi_i |_n$. Again using (3.7) in (3.2) one obtains

$$\partial \ln L/\partial \pi_i \big|_{n+1} = g_i(\boldsymbol{p}^{(n)}, \hat{\nu}^{(M-1)}) - g_i(\boldsymbol{p}^{(n+1)}, \hat{\nu}^{(M-1)})$$

which is of the same sign as Δp_i in that $g_i(\boldsymbol{p}, \hat{\nu})$ is monotone decreasing in p_i. Now $\pi_i \partial \ln L/\partial \pi_i$ is monotone decreasing in π_i so that $\partial \ln L/\partial \pi_i$ has the same sign for all π_i between $p_i^{(n)}$ and $p_i^{(n+1)}$. Thus the change in the likelihood

$$\Delta \ln L = \Delta p_i \cdot \partial \ln L/\partial \pi_i \big|_\epsilon \geq 0$$

with equality if and only if $\Delta p_i = 0$, where $\partial \ln L/\partial \pi_i |_\epsilon$ denotes $\partial \ln L/\partial \pi_i$ at

$(p_1^{(n)}, \cdots, p_{i-1}^{(n)}, p_i^{(n)} + \epsilon \Delta p_i, p_{i+1}^{(n)}, \cdots, p_t^{(n)}, \hat{v}^{(M-1)})$ for $0 < \epsilon < 1$. A similar argument can be used to show that under (3.8)

$$\Delta \ln L = \Delta \hat{v} \cdot \partial \ln L / \partial \nu \big|_\epsilon \geq 0$$

with equality if and only if $\Delta \hat{v} = \{\hat{v}^{(M)} - \hat{v}^{(M-1)}\} = 0$, where $\partial \ln L / \partial \nu \big|_\epsilon$ denotes $\partial \ln L / \partial \nu$ at $(p^{(Mt)}, \hat{v}^{(M-1)} + \epsilon \Delta \hat{v})$ for $0 < \epsilon < 1$.

Rao and Kupper [17] discuss the problem of obtaining the maximum likelihood estimates of (π, θ) in the case $r_{ij} = r$, $i < j$, $i, j = 1, \cdots, t$. The extension of their presentation to unbalanced paired comparison experiments is straightforward.

4. SOME PROPERTIES OF THE ESTIMATORS

The maximum likelihood estimates p of the worth parameters π can be used to rank the set of t treatments. This is true for the Bradley-Terry model and for each of the two generalizations under discussion.

It was noted by Ford [11] that in the case of a balanced paired comparison experiment in which ties are not permitted, e.g., major-league baseball, the ranking obtained from the total percentages of wins is the same as that obtained from the maximum likelihood estimates of the parameters of the Bradley-Terry model. In the event that ties are permitted a common ranking system, used for example in hockey and soccer competition, is based on the points accumulated when a team is awarded 2, 1 or 0 points for a win, tie or loss, respectively. This is precisely the ranking based on $s = (s_1, \cdots, s_t)$. For a balanced paired comparison experiment this ranking is in agreement with that obtained from p for the model (2.6). This follows in that at a solution to Equations 3.2 and 3.3

$$s_i - s_k = r \sum_{j \neq i} \frac{2p_i + \hat{v}\sqrt{p_i p_j}}{p_i + p_j + \hat{v}\sqrt{p_i p_j}} - r \sum_{j \neq k} \frac{2p_k + \hat{v}\sqrt{p_k p_j}}{p_k + p_j + \hat{v}\sqrt{p_k p_j}}$$

$$= r(\sqrt{p_i} - \sqrt{p_k}) \sum_{j=1}^{t} \frac{2(\sqrt{p_i} + \sqrt{p_k})p_j + \hat{v}\sqrt{p_i p_k p_j} + \hat{v}p_j^{3/2}}{(p_i + p_j + \hat{v}\sqrt{p_i p_j})(p_k + p_j + \hat{v}\sqrt{p_k p_j})}$$

where the sum on the right hand remains positive whenever $p_j > 0$, $j = 1, \cdots, t$.

The ranking based on s does not always agree with that obtained from p for the Rao-Kupper model (2.3). This is not surprising in that the statistic (s, T) is sufficient for (π, ν) of the model (2.6), while the statistic $b = (b_{ij}, i \neq j, i, j = 1, \cdots, t)$, where $b_{ij} = w_{ij} + t_{ij}$, is sufficient for (π, θ) of the Rao-Kupper model. For the former model the sufficient statistic depends only on the net performance as measured by the scores $s_1, \cdots, s_t$, while for the latter model, the performance within each pair must be taken into account.

The asymptotic distribution of the maximum likelihood estimates $(p, \hat{v})$ of (π, ν) is obtained by appealing to Theorem 3 of Bradley and Gart [6]. Let $\mu_{ij} = r_{ij}/N$, $i, j = 1, \cdots, t$, where $N = \sum \sum_{i<j} r_{ij}$, and let $N \to \infty$ with all μ_{ij} held fixed. Then $\sqrt{N}(\hat{v} - \nu)$, $\sqrt{N}(p_1 - \pi_1)$, $\cdots$, $\sqrt{N}(p_{t-1} - \pi_{t-1})$ having a limiting multivariate normal distribution with mean vector zero and dispersion matrix $\Sigma = C^{-1}$, where C is obtained as described next.

Define $d_{ij} = 1/(\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j})$, $i, j = 1, \cdots, t$, and let

$$\lambda_{ii} = \frac{1}{\pi_i} \sum_j \mu_{ij}\{1 + (\nu/4)\sqrt{\pi_j/\pi_i}\}d_{ij} - \sum_j \mu_{ij}\{1 + (\nu/2)\sqrt{\pi_j/\pi_i}\}^2d_{ij}{}^2,$$

$$\lambda_{ij} = -\mu_{ij}\{1 + (\nu/4)(\sqrt{\pi_i/\pi_j} + \sqrt{\pi_j/\pi_i})\}d_{ij}{}^2, \qquad i \neq j,$$

and

$$\gamma_i = -\frac{1}{2} \sum_j \mu_{ij}(\pi_i - \pi_j)\sqrt{\pi_j/\pi_i}\,d_{ij}{}^2,$$

for $i, j = 1, \cdots, t$. It then follows that

$$C = \begin{bmatrix} c_{00} & \boldsymbol{\gamma}^{*\prime} \\ \boldsymbol{\gamma}^* & \boldsymbol{\Lambda}^* \end{bmatrix} \tag{4.1}$$

where

$$c_{00} = \sum_{i<j}\sum \mu_{ij}\{(\sqrt{\pi_i\pi_j}/\nu)d_{ij} - \pi_i\pi_jd_{ij}{}^2\},$$

and where through the use of the chain rule for differentiation

$$\boldsymbol{\Lambda}^* = [\lambda_{ij}{}^*; i, j = 1, \cdots, t-1]$$

$$= [\lambda_{ij} - \lambda_{it} - \lambda_{jt} + \lambda_{tt}; i, j = 1, \cdots, t-1]$$

and

$$\boldsymbol{\gamma}^{*\prime} = [\gamma_i{}^*; i = 1, \cdots, t-1]$$

$$= [\gamma_i - \gamma_t; i = 1, \cdots, t-1].$$

In that

$$\sqrt{N}(p_t - \pi_t) = -\sum_{j=1}^{t-1} \sqrt{N}(p_j - \pi_j),$$

the following result is obtained: $\sqrt{N}(\hat{\nu}-\nu)$ and $\sqrt{N}(p-\pi)$ have, as a limiting distribution, the singular normal distribution of t dimensions in a space of $(t+1)$ dimensions with mean vector zero and variance covariance matrix $\boldsymbol{\Sigma}^+ = [\sigma_{ij}; i, j = 0, \cdots, t]$ where

$$\boldsymbol{\Sigma}^+ = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\delta}^+ \\ \boldsymbol{\delta}^{+\prime} & \sigma_{tt} \end{bmatrix},$$

$$\boldsymbol{\delta}^{+\prime} = [\sigma_{it}; i = 0, \cdots, t-1] = -\left[\sum_{j=1}^{t-1}\sigma_{ij}; i = 0, \cdots, t-1\right],$$

$$\sigma_{tt} = \sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\sigma_{ij}.$$

5. THE TEST OF EQUAL PREFERENCES

A large sample test of the hypothesis that the t treatments are equally preferred can be based on the likelihood ratio statistic Λ_1. The limiting distribu-

tion of the statistic $-2 \ln \Lambda_1$ can be obtained for both the null hypothesis and local alternatives by appealing to the widely used result of Wald [18], when extended to associated populations as discussed by Bradley and Gart [6].

The hypothesis

$$H_0: \pi_i = 1/t, \qquad i = 1, \cdots, t, \nu \text{ unspecified}$$

is to be tested against alternatives of the form $\pi_i = 1/t + \delta_i/\sqrt{N}$. Using the fact that the value ν required to maximize the likelihood under H_0 is $\hat{\nu}^0 = 2T$ $/(N-T)$, one obtains the statistic

$$S_1 = -2 \ln \Lambda_1$$
$$= 2[\ln L(p, \hat{\nu}) - (N - T) \ln (N - T) - T \ln 2T + N \ln 2N].$$

It then follows that under H_0 the statistic S_1 has a limiting central chi-square distribution with $(t-1)$ degrees of freedom. Under local alternatives the limiting distribution is the noncentral chi-square distribution with noncentrality parameter

$$\bar{\lambda}_1{}^2 = \frac{1}{2(2 + \nu)} \delta' E \delta$$

where

$$\delta' = (\delta_1, \cdots, \delta_{t-1}) \quad \text{and} \quad E = [e_{ij}; i, j = 1, \cdots, t - 1]$$

with

$$e_{ii} = t^2 \left(\sum_k \mu_{ki} + \sum_k \mu_{kt} + 2\mu_{it} \right),$$

$$e_{ij} = t^2 \left(\mu_{it} + \mu_{jt} - \mu_{ij} + \sum_k \mu_{kt} \right), \qquad i \neq j.$$

The noncentrality parameter follows from the form given by Bradley and Gart [6] by noting that under the hypothesis of equal preferences, the matrix C given by (4.1) becomes

$$C = \begin{bmatrix} c_{00}{}^0 & 0' \\ 0 & \Lambda^{*0} \end{bmatrix}$$

where $c_{00}{}^0 = 2/\nu(2+\nu)$

$$\Lambda^{*0} = \frac{1}{2(2 + \nu)} E.$$

Rao and Kupper [17] have proposed using the statistic $S_2 = -2 \ln \Lambda_2$, where Λ_2 is the likelihood ratio statistic, for testing the hypothesis of equal preferences in the model (2.3). By extending their discussion to unbalanced paired comparison experiments and again appealing to Wald [18] and Bradley and Gart [6], it follows that under H_0 the statistic S_2 has a limiting central chi-square distribution, while under local alternatives the limiting distribution is the non-

central chi-square distribution with noncentrality parameter

$$\bar{\lambda}_2{}^2 = \frac{2\theta^2}{(1+\theta)^3}\, \delta^{*\prime} E \delta^*,$$

with $\delta^{*\prime} = (\delta_1{}^*, \cdots, \delta_{t-1}{}^*)$ where $\delta_i{}^* = \lim_{N \to \infty} \sqrt{N}(\pi_i - 1/t)$.

The asymptotic relative efficiency, in the sense of Pitman, of the tests of equal preferences based on S_1 and S_2 is of value in comparing the performance of the model (2.6) with the Rao-Kupper model (2.3). This will now be obtained by applying the following result due to Hannan [13]; if two test statistics have, under the alternative hypothesis, noncentral chi-square distributions with the same degrees of freedom, the asymptotic relative efficiency of one test with respect to the other is equal to the ratio of the two noncentrality parameters after the alternatives have been set equal.

The models (2.6) and (2.3) are identical under the null hypothesis of equal preferences when $\nu = \theta - 1$. Under the local alternatives $\pi_i = 1/t + \delta_i/\sqrt{N}$, the probabilities of the model (2.6) become

$$p(\ell \mid i, j) = 1/(2+\nu) + (-1)^{\kappa(\ell, j)} \Delta_{ijN}/\sqrt{N}, \qquad \ell = i, j$$

$$p(0 \mid i, j) = \nu/(2+\nu) + o(1)/\sqrt{N},$$

where

$$\Delta_{ijN} = t(\delta_i - \delta_j)/2(2+\nu) + o(1),$$

and where $\kappa(\cdot, \cdot)$ is the Kronecker delta. Similarly, under the local alternatives $\pi_i = 1/t + \delta_i{}^*/\sqrt{N}$, the probabilities of the model (2.3) become

$$p^*(\ell \mid i, j) = 1/(1+\theta) + (-1)^{\kappa(\ell, j)} \Delta_{ijN}{}^*/\sqrt{N}, \qquad \ell = i, j$$

$$p^*(0 \mid i, j) = (\theta - 1)/(\theta + 1) + o(1)/\sqrt{N},$$

where

$$\Delta_{ijN}{}^* = t(\delta_i{}^* - \delta_j{}^*)/(1+\theta)^2 + o(1).$$

Thus, to make the models identical under local alternatives it is necessary that $(\delta_i - \delta_j)/2(2+\nu) = \theta(\delta_i{}^* - \delta_j{}^*)/(1+\theta)^2$, a result which holds for all (i, j) if and only if $\delta_i{}^* = \delta_i(1+\theta)^2/2\theta(2+\nu)$ for all i, and that $\nu = \theta - 1$. It then follows that $\bar{\lambda}_1{}^2 = \bar{\lambda}_2{}^2$ so that, by the criterion of Hannan, the statistics S_1 and S_2, which correspond to the models (2.6) and (2.3) respectively, have the same asymptotic efficiency in testing the hypothesis of equal preferences.

6. THE APPROPRIATENESS OF THE MODEL

The most general model available for paired comparisons is one for which the probabilities $\bar{p}(i \mid i, j)$, $\bar{p}(j \mid i, j)$ and $\bar{p}(0 \mid i, j)$ are unspecified except for the requirement that they sum to unity for each treatment pair (i, j). The maximum likelihood estimates of $\bar{p}(k \mid i, j)$ are then given by the relative frequencies $f(k \mid i, j)/r_{ij}$, $k = i, j, 0$, a well-known result for the multinomial distribution. Thus the fit of this general model to the experimental data is exact.

To test the appropriateness or the goodness of fit of the model (2.6), one

may consider the following hypotheses:

$$H_0: \bar{p}(k \mid i,j) = p(k \mid i,j) \quad \text{for all } (i,j),$$

$$H_1: \bar{p}(k \mid i,j) \neq p(k \mid i,j) \quad \text{for some } (i,j),$$

where $p(k \mid i,j)$, $k = i, j, 0$, are given in (2.6). The likelihood ratio statistic can then be used for the test of H_0 against H_1.

The logarithm of the likelihood function is given by

$$\ln L\{\bar{p}(k \mid i,j)\} = \sum_{i<j} \sum_k f(k \mid i,j) \ln \bar{p}(k \mid i,j). \tag{6.1}$$

Under H_0, (6.1) has the form (3.1). Thus for the test of goodness of fit the statistic is

$$U_1 = -2 \ln \Lambda_1 = 2 \sum_{i<j} \sum_k f(k \mid i,j) \ln\{f(k \mid i,j)/\hat{f}(k \mid i,j)\}, \tag{6.2}$$

where

$$\hat{f}(k \mid i,j) = r_{ij}\hat{p}(k \mid i,j)$$

are the estimates of the expected frequencies obtained from $\hat{p}(k \mid i,j)$, the probabilities $p(k \mid i, j)$ of H_0 when (π, ν) are replaced by their maximum likelihood estimates $(p, \hat{\nu})$. The statistic U_1 has, under H_0, an asymptotic central chi-square distribution with $t(t-2)$ degrees of freedom.

It follows, through adaptation of the demonstration given by Kullback [14, pp. 113-4] or by Bradley [2], that

$$U_1 \simeq \sum_{i<j} \sum_k \{f(k \mid i,j) - \hat{f}(k \mid i,j)\}^2/\hat{f}(k \mid i,j), \tag{6.3}$$

the usual chi-square goodness of fit statistic. The accuracy of the approximation depends directly on how close each expected frequency $\hat{f}(k \mid i, j)$ is to the observed frequency $f(k \mid i, j)$.

A parallel development for the Rao-Kupper model (2.3) yields a statistic U_2 of the form (6.2) which, by virtue of (6.3) is comparable to the statistic which they [17] propose.

7. NUMERICAL EXAMPLES

Two examples will now be given to demonstrate the procedures that have been developed for the model (2.6). The data for Example 1 have been used in a number of earlier studies. Example 2 serves to introduce a set of data for an unbalanced paired comparison experiment. The entire data analysis requires only a fraction of a minute on a high speed computer and hence is performed economically.

Example 1. Rao and Kupper [17] and Glenn and David [12] have illustrated the use of their respective models with data taken from a study by Fleckenstein, Freund and Jackson [10] who describe a balanced paired comparison experiment in which $t = 5$ and $r = 30$. The results of that experiment are summarized in the following tabulation:

Treatment i	w_i	t_i	s_i
1	57	17	131
2	41	21	103
3	81	17	179
4	16	14	46
5	62	17	141

The total number of ties $T = 43$. The maximum likelihood estimates were obtained using the iterative scheme described in Section 3. The iterates converged to within a tolerance of 0.001 in 14 iterations. The final maximum likelihood estimates are given in the following tabulation together with those obtained for model (2.3):

Model		p_1	p_2	p_3	p_4	p_5
(2.6)	$\hat{\nu} = 0.404$	0.183	0.110	0.454	0.034	0.219
(2.3)[a]	$\hat{\theta} = 1.452$	0.196	0.124	0.410	0.047	0.223

[a] See correction to Rao and Kupper [17]. The remaining calculations published by Rao and Kupper were based on their incorrect parameter estimates and hence are in error.

The tests of equal preferences gave $S_1 = 84.8$ and $S_2 = 84.7$ whose critical levels are less than 0.0005. The tests of goodness of fit gave $U_1 = 10.7$ and $U_2 = 10.8$ which, with 15 degrees of freedom, have critical levels exceeding 0.75.

Example 2. The data used in this example have been extracted from those which were obtained in a large scale experiment in which each respondent was asked to test two brands of chocolate milk pudding.[2] Altogether there were six brands of pudding; the frequencies of comparison, preference and no preference are given in the table.

DATA FOR EXAMPLE 2

Pair i	j	r_{ij}	w_{ij}	w_{ji}	t_{ij}
1	2	57	19	22	16
1	3	47	16	19	12
2	3	48	19	19	10
1	4	54	18	23	13
2	4	51	23	19	9
3	4	54	19	20	15
1	5	50	13	19	18
2	5	48	16	20	12
3	5	48	16	15	17
4	5	47	17	14	16
1	6	51	18	21	12
2	6	54	22	20	12
3	6	41	13	18	10
4	6	51	14	19	18
5	6	44	11	21	12

The statistic (s, T) and the maximum likelihood estimates $(p, \hat{\nu})$ obtained from the iterative scheme (3.7), (3.8) are given in the following tabulation (the iteration process required seven iterations to converge within a tolerance of 0.001):

[2] These data have been obtained through the courtesy of Mavis Carroll, John Heimlich, and the General Foods Corporation.

Brand

	1	2	3	4	5	6	
s	239	263	236	257	233	262	$T = 202$
p	0.139	0.173	0.162	0.165	0.159	0.202	$\hat{v} = 0.747$

The test of equal preferences gave $S_1 = 4.08$ which, with 5 degrees of freedom, is not significant at $\alpha = 0.50$. The test of goodness of fit gave $U_1 = 15.8$ whose critical level exceeds 0.88.

REFERENCES

[1] Bradley, R. A., "Another Interpretation of a Model for Paired Comparisons," *Psychometrika*, 30 (1965), 315–8.

[2] ――――, "Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons," *Biometrics*, 10 (1954), 375–90.

[3] ――――, "Rank Analysis of Incomplete Block Designs II. Additional Tables for the Method of Paired Comparisons," *Biometrika*, 41 (1954), 502–37.

[4] ――――, "Rank Analysis of Incomplete Block Designs III. Some Large-Sample Results on Estimation and Power for a Method of Paired Comparisons," *Biometrika*, 42 (1955), 450–70.

[5] ――――, "Some Statistical Methods in Taste Testing and Quality Evaluation," *Biometrics*, 9 (1953), 22–38.

[6] ―――― and Gart, J. J., "The Asymptotic Properties of ML Estimators When Sampling from Associated Populations," *Biometrika*, 49 (1962), 205–14.

[7] ―――― and Terry, M. E., "Rank Analysis of Incomplete Block Designs I. The Method of Paired Comparisons," *Biometrika*, 39 (1952), 324–45.

[8] Brunk, H. D., "Mathematical Models for Ranking from Paired Comparisons," *Journal of the American Statistical Association*, 55 (September 1960), 503–20.

[9] David, H. A , *The Method of Paired Comparisons*, New York: Hafner, 1963.

[10] Fleckenstein, J., Freund, R. A. and Jackson, J. E., "A Paired Comparison Test of Typewriter Carbon Papers," *Tappi*, 41 (1958), 128–30.

[11] Ford, L. R., Jr., "Solution of a Ranking Problem from Binary Comparisons," *American Mathematical Monthly*, 54 (1957), 28–33.

[12] Glenn, W. A. and David, H. A., "Ties in Paired-Comparison Experiments Using a Modified Thurstone-Mosteller Model," *Biometrics*, 16 (1960), 86–109.

[13] Hannan, E. J., "The Asymptotic Powers of Certain Tests Based on Multiple Correlations," *Journal of the Royal Statistical Society*, B, 18 (1956), 227–33.

[14] Kullback, S., *Information Theory and Statistics*, New York: John Wiley & Sons, Inc. 1959.

[15] Lehmann, E. L., "The Power of Rank Tests," *Annals of Mathematical Statistics*, 24 (1953), 23–43.

[16] Luce, R. D., *Individual Choice Behavior*, New York: John Wiley & Sons, Inc., 1959, Chapter 1.

[17] Rao, P. V., and Kupper, L. L., "Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model," *Journal of the American Statistical Association*, 62 (March 1967), 194–204. Corrigenda, 63 (December 1968), 1550.

[18] Wald, A., "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54 (1943), 426–82.

[19] Zermelo, E., "Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, 29 (1929), 436–60.