A BAYESIAN APPROACH TO PAIRED COMPARISON EXPERIMENTATION

BU-426-M by August, 1972

R. R. Davidson and D. L. Solomon

Cornell University

Abstract

We have formulated a Bayesian approach to paired comparison experimentation under the multi-binomial and Bradley-Terry models and have provided two estimators of the model parameters (assuming a natural conjugate prior distribution) for each model. Furthermore, in the Bradley-Terry case, if the experimental design and prior distribution satisfy the criterion of posterior balance, then for a large class of acceptable loss functions, a Bayes ranking of the worth parameters is explicitly calculated and shown to coincide with the ranking determined by each of the two estimators. Thus, if the objective is ranking the worth parameters by their estimates, which are not easily calculated, one can easily calculate the Bayes ranking with the assurance that it agrees with the ranking provided by these estimates.

A BAYESIAN APPROACH TO PAIRED COMPARISON EXPERIMENTATION

By R. R. Davidson and D. L. Solomon
Cornell University

SUMMARY

We have formulated a Bayesian approach to paired comparison experimentation under the multi-binomial and Bradley-Terry models and have provided two estimators of the model parameters (assuming a natural conjugate prior distribution) for each model. Furthermore, in the Bradley-Terry case, if the experimental design and prior distribution satisfy the criterion of posterior balance, then for a large class of acceptable loss functions, a Bayes ranking of the worth parameters is explicitly calculated and shown to coincide with the ranking determined by each of the two estimators. Thus, if the objective is ranking the worth parameters by their estimates, which are not easily calculated, one can easily calculate the Bayes ranking with the assurance that it agrees with the ranking provided by these estimates.

Some key words: Paired comparisons, preference testing, tournaments, prior-posterior analysis, Bayes estimators, Bayes rankings.

1. INTRODUCTION

The testing of a set of items for preference on overall suitability or on specific characteristics often requires of the respondent the ability to make very fine sensory discriminations based on complex physiological processes. To remove some of the confusion associated with simultaneously comparing several objects, the method of paired comparisons has been widely employed. In paired comparison experimentation, expressions of preference are obtained for all possible pairs of

objects in the set of interest. The data from such experiments are used, among other things, to rank the objects in order of preference. League competitions and round-robin tournaments provide additional examples of situations in which rankings are based on the performance of objects when they meet in pairs. The method of paired comparisons has been discussed in a monograph by David (1963).

A number of mathematical models for paired comparisons have been introduced, the most notable being the Thurstone-Mosteller and the Bradley-Terry models. Less structured models under which the responses to each pair are considered to arise from distinct binomial experiments have also been used. Models of this type are referred to as multi-binomial models. The preference probabilities of the multi-binomial model can be required to satisfy transitivity constraints of varying severity.

The present paper undertakes a Bayesian approach to the analysis of data arising from paired comparison experiments. Two statistical models are considered: the multi-binomial model (Section 2) and the Bradley-Terry model (Sections 3 and 4). The objectives are (i) to provide estimators of the preference probabilities of the multi-binomial model and of the worth parameters of the Bradley-Terry model, and (ii) to provide a ranking of the objects in the set of interest.

For each statistical model, the natural conjugate family of priors can be used to represent a wide variety of prior beliefs. Furthermore, the prior parameters are shown to have meaningful interpretations in the sense of "equivalent prior sample information" as discussed by Winkler (1967). The prior-posterior analysis for the multi-binomial model is a direct extension of that for the single binomial. While less tractable, the prior-posterior analysis for the Bradley-Terry model has a number of interesting features which result from properties noted by Ford (1957).

In the estimation problem, two approaches to using the posterior distribution are considered. In the first approach the modal point of the posterior is used as the estimator of the parameters. This is analogous to the use of the maximum likelihood principle in the classical analysis of paired comparison data. The second approach is to obtain the estimator which, for a specified loss, minimizes the Bayes risk.

The ranking problem is a more complicated one and has been considered (Section 4) only for the Bradley-Terry model where either the mode of the posterior or the mean of the posterior (the Bayes rule under quadratic loss) can be used to provide a ranking of the objects. In certain "balanced" situations an intuitive ranking based on the posterior parameters can be shown to be the Bayes ranking for a large class of loss functions. This development parallels the discussion of optimal ranking procedures given by Bühlmann and Huber (1963). In the presence of such balance it is shown that the three rankings referred to above are identical.

2. THE MULTI-BINOMIAL MODEL

One can formulate probability models for paired comparisons which have varying degrees of structure. A relatively unstructured model can be described as follows. Suppose that t objects are to be tested in pairs with n_{ij} responses being obtained to the pair (i,j). Each response consists of an expression of preference for either object i or object j; expressions of no preference are not permitted. Let $\theta = [\theta_{ij}; i,j = 1, \cdots, t]$, where θ_{ij} is the probability of preference for object i over object j, $\theta_{ij} + \theta_{ji} = 1$, and where for convenience $\theta_{ii} \equiv \frac{1}{2}$. Also set $n = [n_{ij}; i,j = 1, \cdots, t]$ and $w = [w_{ij}; i,j = 1, \cdots, t]$, where w_{ij} is the observed frequency of preference for object i over object j, $w_{ij} + w_{ji} = n_{ij}$, and $n_{ii} = w_{ii} \equiv 0$. Then under the assumption that all pairwise comparisons are

stochastically independent, the statistical model is defined by the probability function

$$p_{W|\underset{\sim}{\theta}}(\underset{\sim}{w}|\underset{\sim}{\theta}) = \prod_{i<j} \binom{n_{ij}}{w_{ij}} \theta_{ij}^{W_{ij}} \theta_{ji}^{W_{ji}} . \qquad (2.1)$$

The probability function (2.1) is a product of independent binomial probability functions and will be referred to as the multi-binomial model. Under this model we effectively analyze the data $\underset{\sim}{w}$ as though it had arisen from $\binom{t}{2}$ separate experiments. For instance, in estimating θ_{ij} we would only use the observed frequency w_{ij} obtained in the comparison of the pair (i,j). Although use of the model (2.1) in which there is no structure on $\underset{\sim}{\theta}$ $\left(\text{other than } \underset{\sim}{\theta} + \underset{\sim}{\theta}^T \equiv \underset{\sim}{1}(t \times t)\right)$ is unrealistic in most situations, it is instructive to examine possible prior-posterior analyses for $\underset{\sim}{\theta}$. The development will be of value in understanding the corresponding analyses for models which impose additional structure on $\underset{\sim}{\theta}$.

The prior-posterior analysis is a straightforward extension of that for the single binomial model. Following the spirit of Raiffa and Schlaifer (1961, Section 3.2), it follows directly that the conjugate prior distribution for $\underset{\sim}{\bar{\theta}}$ is

$$p_{\underset{\sim}{\bar{\theta}}}(\underset{\sim}{\theta}) = K(\underset{\sim}{w}^o) \prod_{i<j} \theta_{ij}^{w_{ij}^o} \theta_{ji}^{w_{ji}^o} \equiv g(\underset{\sim}{\theta} : \underset{\sim}{w}^o), \quad \underset{\sim}{\theta} \in \Delta, \qquad (2.2)$$

where $\Delta = \{\underset{\sim}{\theta}: 0 \leq \theta_{ij} \leq 1, \theta_{ij} + \theta_{ji} = 1; i \neq j, i,j = 1, \cdots, t\}$, $\underset{\sim}{w}^o = [w_{ij}^o; i,j = 1, \cdots, t]$ is a matrix of prior parameters, $w_{ij}^o > -1$, $w_{ii}^o \equiv 0$, and where

$$K(\underset{\sim}{w}^o) = \prod_{i<j} \Gamma(w_{ij}^o + w_{ji}^o + 2)/\Gamma(w_{ij}^o + 1)\Gamma(w_{ji}^o + 1) . \qquad (2.3)$$

Under (2.2), the elements of $\underset{\sim}{\bar{\theta}}$ corresponding to different pairs are independently

distributed with $\bar{\theta}_{ij} \sim \text{Beta}(w^o_{ij} + 1, w^o_{ji} + 1)$. Although this independence may be undesirable, it is compatible with the nature of the statistical model.

Setting $\underset{\sim}{n}^o = [n^o_{ij}; \ i,j = 1, \cdots, t]$ with $n^o_{ij} = w^o_{ij} + w^o_{ji} = n^o_{ji}$, we can obtain an equivalent form of (2.2) in terms of the lower halves of the matrices $\underset{\sim}{w}^o$ and $\underset{\sim}{n}^o$. The prior (2.2) is properly defined and continuous over the interior of Δ whenever $w^o_{ij} > -1$ for all $i,j = 1, \cdots, t$. Moreover, when $\underset{\sim}{w}^o$ is nonnegative, (2.2) has a unique modal point in Δ at $\underset{\sim}{\theta}^*$ where $\theta^*_{ij} = w^o_{ij}/n^o_{ij}$. If $w^o_{ij} = 0$ for some $i \neq j$ then $\underset{\sim}{\theta}^*$ is on the boundary of Δ, but if $w^o_{ij} > 0$ for all $i \neq j$ then (2.2) is zero on the boundary of Δ and $\underset{\sim}{\theta}^*$ is in the interior.

When they are nonnegative, the parameters $\underset{\sim}{w}^o$ can be interpreted in the sense of "equivalent prior sample information" as described by Winkler (1967). In specifying (w^o_{ij}, w^o_{ji}) or equivalently (w^o_{ij}, n^o_{ij}) for each pair (i,j), the experimenter can be viewed as saying that he believes that of n^o_{ij} expression of preference for the pair (i,j), w^o_{ij} will favor object i. The ratio w^o_{ij}/n^o_{ij} represents the experimenter's prior assessment of the "most probable" value of θ_{ij} and corresponds to the modal point of the prior distribution, while n^o_{ij} represents the degree of confidence that the experimenter has in his expressed belief. Larger values of n^o_{ij} indicate greater confidence on the part of the experimenter, and stronger indications from the data will be required in order to substantially alter the experimenter's prior beliefs about θ_{ij}.

The posterior distribution for $\bar{\underset{\sim}{\theta}}$ is given by

$$P_{\bar{\theta}|W}(\underset{\sim}{\theta}|\underset{\sim}{w}) = K(\underset{\sim}{w}^1) \prod_{\substack{i,j \\ i<j}} \theta^{w^1_{ij}}_{ij}\theta^{w^1_{ji}}_{ji} = g(\underset{\sim}{\theta} : \underset{\sim}{w}^1), \ \underset{\sim}{\theta} \in \Delta \qquad (2.4)$$

where $K(\cdot)$ is given by (2.3), and where the parameters $\underset{\sim}{w}^1$ of the posterior distribution are obtained by combining the data $\underset{\sim}{w}$ with the parameters $\underset{\sim}{w}^o$ representing

the experimenter's prior beliefs, $\underset{\sim}{w}^1 = \underset{\sim}{w} + \underset{\sim}{w}^o$. Also $\underset{\sim}{n}^1 = \underset{\sim}{w}^1 + \underset{\sim}{w}^{1^T} = \underset{\sim}{n} + \underset{\sim}{n}^o$.

Inferences about the preference probabilities $\underset{\sim}{\theta}$ can be made on the basis of the posterior distribution (2.4). In particular, we will consider the estimation of $\underset{\sim}{\theta}$. One approach would estimate $\underset{\sim}{\theta}$ by the mode $\underset{\sim}{\theta}^*(\underset{\sim}{w})$ of the posterior distribution, $\theta_{ij}^*(\underset{\sim}{w}) = w_{ij}^1/n_{ij}^1$. This is analogous to the usual maximum likelihood estimator $\hat{\underset{\sim}{\theta}}$, $\hat{\theta}_{ij} = w_{ij}/n_{ij}$, obtained under the statistical model. Alternatively, one might adopt a decision theoretic approach and estimate $\underset{\sim}{\theta}$ by the Bayes rule corresponding to the assumed loss function. Under quadratic loss, the Bayes estimator for $\underset{\sim}{\theta}$ is the mean $\overline{\underset{\sim}{\theta}}(\underset{\sim}{w}) = E[\underset{\sim}{\theta}|\underset{\sim}{W} = \underset{\sim}{w}]$ of the posterior distribution, $\overline{\theta}_{ij}(\underset{\sim}{w}) = (w_{ij}^1 + 1)/(n_{ij}^1 + 2)$.

Let us now turn attention to models which impose additional structure on the preference probabilities $\underset{\sim}{\theta}$. Such structure can be of two types: (i) a specific parametric form for the preference probabilities $\underset{\sim}{\theta}$, or (ii) a set of constraints on the space Δ for $\underset{\sim}{\theta}$. The models due to Thurstone-Mosteller and Bradley-Terry are examples of parametric structure. The second of these models will be treated in subsequent sections of this paper. Transitivity constraints provide an example of the second type of structure and will now be discussed.

When the preference probabilities $\underset{\sim}{\theta}$ are unconstrained, a problem arises in defining a ranking of the objects based on $\underset{\sim}{\theta}$. This problem can be resolved by requiring $\underset{\sim}{\theta}$ to satisfy a transitivity constraint. Two such conditions have been used:

weak stochastic transitivity (C_1): for every triad of objects (i,j,k)

$$\theta_{ij} \geq \tfrac{1}{2} \text{ and } \theta_{jk} \geq \tfrac{1}{2} \text{ imply } \theta_{ik} \geq \tfrac{1}{2} , \qquad (2.5)$$

strong stochastic transitivity (C_2): for every triad of objects (i,j,k)

$$\theta_{ij} \geq \tfrac{1}{2} \text{ and } \theta_{jk} \geq \tfrac{1}{2} \text{ imply } \theta_{ik} \geq \max(\theta_{ij}, \theta_{jk}). \qquad (2.6)$$

The stronger condition is not necessarily warranted in all situations. For a discussion of this point see David (1963, p. 13) and Luce (1961). Conditions C_1 and C_2 each lead to a well defined ranking of every triad and hence to a true ranking of the set of t objects. The ranking defined by condition C_2 is precisely that determined by the row sums of $\underset{\sim}{\theta}$, cf. David (1963, p. 19).

The transitivity constraints C_1 and C_2 lead to corresponding modification of the prior and posterior distributions. Let

$$\Delta_i = \{\underset{\sim}{\theta} : \underset{\sim}{\theta} \in \Delta \text{ and satisfies } C_i\}, \quad i = 1, 2.$$

Then if φ_i denotes the probability measure under (2.2) of the region Δ_i, the prior and posterior distributions are given by

$$\left.\begin{array}{l} P_{\underset{\sim}{\theta}}(\underset{\sim}{\theta}) \quad\quad = \varphi_i^{-1}\; g(\underset{\sim}{\theta} : \underset{\sim}{w}^0) \\[3mm] P_{\underset{\sim}{\theta}|\underset{\sim}{W}}(\underset{\sim}{\theta}|\underset{\sim}{w}) = \varphi_i^{-1}\; g(\underset{\sim}{\theta} : \underset{\sim}{w}^1) \end{array}\right\}\; \underset{\sim}{\theta} \in \Delta_i$$

respectively, $i = 1, 2$, where $g(\underset{\sim}{\theta} : \cdot)$ is defined by (2.4).

As for the estimation of $\underset{\sim}{\theta}$ on the basis of the posterior distribution, the constraints C_1 and C_2 play a major role. To obtain the modal estimator we determine whether the ratios w_{ij}^1/n_{ij}^1 satisfy the constraint C_i; if they do then these ratios comprise the modal estimator $\underset{\sim}{\theta}^*(\underset{\sim}{w})$; if not then $\underset{\sim}{\theta}^*(\underset{\sim}{w})$ is on the boundary of Δ_i. The search for the Bayes rule corresponding to a specified loss function poses greater problems. In keeping with the spirit of the constraint C_i, we would restrict attention to estimators taking values in Δ_i. Since Δ_i is not convex the resulting estimation problems require complicated optimization techniques.

3. PRIOR AND POSTERIOR ANALYSIS UNDER THE BRADLEY-TERRY MODEL

The model for paired comparisons presented by Bradley and Terry (1952) imposes a parametric structure on the preference probabilities θ. It is assumed that the responses to the objects may be described in terms of an underlying continuum on which the "worths" of the objects can be relatively located. If we let π_i denote the "worth", an index of relative preference, of the ith object, $\pi_i \geq 0$, then the Bradley-Terry model postulates that

$$\theta_{ij} = \pi_i/(\pi_i + \pi_j) \quad \text{for all } (i,j). \tag{3.1}$$

Since θ is invariant to multiplication of $\pi = (\pi_1, \cdots, \pi_t)$ by a nonzero constant, we will assume without loss of generality that $\sum_{i=1}^{t} \pi_i = 1$. This same model had been proposed earlier by Zermelo (1929), and was subsequently independently presented by Ford (1957). A summary of various representations of the model (3.1) is given in Bradley (1965).

The Bradley-Terry model has the advantage that the parameterization $\pi = (\pi_1, \cdots, \pi_t)$ is in terms of the set of t objects of interest, whereas the parameterization θ of the multi-binomial model is in terms of the $\binom{t}{2}$ paired comparisons. Moreover, under the model (3.1), the preference probabilities θ satisfy the strong transitivity condition (2.6): for

$$\theta_{ik} = \theta_{ij}\theta_{jk}/(\theta_{ij}\theta_{jk} + \theta_{ji}\theta_{kj}),$$

and when $\theta_{ij} \geq \frac{1}{2}$ then $\theta_{ik} \geq \theta_{ij}\theta_{jk}/\theta_{ij}(\theta_{jk} + \theta_{kj}) = \theta_{jk}$, and similarly, when $\theta_{jk} \geq \frac{1}{2}$ then $\theta_{ik} \geq \theta_{ij}$. The true ranking of the objects determined by θ is precisely the ranking determined by π.

When the Bradley-Terry model is assumed to hold, the statistical model (2.1) becomes

$$P_{\underset{\sim}{W}|\underset{\sim}{\Pi}}(\underset{\sim}{w}|\underset{\sim}{\pi}) = \prod_{i<j} \binom{n_{ij}}{w_{ij}} \pi_i^{w_{ij}} \pi_j^{w_{ji}} / (\pi_i + \pi_j)^{n_{ij}} \qquad (3.2)$$

where $\underset{\sim}{w} = [w_{ij}; i,j = 1, \cdots, t]$ and $\underset{\sim}{n} = [n_{ij}; i,j = 1, \cdots, t]$ are defined as for the model (2.1). If we let $\underset{\sim}{s} = (s_1, \cdots, s_t)$ where $s_i = \sum_{j=1}^{t} w_{ij}$, the total number of preferences for the ith object, then (3.2) can be written

$$P_{\underset{\sim}{W}|\underset{\sim}{\Pi}}(\underset{\sim}{w}|\underset{\sim}{\pi}) = \left(\prod_{i=1}^{t} \pi_i^{s_i} \right) \prod_{i<j} \binom{n_{ij}}{w_{ij}} (\pi_i + \pi_j)^{-n_{ij}} . \qquad (3.3)$$

The statistic $(\underset{\sim}{s}, \underset{\sim}{n})$ is sufficient for the statistical model.

The conjugate prior distribution for $\underset{\sim}{\pi}$ obtained from (3.2) has density

$$P_{\underset{\sim}{\Pi}}(\underset{\sim}{\pi}) = C(\underset{\sim}{w}^0) \prod_{i<j} \pi_i^{w_{ij}^0} \pi_j^{w_{ji}^0} / (\pi_i + \pi_j)^{n_{ij}^0} \equiv f(\underset{\sim}{\pi} : \underset{\sim}{w}^0), \ \underset{\sim}{\pi} \in \Omega, \qquad (3.4)$$

where $\Omega = \{\underset{\sim}{\pi} : \pi_i \geq 0, i = 1, \cdots, t, \sum_{i=1}^{t} \pi_i = 1\}$, and where $\underset{\sim}{w}^0 = [w_{ij}^0; i,j = 1, \cdots, t]$ is a matrix of prior parameters $w_{ij}^0 \geq 0$, and $\underset{\sim}{n}^0 = \underset{\sim}{w}^0 + \underset{\sim}{w}^{0T}$ as in Section 2. Alternatively, if we let $\underset{\sim}{s}^0 = (s_1^0, \cdots, s_t^0)$ be the vector of row sums of $\underset{\sim}{w}^0$, $s_i^0 = \sum_{j=1}^{t} w_{ij}^0$, then (3.4) can be rewritten as

$$P_{\underset{\sim}{\Pi}}(\underset{\sim}{\pi}) = K(\underset{\sim}{s}^0, \underset{\sim}{n}^0) \left(\prod_{i=1}^{t} \pi_i^{s_i^0} \right) / \prod_{i<j} (\pi_i + \pi_j)^{n_{ij}^0} \equiv h(\underset{\sim}{\pi} : \underset{\sim}{s}^0, \underset{\sim}{n}^0), \ \underset{\sim}{\pi} \in \Omega. \qquad (3.5)$$

The form of the constant $K(\cdot, \cdot)$ will be described later. Each $(\underset{\sim}{s}^0, \underset{\sim}{n}^0)$ determines exactly one prior distribution so that all matrices $\underset{\sim}{w}^0$ with the same $\underset{\sim}{s}^0 = \underset{\sim}{w}^0 \underset{\sim}{1}$ and $\underset{\sim}{n}^0 = \underset{\sim}{w}^0 + \underset{\sim}{w}^{0T}$ represent equivalent prior beliefs.

The prior distribution is properly defined and continuous on Ω whenever $\underset{\sim}{w}^o$ is nonnegative. This is clear from (3.4). By the definition of $\underset{\sim}{s}^o$ and $\underset{\sim}{n}^o$ in terms of $\underset{\sim}{w}^o$, the prior distribution in the form (3.5) requires that $s_i^o \leq \sum_{j=1}^{t} n_{ij}^o$, $i = 1, \cdots, t$, and that $\sum_{i=1}^{t} s_i^o = \sum_{i<j} n_{ij}^o$.

In eliciting the prior beliefs of an experimenter, the prior parameterization $\underset{\sim}{w}^o$ is likely to be more effective than $(\underset{\sim}{s}^o, \underset{\sim}{n}^o)$. The problem with asking the experimenter to specify $(\underset{\sim}{s}^o, \underset{\sim}{n}^o)$ is that the conditions referred to in the preceding paragraph must be satisfied.

The interpretation of $\underset{\sim}{w}^o$ is identical with that under the conjugate prior (2.2) for the multi-binomial model, namely that the experimenter believes that of n_{ij}^o responses to the pair (i,j) in a conceptual experiment, w_{ij}^o will favor object i. As suggested before, n_{ij}^o represents the strength of the experimenter's belief as to the outcome for the pair (i,j).

Suppose now that $\underset{\sim}{w}^o$ satisfies the following assumption introduced by Ford (1957).

Assumption: In every partition of the indices $\{1, \cdots, t\}$ into two nonempty sets I and J, there exists $i \in I$, $j \in J$ such that $w_{ij}^o > 0$. Equivalently, for each (i,j) there exists a sequence of indices $i_0, i_1, \cdots, i_m$ with $i_0 = i$ and $i_m = j$ such that $w_{i_k i_{k+1}} > 0$ for $k = 0, 1, \cdots, (m - 1)$.

This assumption essentially means that we are assuming a priori that the set of objects cannot be partitioned into two sets where the objects in the first set are dominated by the objects in the second. Under this assumption, Ford shows that $h(\underset{\sim}{\pi} : \underset{\sim}{s}^o, \underset{\sim}{n}^o)$ has a unique mode $\underset{\sim}{\pi}^*$ in the interior of Ω, which is obtained as the solution to the equations

$$\pi_i^* \sum_{j=1}^{t} n_{ij}^o / (\pi_i^* + \pi_j^*) = s_i^o, \quad i = 1, \cdots, t.$$

If we now let $w^*_{ij} = n^0_{ij}\pi^*_i/(\pi^*_i + \pi^*_j)$ it follows that $\underset{\sim}{w}^0$ and $\underset{\sim}{w}^*$ are equivalent prior parameters. Thus the specified $\underset{\sim}{w}^0$ can be interpreted as a specification of $\underset{\sim}{n}^0$ and the preference probabilities $\pi^*_i/(\pi^*_i + \pi^*_j)$, which correspond to the a priori most probable $\underset{\sim}{\pi}$. Furthermore, s^0_i can be interpreted as the total number of preferences for the ith object in the conceptual experiment which corresponds to $\underset{\sim}{n}^0$ and the prior mode $\underset{\sim}{\pi}^*$, $i = 1, 2, \cdots, t$.

It might seem that the Dirichlet distribution with density

$$p_{\underset{\sim}{\pi}}(\underset{\sim}{\pi}) = D^{-1}(a_1, \cdots, a_t)\pi_1^{a_1-1} \cdots \pi_t^{a_t-1}, \quad \underset{\sim}{\pi} \in \Omega, \ a_i > -1$$

where

$$D(a_1, \cdots, a_t) = \underset{i}{\Pi} \ \Gamma(a_i)/\Gamma(\underset{i}{\Sigma}a_i)$$

would provide a reasonable family of prior distributions. This corresponds to a member of the conjugate family only if $\underset{i}{\Sigma}a_i = t$, in which case $\underset{\sim}{s}^0 = \underset{\sim}{a} - \underset{\sim}{1}$ and $\underset{\sim}{n}^0 = \underset{\sim}{0}$. Moreover, $\underset{\sim}{w}^0 \equiv \underset{\sim}{0}$ is the only nonnegative $\underset{\sim}{w}^0$ which yields a Dirichlet prior, namely $\underset{\sim}{a} \equiv \underset{\sim}{1}$, the prior which is uniform on Ω.

An interesting feature of the conjugate prior (3.5), however, is that it can be expressed as an infinite mixture of Dirichlet distributions. To exhibit this we note that if $m \geq 0$, $\underset{\sim}{u} = (u_1, \cdots, u_t)$ is a vector of nonnegative integers, and $\mathcal{U}_{ij} = \{\underset{\sim}{u} : u_i = u_j = 0\}$, then for the pair (i,j)

$$(\pi_i + \pi_j)^{-m} = \sum_{u=0}^{\infty} \binom{m + u - 1}{u} \left(\sum_{k \neq i,j} \pi_k\right)^u$$

$$= \sum_{\underset{\sim}{u} \in \mathcal{U}_{ij}} \binom{m + u - 1}{u} \binom{u}{u_1, \cdots, u_t} \pi_1^{u_1} \cdots \pi_t^{u_t}$$

(3.6)

where the last equality follows from the multinomial expansion of $\left(\sum_{k \neq i,j} \pi_k\right)^u$ when modified to include π_i and π_j with zero exponents. In that we are to consider the

product of the expansions (3.6) over all (i,j) with $i < j$, we will identify the m, u, and $\underset{\sim}{u}$ which correspond to the pair (i,j) by m_{ij}, $u_{\cdot,ij}$, and $\underset{\sim}{u}_{ij}$, respectively. Each term in the expansion of $\underset{i<j}{\Pi} (\pi_i + \pi_j)^{-m_{ij}}$ is the product of $\binom{t}{2}$ terms, one from each of the sums (3.6) which correspond to (i,j), $i < j$. If we let $\mathscr{G} = \underset{i<j}{X}\, \mathcal{U}_{ij}$, then each term in the expansion corresponds to a $g = (\underset{\sim}{u}_{12}, \underset{\sim}{u}_{13}, \cdots, \underset{\sim}{u}_{(t-1)t}) \in \mathscr{G}$.

For each vector $\underset{\sim}{\ell} = (\ell_1, \cdots, \ell_t)$ of nonnegative integers, let $\mathscr{G}_{\underset{\sim}{\ell}} = \{g : g \in \mathscr{G},\ \underset{i<j}{\Sigma}\, \underset{\sim}{u}_{ij} = \underset{\sim}{\ell}\}$. The sets $\mathscr{G}_{\underset{\sim}{\ell}}$ define a partition of $\mathscr{G}$. Thus

$$\underset{i<j}{\prod}{}' (\pi_1 + \pi_j)^{-m_{ij}} = \sum_{\underset{\sim}{\ell}} c(\underset{\sim}{\ell},\underset{\sim}{m})\pi_1^{\ell_1} \cdots \pi_t^{\ell_t}$$

where

$$c(\underset{\sim}{\ell},\underset{\sim}{m}) = \sum_{g \in \mathscr{G}_{\underset{\sim}{\ell}}}\ \prod_{i<j} \binom{m_{ij} + u_{\cdot,ij} - 1}{u_{\cdot,ij}}\binom{u_{\cdot,ij}}{u_{1,ij}, \cdots, u_{t,ij}} \qquad (3.7)$$

and where $\displaystyle\sum_{\underset{\sim}{\ell}}$ denotes summation over all vectors of nonnegative integers.

We can now write $h(\underset{\sim}{\pi} : \underset{\sim}{r},\underset{\sim}{m})$ of (3.5) as

$$h(\underset{\sim}{\pi} : \underset{\sim}{r},\underset{\sim}{m}) = K(\underset{\sim}{r},\underset{\sim}{m}) \sum_{\underset{\sim}{\ell}} c(\underset{\sim}{\ell},\underset{\sim}{m})\pi_1^{r_1+\ell_1} \cdots \pi_t^{r_t+\ell_t} \qquad (3.8)$$

where

$$\left[K(\underset{\sim}{r},\underset{\sim}{m})\right]^{-1} = \sum_{\underset{\sim}{\ell}} c(\underset{\sim}{\ell},\underset{\sim}{m})\, D(r_1 + \ell_1 + 1, \cdots, r_t + \ell_t + 1). \qquad (3.9)$$

One property of the coefficients $c(\underset{\sim}{\ell},\underset{\sim}{m})$ which will be of importance in Section 4 is the following: if $m_{ij} = m$ for all (i,j), then $c\left(\underset{\sim}{\ell}, m(\underset{\sim}{1} - \underset{\sim}{I})\right) = c\left(\underset{\sim}{\ell}_\sigma, m(\underset{\sim}{1} - \underset{\sim}{I})\right)$ for any permutation $\underset{\sim}{\ell}_\sigma$ of $\underset{\sim}{\ell}$.

The posterior distribution for $\underset{\sim}{\Pi}$ is given by the density

$$p_{\underset{\sim}{\Pi}|\underset{\sim}{W}}(\underset{\sim}{\pi}|\underset{\sim}{w}) = f(\underset{\sim}{\pi} : \underset{\sim}{w}^1) = h(\underset{\sim}{\pi} : \underset{\sim}{s}^1, \underset{\sim}{n}^1), \quad \underset{\sim}{\pi} \in \Omega \qquad (3.10)$$

where $f(\underset{\sim}{\pi} : \cdot)$ and $h(\underset{\sim}{\pi} : \cdot, \cdot)$ are defined by (3.4) and (3.5), respectively. The parameters $\underset{\sim}{w}^1$ or $(\underset{\sim}{s}^1, \underset{\sim}{n}^1)$ of the posterior distribution are obtained by combining the data $\underset{\sim}{w}$ or $(\underset{\sim}{s}, \underset{\sim}{n})$ with the prior parameters $\underset{\sim}{w}^0$ or $(\underset{\sim}{s}^0, \underset{\sim}{n}^0)$. Specifically, $\underset{\sim}{w}^1 = \underset{\sim}{w} + \underset{\sim}{w}^0$, $\underset{\sim}{s}^1 = \underset{\sim}{s} + \underset{\sim}{s}^0$, and $\underset{\sim}{n}^1 = \underset{\sim}{n} + \underset{\sim}{n}^0$. Furthermore, $s_i^1 = \sum_{j=1}^{t} w_{ij}^1$ and $\underset{\sim}{n}^1 = \underset{\sim}{w}^1 + \underset{\sim}{w}^{1T}$. The vector $\underset{\sim}{s}^1$ will play a central role in the next section and will be referred to as the posterior scores.

4. ESTIMATION AND RANKING PROCEDURES

The selection of an estimator for the vector of worth parameters $\underset{\sim}{\pi}$ of the Bradley-Terry model is of central interest. Such an estimator can be used to estimate the preference probabilities as defined by (3.1). A second feature of an estimator of $\underset{\sim}{\pi}$ is that it can be used to provide a ranking of the set of objects. The properties of such rankings will be discussed, but first let us proceed to a discussion of two methods of estimation based on the posterior distribution (3.10).

In the classical analysis of data which are assumed to arise under the Bradley-Terry model, the method of maximum likelihood has been used to provide an estimator for $\underset{\sim}{\pi}$. In particular, Ford (1957) has shown that if the observed frequencies of preference $\underset{\sim}{w}$ satisfy the partitioning assumption stated in Section 3, then the likelihood $f(\underset{\sim}{\pi} : \underset{\sim}{w}) = h(\underset{\sim}{\pi} : \underset{\sim}{s}, \underset{\sim}{n})$ has a unique maximum in the interior of Ω. By analogy, we can use the mode $\underset{\sim}{p}^*(\underset{\sim}{w})$ of the posterior distribution $f(\underset{\sim}{\pi} : \underset{\sim}{w}^1) = h(\underset{\sim}{\pi} : \underset{\sim}{s}^1, \underset{\sim}{n}^1)$ to estimate $\underset{\sim}{\pi}$. The mode exists and is unique provided $\underset{\sim}{w}^1$ satisfies the partitioning assumption, a condition which holds for $\underset{\sim}{w}^1$ if it holds for either $\underset{\sim}{w}$ or $\underset{\sim}{w}^0$, and is obtained as a solution to the equations in the posterior parameters

$(\underset{\sim}{s}^1, \underset{\sim}{n}^1),$

$$p_i^* \sum_j n_{ij}^1 / (p_i^* + p_j^*) = s_i^1, \quad i = 1, \cdots, t. \tag{4.1}$$

Ford also describes an iterative scheme for solving the above equations which he

shows converges monotonically to the unique solution $\underset{\sim}{p}^*(\underset{\sim}{w})$. If the partitioning

assumption does not hold for $\underset{\sim}{w}^1$, we can partition the objects into several subsets

so that in each the assumption holds for $\underset{\sim}{w}$ and hence for $\underset{\sim}{w}^1$.

The second estimator for $\underset{\sim}{\pi}$ to be considered is the Bayes estimator under a

quadratic loss function, namely $\underset{\sim}{\bar{p}}(\underset{\sim}{w}) = E(\underset{\sim}{\Pi} | \underset{\sim}{W} = \underset{\sim}{w})$, the posterior mean of $\underset{\sim}{\Pi}$.

Setting $N^1 = \Sigma s_i^1 = \underset{i<j}{\Sigma} n_{ij}^1$ and $\ell_\bullet = \Sigma_i \ell_i$, it follows that

$$\bar{p}_j(\underset{\sim}{w}) = K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \underset{\underset{\sim}{\ell}}{\sum} C(\underset{\sim}{\ell}, \underset{\sim}{n}^1) \left[(s_j^1 + \ell_j + 1)/(N^1 + \ell_\bullet + t) \right]$$

$$\tag{4.2}$$

$$\cdot D(s_1^1 + \ell_1 + 1, \cdots, s_t^1 + \ell_t + 1), j = 1, \cdots, t,$$

where $C(\underset{\sim}{\ell}, \underset{\sim}{n}^1)$ and $K(\underset{\sim}{s}^1, \underset{\sim}{n}^1)$ are defined in (3.7) and (3.9) and where $\underset{\underset{\sim}{\ell}}{\sum}$ denotes

summation over all vectors of nonnegative integers. Note that $0 < \bar{p}_j(\underset{\sim}{w}) < 1$,

$j = 1, 2, \cdots, t$ and $\sum_{j=1}^{t} \bar{p}_j(\underset{\sim}{w}) \equiv 1$. Expression (4.2), although not in closed form,

will be used in examining properties of the ranking determined by $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$.

In that the worth parameters $\underset{\sim}{\pi}$ of the Bradley-Terry model determine the true

ranking of the objects, it is reasonable to use an estimator of $\underset{\sim}{\pi}$ to provide an

estimated ranking. In particular, we can use the rankings determined by $\underset{\sim}{p}^*(\underset{\sim}{w})$ and

$\underset{\sim}{\bar{p}}(\underset{\sim}{w})$, rankings which are meaningful for any set of posterior parameters $(\underset{\sim}{s}^1, \underset{\sim}{n}^1)$.

When $n_{ij}^1 = n$ for all (i,j), a condition which can be referred to as posterior

balance, the posterior scores $\underset{\sim}{s}^1$ also determine a meaningful ranking. The condition

of posterior balance depends only on the prior parameter $\underset{\sim}{n}^0$ and on the design $\underset{\sim}{n}$ and does not depend on the observed responses. If an experimenter, after specifying $\underset{\sim}{n}^0$, desires the same posterior information, as measured by $\underset{\sim}{n}^1$, about each pair (i,j), $i < j$, then he could select $\underset{\sim}{n}$ to achieve posterior balance. It will now be demonstrated that under posterior balance, the ranking determined by $\underset{\sim}{s}^1$ is a Bayes ranking for a wide class of loss functions, and furthermore that the rankings determined by $\underset{\sim}{s}^1$, $\underset{\sim}{p}^*(\underset{\sim}{w})$, and $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$ are identical. When posterior balance does not hold, the ranking determined by $\underset{\sim}{s}^1$ is obviously inappropriate, but the other two rankings are still meaningful.

A decision theoretic approach to the problem of ranking the objects in a balanced paired comparison experiment has been taken by Bühlmann and Huber (1963). When consideration is restricted to ranking procedures which are permutation invariant, they have shown that the ranking based on the observed scores $\underset{\sim}{s}$ is uniformly best, with respect to an "acceptable" loss function, if and only if the underlying probability structure is given by the Bradley-Terry model. An "acceptable" loss function has the property that the loss does not decrease when the ranking is made worse by interchanging two objects.

More specifically, let D denote the set of all possible rankings $\underset{\sim}{d} = (d_1, \cdots, d_t)$, where d_i is the rank of the ith object, and where $d_i < d_j$ means i is ranked ahead of j. A ranking in which there are no ties is represented by a $\underset{\sim}{d}$ which is a permutation of the integers $(1, 2, \cdots, t)$. In representing a ranking with ties, the rank of each object in a tied group is the average of the ranks associated with the tied group. Clearly, the number of rankings in D is finite. For every permutation $\sigma \in \Sigma$, the group of permutations of the integers $(1, 2, \cdots, t)$, let $\underset{\sim}{d}_\sigma = (d_{\sigma(1)}, \cdots, d_{\sigma(t)})$ and $\underset{\sim}{\pi}_\sigma = (\pi_{\sigma(1)}, \cdots, \pi_{\sigma(t)})$. For any specific pair (i,j), let α be the permutation which interchanges the ith and j^{th} elements of $\underset{\sim}{d}$; that is, $d_{\alpha(i)} = d_j$, $d_{\alpha(j)} = d_i$, and $d_{\alpha(k)} = d_k$ when $k \neq i,j$.

Definition: A loss function $L(\underset{\sim}{\pi},\underset{\sim}{d})$ defined on $\Omega \times D$ is said to be <u>acceptable</u> if

(i) $L(\underset{\sim}{\pi},\underset{\sim}{d}) = L(\underset{\sim}{\pi}_\sigma,\underset{\sim}{d}_\sigma)$ for every permutation $\sigma \in \Sigma$,

(ii) $L(\underset{\sim}{\pi},\underset{\sim}{d}) \leq L(\underset{\sim}{\pi},\underset{\sim}{d}_\alpha)$ when $\pi_i \leq \pi_j$ and $d_i \geq d_j$.

When the prior structure of Section 3 is imposed, the arguments presented by Bühlmann and Huber (1963) can be modified to give the following:

<u>Theorem 1.</u> <u>If</u> $n^1_{ij} = n^1$ <u>for</u> <u>all</u> <u>pairs</u> (i,j), <u>then</u> <u>for</u> <u>any</u> <u>acceptable</u> <u>loss</u> <u>function</u> $L(\underset{\sim}{\pi},\underset{\sim}{d})$ <u>the</u> <u>ranking determined by the posterior scores</u> $\underset{\sim}{s}^1 = \underset{\sim}{s} + \underset{\sim}{s}^0$ <u>is Bayes with</u> <u>respect</u> <u>to</u> <u>the</u> <u>prior</u> $h(\underset{\sim}{\pi} : \underset{\sim}{s}^0,\underset{\sim}{n}^0)$.

<u>Proof</u>: The set D of nonrandomized rankings is finite so that there exists a ranking $d^* \in D$ which minimizes the Bayes risk or equivalently the posterior risk

$$r(\underset{\sim}{d}: \underset{\sim}{s}^1,\underset{\sim}{n}^1) \equiv K(\underset{\sim}{s}^1,\underset{\sim}{n}^1) \int_\Omega L(\underset{\sim}{\pi},\underset{\sim}{d}) \prod_{i=1}^{t} \pi_i^{s^1_i} \Big/ \prod_{i<j} (\pi_i + \pi_j)^{n^1} (d\underset{\sim}{\pi}). \tag{4.3}$$

If we let $\Omega_\sigma = \{\underset{\sim}{\pi}: \pi_{\sigma(1)} \leq \pi_{\sigma(2)} \leq \cdots \leq \pi_{\sigma(t)}\}$, where by convention we take $\sigma(i) < \sigma(i+1) < \cdots < \sigma(i+k)$ when for some $k, \pi_{\sigma(i)} = \cdots = \pi_{\sigma(i+k)}$, then $\{\Omega_\sigma, \sigma \in \Sigma\}$ constitutes a partition of Ω into $t!$ disjoint convex regions. Let Ω_e correspond to the identity $e = (1, \cdots, t)$ of Σ. For each $\sigma \in \Sigma$, $\underset{\sim}{\pi} \in \Omega_\sigma$ implies $\underset{\sim}{\eta} \equiv \underset{\sim}{\pi}_\sigma \in \Omega_e$. If $\nu = \sigma^1$, then applying the transformation σ to Ω_σ gives

$$r(\underset{\sim}{d}: \underset{\sim}{s}^1,\underset{\sim}{n}^1) = K(\underset{\sim}{s}^1,\underset{\sim}{n}^1) \sum_\sigma \int_{\Omega_\sigma} L(\underset{\sim}{\pi},\underset{\sim}{d}) \prod_{i=1}^{t} \pi_i^{s^1_i} \Big/ \prod_{i<j} (\pi_i + \pi_j)^{n^1} (d\underset{\sim}{\pi})$$

$$= K(\underset{\sim}{s}^1,\underset{\sim}{n}^1) \sum_\sigma \int_{\Omega_e} L(\underset{\sim}{\eta}_\nu,\underset{\sim}{d}) \prod_{i=1}^{t} \eta_{\nu(i)}^{s^1_i} \Big/ \prod_{i<j} (\eta_i + \eta_j)^{n^1} (d\underset{\sim}{\eta}) \tag{4.4}$$

$$= K(\underset{\sim}{s}^1,\underset{\sim}{n}^1) \int_{\Omega_e} \sum_\nu L(\underset{\sim}{\eta}_\nu,\underset{\sim}{d}) \prod_{i=1}^{t} \eta_{\nu(i)}^{s^1_i} \Big/ \prod_{i<j} (\eta_i + \eta_j)^{n^1} (d\underset{\sim}{\eta})$$

where the last equality follows in that $\sum\limits_{\sigma}$ and $\sum\limits_{\nu}$ denote summation over all permutations in Σ.

Consider the fixed but arbitrary pair (i,j) and let $\underset{\sim}{d}_\alpha$ be the ranking obtained from $\underset{\sim}{d}$ by interchanging the ith and j^{th} elements. Noting that $\alpha^1 = \alpha$ and using the invariance $\big($condition (i)$\big)$ of an acceptable loss function, we obtain

$$r(\underset{\sim}{d}: \underset{\sim}{s}^1, n^1) = K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \int\limits_{\Omega_e} \sum\limits_{\nu} L(\underset{\sim}{\eta}_{\nu\alpha}, \underset{\sim}{d}) \prod\limits_{i=1}^{t} \eta_{\nu\alpha(i)}^{s_i^1} \Big/ \prod\limits_{i<j} (\eta_i + \eta_j)^{n^1} (d\underset{\sim}{\eta})$$

$$\tag{4.5}$$

$$= K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \int\limits_{\Omega_e} \sum\limits_{\nu} L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}_\alpha) \prod\limits_{i=1}^{t} \eta_{\nu\alpha(i)}^{s_i^1} \Big/ \prod\limits_{i<j} (\eta_i + \eta_j)^{n^1} (d\underset{\sim}{\eta}).$$

Combining the last expressions of (4.4) and (4.5), it follows as in Bühlmann and Huber (1963, p. 509) that

$$r(\underset{\sim}{d}: \underset{\sim}{s}^1, n^1) = \tfrac{1}{2} K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \int\limits_{\Omega_e} \sum\limits_{\nu} \Big[L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}) \eta_{\nu(i)}^{s_i^1} \eta_{\nu(j)}^{s_j^1} + L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}_\alpha) \eta_{\nu(j)}^{s_i^1} \eta_{\nu(i)}^{s_j^1} \Big]$$

$$\tag{4.6}$$

$$\cdot \prod\limits_{k \neq i,j} \eta_{\nu(k)}^{s_k^1} \Big/ \prod\limits_{i<j} (\eta_i + \eta_j)^{n^1} (d\underset{\sim}{\eta}).$$

Now suppose that $s_i^1 \geq s_j^1$ and $d_i \leq d_j$ (so that i is ranked ahead of j). For each $\underset{\sim}{\eta} \in \Omega_e$ and each $\nu \in \Sigma$ we have

$$L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}) \geq L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}_\alpha) \quad \text{and} \quad \eta_{\nu(i)}^{s_i^1} \eta_{\nu(j)}^{s_j^1} \leq \eta_{\nu(j)}^{s_i^1} \eta_{\nu(i)}^{s_j^1} \quad \text{when} \quad \nu(i) < \nu(j)$$

$$L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}) \leq L(\underset{\sim}{\eta}_\nu, \underset{\sim}{d}_\alpha) \quad \text{and} \quad \eta_{\nu(i)}^{s_i^1} \eta_{\nu(j)}^{s_j^1} \geq \eta_{\nu(j)}^{s_i^1} \eta_{\nu(i)}^{s_j^1} \quad \text{when} \quad \nu(i) > \nu(j)$$

where the inequalities on $L(\underset{\sim}{\eta}, \underset{\sim}{d})$ follow from condition (ii) of an acceptable loss function. The other inequalities follow from the fact that if $a_2 \geq a_1 > 0$ and

$b_2 \geq b_1$ then $a_1^{b_1} a_2^{b_2} \geq a_1^{b_2} a_2^{b_1}$, or equivalently, by use of logarithms, $a_1 b_1 + a_2 b_2 \geq a_1 b_2 + a_2 b_1$. By applying the latter fact to the bracketed term in (4.6) for each $\eta \in \Omega_e$ and $\nu \in \Sigma$, we conclude that $r(\underset{\sim}{d} : \underset{\sim}{s}^1, n^1) \leq r(\underset{\sim}{d}_\alpha : \underset{\sim}{s}^1, n^1)$.

It then follows that if $sgn(s_i^1 - s_j^1) = sgn(d_j^* - d_i^*)$ for each pair (i,j), then the ranking $\underset{\sim}{d}^*$ minimizes the posterior risk. That is, the ranking determined by $\underset{\sim}{s}^1$ is Bayes.

The equivalence of the rankings determined by $\underset{\sim}{s}^1$ $\underset{\sim}{p}^*(\underset{\sim}{w})$, and $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$ will now be established.

Theorem 2. If $n_{ij}^1 = n^1$ for all pairs (i,j), then the rankings determined by the posterior mode $\underset{\sim}{p}^*(\underset{\sim}{w})$ and the posterior mean $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$ are identical with the Bayes ranking determined by the posterior scores $\underset{\sim}{s}^1$.

Proof: (i) posterior mode $\underset{\sim}{p}^*(\underset{\sim}{w})$. As in Ford (1957), it follows from (4.1) that

$$s_i^1 - s_j^1 = n^1 \left[p_i^* \sum_{k \neq i} (p_i^* + p_k^*)^1 - p_j^* \sum_{k \neq j} (p_j^* + p_k^*)^1 \right]$$

$$= n^1 (p_i^* - p_j^*) \sum_{k=1}^{t} p_k^* \Big/ (p_i^* + p_k^*)(p_j^* + p_k^*).$$

Thus, $sgn(p_i^* - p_j^*) = sgn(s_i^1 - s_j^1)$ for each (i,j) and the rankings determined by $\underset{\sim}{p}^*(\underset{\sim}{w})$ and $\underset{\sim}{s}^1$ are identical.

(ii) posterior mean $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$. From (4.2) we obtain

$$\bar{p}_i - \bar{p}_j = (s_i^1 - s_j^1) K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \sum_{\underset{\sim}{\ell}} C(\underset{\sim}{\ell}, \underset{\sim}{n}^1)(N^1 + \ell_\cdot + t)^1 D(s_1^1 + \ell_1 + 1, \cdots, s_t^1 + \ell_t + 1)$$

$$\tag{4.7}$$

$$+ K(\underset{\sim}{s}^1, \underset{\sim}{n}^1) \sum_{\underset{\sim}{\ell}} C(\underset{\sim}{\ell}, \underset{\sim}{n}^1)(N^1 + \ell_\cdot + t)^1 (\ell_i - \ell_j) D(s_1^1 + \ell_1 + 1, \cdots, s_t^1 + \ell_t + 1).$$

If $\underset{\sim}{n}^1 = n^1(\underset{\sim}{1} - \underset{\sim}{I})$, then $C(\underset{\sim}{\ell},\underset{\sim}{n}^1) = C(\underset{\sim}{\ell}_\alpha,\underset{\sim}{n}^1)$ where $\underset{\sim}{\ell}_\alpha$ is obtained from $\underset{\sim}{\ell}$ by interchanging the ith and j^{th} elements. Then by pairing terms which arise from $\underset{\sim}{\ell}$ and $\underset{\sim}{\ell}_\alpha$, the second member of (4.7) becomes

$$K(\underset{\sim}{s}^1,\underset{\sim}{n}^1) \sum_{\substack{\underset{\sim}{\ell} \\ \ell_i > \ell_j}} C(\underset{\sim}{\ell},\underset{\sim}{n}^1) \prod_{k \neq i,j} \frac{\Gamma(s_k^1 + \ell_k + 1)}{\Gamma(N^1 + \ell_. + t + 1)} \gamma(s_i^1,s_j^1,\ell_i,\ell_j)$$

where

$$\gamma(s_i^1,s_j^1,\ell_i,\ell_j) = (\ell_i - \ell_j)\left[\Gamma(s_i^1 + \ell_i + 1)\Gamma(s_j^1 + \ell_j + 1) - \Gamma(s_i^1 + \ell_j + 1)\Gamma(s_j^1 + \ell_i + 1)\right] .$$

Noting that $g(x) = \ell n\ \Gamma(x)$ is continuous and convex, it follows from inequality #363 of Hardy, Littlewood, and Polya (1934, p. 261) that for each term in the sum $(\ell_i > \ell_j)$,

$$\text{sgn}\left[\gamma(s_i^1,s_j^1,\ell_i,\ell_j)\right] = \text{sgn}(s_i^1 - s_j^1).$$

Thus, $\text{sgn}(\bar{p}_i - \bar{p}_j) = \text{sgn}(s_i^1 - s_j^1)$ for each (i,j) and the rankings determined by $\underset{\sim}{\bar{p}}(\underset{\sim}{w})$ and $\underset{\sim}{s}^1$ are identical.

5. DISCUSSION

We have formulated a Bayesian approach to paired comparison experimentation under the multi-binomial and Bradley-Terry models and have provided two estimators of the model parameters (assuming a natural conjugate prior distribution) for each model. Furthermore, in the Bradley-Terry case, if the experimental design and prior distribution satisfy the criterion of posterior balance, then for a large class of acceptable loss functions, a Bayes ranking of the worth parameters is explicitly calculated and shown to coincide with the ranking determined by each of

the two estimators. Thus, if the objective is ranking the worth parameters by their estimates, which are not easily calculated, one can easily calculate the Bayes ranking with the assurance that it agrees with the ranking provided by these estimates.

An interesting and unsolved problem is the nature of the Bayes ranking when posterior balance does not obtain, and the extent to which this ranking agrees with those determined by the estimators given here of the parameters of the Bradley-Terry model. A second problem of interest is the formulation of optimality criteria for the design parameters $\underset{\sim}{n}$ of the paired comparison experiment. One possibility, when estimation is the goal, is to choose an $\underset{\sim}{n}$ which tends to yield a tight posterior distribution, and is subject to a constraint such as $\sum_{i<j} n_{ij} = N$. Another is to propose a cost associated with each statement of preference and minimize the sum of this cost and the Bayes risk of the Bayes rule for the given loss function.

REFERENCES

BRADLEY, R. A. (1965). Another interpretation of a model for paired comparisons. Psychometrika 30, 315-18.

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

BÜHLMANN, H. and HUBER, P. J. (1963). Pairwise comparison and ranking in tournaments. Ann. Math. Statist. 34, 501-10.

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

FORD, L. R., Jr. (1957). Solution of a ranking problem from binary comparisons. Am. Math. Monthly 54, 28-33.

HARDY, G. H., LITTLEWOOD, J. E., and POLYA, G. (1934). Inequalities. London: Cambridge University Press.

LUCE, R. D. (1961). A choice theory analysis of similarity judgments. Psychometrika 36, 151-63.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Boston: Harvard University Press.

WINKLER, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *J. Am. Statist. Ass.* 62, 776-800.

ZERMELO, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximum problem der Wahrscheinlichkeitsrechnung. *Math. Zeit.* 29, 436-60.