PAIRWISE COMPARISON AND RANKING: OPTIMUM PROPERTIES OF THE ROW SUM PROCEDURE

BY PETER J. HUBER[1]

University of California, Berkeley

1. Introduction and summary. The present paper is concerned with those pairwise comparison experiments for which ranking in descending order of the "row sums" is the appropriate ranking procedure.

More precisely, assume that n items are compared pairwise, omitting no pairs, and that the result of the comparison between item i and j is expressed as a real number x_{ij} satisfying

$$x_{ij} + x_{ji} = 0 \qquad \text{for all } i, j. \tag{1}$$

x_{ij} might be some measured difference between i and j, or it might take the values 1, 0, -1 according as item i is judged to be superior, equal or inferior to item j, or it might be a statistic summarizing the results of several comparisons between the two items, etc.

It will be shown that ranking in descending order of the scores

$$s_i = \sum_{j=1}^{n} x_{ij} \tag{2}$$

(row sum procedure) uniformly minimizes the risk among all permutation invariant procedures, and for all "reasonable" loss functions, provided the x_{ij} ($i < j$) are independent random variables distributed according to an exponential distribution of the type

$$P(x_{ij} \leqq t) = c(\vartheta_i - \vartheta_j) \int_{-\infty}^{t} e^{(\vartheta_i - \vartheta_j)\tau} \mu(d\tau), \tag{3}$$

where μ is a symmetric probability measure on the real line. Then $s = (s_1, \cdots, s_n)$ is a sufficient statistic for the joint distribution of the x_{ij}. Under suitable regularity conditions, the converse also holds: if the distributions of the x_{ij} are not of this form, then s is not sufficient, and the row sum procedure is not optimal.

Thus, the model just described seems to constitute the natural domain of the row sum ranking procedure. Fortunately, this domain contains the important case where the x_{ij} ($i < j$) are independent normal variables with mean $\vartheta_i - \vartheta_j$ and equal variance σ^2. Another particular case—the case of tournaments without draws—where the x_{ij} can take only two values, has been treated in the joint paper [1]. Incidentally, the present paper grew out of an attempt to cover the case of tournaments with draws (see example (iv) in Section 4 below).

Received November 27, 1962.

[1] This research was performed while the author was a Fellow of the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California, Berkeley.

Although one cannot expect strict optimality properties, one may still ask whether the row sum procedure has "nice" (e.g., minimax) properties even outside the above model. A modest result in this direction is the following. Take a particular joint distribution of the x_{ij} belonging to the above model, and consider the class of all joint distributions of the x_{ij} that lead (up to permutations) to the same joint distribution of the s_i as this particular one. Then this latter will be least favorable for the ranking problem, and the row sum procedure will have minimax properties in each such class (modified minimax principle of Wesler [2]). For instance, if the x_{ij} $(i < j)$ are independent normal variables with mean ϑ_{ij} and equal variance, then the joint distribution of the s_i depends only on the $\vartheta_i = (1/n) \sum_{j=1}^{n} \vartheta_{ij}$, and the linear model $\vartheta'_{ij} = \vartheta_i - \vartheta_j$ is least favorable, if the ϑ_i are kept fixed (up to permutations).

There is an immediate generalization of the results to the case of comparison in k-tuples instead of in pairs, as indicated in the last section of this paper.

2. Setup of the decision problem. We have to specify a sample space, a set of theories, a set of decisions and a loss function:

The sample space $\mathfrak{X}$ is the $n(n - 1)/2$-dimensional Euclidean space with coordinates x_{ij} $(i < j)$; we agree to put $x_{ji} = -x_{ij}$ and $x_{ii} = 0$ according to (1). Let the group $\mathfrak{S}_n$ of all permutations of n items act on $\mathfrak{X}$ such that the permutation $\sigma \in \mathfrak{S}_n$ transforms a point $x \in \mathfrak{X}$ into the point x^σ having coordinates

$$x_{ij}^\sigma = x_{\sigma(i)\sigma(j)} . \tag{4}$$

In the sequel, invariance always means invariance with respect to this group $\mathfrak{S}_n$.

Let $\mathfrak{Y}$ be the space of the statistic s; $\mathfrak{Y}$ is the $(n - 1)$-dimensional hyperplane in n-space defined by $\sum s_i = 0$. Put $s_i^\sigma = s_{\sigma(i)}$.

If the x_{ij} $(i < j)$ are independent and distributed according to (3), and if $\nu = \prod_{i<j} \mu$ is the product probability measure on $\mathfrak{X}$, then the joint distribution of the x_{ij} is of the form

$$dP_\vartheta = c(\vartheta) \exp \left(\sum_i \vartheta_i s_i\right) d\nu \tag{5}$$

as a straightforward computation shows. Define $\vartheta_i^\sigma = \vartheta_{\sigma(i)}$, then the norming factor $c(\vartheta) = \prod_{i<j} c(\vartheta_i - \vartheta_j)$ is invariant, since ν is.

As set of theories we take the family Θ consisting of the $n!$ joint distributions on $\mathfrak{X}$ generated by applying $\mathfrak{S}_n$ to a particular distribution (5). Obviously, $s = (s_1 , \cdots , s_n)$ constitutes a sufficient statistic for Θ (as well as for the whole parametric family of distributions depending on the parameter ϑ). Denote by Q_ϑ the joint distribution of s.

The decision space D is the set of all possible rankings. Let $\mathfrak{S}_n$ act on D as follows: if, under ranking d, item i has rank d_i , then, under ranking d^σ, it will have rank $d_i^\sigma = d_{\sigma(i)}$.

It will be assumed that the "true" ranking is in descending order of the ϑ_i , and that departures from the true ranking are punished by some real-valued loss;

denote by $L(\vartheta, d)$ the loss incurred when ϑ is the true value of the parameter and decision d is taken.

DEFINITION. A loss function L is called acceptable if it satisfies the following two conditions:

(i) L is invariant: for all ϑ, d and σ

$$L(\vartheta^\sigma, d^\sigma) = L(\vartheta, d).$$

(ii) L does not decrease if the ranking is made worse by interchanging two items. More precisely, assume that $\vartheta_i \geqq \vartheta_j$ and let α be the transposition that interchanges i and j. If d ranks item i before item j, $d_i < d_j$, then

$$L(\vartheta, d) \leqq L(\vartheta, d^\alpha).$$

Property (ii) implies in particular that ranking in descending order of the ϑ_i minimizes the loss, as it should do. (Starting from any ranking, one may decrease the loss by successively interchanging two items, until they are ranked in descending order of the ϑ_i.)

3. Solution of the decision problem. The following theorem is stated in a slightly more general form than we would really need:

THEOREM 1. *Assume that the joint distribution of the* s_i *is of the form* $Q_\vartheta(ds) = c(\vartheta)f(\vartheta, s)\lambda(ds)$, *where* λ *is some permutation invariant measure on* $\mathcal{Y}$, *and* f *is a density function satisfying:*

(i) $f(\vartheta^\sigma, s^\sigma) = f(\vartheta, s)$,

(ii) $f(\vartheta, s) \geqq f(\vartheta^\alpha, s)$ *whenever* $\vartheta_i \geqq \vartheta_j$, $s_i \geqq s_j$ *and* α *is the transposition interchanging items* i *and* j.

If L *is an acceptable loss function, then ranking in descending order of the* s_i, *breaking ties at random, has minimal risk among all invariant ranking procedures that depend on* x *only through* s.

REMARK 1. The assumptions of the above theorem are in particular satisfied if the x_{ij} $(i < j)$ are independent random variables distributed according to (3): Q_ϑ can then be written as $Q_\vartheta(ds) = c(\vartheta) \exp \{\sum_i \vartheta_i s_i\}\lambda(ds)$. Invariance of λ follows from invariance of ν, (i) is obvious, and (ii) follows from the identity $(s_i\vartheta_i + s_j\vartheta_j) - (s_i\vartheta_j + s_j\vartheta_i) = (s_i - s_j)(\vartheta_i - \vartheta_j)$, and monotonicity of exp. Since s then is sufficient, we have the

COROLLARY. *If the* x_{ij} $(i < j)$ *are independent random variables distributed according to* (3), *and the loss function* L *is acceptable, then the row sum ranking procedure has minimal risk among all invariant procedures.* (*Since* $\mathfrak{S}_n$ *is finite, this implies also admissibility among all procedures.*)

REMARK 2. The definition of acceptable loss functions is broad enough to cover cases where one is not interested in a complete ranking, but only, say, in selecting the item i with the highest ϑ_i. For instance, take the loss $L(\vartheta, d)$ to be 0 or 1 according as the item with the highest ϑ_i is ranked first or not, then L is acceptable, and Theorem 1 implies that selecting (= ranking first) the item i with the highest s_i minimizes the risk (in other words, maximizes the probability of selecting the best i).

PROOF OF THEOREM 1. Let φ be a randomized decision procedure, that is, a family of nonnegative measurable functions φ_d on $\mathcal{Y}$ indexed by $d \varepsilon D$, such that $\sum_{d \varepsilon D} \varphi_d = 1$. The risk then is

$$R(\vartheta, \varphi) = \int \sum_d L(\vartheta, d)\varphi_d(s)Q_\vartheta(ds).$$

Instead of minimizing $R(\vartheta, \varphi)$ among all permutation invariant procedures, it is more convenient to minimize the Bayes risk $\bar{R}(\vartheta, \varphi)$ corresponding to the uniform a priori distribution on the orbit $\mathfrak{S}_n(\vartheta) = \{\vartheta^\sigma \mid \sigma \varepsilon \mathfrak{S}_n\} = \Theta$. Since, for invariant procedures, the risk R is constant on orbits, this will lead to the same results. This Bayes risk computes as

$$\bar{R}(\vartheta, \varphi) = (1/n!) \sum_\sigma \int \sum_{d \varepsilon D} L(\vartheta^\sigma, d)\varphi_d(s)Q_\vartheta(ds) = \int \sum_d \varphi_d(s)a_d(s)\lambda(ds)$$

where $a_d(s) = [c(\vartheta)/n!]\sum_\sigma L(\vartheta^\sigma, d)f(\vartheta^\sigma, s)$.

Let $a_{\min}(s) = \inf_d a_d(s)$. Then a procedure φ will minimize $\bar{R}(\vartheta, \varphi)$ iff $\varphi_d(s) = 0$ whenever $a_d(s) > a_{\min}(s)$. Now I shall show that the minimum is reached, $a_d(s) = a_{\min}(s)$, if d corresponds to ranking in descending order of the s_i. More precisely, let α be the transposition that interchanges i and j, then I shall show that $a_d(s) \leqq a_{d^\alpha}(s)$ whenever $s_i \geqq s_j$ and $d_i < d_j$. This implies (similarly as above for L) that the minimum of $a_d(s)$ is reached for ranking in descending order of the s_i, and that it does not matter how ties between the s_i are dealt with. In order to prove this, compute

$$\begin{aligned} a_d(s) - a_{d^\alpha}(s) &= [c(\vartheta)/n!]\sum_\sigma f(\vartheta^\sigma, s)(L(\vartheta^\sigma, d) - L(\vartheta^\sigma, d^\alpha)) \\ &= [c(\vartheta)/n!]\sum_\sigma f(\vartheta^{\sigma\alpha}, s)(L(\vartheta^{\sigma\alpha}, d) - L(\vartheta^{\sigma\alpha}, d^\alpha)) \\ &= [c(\vartheta)/n!]\sum_\sigma f(\vartheta^{\sigma\alpha}, s)(L(\vartheta^\sigma, d^\alpha) - L(\vartheta^\sigma, d)). \end{aligned}$$

By taking the arithmetic mean of the first and the last of these three sums, one obtains

$$a_d(s) - a_{d^\alpha}(s) = [\tfrac{1}{2}c(\vartheta)/n!]\sum_\sigma (L(\vartheta^\sigma, d) - L(\vartheta^\sigma, d^\alpha))(f(\vartheta^\sigma, s) - f(\vartheta^{\sigma\alpha}, s)).$$

Now assume that $s_i \geqq s_j$ and $d_i < d_j$. Then the assumptions made about L and f imply that the two factors inside the sum have opposite signs, hence $a_d(s) - a_{d^\alpha}(s) \leqq 0$.

Thus, ranking in descending order of the s_i minimizes $\bar{R}(\vartheta, \varphi)$ among all procedures which depend on x only through s. Ties between the s_i may be treated in an arbitrary way, but if they are broken up at random (with equal probabilities), the corresponding φ is invariant. This terminates the proof.

Now I shall establish the curious fact that for independent x_{ij} the exponential distributions (3) are essentially the only distributions for which s is sufficient.

THEOREM 2. *Assume that there are at least* 3 *items, that the* x_{ij} $(i < j)$ *are inde-*

pendent random variables taking values in a fixed symmetric interval I (possibly the whole real line), whose distributions have strictly positive densities f_{ij} in I, either

(A) *with respect to Lebesgue measure, or*

(B) *with respect to counting measure on a symmetric lattice.*

Let $F(x) = \prod_{i<j} f_{ij}(x_{ij})$ be the joint density of the x_{ij}, and let $\mathcal{P}$ be the class consisting of the $n!$ densities related to F by a permutation of the n items.

If $s = (s_1, \cdots, s_n)$ is sufficient for $\mathcal{P}$, i.e., if $F(x) = G(s)H(x)$, a.e. *in I where G depends on x only through s and H is invariant under $\mathfrak{S}_n$, then the f_{ij} must have the form $f_{ij}(t) = c_{ij}h(t) \exp\{(\vartheta_i - \vartheta_j)t\}$,* a.e. *in I where $c_{ij} = c_{ji}$ and ϑ_i are some constants, and h is even: $h(-t) = h(t)$.*

Proof. Let I be the interval $(-m, m)$ and denote by I_u the translated interval $(-m - u, m - u)$. According to (1), define $f_{ji}(t) = f_{ij}(-t)$. Let x be a generic point of $\mathfrak{X}$ (a lattice point in case (B)), let i, j, k be a fixed triple of items, let u be a fixed real number (in case (B), u (and v below) must be lattice translations), and denote by x' the point having the same coordinates as x except for $x'_{ij} = x_{ij} + u$, $x'_{jk} = x_{jk} + u$, $x'_{ki} = x_{ki} + u$. To avoid an overabundance of indices, we shall write t, y, z instead of x_{ij}, x_{jk}, and x_{ki}, respectively.

We have $s(x) = s(x')$, and hence

$$\frac{F(x')}{F(x)} = \frac{f_{ij}(t')f_{jk}(y')f_{ki}(z')}{f_{ij}(t)f_{jk}(y)f_{ki}(z)} = \frac{H(x')}{H(x)} \qquad \text{a.e.} \tag{6}$$

for $t, y, z \in I \cap I_u$. $H(x')/H(x)$ is invariant under permutations. Hence, if we divide (6) by the corresponding expression with i and j interchanged, we obtain

$$\frac{f_{ij}(t+u)f_{ji}(t)}{f_{ji}(t+u)f_{ij}(t)} \frac{f_{jk}(y+u)f_{ik}(y)}{f_{ik}(y+u)f_{jk}(y)} \frac{f_{ki}(z+u)f_{kj}(z)}{f_{kj}(z+u)f_{ki}(z)} = 1 \qquad \text{a.e.} \tag{7}$$

The variable t does not occur outside the first factor in (7), hence this factor is equivalent to a constant depending on i, j and u:

$$\frac{f_{ij}(t+u)f_{ji}(t)}{f_{ji}(t+u)f_{ij}(t)} = \gamma_{ij}(u) \qquad \text{a.e. for} \qquad t \in I \cap I_u. \tag{8}$$

If one replaces u by v and t by $t + u$ in the above expression, one obtains

$$\frac{f_{ij}(t+u+v)f_{ji}(t+u)}{f_{ji}(t+u+v)f_{ij}(t+u)} = \gamma_{ij}(v) \qquad \text{a.e. for} \qquad t \in I \cap I_u \cap I_{u+v}. \tag{9}$$

Multiplying (8) and (9) yields the functional equation

$$\gamma_{ij}(u+v) = \gamma_{ij}(u)\gamma_{ij}(v) \tag{10}$$

valid for those pairs u, v for which $I \cap I_u \cap I_{u+v}$ has strictly positive measure. It may be concluded from (8) that γ_{ij} is a measurable function of u, thus (10) yields that it is exponential $\gamma_{ij}(u) = \exp(2\vartheta_{ij}u)$ for some constant ϑ_{ij}. Moreover, it follows that

$$f_{ij}(t)/f_{ji}(t) = \text{const.} \exp(2\vartheta_{ij}t) \qquad \textbf{a.e.} \tag{11}$$

If we replace t by $-t$, the left side of (11) is turned into its reciprocal, hence const. = 1, and similarly, it follows that $\vartheta_{ij} = -\vartheta_{ji}$.

Put $\varphi_{ij}(t) = [f_{ij}(t)f_{ji}(t)]^{\frac{1}{2}}$ then we have $\varphi_{ij}(t) = \varphi_{ji}(t) = \varphi_{ij}(-t)$ and $f_{ij}(t) = \varphi_{ij}(t) \exp(\vartheta_{ij}t)$ a.e. Insert this into (7) to obtain

$$\exp 2(\vartheta_{ij} + \vartheta_{jk} + \vartheta_{ki})u \cdot \frac{\varphi_{jk}(y+u)\varphi_{ik}(y)}{\varphi_{ik}(y+u)\varphi_{jk}(y)} \frac{\varphi_{ik}(z+u)\varphi_{jk}(z)}{\varphi_{jk}(z+u)\varphi_{ik}(z)} = 1 \tag{12}$$

$$\text{a.e.} \quad \text{for } y, z \in I \cap I_u .$$

I assert now that all three factors of (12) are separately a.e. equal to 1. Proof: y and z do not occur outside the second and third factor, respectively. Hence these factors do not really depend on y and z, and are a.e. reciprocal to each other. But, replacing z by $-z - u$ and using the symmetry properties of the φ's, one obtains that the second and the third factor are also a.e. equal, which proves the assertion.

Hence, the first factor yields $\vartheta_{ij} + \vartheta_{jk} + \vartheta_{ki} = 0$. If we put, e.g., $\vartheta_1 = 0$, $\vartheta_i = \vartheta_{i1}$ ($i \neq 1$), we may conclude that $\vartheta_{ij} = \vartheta_i - \vartheta_j$ for all i, j.

Furthermore, the second factor of (12) yields

$$\varphi_{ik}(y+u)/\varphi_{ik}(y) = \varphi_{jk}(y+u)/\varphi_{jk}(y) \qquad \text{a.e.}$$

This implies that the function $\varphi_{ik}(y+u)/\varphi_{ik}(y)$ is a.e. the same for all pairs (i, k) of indices. Hence, φ_{ij} must have the form $\varphi_{ij}(t) = c_{ij}h(t)$ a.e. for some constants $c_{ij} = c_{ji}$ and some function h satisfying $h(t) = h(-t)$.

Putting things together, we obtain the final result

$$f_{ij}(t) = c_{ij}h(t) \exp\{(\vartheta_i - \vartheta_j)t\} \qquad \text{a.e.}$$

It may be convenient to make this representation of f_{ij} unique by norming h and ϑ such that $\int h = 1$ and $\sum_i \vartheta_i = 0$.

Now I want to show that the row sum procedure is not optimal, if s is not sufficient, under the following regularity conditions slightly stronger than those of Theorem 2: The x_{ij} ($i < j$) are independent, and have (A) strictly positive continuous densities with respect to Lebesgue measure on the real line, or (B) strictly positive densities with respect to counting measure on a symmetric lattice containing 0. There are at least four items, and the loss is 0 for a unique "correct" and 1 for all "incorrect" rankings. In the case of exponential families of the type (3) we take, of course, "correct" ranking to be in descending order of the ϑ_i, but otherwise it might be defined in an arbitrary permutation invariant fashion.

The "correct" ranking then sets up a 1–1 correspondence between D and the set $\mathcal{P}$ (see Theorem 2) of joint distributions of the x_{ij}, and one may conclude from the first few steps of the proof of Theorem 1, that a ranking procedure is optimal among invariant procedures if and only if it selects with probability 1 a distribution $P \in \mathcal{P}$ having maximal a posteriori probability (for the uniform a priori distribution on $\mathcal{P}$).

Proposition. *Under the above assumptions, the following four statements are equivalent:*

(i) *s is sufficient for $\mathcal{P}$.*

(ii) *For every 4 items i, j, k, l and every t, y, z we have: $f_{ij}(t)f_{jk}(y)f_{ki}(z)f_{il}(z-t)\cdot f_{jl}(t-y)f_{kl}(y-z)$ is symmetric in i, j, k, l.*

(iii) *The row sum procedure is optimal among invariant procedures.*

(iv) *There is an optimal invariant procedure that depends only on s.*

Proof. We shall establish the following implications:

$$\text{(ii)} \Leftrightarrow \text{(i)} \Rightarrow \text{(iii)} \Rightarrow \text{(iv)} \Rightarrow \text{(ii)}.$$

(i) $\Rightarrow$ (ii): follows immediately from Theorem 2.

(ii) $\Rightarrow$ (i): (ii) implies (7), hence the conclusion of Theorem 2 holds, which implies sufficiency.

(i) $\Rightarrow$ (iii): is an immediate consequence of Theorem 2 and Theorem 1.

(iii) $\Rightarrow$ (iv): is obvious.

(iv) $\Rightarrow$ (ii): Assume that (ii) does not hold for some i, j, k, l, t, y, z. Consider the following sample point x: $x_{ij} = t$, $x_{jk} = y$, $x_{ki} = z$, $x_{il} = z - t$, $x_{jl} = t - y$, $x_{kl} = y - z$, $x_{rs} = 0$ for all other independent comparisons. Then $s = 0$. But the a posteriori distribution on $\mathcal{P}$ (for the uniform a priori distribution) is not uniform, since (ii) does not hold, hence some $P_0 \varepsilon \mathcal{P}$ will have an a posteriori probability strictly inferior to the maximal a posteriori probability. Since the f_{ij} are continuous, this will hold in some neighborhood

$$U = \{x' \,\|\, x'_{ij} - x_{ij} \,| < \epsilon \quad \text{for all} \quad i, j\} \quad \text{of} \quad x.$$

Hence, in U, an optimal procedure will almost never select the ranking d_0 corresponding to P_0. But this is impossible: let φ be an invariant procedure depending only on s, and let $V = \{s \,\|\, s_i \,| < \epsilon \text{ for all } i\}$. Because of invariance, $\{s \,\varepsilon\, V \mid \varphi_{d_0}(s) > 0\}$ cannot be a null set, hence also its counter image in U is not a null set, which establishes the contradiction.

A similar but simpler proof works also in the discrete case (B), provided 0 is contained in the symmetric lattice of possible values of x_{ij}. If 0 is not contained in the lattice, or if I is not the whole real line, the proposition is probably still true, although it has been proved so far only for the case where x_{ij} can take only two values (cf. [1]).

4. Examples.

(i) *Normal case.* Assume that the x_{ij} $(i < j)$ are independent normal random variables with mean $\vartheta_i - \vartheta_j$ and equal variance σ^2. This corresponds to the following model: each item has a performance y_i that is normal with mean ϑ_i and variance $\sigma^2/2$, and the x_{ij} are independent observations of $y_i - y_j$. The probability density of x_{ij} with respect to Lebesgue measure then is

$$f_{ij}(t) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{(\vartheta_i - \vartheta_j)^2}{2\sigma^2}\right\} \exp\left\{-\frac{t^2}{2\sigma^2}\right\} \exp\left\{\left(\frac{\vartheta_i}{\sigma^2} - \frac{\vartheta_j}{\sigma^2}\right) t\right\}.$$

Thus row sum ranking is optimal.

(ii) *A case of dependent* x_{ij}. Assume that the x_{ij} $(i < j)$ are multivariate normal with mean $\vartheta_i - \vartheta_j$, variance σ^2, cov $(x_{ij}, x_{kl}) = 0$ if all 4 indices are distinct, cov $(x_{ij}, x_{il}) = \tau$ for $j \neq l$. Here s is sufficient and Theorem 1 still applies. Instead of computing the joint distribution of the x_{ij}, it is probably simpler to prove this in the following way: if the constant k is chosen suitably, the variables $y_{ij} = x_{ij} + k(s_i - s_j)$ are independent normal with mean $(1 + kn)(\vartheta_i - \vartheta_j)$ and equal variance. Hence (cf. (i)) ranking in descending order of $\sum_j y_{ij} = (1 + kn)s_i$ is optimal. Now use the fact that x_{ij} and y_{ij} determine each other uniquely, and thus constitute equivalent experiments.

(iii) *Tournaments.* Assume that x_{ij} can take only the two values $+1$ and -1 with probabilities p_{ij} and $p_{ji} = 1 - p_{ij}$, respectively; row sum ranking is optimal for the Bradley-Terry model $p_{ij} = p_i/(p_i + p_j)$, because if $\vartheta_i = (\frac{1}{2}) \log p_i$, then

$$p_{ij} = 1/[1 + \exp - 2(\vartheta_i - \vartheta_j)] = \exp (\vartheta_i - \vartheta_j)/[2 \cosh (\vartheta_i - \vartheta_j)].$$

This case was treated in [1] under a slightly different notation.

(iv) *Tournaments with draws.* Assume that x_{ij} can take the values 1, 0, -1 with probabilities p_{ij}, q_{ij}, and p_{ji}, respectively, $p_{ij} + q_{ij} + p_{ji} = 1$. If $p_{ij}:q_{ij}:p_{ji} = \exp (\vartheta_i - \vartheta_j):k:\exp (\vartheta_j - \vartheta_i)$ for some fixed $k \geqq 0$ (in other words, $p_{ij}/p_{ji} = p_i/p_j$ and the probability of a draw is k times the geometric mean of p_{ij} and p_{ji}), then row sum ranking is optimal.

(v) *Binomial and similar cases.* More generally, assume that x_{ij} can take the values $t = 0, \pm 1, \cdots, \pm m$, with probabilities

$$P(x_{ij} = t) = c(\vartheta_i - \vartheta_j)h(t) \exp (\vartheta_i - \vartheta_j)t, \quad h(-t) = h(t),$$

(Case (iv) corresponds to $h(1) = h(-1) = 1, h(0) = k$). For instance, if x_{ij} is binomial:

$$P(x_{ij} = t) = \binom{2m}{m+t} p^{m+t}(1 - p)^{m-t} = (p(1 - p))^m \binom{2m}{m+t} \left(\frac{p}{1-p}\right)^t;$$

then row sum ranking is optimal, if $p = 1/(1 + \exp \{-(\vartheta_i - \vartheta_j)\})$.

5. Minimax properties. In the preceding sections, no general definition of the "true" ranking was given that would be applicable to any arbitrary underlying probability model for the x_{ij}; of course, any such definition would include some arbitrariness. For instance, if only two items are compared, should one consider item 1 to be better than item 2, if (a) the expectation, or if (b) the median of x_{12} is greater than 0?

For the situation considered in Theorem 1 however, ranking in descending (or ascending) order of the ϑ_i seems to constitute the natural true ranking. Another situation, where one has a natural true ranking, occurs if every two items are comparable in the sense that item i is considered to be better than item j, if for all $k \neq i, j$, x_{ik} is stochastically greater than x_{jk}, and x_{ij} is stochastically greater than x_{ji}.

One of the major difficulties encountered in establishing properties of our ranking procedure outside the situation of Theorem 1 lies just in the lack of a universally acceptable definition of "true" ranking.

However, one small but nevertheless important extension is readily accessible. Whenever we have the situation envisaged in Theorem 1, it seems reasonable to define "correct ranking" to be in descending order of the ϑ_i, even if s is not sufficient. Of course, there might be also some other reasonable true rankings (especially as we have left undefined the term "reasonable"), if s is not sufficient.

After having agreed upon this, we consider the class of all joint distributions of the x_{ij} that lead, up to a permutation, to the same fixed joint distribution of the s_i of the above form. If this class contains a distribution of the x_{ij} for which s is sufficient, then this distribution must be least favorable for the ranking problem, since our ranking procedure leads to a constant risk over the whole class and for this particular distribution the risk cannot be improved by utilizing the full information contained in the x_{ij}. Hence, ranking in descending order of the s_i has minimax properties for this class of distributions. The importance of this seemingly trivial remark may be inferred from the following example.

Assume that the x_{ij} $(i < j)$ are independent normal random variables with mean ϑ_{ij} and common variance 1. Put $x_{ji} = -x_{ij}$ and $\vartheta_{ji} = -\vartheta_{ij}$; furthermore, put $\vartheta_i = (1/n) \sum_j \vartheta_{ij}$. I assert now: *The joint distribution of s depends on the ϑ_{ij} only through the ϑ_i.*

Proof. Since s is multivariate normal, it suffices to compute expectations: $Es_i = n\vartheta_i$, and covariance matrix: var $(s_i) = n - 1$, cov $(s_i, s_j) = -1$ for $i \neq j$. Both depend only on the ϑ_i.

Hence, if the $\vartheta_i = (1/n) \sum_j \vartheta_{ij}$ are kept fixed up to permutations, then the row sum procedure leads to a fixed risk for all values of the ϑ_{ij}, and this risk cannot be improved by using the full information contained in the x_{ij} in case we have the linear model $\vartheta_{ij} = \vartheta_i - \vartheta_j$. In other words, the linear model is a least favorable configuration.

6. Performance of the row sum procedure. Once one has decided to use the row sum procedure, the expression for the risk as given in Section 3 can be simplified.

Define, for any n-tuple $t = (t_1, \cdots, t_n)$ of real numbers $L(\vartheta, t) = L(\vartheta, d)$ where d is ranking in descending order of the t_i. If there are ties between the t_i such that different rankings are possible, then define $L(\vartheta, t)$ to be the arithmetic mean of the corresponding losses. Then, the risk of the row sum procedure is $R(\vartheta) = E_\vartheta L(\vartheta, s)$.

In the particular case where the x_{ij} $(i < j)$ are independent normal variables with mean $\vartheta_i - \vartheta_j$ and equal variance, say 1, one can simplify further, because then the risk can be computed as $R(\vartheta) = E_\vartheta L(\vartheta, s) = E_\vartheta L(\vartheta, t)$ where the t_i are independent normal random variables with mean $n\vartheta_i$, and variance n.

Proof. Let $\bar{t} = (1/n)t_i$, and put $s_i' = t_i - \bar{t}$. Then s and s' have the same expectations and the same covariance matrix, hence they have the same joint distribution. Since $L(\vartheta, s') = L(\vartheta, t)$, the assertion follows.

7. Comparison in k-tuples. Assume that some judge looks at every k-tuple $T = \{i_1, \cdots, i_k\}$ of items and determines a "local" score $x_{T,i}$ of item i inside the k-tuple T, such that $\sum_{i \varepsilon T} x_{T,i} = 0$. For convenience, put $x_{T,i} = 0$ whenever $i \not\varepsilon T$, so we have the analogue of (1):

$$(1') \qquad \sum_{i=1}^{n} x_{T,i} = 0, \qquad \text{for all } T.$$

Put

$$(2') \qquad s_i = \sum_T x_{T,i}.$$

We are concerned with the optimality properties of ranking in descending order of the s_i.

Let $\mathcal{Z}$ be the $(k - 1)$-dimensional hyperplane $\sum_{j=1}^{n} z_j = 0$ in k-space, and let λ be a permutation invariant measure on $\mathcal{Z}$. Assume that the local scores $x_T = (x_{T,i_1}, \cdots, x_{T,i_k}) \varepsilon \mathcal{Z}$ have densities of the form

$$(3') \qquad f_T(x_T) = c_T(\vartheta) \exp \{\sum_{i \varepsilon T} \vartheta_i x_{T,i}\}$$

with respect to λ, and are independent of each other. Then, $s = (s_1, \cdots, s_n)$ forms a sufficient statistic. In fact, the joint distribution of the $x_{T,i}$ is of the form $c(\vartheta) \exp \{\sum_i \vartheta_i s_i\} \, d\nu$ where $c(\vartheta) = \prod_T c_T(\vartheta)$ and ν is the product measure $\prod_T \lambda_T$ on $\prod_T \mathcal{Z}_T$ (where $\lambda_T = \lambda$ and $\mathcal{Z}_T = \mathcal{Z}$ for all T). Both $c(\vartheta)$ and ν are permutation invariant, the operation of $\mathfrak{S}_n$ being defined in the obvious way. Hence, Theorem 1 applies, and ranking in descending order of the s_i is optimal.

REFERENCES

[1] Bühlmann, H. and Huber, P. J. (1963). Pairwise comparison and ranking in tournaments. *Ann. Math. Statist.* **34** 501–510.

[2] Wesler, O. (1959). Invariance theory and a modified minimax principle. *Ann. Math. Statist.* **30** 1–20.