JOURNAL OF MATHEMATICAL PSYCHOLOGY 35, 294–318 (1991)

Probability Models on Rankings

DOUGLAS E. CRITCHLOW, MICHAEL A. FLIGNER, AND JOSEPH S. VERDUCCI

The Ohio State University

This paper investigates many of the probability models on permutations that have been proposed in the statistical and psychological literature. The various models are categorized into the following general classes: (1) Thurstone order statistics models, (2) ranking models induced by paired comparisons, (3) ranking models based on distances between permutations, and (4) multistage ranking models. Several thematic properties of ranking models are introduced that provide the basis for a systematic study of each of the classes of models. These properties are label-invariance, reversibility, strong unimodality, complete consensus, and L-decomposability. Next, several important subclasses of the four general classes are explored, including a determination of the pairwise intersections of the different classes of models. The paper concludes with an illustration of many of the models on a set of ranked "word association" data.

1. INTRODUCTION

Suppose that k items are ranked from best (1) to worst (k) according to some nondeterministic process. For example, the observed ranking could represent an arbitrarily chosen child's preferences for $k=5$ types of crackers, or the order of finish of $k=8$ horses in tomorrow's race. If the items are labelled by the integers $1, ..., k$, then the ranking may be thought of as a *random permutation* of these integers. This paper investigates many of the probability models on permutations that have been proposed in the statistical and psychological literature, interrelates their properties, and compares the types of analyses that can be made with the different models.

The uniform distribution, under which all $k!$ permutations are equally likely, is the simplest probability model and has been studied widely because it corresponds to the distribution of rankings under various null hypotheses. However, Kendall (1950) remarks that the uniform model is often inadequate for many practical ranking situations, and that "it is necessary to find some way of specifying a population with a tractable number of parameters in the non-null case." The various approaches to constructing non-uniform models include methods based on (1) order statistics, (2) paired comparisons, (3) distances between permutations, and (4) stagewise decompositions of the ranking process.

Some widely used examples of non-uniform ranking models resulting from these

Reprint requests should be addressed to Dr. Douglas E. Critchlow, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210.

0022-2496/91 $3.00

methods include the *Thurstone* (1927)–*Mosteller* (1951)–*Daniels* (1950) *model*, which assigns to each ranking the probability of the corresponding ordering of independent normal random variables with a common variance but different means, and *Mallows'* (1957) *ϕ-model*, which assigns probability to a ranking in inverse relation to its tau-distance from some central ranking. These brief examples will be discussed in detail later, but at present should help to motivate the general development of ranking models with respect to certain thematic properties.

The thematic properties for probability models on rankings are formulated in Section 2. In Section 3, various approaches to model construction are studied in terms of these properties. Section 4 explores the extent to which the various approaches produce common models. In Section 5, many of the models are illustrated on a set of ranked data.

The focus of the present paper is mainly on probability models for rankings, rather than on choice probabilities or probability models for paired comparisons. These latter topics are considered only to the extent necessary to describe or develop certain classes of ranking models.

2. Properties of Ranking Models

In this section the properties of label-invariance, reversibility, strong unimodality, complete consensus, and L-decomposability are motivated and defined. Invariance and reversibility are properties associated with classes of probability models, whereas the other properties are associated with individual probability models. The basic notation is as follows.

Once the k items are labelled arbitrarily as item 1 to item k, a *ranking* of the k items corresponds to a permutation function π from $\{1, ..., k\}$ onto $\{1, ..., k\}$, where $\pi(i)$ is the rank assigned to item i, $i = 1, ..., k$. Note that in this formulation tied ranks are not permitted. An alternative description of any ranking π is given by the associated *ordering* of the k items, denoted by the bracketed vector $\langle \pi^{-1}(1), ..., \pi^{-1}(k) \rangle$, where $\pi^{-1}(r)$ is the item assigned rank r, $r = 1, ..., k$. For example, $\langle 2\ 3\ 1 \rangle$ denotes the ordering in which item 2 is ranked as best, item 3 is ranked as second best, and item 1 is ranked as worst. With composition of rankings defined by $(\pi \circ \sigma)(i) = \pi[\sigma(i)]$, the set S_k of all rankings on $\{1, ..., k\}$ constitutes the permutation group. A probability mass function $P(\pi)$, $\pi \in S_k$, represents a *probability model* on rankings, whereas $\mathscr{P}$ denotes a *class* of such models, usually indexed by a set of parameters.

Frequently used permutations are the identity $e = \langle 1, ..., k \rangle$ and the transposition τ_{ij} defined by $\tau_{ij}(i) = j$, $\tau_{ij}(j) = i$, and $\tau_{ij}(m) = m$ for all $m \neq i, j$. Note that $\pi \circ \tau_{ij}$ is the permutation that agrees with π except that the ranks assigned to items i and j are exchanged, while $\tau_{rs} \circ \pi$ is the permutation that agrees with π except that the item $\pi^{-1}(r)$ now receives rank s, and the item $\pi^{-1}(s)$ now receives rank r.

Label-invariance. Suppose that a probability model $P \in \mathscr{P}$ is appropriate for a certain analysis. A relabelling of the items generally induces another probability model. Since items are initially labelled arbitrarily, it is natural to include among $\mathscr{P}$ all models obtained from P by relabelling items. Formally, the class $\mathscr{P}$ is *label-invariant* if for every $P \in \mathscr{P}$, $\pi \in S_k$, and "relabelling permutation" $v \in S_k$, the probability model P_v defined by $P_v(\pi) = P(\pi \circ v)$ is also in $\mathscr{P}$. As is easily verified, all of the classes of models presented in this paper possess the label-invariance property.

Reversibility. In contrast to item labels, which are arbitrary, the ranks given to the items have a natural linear ordering from best to worst. However, it is sometimes reasonable to consider a reversal of the natural ordering, so that if a particular strategy of ranking items from best to worst produces a probability model $P \in \mathscr{P}$, then the same strategy applied to ranking the items from worst to best "should" produce a probability model in the same class $\mathscr{P}$. Formally, define the reversing permutation γ by $\gamma(j) = (k+1) - j$, $j = 1, ..., k$, so that $\gamma \circ \pi$ corresponds to the permutation that assigns rank 1 to the worst item, rank 2 to the next worst item, and so forth. The class $\mathscr{P}$ is *reversible* if for every $P \in \mathscr{P}$ and $\pi \in S_k$, P^γ defined by $P^\gamma(\pi) = P(\gamma \circ \pi)$ is also in $\mathscr{P}$. This concept of reversibility was introduced by Luce (1959), pp. 69–70), and further examined by Pendergrass and Bradley (1960) and Marley (1968, 1982). Unlike label-invariance, the property of reversibility is not shared by all classes of ranking models that have been proposed in the literature.

Strong Unimodality. A ranking π_0 is a *modal ranking* if it uniquely maximizes P. A probability model on rankings is strongly unimodal if it has a modal ranking π_0, and the probability $P(\pi)$ is nonincreasing as π moves farther from π_0 along a certain type of path, defined as follows: A step along such a path moves from an arbitrary permutation π to the permutation $\pi \circ \tau_{ij}$, where items i and j are assigned adjacent ranks by π. The step is away from π_0 if π and π_0 initially agree on the relative ordering of items i and j, and toward π_0 otherwise. Formally, a model is *strongly unimodal* with modal ranking π_0, if for every pair of items i and j such that $\pi_0(i) < \pi_0(j)$ and any permutation π such that $\pi(i) = \pi(j) - 1$, $P(\pi) \geqslant P(\pi \circ \tau_{ij})$ with strict inequality in the case that $\pi = \pi_0$.

Complete Consensus. In many models, the items may be ordered in a stronger sense than that implied by either the existence of a modal ranking π_0 or even strong unimodality. For any two items i and j, suppose that $\pi_0(i) < \pi_0(j)$, and that $P[\pi(i) < \pi(j)]$, the probability that item i is preferred to item j, exceeds 0.5. If this probability continues to exceed 0.5 given any fixed assignment of ranks to the other $k-2$ items, then item i is *strongly preferred* to item j. If every pair of items can be related by strong preference, then the population has a complete consensus.

Formally, a model has the property of *complete consensus* with consensus ordering π_0, if for every pair of items i and j such that $\pi_0(i) < \pi_0(j)$, and any permutation π such that $\pi(i) < \pi(j)$, $P(\pi) \geqslant P(\pi \circ \tau_{ij})$ with strict inequality in the case that $\pi = \pi_0$. From the definitions, it follows immediately that complete consensus implies strong unimodality. The notion of complete consensus is discussed by Henery (1981) and Fligner and Verducci (1988).

Whenever a model exhibits strong unimodality or complete consensus, label-invariance implies that the modal ranking π_0 may be set equal to the identity permutation e, without loss of generality. This convention often simplifies notation.

L-decomposability. The idea of L-decomposability is motivated by Luce's (1959) ranking postulate, defined in (1) below, coupled with the notion of choice probabilities. In the ranking postulate, the ranking process is assumed to be composed of a sequence of stages: at stage 1 the most preferred item is chosen, at stage 2 the most preferred of the remaining items is chosen, and so forth until a full ranking is produced. The probability of the ensuing ranking is assumed to be the product of the choice probabilities across the various stages. The choice probabilities at the rth stage must depend only on the set of items remaining at that stage, and not on the relative ordering of the $r-1$ items previously selected.

Formally, choice probabilities constitute a family of probability measures $\{P_B\}$ where B ranges over all subsets of the item labels $\{1, ..., k\}$. For each $i \in B$, $P_B(i)$ is the probability that item i is chosen as the most preferred item among those listed in B. With this notation, Luce's ranking postulate states that for every ranking π with associated ordering $\langle i_1, ..., i_k \rangle$,

$$P(\pi) = P_{\{i_1, ..., i_k\}}(i_1) \cdot P_{\{i_2, ..., i_k\}}(i_2) \cdot \cdots \cdot P_{\{i_{k-1}, i_k\}}(i_{k-1}). \tag{1}$$

Any ranking model $P(\pi)$ which can be expressed in the form (1) for some collection of choice probabilities is said to be *L-decomposable.*

Luce (1959) begins with a particular family of choice probabilities—namely, those satisfying his Choice Axiom (see Equation (5) below). From these specific choice probabilities, Luce then obtains ranking probabilities $P(\pi)$ via (1). The notion of L-decomposability is a slight generalization: *given* the ranking probabilities $P(\pi)$, does there exist *some* collection of choice probabilities—not necessarily satisfying the Choice Axiom—which reproduce these ranking probabilities via (1)? If so, the ranking model is said to be L-decomposable.

At first glance, it may appear that all ranking models are L-decomposable, since the law of multiplication for conditional probabilities (always) asserts that

$$\begin{aligned} P(\pi) = {} & P\{\pi^{-1}(1) = i_1\} \cdot P\{\pi^{-1}(2) = i_2 \mid \pi^{-1}(1) = i_1\} \\ & \cdot P\{\pi^{-1}(3) = i_3 \mid \pi^{-1}(1) = i_1, \pi^{-1}(2) = i_2\} \\ & \cdot \cdots \cdot P\{\pi^{-1}(k-1) = i_{k-1} \mid \pi^{-1}(1) = i_1, ..., \pi^{-1}(k-2) = i_{k-2}\}. \end{aligned} \tag{2}$$

However, consider for example the conditional probability $P\{\pi^{-1}(3) = i_3 \mid \pi^{-1}(1) = i_1, \pi^{-1}(2) = i_2\}$ in the right hand side above. This probability may depend on the relative ordering of items i_1 and i_2, since in general $P\{\pi^{-1}(3) = i_3 \mid \pi^{-1}(1) = i_1, \pi^{-1}(2) = i_2\}$ need not equal $P\{\pi^{-1}(3) = i_3 \mid \pi^{-1}(1) = i_2, \pi^{-1}(2) = i_1\}$. On the other hand, L-decomposability requires that the preceding conditional probability not depend on the relative ordering of i_1 and i_2. This is an immediate consequence of the following theorem, the proof of which is given in the Appendix.

THEOREM 1. *In order for a ranking model to be L-decomposable, a necessary and sufficient condition is that for each $r=3, ..., k$ and for each $i_1, ..., i_r$, the conditional probability*

$$P[\pi^{-1}(r)=i_r \mid \pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}]. \tag{3}$$

is a symmetric function of $i_1, ..., i_{r-1}$.

In the course of proving Theorem 1, it is also shown that if a ranking model is L-decomposable, then the choice probabilities which give rise to the ranking probabilities via (1) are uniquely determined. In fact, the choice probabilities can be calculated directly from the ranking probabilities, as shown by Eq. (25) in the Appendix. This generalizes a result of Luce (1959, pp. 72–73).

3. RANKING MODELS

In this section, most of the ranking models proposed in the literature are categorized into order statistics models, models induced from paired comparisons, distance-based models, and stagewise models. Models in each category are developed and their properties are described.

3.1. *Thurstone Order Statistics Models*

In his research on psychophysical phenomena, such as determining the loudness of sounds, Thurstone (1927) postulates a process wherein the actual perception of each stimulus is a random variable. The relative ordering of these random variables determines the subject's ordering of the stimuli. Thurstone's original method considers only paired comparisons of the stimuli, but Daniels (1950) extends the method to experiments where the data are full orderings of the k stimuli.

Formally, let X_i be random variables with arbitrary continuous distributions F_i, where $i=1, ..., k$ indexes items (stimuli). A random ranking π is defined by setting $\pi(i)$ equal to the rank, from smallest to largest, of X_i among $\{X_1, ..., X_k\}$. Thus the ranking π with ordering $\langle i_1, ..., i_k\rangle$ is assigned the probability

$$P(\pi)=P\{X_{i_1}< \cdots <X_{i_k}\}. \tag{4}$$

If the X_i are allowed to have arbitrary dependencies, it is easy to show that any probability distribution on rankings can be expressed in the form (4). On the other hand, for $k \geqslant 4$, the assumption that the X_i are independent leads to a proper subset of the class of all probability distributions on rankings. For example, the probability distribution defined by $P(\langle 1\,2\,3\,4\rangle)=p \in (0, 1)$ and $P(\langle 3\,4\,1\,2\rangle)=1-p$ cannot be written in the form (4) with independent X_i. The class of probability models defined by (4) with X_i independent is referred to as the class of *order statistics models.*

The most widely studied type of order statistics models are those for which the distributions $F_i(x) = F(x - \mu_i)$ constitute a location family with shifts μ_i $(i = 1, ..., k)$. Such models are referred to as *Thurstone models* (see, for example, Yellott (1977)). Note that the class $\mathscr{P}_F$ of all Thurstone models with a particular F contains probability measures $P_{\mathbf{\mu}}$ that depend only on the difference vector $\mathbf{\mu} = [\mu_2 - \mu_1, ..., \mu_k - \mu_1]$; thus the class $\mathscr{P}_F$ has $k - 1$ free parameters. Although the class $\mathscr{P}_F$ is based explicitly upon a location family, it might just as well have arisen from the scale family obtained by the transformation $Y = \exp(X)$, since the ordering of the Y's is the same as that of the X's.

The class $\mathscr{P}_F$ relates to the properties of Section 2 as follows. Without loss of generality, assume that F has median 0. If the distribution F is symmetric, then the class $\mathscr{P}_F$ is reversible since $\pi \sim P_{\mathbf{\mu}} \in \mathscr{P}_F$ implies that $\gamma \circ \pi \sim P_{-\mathbf{\mu}} \in \mathscr{P}_F$, with γ defined as in the discussion of reversibility in Section 2. Not all Thurstone models have unique modal rankings (see Savage (1956) for distributions F that lead to counterexamples), even when the μ_i are distinct. However, if the μ_i are distinct and the likelihood ratio $F'(x - \mu_i)/F'(x - \mu_j)$ is a non-increasing function of x for $\mu_i < \mu_j$, Henery (1981) shows that a Thurstone model has complete consensus, and therefore is also strongly unimodal (see also Savage (1956, 1957)). Because the exact calculation of $P(\pi)$ involves a multiple integral, it is difficult to verify L-decomposability, except in the special case of the Gumbel distribution $F(x) = 1 - \exp(-\exp x)$ where the integral has a closed form.

A special property of the order statistics model not shared by the other classes of models deserves mention. Namely, as Henery (1981) points out, the relative order of any subset of the items is independent of the ordering of any disjoint subset. This follows directly from the independence of the X_i although it is true more generally, since assuming that the $X_i - \mu_i$ are exchangeable leads to the same model (Mosteller (1951)).

Two well-studied cases of the Thurstone model are the Thurstone (1927)–Mosteller (1951)–Daniels (1950) model where F is normal, and the Luce model where F is Gumbel. Brook and Upton (1974) use the Thurstone–Mosteller–Daniels model to analyze voters' rankings of candidates for public office. Since $k = 3$ in their example, they are able to integrate the multiple integral for $P(\pi)$ numerically to obtain maximum likelihood estimates. Henery (1981) offers an approximation to the integral that can be used in more general cases.

A ranking model equivalent to the Thurstone model with Gumbel F is derived by Luce (1959). Luce's approach is to construct choice probabilities P_B on the basis of his Choice Axiom, which implies that for each item i there is a value p_i such that

$$P_B(i) = p_i \Big/ \sum_{j \in B} p_j, \tag{5}$$

for any subset B of $\{1, ..., k\}$. For identifiability, the p_i may be scaled to sum to 1, so that p_i becomes the probability that item i is most preferred among the full set of items. In conjunction with the ranking postulate (1), the choice probabilities (5) can be shown to yield the Thurstone model with Gumbel F and $\mu_i = -\log p_i$. (See, for example, Yellott (1977), who attributes this result to Holman and Marley.)

Explicitly, for every ranking π with associated ordering $\langle i_1, ..., i_k \rangle$, the Luce model is given by

$$P(\pi) = \prod_{r=1}^{k-1} \left(p_{i_r} \Big/ \sum_{j \in B} p_j \right), \tag{6}$$

where $B_r = \{i_r, ..., i_k\}$ is the set of items remaining at stage r.

Pendergrass and Bradley (1960) independently derive the Luce model for the case $k=3$, and they observe that the class of such models is not reversible. This lack of reversibility does not contradict the previous discussion since the Gumbel distribution is asymmetric. The Luce model also coincides with the first order model in Plackett's (1975) system of logistic models. Henery (1983) and Stern (1987) investigate order statistics models formed from scale families of gamma distributions with a fixed shape parameter. As previously mentioned such models based on scale families are equivalent to Thurstone models. In particular, if X has an exponential distribution (gamma with shape parameter 1), then $Y = \log X$ is Gumbel, and so the scale family of exponential distributions studied by Henery and Stern also leads to the Luce model (6). Harville (1973) modifies the Luce model to accommodate rankings of only the q best items ($q < k-1$) and applies that model to horse racing.

In addition to the order statistics model (4), there are other methods for constructing ranking models based on independent $X_1, ..., X_k$. For example, Thurstone (1927) defines the paired comparison probability that item i is preferred to item j by $p_{ij} = P(X_i < X_j)$. These, as well as more general paired comparison probabilities, can then be employed to construct a probability model on rankings, by utilizing the general method of the next subsection.

3.2. *Ranking Models Induced by Paired Comparisons*

Babington Smith (1950) suggests a conditioning argument for inducing a probability model on rankings from a set of arbitrary paired comparison probabilities. This method naturally links the properties of Section 2 to existing notions of stochastic ordering in paired comparison models. Although the Babington Smith model is indexed by $C(k; 2) \equiv k(k-1)/2$ parameters and is difficult to work with, simplifications proposed by Mallows (1957) lead to some analytically tractable models.

For each pair of items $i < j$, let p_{ij} be the probability that item i is preferred to item j ($i \to j$) in a paired comparison of these two items. Imagine a tournament in which all of the $C(k; 2)$ possible paired comparisons are made independently. If the results of this tournament contain no circular triads ($h \to i \to j \to h$), then the tournament corresponds to a unique ranking π of the items; otherwise the entire tournament is repeated until a unique ranking is obtained. The probability of any resulting π is thus given by

$$P(\pi) = \text{constant} \prod_{\{(i,j) : \pi(i) < \pi(j)\}} p_{ij}, \tag{7}$$

where p_{ij} for $i>j$ is defined by $p_{ij} \equiv 1-p_{ji}$, and the constant is chosen to make the probabilities sum to 1. Paired comparison models that constrain the $\{p_{ij}\}$ lead to important subclasses of the general Babington Smith model (7).

With regard to the properties of Section 2, the class of all models P of the form (7) is reversible, since P^{γ} also has the form (7) with each p_{ij} replaced by $1-p_{ij}$ (Marley (1968)). Similarly, all of the subclasses considered later in this subsection are reversible.

Strong unimodality and complete consensus depend on whether the paired comparison probabilities are weakly or strongly stochastically transitive. By definition (David (1988), pp. 5–6), the $\{p_{ij}\}$ are *weakly stochastically transitive* if $p_{ij} \geqslant 0.5$ and $p_{jm} \geqslant 0.5$ imply that $p_{im} \geqslant 0.5$; the $\{p_{ij}\}$ are *strongly stochastically transitive* if $p_{ij} \geqslant 0.5$ and $p_{jm} \geqslant 0.5$ imply that $p_{im} \geqslant \max\{p_{ij}, p_{jm}\}$. The following two theorems, whose proofs appear in the Appendix, make the connections explicit.

THEOREM 2. *Let $\{p_{ij}\}$ be a set of paired comparison probabilities such that no $p_{ij}=0.5$. Then the $\{p_{ij}\}$ are weakly stochastically transitive if and only if the associated model* (7) *is strongly unimodal.*

THEOREM 3. *Let $\{p_{ij}\}$ be a set of paired comparison probabilities such that no $p_{ij}=0.5$. If the $\{p_{ij}\}$ are strongly stochastically transitive, then the associated model* (7) *has complete consensus.*

The converse of Theorem 3 is false; for example when $k=3$, the paired comparison probabilities $p_{13}=0.6$, $p_{12}=p_{23}=0.7$ are not strongly stochastically transitive, yet produce a model having complete consensus. Note that if the probabilities p_{ij} are formed from a Thurstone model by $p_{ij}=P(X_i<X_j)$, then the $\{p_{ij}\}$ are strongly stochastically transitive.

Finally, to verify L-decomposability, it is straightforward to check that (3) is a symmetric function of $i_1, \ldots, i_{r-1}$, for all models of the form (7).

In a seminal paper, Mallows (1957) proposes four simple subclasses of the Babington Smith model (7), each having significantly fewer than $C(k; 2)$ parameters. The last three of these subclasses coincide with distance-based models discussed in Subsection 3.3. Figure 1 summarizes the relationships among these subclasses of the Babington Smith models, and serves as a guide for the remainder of this subsection.

General Babington Smith Models

Mallows-Bradley-Terry Models $\qquad$ $H(\theta, \phi)$

$H(\theta)$ = Mallows-Bradley-Terry $\cap$ $H(\theta, \phi)$ $\qquad$ $H(\phi)$

FIG. 1. Subclasses of the Babington Smith models.

The Class of Mallows–Bradley–Terry Models. One way to reduce the number of parameters in (7) is to assume that the paired comparison probabilities have the Bradley–Terry (1952) form

$$p_{ij} = v_i/(v_i + v_j)$$

for some nonnegative parameters $v_1, ..., v_k$. Note that if the v_i are multiplied by an arbitrary positive constant c, the p_{ij} remain the same; thus, for identifiability, c may be chosen so that the v_i sum to 1. Mallows (1957) suggests substituting the Bradley–Terry paired comparison probabilities into (7), which leads to the following ranking model. For any ranking π with associated ordering $\langle i_1, ..., i_k \rangle$,

$$P(\pi) = C(\mathbf{v}) \prod_{r=1}^{k-1} (v_{i_r})^{k-r}, \tag{8}$$

where $\mathbf{v} = (v_1, ..., v_k)$ and $C(\mathbf{v})$ is chosen to make the probabilities sum to 1. Since the Bradley–Terry paired comparison probabilities are strongly stochastically transitive, this subclass has complete consensus.

Martin-Löf (1973) independently derives model (8) as the maximal entropy model, for which the sufficient statistic is the vector whose ith component is the sample sum of the ranks associated with the ith item. In the case that $k = 3$, the model is investigated by Pendergrass and Bradley (1960) and by Fienberg and Larntz (1976), who develop it as a log-linear model.

The Class $H(\theta, \phi)$. An alternative subclass of the Babington Smith model, also suggested by Mallows (1957), has just two continuous parameters. It is derived from the general Babington Smith model (7) by making an additional assumption. Namely, rankings which are the same distance from an assumed modal ranking π_0, in two special metrics, should have the same probability. The first metric distance between an arbitrary ranking π and the modal ranking π_0 is the tau metric, given by

$$T(\pi, \pi_0) = \sum_{i<j} I\{[\pi(i) - \pi(j)][\pi_0(i) - \pi_0(j)] < 0\}, \tag{9}$$

where $I(\cdot)$ is the indicator function: $I(A) = 1$ if the event A occurs, and $= 0$ otherwise. The metric T counts the numbers of pairs of items that are ranked in reverse order by π and π_0, and is related to Kendall's (1938) correlation coefficient. The second metric $R(\pi, \pi_0)$, given by

$$R^2(\pi, \pi_0) = \sum_{i=1}^{k} [\pi(i) - \pi_0(i)]^2, \tag{10}$$

is related to Spearman's (1904) correlation coefficient. Mallows shows that the resulting ranking model must have the form

$$P(\pi \mid \theta, \phi, \pi_0) = C(\theta, \phi)\, \theta^{R^2(\pi, \pi_0)} \phi^{T(\pi, \pi_0)}, \tag{11}$$

where θ, $\phi > 0$, and $C(\theta, \phi)$ is chosen to make the probabilities sum to 1. The class of models of the form (11) is denoted by $H(\theta, \phi)$. It is easily verified that the class $H(\theta, \phi)$ is label-invariant, so that for notational convenience it may be assumed that $\pi_0 = e$ without loss of generality. For $\pi_0 = e$, Mallows (1957) shows that the particular paired comparison probabilities p_{ij} that generate the model (11) via (7) are

$$p_{ij} = \theta^{i-j}\phi^{-1}/(\theta^{i-j}\phi^{-1} + \theta^{j-i}\phi). \tag{12}$$

The Class $H(\theta)$. Mallows (1957) further proves that the intersection of the Mallows–Bradley–Terry and $H(\theta, \phi)$ classes is the one parameter class $H(\theta)$ consisting of all models of the form

$$P(\pi \mid \theta, \pi_0) = C(\theta)\, \theta^{R^2(\pi, \pi_0)},$$

which is obtained by setting $\phi = 1$ in Eq. (11). The corresponding paired comparison probabilities are given by setting $\phi = 1$ in Eq. (12); these may be written in the Bradley–Terry form $p_{ij} = v_i/(v_i + v_j)$ by letting $v_i = \theta^{2i}$.

The Class $H(\phi)$. Alternatively, setting $\theta = 1$ in Equation (11) yields Mallows' well-known *ϕ-model*

$$P(\pi \mid \phi, \pi_0) = C(\phi)\, \phi^{T(\pi, \pi_0)}, \tag{13}$$

which has been studied by Feigin and Cohen (1978) and others. Since $T(\pi, e) + T(\pi, \gamma) = C(k; 2)$ for all π, it follows that $\phi^{T(\pi, e)}$ and $(\phi^{-1})^{T(\pi, \gamma)}$ are proportional; hence by relabelling the items it can be assumed that $0 < \phi \leqslant 1$ and that $\pi_0 = e$. The ϕ-model then has the interpretation that the probability $P(\pi)$ decreases according to increasing tau-distance from the modal ranking e.

The corresponding paired comparison probabilities have the simple form $p_{ij} = (\phi + 1)^{-1} \geqslant 0.5$ for $i < j$. Thus the probability that $i \to j$ depends only upon whether $i < j$ or $i > j$. It follows that the intersection of the $H(\phi)$ and Mallows–Bradley–Terry classes is the uniform distribution.

3.3. *Distance-Based Ranking Models*

Mallows' ϕ-model (13) may be written as

$$P(\pi \mid \lambda, \pi_0) = C(\lambda) \exp\{-\lambda T(\pi, \pi_0)\}, \tag{14}$$

where $\lambda = -\log \phi$ and $C(\lambda)^{-1} = \sum_{\pi \in S_k} \exp\{-\lambda T(\pi, \pi_0)\}$. Thus for fixed π_0, (14) has the form of an exponential family. Diaconis (1988, p. 104) uses this representation to motivate a general class of *metric-based ranking models,*

$$P(\pi \mid \lambda, \pi_0) = C(\lambda) \exp\{-\lambda d(\pi, \pi_0)\}, \tag{15}$$

where d is an arbitrary metric on the set S_k of all permutations. In addition to satisfying the usual metric axioms of positivity, symmetry, and the triangle inequality, d is assumed to be *right-invariant*:

$$d(\pi \circ \sigma, \nu \circ \sigma) = d(\pi, \nu) \qquad \text{for all} \quad \pi, \nu, \sigma \in S_k.$$

For a fixed metric d, the model (15) has just one discrete parameter π_0, and a single continuous parameter λ. The parameter $\pi_0 \in S_k$ is the modal ranking, and $\lambda \geqslant 0$ is a dispersion parameter. When $\lambda = 0$, (15) is the uniform distribution, and as λ increases the distribution becomes more peaked about π_0.

Diaconis (1988, pp. 112–119) considers the following metrics on S_k to be among the most widely used in applied scientific and statistical problems:

$$T(\pi, \nu) = \sum_{i<j} I\{[\pi(i) - \pi(j)][\nu(i) - \nu(j)] < 0\}$$

is the Kendall's tau metric;

$$R(\pi, \nu) = \left(\sum_i [\pi(i) - \nu(i)]^2\right)^{1/2}$$

is the Spearman's rho metric;

$$F(\pi, \nu) = \sum_i |\pi(i) - \nu(i)|$$

is the Spearman's footrule metric;

$$H(\pi, \nu) = \sum_i I\{\pi(i) \neq \nu(i)\}$$

is the Hamming metric;

$$C(\pi, \nu) = \text{minimal number of transpositions needed to transform } \pi \text{ into } \nu$$

is Cayley's metric; and

$$U(\pi, \nu) = k - (\text{the maximal number of items ranked in the same relative order by } \pi \text{ and } \nu)$$

is Ulam's metric. All six metrics are right-invariant. The Hamming metric is used in coding theory to measure the distance between two binary sequences (Berlekamp (1968)), and Ulam's metric is used in DNA research to measure the distance between two strings of molecules (Ulam (1972, 1981), Gordon (1979)). Diaconis (1988, pp. 102–112) discusses a plethora of other potential statistical applications, although attention here is restricted to ranking models.

Rukhin (1972), Diaconis (1988, p. 121), and Critchlow (1985, pp. 27–32) also describe a "fixed vector" approach for creating right-invariant metrics. An application of this approach leads to the generalization of Spearman's footrule metric F

$$F_h(\pi, \nu) = \sum_i |h[\pi(i)] - h[\nu(i)]|,$$

where h: $\{1, ..., k\} \to \mathbf{R}$ is a strictly increasing function that rescales the ranks. In conjunction with the general metric-based ranking model (15), the F_h metric leads to the model

$$P(\pi \mid \lambda, \pi_0) = C(\lambda) \exp\{-\lambda F_h(\pi, \pi_0)\}. \tag{16}$$

Critchlow (pp. 111–116) uses a variant of model (16) in an analysis of ranked data.

As already noted in Eq. (14), Mallows' ϕ-model is a metric-based ranking model, with $d = T$. The models belonging to the classes $H(\theta)$ and $H(\theta, \phi)$ of Subsection 3.3 have similar representations to (15), namely as $P(\pi \mid \lambda, \pi_0) = C(\lambda) \exp\{-\lambda R^2(\pi, \pi_0)\}$ and $P(\pi \mid \lambda, \kappa, \pi_0) = C(\lambda, \kappa)$ $\exp\{-\lambda(R^2 + \kappa T)(\pi, \pi_0)\}$, respectively; yet they are not quite metric-based models, since R^2 does not satisfy the triangle inequality. The class of metric-based models may be enlarged to include such models by dropping the triangle inequality requirement for d. The resulting function d is called a *right-invariant distance* rather than a true metric, and the resulting model is a *distance-based ranking model*. Not much is lost in this generalization, although the triangle inequality property has a desirable interpretation with respect to the behavior of likelihood ratio tests for π_0 (see Critchlow (1985), pp. 101–102).

Properties of distance-based ranking models follow naturally from analogous features of the distance d. Let $\mathscr{P}_d$ denote the class of models associated with a fixed distance d.

Label-invaraince of the class $\mathscr{P}_d$ follows from the right-invariance of d, since for every $P \in \mathscr{P}_d$ and $\nu \in S_k$,

$$P_\nu(\pi) = P(\pi \circ \nu) = C(\lambda) \exp\{-\lambda d(\pi \circ \nu, \pi_0)\} = C(\lambda) \exp\{-\lambda d(\pi, \pi_0 \circ \nu^{-1})\},$$

where ν^{-1} is the permutation which satisfies $\nu \circ \nu^{-1} = e$. Thus $P_\nu \in \mathscr{P}_d$ has modal ranking $\pi_0 \circ \nu^{-1}$ and the same dispersion parameter as P. Similarly, $\mathscr{P}_d$ is reversible if the distance d satisfies

$$d(\gamma \circ \pi, \gamma \circ \nu) = d(\pi, \nu) \qquad \text{for all} \quad \pi, \nu \in S_k, \tag{17}$$

with the reversing permutation γ defined by $\gamma(j) = (k+1) - j$, $j = 1, ..., k$, as in Section 2. It is easy to check that the distances T, R, R^2, F, H, C and U satisfy property (17), and hence give rise to reversible classes.

A distance-based model displays complete consensus if and only if d possesses the *transposition property*: $d(\pi, \nu) \leqslant d(\pi \circ \tau_{ij}, \nu)$ for any permutations π, $\nu \in S_k$ and items i, j such that $\pi(i) < \pi(j)$ and $\nu(i) < \nu(j)$. This is immediate from the definition of complete consensus. The transposition property for distances between permutations is defined and studied in a different context by Critchlow (1985, pp. 50–53), who proves that this property holds for various metrics; see also Hollander, Proschan, and Sethuraman (1977). Since the distances T, R, R^2, F, and F_h satisfy the transposition property, the models based on these distances exhibit complete consensus.

Clearly, a necessary and sufficient condition for strong unimodality is the *weak transposition property*: $d(\pi, \nu) \leqslant d(\pi \circ \tau_{ij}, \nu)$ for any permutations $\pi, \nu \in S_k$ and items i, j such that $\pi(i) < \pi(j) - 1$. It can be shown that the distances T, R, R^2, F, U, and F_h satisfy the weak transposition property.

Finally, a condition for L-decomposability is given by the following theorem, whose proof appears in the Appendix.

THEOREM 4. *A distance-based ranking model* (15) *is L-decomposable if and only if the distance d is* additively decomposable, *in the sense that for each* $r = 2, ..., k$, *there exist functions* f_r *and* g_r *such that*

$$d(\pi, e) = f_r[\pi^{-1}(1), ..., \pi^{-1}(r-1)] + g_r[\pi^{-1}(r), ..., \pi^{-1}(k)]. \tag{18}$$

In particular, the model based on T is L-decomposable, since T satisfies (18) with

$$f_r[\pi^{-1}(1), ..., \pi^{-1}(r-1)] = \sum_{s < t \leqslant r-1} I\{\pi^{-1}(s) > \pi^{-1}(t)\} + \sum_{s=1}^{r-1} \pi^{-1}(s) - r(r-1)/2$$

and

$$g_r[\pi^{-1}(r), ..., \pi^{-1}(k)] = \sum_{r \leqslant s < t} I\{\pi^{-1}(s) > \pi^{-1}(t)\}.$$

Moreover, condition (18) holds if $d(\pi, e) = \sum_r a_r[\pi^{-1}(r)]$ for some functions $a_1, ..., a_k$, so that the models based on F, R^2, H, and F_h are L-decomposable. However, the models based on R, C, and U are not L-decomposable.

Critchlow (1985) develops distance-based models for *partially ranked data*, such as rankings of only the q best items, where $q < k - 1$. The set of all partial rankings of the best q items is identified with the quotient space S_k/S_{k-q} of the permutation group, consisting of all right cosets of the subgroup $S_{k-q} = \{\pi \in S_k \mid \pi(i) = i, i = 1, ..., q\}$. Any metric d on S_k induces a *Hausdorff metric* d^* on the quotient space S_k/S_{k-q}, which serves as the metric for these models. In particular, Critchlow gives the Hausdorff extensions of T, R, F, H, C, and U.

Fligner and Verducci (1986) introduce multiparameter extensions of (15) when d can be decomposed into a sum of $k-1$ functions

$$d(\pi, \nu) = \sum_{r=1}^{k-1} d_r(\pi, \nu),$$

and where the range of $d_r(\pi, \nu)$ does not depend on the value of $d_s(\pi, \nu)$ for $s \neq r$. The metrics T and C have such decompositions, and for $d = T$ the resulting model is a special case of the multistage models discussed in the next subsection.

3.4. *Multistage Ranking Models*

Multistage models (Fligner and Verducci (1988)) decompose the ranking process into a sequence of $k-1$ stages, in a manner similar to the decomposition (6) for Luce's model. However, unlike Luce's choice probabilities (5), the set $\{P_{B_r}(i) : i \in B_r\}$ of choice probabilities at stage r is a *fixed* set of probabilities that does not depend on the item indices in B_r. That is, the set of choice probabilities at a given stage depends only on the stage, rather than on the items remaining at that stage. Such might be the case if a subject expends differing amounts of effort at the various stages, or if the choice probabilities substantially depend on the number of items in the choice set. Additionally, multistage models are more tractable than Luce models in terms of inference.

The complete description of a multistage ranking model is as follows. Let $\{p(m, r) : m = 0, ..., k-r\}$ denote the fixed set of choice probabilities at stage r, $r = 1, ..., k-1$. These probabilities are assigned to the items indexed in B_r by assessing the correctness of the choice made at stage r with respect to a central ranking π_0. Specifically, let $V_r = m$ if, at stage r, the $(m+1)$st best of the items in B_r (according to π_0) is selected, so that m may be thought of as the "number of mistakes" made at stage r. For example, if π_0 and π correspond to the orderings $\langle 3\ 1\ 2\ 4 \rangle$ and $\langle 3\ 4\ 2\ 1 \rangle$, respectively, then $V_1 = 0$ since the best item (item 3) is selected at the first stage; $V_2 = 2$ since the third best (item 4) of the three remaining items is selected at the second stage; and $V_3 = 1$ since the second best (item 2) of the two remaining items is selected at the last stage. Then $p(m, r) \equiv P(V_r = m)$ and the *general multistage model* is a $C(k; 2)$ parameter model given by

$$P(\pi) = \prod_{r=1}^{k-1} p(V_r, r). \tag{19}$$

Note that for any π, the corresponding vector $\mathbf{V} = (V_1, ..., V_{k-1})$ is related to the tau-distance between π and π_0 by

$$T(\pi, \pi_0) = \sum_{r=1}^{k-1} V_r.$$

It is easily shown that the class of models defined by (19) is L-decomposable but not reversible. The properties of strong unimodality and complete consensus require further assumptions on the choice probabilities at each stage. These assumptions yield useful subclasses of the general multistage model, described in Figure 2.

For a fixed central ranking π_0, the condition that for each r,

$$p(0, r) > p(1, r) \geqslant \cdots \geqslant p(k-r, r)$$

is necessary and sufficient for strong unimodality, with π_0 as the mode. Thus, at each stage, an item more preferred according to π_0 is at least as likely to be chosen

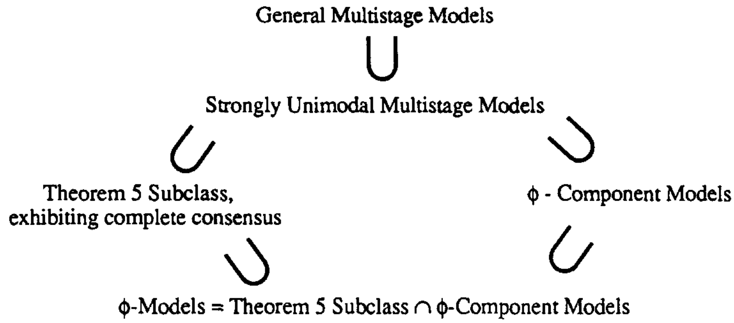

FIG. 2. Subclasses of the multistage models.

as one that is less preferred. An important consequence of strong unimodality for model (19), which is not true for general ranking models, is that the mode π_0 orders the items according to their expected ranks. Namely, if $\pi_0(i) < \pi_0(j)$, then $E[\pi(i)] < E[\pi(j)]$. For proofs, see Fligner and Verducci (1988).

Complete consensus requires conditions that constrain the choice probabilities across stages; for example,

THEOREM 5 (Fligner and Verducci (1988)). *For model* (19), *suppose that there exists a strictly decreasing positive function g, with* log *g concave, such that*

$$p(m, r) = c_r\, g(m) \qquad \textit{for} \quad m = 0, ..., k-r \qquad \textit{and} \qquad r = 1, ..., k-1,$$

where

$$c_r^{-1} = \sum_{m=0}^{k-r} g(m).$$

Then model (19) *has a complete consensus.*

The ϕ-component Model. The conditions of Theorem 5 imply that the ratio $p(m, r)/p(m+1, r)$, $r < k - m$, is independent of the stage r. An alternative simplification of the strongly unimodal multistage model results from assuming that this ratio is independent of the number m of mistakes. Under this assumption, Fligner and Verducci (1986) show that there exist non-negative constants $\lambda_1, ..., \lambda_{k-1}$ such that

$$p(m, r) = c_r \exp(-\lambda_r m), \tag{20}$$

where

$$c_r^{-1} = \sum_{m=0}^{k-r} \exp(-\lambda_r m) = \{1 - \exp[-\lambda_r(k - r + 1)]\}/[1 - \exp(-\lambda_r)].$$

The resulting $k-1$ parameter model is called the *ϕ-component model.* The parameter λ_r completely determines the probabilities $p(m, r)$ for stage r and is inversely related to the expected number of mistakes made at the stage. Fligner and Verducci (1988) show that the ϕ-component model has a complete consensus if $\lambda_r \geqslant \lambda_{r+1}$ for each $r = 1, ..., k-2$.

Mallows' ϕ-model is obtained either by letting $g(m) = \exp(-\lambda m)$ in Theorem 5, or by setting $\lambda_1 = \cdots = \lambda_{k-1}$ in the ϕ-component model. It is easily shown that the ϕ-model is the only multistage model such that the ratio $p(m, r)/p(m+1, r)$ is independent of *both* r and m. Thus the ϕ-model can be motivated as a multistage model, a distance-based model, or a ranking model induced by paired comparisons.

4. Intersections of the Classes of Models

In Subsections 3.1 through 3.4, the following broad classes of ranking models have been studied:

Class I. Order Statistics Models (4), where $X_1, ..., X_k$ are arbitrary independent random variables.

Class II. Babington Smith Models (7), arising from general paired comparison probabilities.

Class III. Distance-Based Models (15), with d an arbitrary distance on S_k.

Class IV. The General Multistage Models (19).

In addition, an important subclass of IV is

Class IVa. The ϕ-Component Models.

As previously noted, the class $H(\phi)$ of Mallows ϕ-models is contained in each of II, III, IV, and IVa. Thus the ϕ-models can be motivated from a multitude of points of view. A natural question, to be discussed below, is whether there are any other models which belong to the intersections of the various classes. The following formulae summarize the relationships among the various classes, which will be demonstrated in this section:

$$\mathrm{II} \cap \mathrm{III} \subset H(\theta, \phi),$$

$$\mathrm{II} \cap \mathrm{IV} = \mathrm{II} \cap \mathrm{IVa} = H(\phi),$$

$$\mathrm{III} \cap \mathrm{IV} = \text{the } k-1 \text{ parameter class satisfying (23)},$$

$$\mathrm{II} \cap \mathrm{IVa} = H(\phi) \text{ with } \phi \neq 1, \text{ and}$$

$$\mathrm{I} \cap H(\phi) = \text{the uniform model.}$$

The pairwise intersections of the classes II, III, and IV will be investigated initially, since these are the easiest to deal with analytically.

II $\cap$ III. This intersection is essentially studied by Mallows (1957). To put the question in the form treated by Mallows, suppose first that $\pi_0 = e$ for a distance-based model. Then for each $\pi \in S_k$,

$$\begin{aligned} P(\pi) &= C(\lambda)\exp\{-\lambda d(\pi, e)\} = C(\lambda)\exp\{-\lambda d(e, \pi)\} \\ &= C(\lambda)\exp\{-\lambda d(\pi^{-1}, e)\} = P(\pi^{-1}), \end{aligned}$$

using the symmetry and right-invariance properties of d. Mallows (1957, pp. 117–118) proves that for a Babington Smith model, $P(\pi) \equiv P(\pi^{-1})$ if and only if P is an $H(\theta, \phi)$ model with $\pi_0 = e$.

Next, consider a general distance-based model with $\pi_0 \neq e$. By right-invariance of d,

$$P(\pi) = C(\lambda)\exp\{-\lambda d(\pi, \pi_0)\} = C(\lambda)\exp\{-\lambda d(\pi \circ \pi_0^{-1}, e)\}.$$

If this model is also to be a Babington Smith model, then so is the distribution of $\pi \circ \pi_0^{-1}$ by label-invariance. The results of the preceding paragraph now imply that

$$P(\pi) = C(\theta, \phi)\, \theta^{R^2(\pi \circ \pi_0^{-1}, e)} \phi^{T(\pi \circ \pi_0^{-1}, e)} = C(\theta, \phi)\, \theta^{R^2(\pi, \pi_0)} \phi^{T(\pi, \pi_0)},$$

which is (11). Hence

$$\mathrm{II} \cap \mathrm{III} \subset H(\theta, \phi).$$

The precise restrictions on θ and ϕ, under which the resulting $H(\theta, \phi)$ model is, in fact, a distance-based model, are rather complicated. They will not be discussed here, except to note that an $H(\theta, \phi)$ model is certainly distance-based if θ and ϕ are both in $(0, 1]$ and not both equal to 1.

II $\cap$ IV. This intersection is $H(\phi)$. To see this, let P be the probability mass function for an arbitrary model in the intersection. Assume without loss of generality that $\pi_0 = e$ for the multistage representation of the model, and, to avoid degenerate cases, that all the paired comparison probabilities p_{ij} are strictly positive in the Babington Smith representation of the model. Let $\Delta_{ij} = p_{ij}/p_{ji}$. For each $r = 1, ..., k-1$ and $m = 1, ..., k-r-1$, let $\alpha(=\alpha_r)$ and $\beta(=\beta_{r,m})$ be the permutations

$$\alpha(r+1) = r-1, \quad \alpha(r-1) = r, \quad \alpha(r) = r+1, \quad \text{and} \quad \alpha(i) = i \quad \forall i \neq r-1, r, r+1$$

and

$$\beta(r+m+1) = r+1, \qquad \beta(i) = i+1 \qquad \forall i = r+1, ..., r+m,$$

and

$$\beta(i) = i \qquad \forall i \neq r+1, ..., r+m+1.$$

From the multistage representation (19), it is easy to check that

$$\frac{P(\tau_{r,r+1})}{P(e)}=\frac{P(\tau_{r-1,r+1})}{P(\alpha)}=\frac{P(\tau_{r,r+2}\circ\beta)}{P(\beta)}=\frac{p(1,r)}{p(0,r)}. \tag{21}$$

On the other hand, the Babington Smith representation (7) implies

$$\frac{P(\tau_{r,r+1})}{P(e)}=\Delta_{r+1,r},\quad \frac{P(\tau_{r-1,r+1})}{P(\alpha)}=\Delta_{r,r-1},\quad \text{and}\quad \frac{P(\tau_{r,r+2}\circ\beta)}{P(\beta)}=\frac{\Delta_{r+1,r}\Delta_{r+m+1,r}}{\Delta_{r+m+1,r+1}}. \tag{22}$$

Comparing (21) and (22) gives

$$\Delta_{r+1,r}=\Delta_{r,r-1} \qquad \text{and} \qquad \Delta_{n,r}=\Delta_{n,r+1}$$

for all $r=1, ..., k$ and $n=r+2, ..., k$. It follows that Δ_{ij} is a fixed constant Δ for all i, j pairs with $i>j$. Hence p_{ij} also depends only on whether $i>j$ or $i<j$. As noted in Subsection 3.2, such $\{p_{ij}\}$ are the paired comparison probabilities which induce the ϕ-model.

III $\cap$ IV. This intersection is the subclass of multistage models whose choice probabilities $p(m, r)$ satisfy

$$\frac{p(m+1,r)}{p(m,r)}=\frac{p(1,r+m)}{p(0,r+m)} \tag{23}$$

for all $m=0, ..., k-r-1$ and $r=1, ..., k-1$. The proof is similar to that given for II $\cap$ IV, and is omitted. Thus the subclass III $\cap$ IV has just $k-1$ parameters $\theta_r \equiv p(1, r)/p(0, r)$, for $r=1, ..., k-1$. The resulting models, which are both distance-based and multistage, appear quite interesting and are currently under investigation.

III $\cap$ IVa. This intersection is $H(\phi)$ with $\phi \neq 1$: that is, the class of all Mallows ϕ-models except for the uniform model. Indeed, let P be an arbitrary model in the intersection. Without loss of generality $\pi_0=e$, and hence $P(\pi)\equiv P(\pi^{-1})$ for all π since P is distance-based. In particular, $P(\alpha)=P(\alpha^{-1})$ with $\alpha(=\alpha_r)$ defined as in the derivation of II $\cap$ IV. Comparison with the special choice probabilities (20) for the ϕ-component models shows that

$$\begin{aligned} P(\alpha)=P(\alpha^{-1}) &\Rightarrow p(2, r-1)\, p(0, r)=p(1, r-1)\, p(1, r) \\ &\Rightarrow \exp(-2\lambda_{r-1})=\exp(-\lambda_{r-1})\exp(-\lambda_r) \\ &\Rightarrow \lambda_{r-1}=\lambda_r \end{aligned}$$

for all $r=2, ..., k-1$. Thus all of the parameters λ_r must be equal in (20), and P must be a ϕ-model. It is easily checked that the only ϕ-model which is *not* distance-based is the uniform model ($\phi=1$), since then the modal ranking is not unique.

Intersections with Class I. It remains to discuss the intersections of II, III, and IV with Class I (the order statistics models). These intersections are relatively difficult to ascertain, because the class of order statistics models is so broad. To illustrate the type of argument which might be used to find the intersections, the special case $\mathrm{I} \cap H(\phi)$ will now be determined. The key to computing $\mathrm{I} \cap H(\phi)$ is a property of order statistics models mentioned in Subsection 3.1, namely, the relative order of any subset of the items is independent of the ordering of any disjoint subset. Thus, if π is a random permutation generated according to an order statistics model,

$$P(\pi(1) < \pi(3) \mid \pi(2) < \pi(4)) = P(\pi(1) < \pi(3) \mid \pi(4) < \pi(2)).$$

On the other hand, if the preceding conditional probabilities are calculated according to a ϕ-model, the result is (after considerable simplification)

$$\phi^7 - \phi^5 - \phi^3 + \phi = 0.$$

This factors as $\phi(\phi^2+1)(\phi-1)^2\,(\phi+1)^2 = 0$, with roots $\phi = 0,\ \pm i,\ \pm 1$. Of these roots, only $\phi = 1$ is permissible for the ϕ-model, and the resulting model is uniform. Hence $\mathrm{I} \cap H(\phi)$ is the uniform model.

5. Example

The Graduate Record Examination Board asked 98 college students to rank five words according to their strength of association with a target word. For the target word "song," the five choices were labelled (1) score, (2) instrument, (3) solo, (4) benediction, and (5) suit. Table 1 lists the possible orderings in the first column, and their observed frequencies in the second column. The 15 orderings with observed frequency one are reported as a single category, as are the 92 orderings with observed frequency zero.

The modal ordering from this table is $\langle 3\ 2\ 1\ 4\ 5 \rangle$, indicating that "solo" is most associated with "song" and "suit" is least associated. This ordering happens to agree with the ordering of the items according to their average ranks, which were 2.72, 2.27, 1.60, 3.71, and 4.69 for items (1) through (5), respectively.

Columns 3 through 8 give the fitted frequencies for the Luce model (6), the Mallows–Bradley–Terry model (8), the Mallows ϕ-model (14), the metric model (15) with footrule metric F, the model (16) based on the generalized footrule metric F_h, and the ϕ-component model (20), respectively. In fitting the F_h-based model, the values $h(i)$, $i = 1, \ldots, 5$, of the rescaling function h are treated as parameters, and are estimated from the data. Without loss of generality, it is assumed that $0 = h(1) < h(2) < h(3) < h(4) < h(5) = 1$.

The distance-based models ϕ and F, as well as their extensions F_h and ϕ-comp, tend to overfit the mode, whereas Luce and MBT tend to underfit it. The only other notable difference among the models is that MBT allocates much smaller

TABLE I

Observed and Fitted Frequencies

Ordering	Observed	Models					
		Luce	MBT	ϕ	F	F_h	ϕ-Comp
⟨32145⟩	19	16.1	14.4	21.6	25.2	24.6	22.7
⟨31245⟩	10	9.6	10.5	7.5	7.1	9.8	7.3
⟨13245⟩	9	4.9	4.6	2.6	2.0	3.6	4.0
⟨32415⟩	8	7.3	7.5	7.5	7.1	4.9	6.4
⟨12345⟩	7	2.6	2.7	0.9	2.0	3.6	1.3
⟨32154⟩	6	4.5	5.2	7.5	7.1	5.0	6.2
⟨23145⟩	6	9.7	8.5	7.5	7.1	8.9	9.6
⟨32451⟩	5	0.8	1.4	2.6	2.0	1.0	1.8
⟨21345⟩	4	3.0	3.7	2.6	2.0	3.6	3.1
⟨31425⟩	3	2.3	4.1	2.6	2.0	1.9	2.1
⟨32541⟩	2	0.6	0.5	0.9	2.0	1.0	0.5
⟨34215⟩	2	3.5	2.9	2.6	2.0	1.9	2.3
⟨23415⟩	2	4.4	4.4	2.6	2.0	1.8	2.7
{15 orderings}	1 each	14.3	17.8	14.3	12.7	11.9	12.8
{92 orderings}	0 each	14.4	9.8	14.7	15.7	14.5	15.2

estimated probability to the unobserved orderings. None of the models adequately accounts for the frequent occurences of the orderings ⟨1 3 2 4 5⟩ and ⟨1 2 3 4 5⟩. Thus the models leave unexplained the pattern the pattern that item 1 tends to be ranked first only when items 4 and 5 are ranked fourth and fifth.

Table 2 summarizes the models in terms of the log-likelihood, the number of parameters, and the maximum likelihood estimates of these parameters. In addition

TABLE 2

Log-likelihood and Parameter Estimates

Model (equation no.)		Log-likelihood	No. of unconstrained continuous parameters	Maximum likelihood estimates
Luce	(6)	-316.40	4	$p_1=0.127$ $p_2=0.257$ $p_3=0.552$ $p_4=0.050$
MBT	(8)	-313.41	4	$v_1=0.186$ $v_2=0.252$ $v_3=0.429$ $v_4=0.098$
ϕ	(14)	-317.73	1	$\lambda=1.06$
F	(15)	-318.06	1	$\lambda=0.634$
F_h	(16)	-313.32	4	$\lambda=2.57$ $h(2)=0.197$ $h(3)=0.376$ $h(4)=0.690$
ϕ-comp.	(20)	-314.77	4	$\lambda_1=0.86$ $\lambda_2=1.14$ $\lambda_3=1.26$ $\lambda_4=1.30$
Multinomial		-277.78	119	
Uniform		-469.17	0	

to the continuous parameters listed in column 4, the distance-based models ϕ and F, as well as their extensions F_h and ϕ-comp, also include a discrete parameter: the modal ranking π_0. The maximum likelihood estimate of π_0 happens to coincide with the sample mode for each of these four models. The overall fits of the six models are comparable; models with four continuous parameters generally fit somewhat better than those with one. All of the models fit much better than the uniform, and not that much worse than the general multinomial, considering its large number of fitted parameters.

The estimated parameters for the various models in Table 2 have differing interpretations, but lead to similar conclusions. For the Luce model, the maximum likelihood estimates indicate that item 3, "solo," has probability 0.552 of being chosen as the word most associated with "song." In contrast item 5 has probability $1-(0.127+0.257+0.552+0.050)=0.014$. The ratio p_5 to p_4 is the largest between successive items in the modal ranking, indicating a pronounced distinction between these two items. Specifically, if items 4 and 5 occupy the last two places, then item 4 is ranked ahead of item 5 with probability $0.050/(0.050+0.014)=0.78$. The estimated parameters of the Mallows–Bradley–Terry model allow computation of the probability that item i is preferred to item j in a pairwise contest, under the assumption that the rankings are conditionally induced from paired comparisons. For example, the probability of item 4 being preferred to item 5 is estimated from Eq. (5) by 0.74, in agreement with the Luce model. For the generalized footrule model, the rescaling values 0.00, 0.20, 0.38, 0.60, 1.00 of the h function also confirm that the clearest distinctions occur in choosing the least preferred items. The same conclusion is suggested by the increasing trend in the parameters of the ϕ-component model.

As noted by Feigin and Cohen (1978), the λ parameter of the ϕ-model may be interpreted as a measure of general concordance. Using their table, the estimated value of λ corresponds to an expected tau-distance of 1.63 from the mode, where the range of possible tau distances is 0 to 10 in this example. Similarly the estimated value of λ in the footrule model corresponds to an expected footrule distance of 2.98 from the mode, where the range of possible footrule distances is the set of even integers from 0 to 12.

In terms of computational burden, maximum likelihood estimates are simple to compute for the ϕ and the ϕ-component models, for any number k of items (Fligner and Verducci (1986)). Currently the other models require considerably more computation, especially the multiparameter models. This computational difficulty increases rapidly with k.

APPENDIX

This appendix presents the proofs of Theorem 1 of Section 2, Theorems 2 and 3 of Subsection 3.2, and Theorem 4 of Subsection 3.3.

Proof of Theorem 1. Suppose that for each r and for each $i_1, ..., i_r$, the conditional probability $P[\pi^{-1}(r)=i_r \mid \pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}]$ is a symmetric function of $i_1, ..., i_{r-1}$. Then given an arbitrary subset of items $\{i_r, i_{r+1}, ..., i_k\}$, let $\{i_1, ..., i_{r-1}\}=\{1, ..., k\} \cap \{i_r, i_{r+1}, ..., i_k\}^c$ be the set of remaining items, and define the choice probability

$$P_{\{i_r, i_{r+1}, ..., i_k\}}(i_r)=P[\pi^{-1}(r)=i_r \mid \pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}]. \qquad (24)$$

The choice probabilities are well-defined, precisely because the right hand side of (24) is a symmetric function of $i_1, ..., i_{r-1}$: thus the right hand side of (24) depends only on i_r and the (unordered) set $\{i_1, ..., i_{r-1}\}$, or equivalently, only on i_r and $\{i_r, i_{r+1}, ..., i_k\}=\{1, ..., k\} \cap \{i_1, ..., i_{r-1}\}^c$. Furthermore, for these choice probabilities, (1) is equivalent to (2), which is the usual law of multiplication for conditional probabilities. Hence (1) holds and the ranking model is L-decomposable.

Conversely, suppose the ranking model $P(\pi)$ is L-decomposable, so that there exist choice probabilities satisfying (1). Given arbitrary $i_1, ..., i_r$, let $\{i_{r+1}, ..., i_k\}=\{1, ..., k\} \cap \{i_1, ..., i_r\}^c$, and let $C_m=\{v \in S_k : v^{-1}(j)=i_j,\ j=1, ..., m\}$ denote the set of permutations that agree with $\langle i_1, ..., i_k\rangle$ about the m most preferred items. Then

$$\begin{aligned} &P\{\pi^{-1}(r)=i_r \mid \pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}\} \\ &=\frac{P\{\pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}, \pi^{-1}(r)=i_r\}}{P\{\pi^{-1}(1)=i_1, ..., \pi^{-1}(r-1)=i_{r-1}\}} \\ &=\frac{\sum_{v \in C_r} P(v)}{\sum_{v \in C_{r-1}} P(v)}=\frac{\prod_{j=1}^{r} P_{\{i_j, ..., i_k\}}(i_j)}{\prod_{j=1}^{r-1} P_{\{i_j, ..., i_k\}}(i_j)} \\ &=P_{\{i_r, ..., i_k\}}(i_r)=P_{\{1, ..., k\} \cap \{i_1, ..., i_{r-1}\}^c}(i_r), \end{aligned} \qquad (25)$$

where the third equality above follows from (1). Thus $P\{\pi^{-1}(r)=i_r \mid \pi^{-1}(1)= i_1, ..., \pi^{-1}(r-1)=i_{r-1}\}$ is a symmetric function of $i_1, ..., i_{r-1}$, since it depends only on the (unordered) set $\{i_1, ..., i_{r-1}\}$. ∎

Proof of Theorem 2. Suppose the $\{p_{ij}\}$ are weakly stochastically transitive, and that no $p_{ij}=0.5$. It is easy to see that there exists a unique π_0 such that $\pi_0(i)<\pi_0(j) \Leftrightarrow p_{ij}>0.5$. As noted in Section 2, without loss of generality the items may be relabeled so that π_0 is the identity permutation e. Then e uniquely maximizes P since $P(e)=\text{constant} \prod_{i<j} p_{ij}$ with each factor $p_{ij}>0.5$, whereas the probability of any other ranking π is obtained by replacing some of the p_{ij} terms in the product by $p_{ji}<0.5$. Also, if $\pi \neq e$ and $i<j$ are any items such that $\pi(i)= \pi(j)-1$, then $P(\pi)/P(\pi \circ \tau_{ij})=p_{ij}/p_{ji}>1$, which completes the proof that the induced Babington Smith model is strongly unimodal.

Conversely, suppose the model is strongly unimodal with modal ranking π_0, and that no $p_{ij}=0.5$. Without loss of generality, $\pi_0=e$. For any items $i<j$, let π be a permutation such that $\pi(i)=\pi(j)-1$. By strong unimodality,

$$1 \leqslant \frac{P(\pi)}{P(\pi \circ \tau_{ij})} = \frac{p_{ij}}{p_{ji}} = \frac{p_{ij}}{1-p_{ij}},$$

whence $p_{ij} \geqslant 0.5$. Thus $p_{ij}>0.5$ whenever $i<j$ (and $p_{ij}<0.5$ whenever $i>j$), so the $\{p_{ij}\}$ are weakly stochastically transitive. ∎

Proof of Theorem 3. Suppose the $\{p_{ij}\}$ are strongly stochastically transitive and no $p_{ij}=0.5$. Then, as argued in the proof of Theorem 2, the induced ranking model is strongly unimodal; without loss of generality $\pi_0=e$ and $p_{ij}>0.5$ for all $i<j$. For any permutation π and any $i<j$ such that $\pi(i)<\pi(j)$,

$$\begin{aligned} \frac{P(\pi)}{P(\pi \circ \tau_{ij})} &= \frac{p_{ij}}{p_{ji}} \prod_{\substack{m: \\ \pi(i)<\pi(m)<\pi(j)}} \left(\frac{p_{im}\, p_{mj}}{p_{jm}\, p_{mi}}\right) \\ &= \frac{p_{ij}}{1-p_{ij}} \prod_{\substack{m: \\ \pi(i)<\pi(m)<\pi(j)}} \left(\frac{p_{im}(1-p_{jm})}{p_{jm}(1-p_{im})}\right). \end{aligned}$$

To show that this expression is $\geqslant 1$ (which is the complete consensus condition), note $p_{ij}/(1-p_{ij})>1$ since $p_{ij}>0.5$; thus it suffices to show that $p_{im} \geqslant p_{jm}$ for each m. There are three cases: (i) if $m<i$ then strong stochastic transitivity implies $p_{mj} \geqslant \max\{p_{mi}, p_{ij}\} \geqslant p_{mi}$, whence $p_{im} \geqslant p_{jm}$; (ii) if $i<m<j$ then $p_{im} \geqslant 0.5 \geqslant p_{jm}$; (iii) if $m>j$ then strong stochastic transitivity implies $p_{im} \geqslant \max\{p_{ij}, p_{jm}\} \geqslant p_{jm}$. Hence the model has complete consensus. ∎

Proof of Theorem 4. Without loss of generality, the items may be relabelled so that $\pi_0=e$. Suppose d is additively decomposable. By Theorem 1, to verify L-decomposability, it is sufficient to show that (3) is a symmetric function of $i_1, \ldots, i_{r-1}$. Using the same notation and reasoning as in Eq. (25),

$$\begin{aligned} &P(\pi^{-1}(r)=i_r \mid \pi^{-1}(1)=i_1, \ldots, \pi^{-1}(r-1)=i_{r-1}) \\ &\quad = \frac{\sum_{v \in C_r} P(v)}{\sum_{v \in C_{r-1}} P(v)} = \frac{\sum_{v \in C_r} \exp[-\lambda d(v, e)]}{\sum_{v \in C_{r-1}} \exp[-\lambda d(v, e)]} \\ &\quad = \frac{\sum_{v \in C_r} \exp(-\lambda\{f_r[v^{-1}(1), \ldots, v^{-1}(r-1)] + g_r[v^{-1}(r), \ldots, v^{-1}(k)]\})}{\sum_{v \in C_{r-1}} \exp[-\lambda\{f_r[v^{-1}(1), \ldots, v^{-1}(r-1)] + g_r[v^{-1}(r), \ldots, v^{-1}(k)]\}} \\ &\quad = \frac{\sum_{v \in C_r} \exp(-\lambda\{f_r[i_1, \ldots, i_{r-1}] + g_r[i_r, v^{-1}(r+1), \ldots, v^{-1}(k)]\})}{\sum_{v \in C_{r-1}} \exp(-\lambda\{f_r[i_1, \ldots, i_{r-1}] + g_r[v^{-1}(r), \ldots, v^{-1}(k)]\}} \\ &\quad = \frac{\sum_{v \in C_r} \exp\{-\lambda g_r[i_r, v^{-1}(r+1), \ldots, v^{-1}(k)]\}}{\sum_{v \in C_{r-1}} \exp\{-\lambda g_r[v^{-1}(r), \ldots, v^{-1}(k)]\}}. \end{aligned}$$

The two sums in the above expression do not depend on the relative order of $i_1, ..., i_{r-1}$.

Conversely, suppose a distance-based ranking model is L-decomposable. From Definition (1) and Eq. (15), for any ranking π with associated ordering $\langle i_1, ..., i_k \rangle$,

$$P(\pi) = C(\lambda) \exp\{-\lambda d(\pi, e)\} = \prod_{j=1}^{k-1} P_{\{i_j, ..., i_k\}}(i_j).$$

Fix $r \in \{2, ..., k\}$. Then

$$C(\lambda) \exp\{-\lambda d(\pi, e)\} = \prod_{j=1}^{r-1} P_{\{i_j, ..., i_k\}}(i_j) \prod_{j=r}^{k-1} P_{\{i_j, ..., i_k\}}(i_j).$$

Taking the natural logarithm of both sides yields

$$d(\pi, e) = \text{constant} - \frac{1}{\lambda} \sum_{j=1}^{r-1} \log P_{\{i_j, ..., i_k\}}(i_j) - \frac{1}{\lambda} \sum_{j=r}^{k-1} \log P_{\{i_j, ..., i_k\}}(i_j).$$

Rewriting the first sum as

$$-\frac{1}{\lambda} \sum_{j=1}^{r-1} \log P_{\{i_1, ..., i_{j-1}\}^c}(i_j)$$

gives the desired decomposition of the form (18). ∎

Acknowledgments

The authors thank Professor A. A. J. Marley and a referee for very helpful suggestions, and Arthur Silverberg and the Educational Testing Service for providing the word association data. The work of the first author was supported in part by National Science Foundation Grant DMS-86-01645.

References

Babington Smith, B. (1950). Discussion on Professor Ross's paper. *Journal of the Royal Statistical Society, Series B*, **12** 153–162.

Berlekamp, E. R. (1968). *Algebraic coding theory*. New York: McGraw–Hill.

Bradley, R. A., & Terry, M. A. (1952). Rank analysis of incomplete block designs, I. *Biometrika*, **39**, 324–345.

Brook, D., & Upton, G. J. G. (1974). Biases in local government elections due to position on the ballot paper. *Applied Statistics*, **23**, 414–419.

Critchlow, D. E. (1985). *Metric methods for analyzing partially ranked data*. Lecture Notes in Statistics, Vol. 34. New York: Springer-Verlag.

Daniels, H. E. (1950). Rank correlation and population models. *Biometrika*, **33**, 129–135.

David, H. A. (1988). *The method of paired comparisons*. Second ed. New York: Oxford Univ. Press.

Diaconis, P. (1988). *Group representations in probability and statistics*. Hayward: Institute of Mathematical Statistics.

Diaconis, P. (1989). A generalization of spectral analysis with application to ranked data. In The Wald Memorial Lectures. *Annals of Statistics*, **17**, 949–979.

Feigin, P., & Cohen, A. (1978). On a model for concordance between judges. *Journal of the Royal Statistical Society, Series B*, **40**, 203–213.

Fienberg, S. E. & Larntz, K. (1976). Log-linear representation for paired and multiple comparison models. *Biometrika*, **63**, 245–254.

FLIGNER, M. A., & VERDUCCI, J. S. (1986). Distance-based ranking models. *Journal of the Royal Statistical Society, Series B*, **83**, 859–869.

FLIGNER, M. A., & VERDUCCI, J. S. (1988). Multi-stage ranking models. *Journal of the American Statistical Association*, **83**, 892–901.

GORDON, A. D. (1979). A measure of the agreement between rankings. *Biometrika*, **66**, 7–15.

HARVILLE, D. A. (1973). Assigning probabilities to the outcomes of multi-entry competitions. *Journal of the American Statistical Association*, **68**, 312–316.

HENERY, R. J. (1981). Permutation probabilities as models for horse races. *Journal of the Royal Statistical Society, Series B*, **43**, 86–91.

HENERY, R. J. (1983). Permutation probabilities for gamma random variables. *Journal of Applied Probability*, **20**, 822–834.

HOLLANDER, M., PROSCHAN, F., & SETHURAMAN, J. (1977). Function decreasing in transposition and their application in ranking problems. *Annals of Statistics*, **5**, 722–733.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

KENDALL, M. G. (1950). Discussion on symposium on ranking methods. *Journal of the Royal Statistical Society, Series B*, **12**, 189.

LUCE, R. D. (1959). *Individual choice behavior.* New York: Wiley.

MALLOWS, C. L. (1957). Non-null ranking models, I. *Biometrika*, **44**, 114–130.

MARLEY, A. A. J. (1968). Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, **5**, 311–332.

MARLEY, A. A. J. (1982). Random utility models with all choice probabilities expressible as 'functions' of the binary choice probabilities. *Mathematical Social Sciences*, **3**, 39–56.

MARTIN-LÖF, P. (1973). *Statistika Modeller.* Lecture notes in Swedish, compiled by Rolf Sundberg, Stockholm University.

MOSTELLER, F. (1951). Remarks on the method of paired comparisons. I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, **16**, 3–9.

PENDERGRASS, R. N., & BRADLEY, R. A. (1960). Ranking in triple comparisons. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 331–351). Palo Alto: Stanford Univ. Press.

PLACKETT, R. L. (1975). The analysis of permutations. *Applied Statistics*, **24**, 193–202.

RUKHIN, A. (1972). Estimation of a rotation parameter on a sphere. *Zapiski Nauchnykh Seminarov LOMI*, **29**, 74–91 [in Russian]. English translation (1975): *Journal of Soviet Mathematics*, **3**, 777–791.

SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics—the two-sample case. *Annals of Mathematical Statistics*, **27**, 590–615.

SAVAGE, I. R. (1957). Contributions to the theory of rank order statistics—the "trend" case. *Annals of Mathematical Statistics*, **28**, 968–977.

SPEARMAN, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.

STERN, H. S. (1987). *Gamma processes, paired comparisons and ranking.* Technbical Report No. 64, Department of Statistics, Stanford University.

THURSTONE, L. L. (1927). A law of comparative judgement. *Psychological Reviews*, **34**, 273–286.

ULAM, S. M. (1972). Some ideas and propects in biomathematics. *Annual Review of Biophysics and Bioengineering*, **1**, 277–292.

ULAM, S. M. (1981). Future applications of mathematics in the natural sciences. In *American mathematical heritage: Algebra and applied mathematics* (pp. 101–114). Texas Tech. University Mathematics Series, Vol. **13**.

YELLOTT, J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, **15**, 109–144.

RECEIVED: July 24, 1989