J. R. Statist. Soc. B (1986)
48, *No.* 3, *pp.* 359–369

Distance based Ranking Models

By M. A. FLIGNER† and J. S. VERDUCCI

Ohio State University, USA

[Received November 1985. Revised April 1986]

SUMMARY

A class of ranking models is proposed for which the probability of a ranking decreases with increasing distance from a modal ranking. Some special distances, namely those associated with Kendall and Cayley, decompose into a sum of independent components under the uniform distribution. These distances lead to multiparameter generalizations whose parameters may be interpreted as information at various stages in a ranking process. Estimation of model parameters is described, and the results are applied to an example of word associations. A censoring argument motivates simple extensions of these models to include partial rankings. The generalized Cayley distance model is illustrated for random arrangements arising from mechanisms other than ranking.

Keywords: CYCLIC STRUCTURE; DISCORDANT PAIRS; EXPONENTIAL FAMILIES; NON-NULL RANKING MODELS; PARTIAL RANKINGS; PERMUTATIONS

1. INTRODUCTION

Suppose that a set of k objects is to ranked independently by each of a panel of n judges, and that each ranking proceeds according to some criterion that presumes a true underlying ordering of the objects. Mallows (1957) introduces two one-parameter models to describe such a ranking process, the parameter reflecting the variability of the rankings about the true or modal ordering. The models are referred to as Mallows' θ and ϕ models, which are based on the correlation coefficients of Spearman (1904) and Kendall (1938), respectively. The ϕ model is further investigated in Feigin and Cohen (1978), and Diaconis (1982) suggests use of the Mallows' models with other distances. Critchlow (1985) considers metric-preserving extensions of these distances as a basis for models on partial rankings.

In Section 2 we examine the Mallows-type model for a general distance, and then focus on those distances for which the model factors into components representing independent stages in the ranking process. This factorization simplifies estimation in the one parameter model and leads to natural $k-1$ parameter extensions. In Section 3 Mallows' ϕ model is reviewed, and its $k-1$ parameter extensions are presented in detail. An example of word associations illustrates the necessary calculations. In Section 4 we consider one and $k-1$ parameter models utilizing Cayley's (1849) distance. In both Sections 3 and 4, censoring-induced models for the partial rankings are developed. Section 5 contains a discussion of how the various parameters may be interpreted as aspects of concordance.

2. GENERALIZED MALLOWS' MODELS

Consider the situation in which each of several judges is to rank k items in decreasing order of preference, thereby producing a ranking $\pi = (\pi(1), \ldots, \pi(k))$, where $\pi(i)$ is the rank given to item i, $i = 1, \ldots, k$. Let Ω be the set of all $k!$ possible rankings π, and let $d(\cdot, \cdot)$ be a discrepancy function on $\Omega \times \Omega$, that is $d(\sigma, \pi) \geqslant 0$ for every $\sigma, \pi \in \Omega$ with equality if and only if $\sigma = \pi$. To ensure that the value of d does not change when the labels of the items are permuted, we assume that the discrepancy functions are right invariant, that is $d(\pi\nu, \sigma\nu) = d(\pi, \sigma)$ for every π,

† *Address for correspondence*: Department of Statistics, The Ohio State University, 141 Cockins Hall, 1958 Neil Ave., Columbus, Ohio 43210, USA.

0035–9246/86/48359 $2.00

σ, $\nu \in \Omega$. Although these are the only properties assumed, many of the discrepancy functions considered are in fact right invariant metrics.

The judges' rankings are assumed to be generated according to the model

$$P_\theta(\pi) = [\exp\{-\theta d(\pi, \pi_0)\}]/\psi(\theta), \quad \pi \in \Omega, \quad \theta \in R, \tag{2.1}$$

where $d(\cdot, \cdot)$ is a discrepancy, function π_0 is a fixed ranking, and $\psi(\theta)$ is a normalizing constant. When $\theta > 0$, the ranking π_0 is the modal ranking and when θ approaches infinity P_θ becomes concentrated at the single ranking π_0. When $\theta = 0$, P_θ is the uniform distribution, and for $\theta < 0$, π_0 is an antimode. The special cases of (2.1) with $d(\cdot, \cdot)$ being Kendall's or Spearman's distance were first investigated by Mallows (1957), and the general probability model (2.1) has been suggested in Diaconis (1982).

Since all discrepancy functions considered are right invariant, $d(\pi, \pi_0) = d(\pi\pi_0^{-1}, e)$ with $e = (1, \ldots, k)$, and we assume in what follows that $\pi_0 = e$. Thus, to simplify the notation, we let $D(\nu) = d(\nu, e)$ denote the discrepancy function which defines(2.1), where in the original notation $\nu = \pi\pi_0^{-1}$. When we consider ν as a random permutation generated according to some probability distribution, then $\mathrm{D} = \mathrm{D}(\nu)$ is a random variable. The use of $D(\nu)$ as a discrepancy function defining (2.1) or as a random variable will be clear from the context. If ν is a random permutation generated according to (2.1) then, with some abuse of terminology we refer to the probability distribution of $D(\nu)$ as the distribution of D under (2.1). Finally, the moment generating function of D under (2.1) is denoted by $M_{D,\theta}(t)$, and for simplicity the moment generating function of D under the uniform distribution is denoted by $M(t) = M_{D,0}(t)$.

The major difficulty in working with the general model (2.1) is that $\psi(\theta)$ may be difficult to evaluate explicitly. A first step to finding a simple form for $\psi(\theta)$ is to relate it to the moment generating function $M(t)$ of D, namely $\psi(\theta) = k!M(-\theta)$. This follows as in Mallows (1957), since

$$\psi(\theta) = \sum_{\pi \in \Omega} \exp\{-\theta D(\pi)\} = k! \sum_{d_i} P_0(\{D = d_i) \exp\{-\theta d_i\} = k! M(-\theta).$$

Because model (2.1) is an exponential family, the moment generating function of D under (2.1) is $M_{D,\theta}(t) = M(t - \theta)/M(-\theta)$. A direct argument (see Lehmann (1983), p. 31) shows that

$$E_\theta[D] = \left.\frac{d[\log\{M(t)\}]}{dt}\right|_{t=-\theta} \tag{2.2}$$

and

$$\mathrm{var}_\theta[D] = \left.\frac{d^2[\log\{M(t)\}]}{dt^2}\right|_{t=-\theta}$$

are the mean and variance of D.

Suppose that each of a sample of n judges produces a ranking π_i according to model (2.1). Let $D_i \equiv D(\pi_i)$ and $\bar{D} = \Sigma D_i/n$. Then the m.l.e. $\hat{\theta}$ of θ is the solution of the equation $E_{\hat{\theta}}[D] = \bar{D}$. In addition, the limiting distribution of $n^{1/2}(\hat{\theta} - \theta)$ as $n \to \infty$ is normal with mean 0 and variance $\{\mathrm{var}_\theta[D]\}/\{(E'_\theta[D]\}^2$. Thus we see that the properties of the model (2.1) are intimately connected with the moment generating function of D under the uniform distribution.

Although no general theory exists for deriving $M(t)$ for any distance, there are some interesting cases in which closed form expressions for $M(t)$ can be found. One such case is the Hamming distance, defined below. Some properties of this distance are discussed in Diaconis (1982). In particular, this distance is right invariant.

Example. (Hamming distance)
Let

$$\varepsilon_i(\pi) = \begin{cases} 1 & \text{if } \pi(i) = i, \\ 0 & \text{otherwise,} \end{cases} \qquad i, = 1, \ldots, k$$

and define

$$X(\pi) = \sum_{i=1}^{k} \varepsilon_i(\pi)$$

to be the number of items that receive the same rank under π and e. The Hamming distance is defined by $D_H(\pi) \equiv k - X(\pi)$.

Suppose that π has the uniform distribution on Ω and that $f(t) = E[t^X]$ so that the jth derivative of $f(t)$ evaluated at $t = 1$ is

$$f^{(j)}(1) = E[X(X-1)\cdot\ldots\cdot\{X-(j-1)\}], \quad j = 1, \ldots, k.$$

The expansion of $f(t)$ in a Taylor series around $t = 1$ is

$$f(t) = \sum_{j=0}^{k} \{f^{(j)}(1)(t-1)^j\}/j! = \sum_{j=0}^{k} E[C(X,j)](t-1)^j,$$

where the binomial coefficient $C(X,j)$ is zero when $X < j$. This coefficient may be re-expressed as

$$C(X,j) = \sum_A \varepsilon_{i_1}\cdot\ldots\cdot\varepsilon_{i_j}$$

where $A = \{(i_1, \ldots, i_j) \mid 1 \leqslant i_i < i_2 < \ldots < i_j \leqslant k\}$. Consequently,

$$E[C(X,j)] = \sum_A E[\varepsilon_{i_1}\cdot\ldots\cdot\varepsilon_{i_j}] = \sum_A \{(k-j)!\}/k! = C(k,j)\{(k-j)!\}/k! = 1/j!.$$

Hence

$$f(t) = \sum_{j=0}^{k} \{(t-1)^j\}/j!,$$

and the moment generating function of D is

$$M(t) = E[\exp\{(k-X)t\}] = \exp(kt) f\{\exp(-t)\} = \exp(kt)\,\Sigma\,[\{\exp(-t)-1\}^j]/j!. \,|$$

The moment generating function $M(t)$ may sometimes be derived via a decomposition of the distance $D(\pi)$. This method is useful for both Kendall's and Cayley's distances discussed in the next two sections. If a distance D can be expressed as

$$D(\pi) = \sum_{j=1}^{p} S_j(\pi), \tag{2.3}$$

where $S_j(\pi)$ are independent with moment generating functions $M_j(t)$ when π is uniformly distributed, then $M(t) = \Pi M_j(t)$. Of course, for this approach to be useful, $M_j(t)$ must not be too complicated. The fact that $\{S_j(\pi);\, j = 1, \ldots, p\}$ continue to be independent under the model (2.1) becomes apparent in the following more general context.

An arbitrary set of random variables $S_j(\pi)$, $j = 1, \ldots, p$, which are independent under the uniform distribution determine a p-parameter family of probability models on Ω, namely

$$P_{\boldsymbol{\theta}}(\pi) = [\exp\{-\Sigma\, \theta_j S_j(\pi)\}]/\{k!\, \Pi\, M_j(-\theta_j)\} \quad \theta_j \in R, j = 1, \ldots, p. \tag{2.4}$$

The family of models (2.4) contains the motivating distance model (2.1) with $D(\pi) = \Sigma S_j(\pi)$ and $\theta_j = \theta$, $j = 1, \ldots, p$. Since model (2.4) is a p-parameter exponential family, the joint moment generating function of $S = (S_1, \ldots, S_p)$ under this model can be obtained easily and shown to be the product of the moment generating functions of the components. This yields the independence of the random variables $S_1, \ldots, S_p$ under model (2.4). The mean and variance of S_j under model (2.4) are then given by formulae (2.2) with $M = M_j$ and $\theta = \theta_j$.

The definition of $D(\pi)$ has been simplified by assuming that $\pi_0 = e$, namely that the items have been relabelled in decreasing order of preference. The dependence of the likelihood on π_0 is now made explicit to illustrate the estimation problem for π_0 unknown. Suppose the true ranking is π_0 and a sample of judges produce rankings π_i, $i = 1, \ldots, n$, where $\pi_i \pi_0^{-1}$ are independent observations from (2.4), and $\bar{S}_j(\pi_0) = \Sigma_i S_j(\pi_i \pi_0^{-1})/n$. The solution $\hat{\theta}_j(\pi_0)$ to

$$\bar{S}_j(\pi_0) = E_{\theta_j}[S_j],$$

is the m.l.e. of θ_j for π_0 known. In some cases to be considered $\hat{\theta}_j(\pi_0)$ can be evaluted in closed form. If π_0 is unknown, then $\hat{\pi}_0$ is the value of π_0 that minimizes

$$\sum \log[M_j\{-\hat{\theta}_j(\pi_0)\}] + n \sum \hat{\theta}_j(\pi_0)\bar{S}_j(\pi_0)$$

and $\hat{\theta}_j$ is $\hat{\theta}_j(\hat{\pi}_0)$. The next two sections illustrate these general methods.

3. MEASURES BASED ON DISCORDANT PAIRS

In this section we begin with a few basic properties of Mallows' ϕ model. This model is extended to a $(k - 1)$ parameter model as in (2.4) based on a multi-stage ranking process, and a numerical example is provided. The model for partial rankings induced by censoring is then derived.

Let $\pi^{-1} = \langle \pi^{-1}(1), \ldots, \pi^{-1}(k) \rangle$ denote the inverse permutation, so that $\pi^{-1}(i)$ is the item receiving rank i. We first define a vector $V = (V_1, \ldots, V_{k-1})$ which is a 1-1 function of π and whose components are independent when π is uniform on Ω. Specifically, if

$$V_j(\pi) = \sum_{l>j} I(\pi^{-1}(j) - \pi^{-1}(l)), \quad j = 1, \ldots, k-1$$

where $I(a) = 1$ if $a > 0$ and 0 otherwise, then V_j has the uniform distribution on the integers 0, 1,. . . ,k-j (see Feller (1968), p. 257), and

$$D_K(\pi) = \sum_{j=1}^{k-1} V_j(\pi) \tag{3.1}$$

is the number of discordant pairs between the rankings π and $\pi_0 = e$. Note, V_1 is the number of adjacent transpositions required to place item 1 in the first position; V_2 is the number of additional transpositions necessary to place item 2 in the second position, and so forth. Thus the distance $D_K(\pi)$ is also the minimum number of adjacent transpositions required to put π in the natural order.

When $d(\pi, \pi_0) = D_K(\pi)$ model (2.1) becomes Mallows' ϕ model, which was motivated from the Smith (1950) model for paired comparisons. In this case, the parameters (θ, π_0) and $(-\theta, \bar{e}\pi_0)$ index the same probability model in (2.1), where $\bar{e} = (k, k-1, \ldots, 1)$. To identify the parameters exactly, it will be assumed that $\theta \geqslant 0$ and π_0 represents the modal ranking.

For Mallows' ϕ model expression (2.2) gives

$$E_\theta[D_K] = \frac{k \exp(-\theta)}{1 - \exp(-\theta)} - \sum_{j=1}^{k} \frac{j \exp(-j\theta)}{1 - \exp(-j\theta)}, \tag{3.2}$$

and

$$\mathrm{var}_\theta[D_K] = \frac{k \exp(-\theta)}{\{1 - \exp(-\theta)\}^2} - \sum_{j=1}^{k} \frac{j^2 \exp(-j\theta)}{\{1 - \exp(-j\theta)\}^2}.$$

Because the m.l.e. $\hat{\theta}$ of θ is the solution to the equation $\bar{D}_K = E_{\hat{\theta}}(D_K)$ and $E_\theta D_K$ is monotone decreasing, a standard line search quickly converges to the m.l.e. Feigin and Cohen (1978) have tabled (3.2) for $k = 1(1), 10$, and their tables can be used to obtain an approximate solution. Note that their parameter θ is our $e^{-\theta}$.

3.1. *The k–1 Parameter Models*

An application of the extension (2.4) to the decomposition (3.1) leads to the model

$$P_{\boldsymbol{\theta}}(\pi) = \left[\exp\left(-\sum_{j=1}^{k-1} \theta_j V_j\right)\right] \Big/ \psi(\boldsymbol{\theta}), \tag{3.3}$$

where

$$\psi(\boldsymbol{\theta}) = \sum_{j=1}^{k-1} [1 - \exp\{-(k-j+1)\theta_j\}]/\{1 - \exp(-\theta_j)\}.$$

The parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k-1})$ can be interpreted by viewing the ranking process as a sequence of independent stages. In the first stage, the judge considers the k items and selects what he views as the best to be ranked first. In the second stage, he considers the remaining $k-1$ items and selects what he views as the best remaining item to be ranked second, and so forth. Selecting the best of the remaining items at the jth stage corresponds to the event $V_j = 0$. Inverting the moment generating function of V_j gives

$$P_{\boldsymbol{\theta}}(V_j = 0) = \{1 - \exp(-\theta_j)\}/[1 - \exp\{-(k-j+1)\theta_j\}]. \tag{3.4}$$

Thus, the probability of making the correct decision at stage j is an increasing function of θ_j. The event $V_j = l, l = 1, \ldots, k-j$, occurs if the judge selects the $(l+1)$'st best of the remaining items at stage j and these probabilities decrease geometrically, that is,

$$P_{\boldsymbol{\theta}}(V_j = l) = [\exp(-l\theta_j)\{1 - \exp(-\theta_j)\}]/[1 - \exp\{-(k-j+1)\theta_j\}],\ l = 0, 1, \ldots, k-j. \tag{3.5}$$

The assumption $\theta_1 = \ldots = \theta_{k-1}$ yields Mallows' model, and implies in a certain sense that the information available to the judge is fixed at the beginning of the ranking process. Allowing the θ_j's to vary allows the judge to reevaluate the items in a new light as he makes the jth decision. Since the pairs $(\boldsymbol{\theta}, \pi_0)$ and $(-\boldsymbol{\theta}, \bar{e}\pi_0)$ yield the same model (3.3), we assume $\theta_1 \geqslant 0$. Thus, large values of the parameters reflect greater agreement between the judges' rankings and the basic ordering π_0. If the judges' stepwise ranking process proceeds instead by first eliminating the worst item and so forth until only the best item remains, a redefinition of the V's allows this process to be modelled as well.

Estimation of $\boldsymbol{\theta}$ follows the ideas of Section 2. Let π_0 be the true ranking, π_i be the ranking of judge i, $i = 1, \ldots, n$, $V_{ij} = V_j(\pi_i \pi_0^{-1})$ and $\bar{V}_j(\pi_0) = (\Sigma_i V_{ij})/n$. Then $\hat{\theta}_j(\pi_0)$ is the solution to the equation

$$\bar{V}_j(\pi_0) = \frac{\exp(-\hat{\theta}_j))}{1 - \exp(-\hat{\theta}_j)} - \frac{(k-j+1)\exp\{-(k-j+1)\hat{\theta}_j\}}{1 - \exp\{-(k-j+1)\hat{\theta}_j\}}. \tag{3.6}$$

This procedure gives the m.l.e. of θ_j when π_0 is known. For π_0 unknown, $\hat{\pi}_0$ and $\hat{\theta}_j$ are obtained simultaneously as in section 2.

Model (3.3) has applications in testing the goodness of fit of Mallows' ϕ model. When k is small, for each ranking the expected number can be estimated under Mallows' ϕ model and compared with the observed number. For large k, Cohen and Mallows (1983) suggest pooling the rankings with a common value of $D_K(\pi)$, and for each value of $D_K(\pi)$, comparing the observed and the expected number under Mallows' ϕ model. As they have noted, this approach is not sensitive to departures from the model when rankings with the same value of D_K have different probabilities. In this case the log likelihood ratio test statistic for H_0:

$\theta_1 = \ldots = \theta_k$ versus the general alternative in model (3.3) could serve as a goodness of fit statistic. This statistic takes the form

$$2n\left[\log \prod_{j=1}^{k-1} \frac{M_j(-\theta)}{M_j(-\theta_j)} + \sum_{j=1}^{k-1} (\hat{\theta} - \hat{\theta}_j)V_j\right], \tag{3.7}$$

where $\hat{\theta}$ is the m.l.e. of the common value of θ_j under H_0. For large n, (3.7) can be referred to a χ^2 distribution with k-2 degrees of freedom.

We now introduce an alternative k-1 parameter model utilizing the V_j's and a coarser discrepancy function which does not incorporate the structure (3.5) on the distribution of the V_j's and has the advantage of closed form expressions for the MLE's. Let $I_j \equiv I(V_j) = 1$ if $V_j > 0$ and 0 otherwise, and consider the model

$$P_{\boldsymbol{\theta}}(\pi) = [\exp(-\Sigma\theta_j I_j)]/\psi(\boldsymbol{\theta}), \tag{3.8}$$

with

$$\psi(\boldsymbol{\theta}) = \prod_{j=1}^{k-1} \{1 + (k-j)\exp(-\theta_j)\}$$

and

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k-1}) \in R^{k-1}.$$

In terms of the stepwise ranking process, the probability of selecting the best of the remaining items at stage j is now given by

$$P_{\boldsymbol{\theta}}(I_j = 0) = P_{\boldsymbol{\theta}}(V_j = 0) = \{1 + (k-j)\exp(-\theta_j)\}^{-1}. \tag{3.9}$$

Let π_0 be the true ranking, π_i be the ranking of judge i ($i = 1, \ldots, n$), $I_{ij} = I_j(\pi_i\pi_0^{-1})$, and $\bar{I}_j = \Sigma I_{ij}/n$. Then $\hat{\theta}_j = -\log[\bar{I}_j/\{(k-j)(1-\bar{I}_j)\}]$. For π_0 unknown, $\hat{\pi}_0$ and $\hat{\theta}_j$ are obtained simultaneously as in Section 2, but the operation is simpler in this case.

To illustrate these models, we examine data collected under the auspices of the Graduate Record Examination Board. A sample of 98 college students were asked to rank five words according to strength of association with a target word. Note that association does not necessarily mean similarity, so that the word "poor" might be more strongly associated with the target word "rich" than would be the word "wealthy."

For the target word "idea," the five choices were (1) thought, (2) play, (3) theory, (4) dream, and (5) attention. Each student independently ranked these words from 1-least associated to 5-most associated. Table 1 lists the frequencies of the sampled rankings for the 15 distinct rankings reported. The modal ranking is $\pi_0 = (5, 1, 4, 3, 2)$, indicating "thought" to be most associated with "idea." The first two observations in the table are suspect, because these two subjects appear to have ranked the items in reverse order.

TABLE 1
Rankings of words associated with "Idea"

Ranking	*Frequency*	*Ranking*	*Frequency*	*Ranking*	*Frequency*
(1 3 4 5 2)	1	(4 2 3 5 1)	2	(5 1 4 2 3)	6
(1 4 2 3 5)	1	(4 3 5 2 1)	1	(5 1 4 3 2)	33
(3 2 5 4 1)	2	(5 1 2 4 3)	5	(5 2 3 4 1)	8
(4 1 2 5 3)	1	(5 1 3 2 4)	2	(5 2 4 1 3)	1
(4 1 5 3 2)	5	(5 1 3 4 2)	18	(5 2 4 3 1)	12

We recoded all the data to conform with the decreasing rank convention; otherwise the stages of the model would be reversed, and the least related item would be chosen first, etc. Table 2 contains the results from fitting Mallows ϕ model, model (3.3), model (3.8), and a one-parameter version of model (3.8) where all parameters are constrained to be equal.

TABLE 2
Summary of model fitting for word rankings

Model	θ	θ_1	θ_2	θ_3	θ_4	*Log-likelihood*
Mallows' ϕ	1.42					−251.27
Model (3.3)		1.82	1.16	1.75	.97	−244.59
Model (3.8)*	2.03					−297.67
Model (3.8)		3.26	1.60	2.33	.97	−274.38

* constrained model

The likelihood ratio test of model (3.3) versus Mallows' ϕ model gives a test statistic $\chi^2 = 13.6$. Similarly the likelihood ratio test of Model (3.8) versus its constrained version yields $\chi^2 = 46.6$. Thus both these models indicate a significantly better fit when the parameters are not identical. Although the θ_j parameters are not directly comparable, even within the same model, equations (3.4) and (3.9) afford a simple interpretation of these parameters in terms of the probability of making the "best" decision at each stage. The probabilities are displayed in Table 3, along with the corresponding chance probabilities.

TABLE 3
Probability of best decision at each stage

Model	*Stage* 1	*Stage* 2	*Stage* 3	*Stage* 4
Model (3.3)	.837	.693	.831	.724
Model (3.8)	.867	.622	.837	.724
Uniform	.200	.250	.333	.500

The probabilities estimated under model (3.8) are close to those of model (3.3), even though model (3.3) fits the data much better. Under either model, it is clear that stage 1 shows the greatest improvement over chance agreement, and stage 4 shows the least.

The data were reanalysed without the two suspect observations. All log-likelihoods increased by a sum of about 20, keeping the goodness-of-fit statistics for comparing models essentially the same. Similarly, the estimates of each parameter increased slightly, leaving the interpretation of each model essentially the same.

3.2. *Partial Rankings*

If a judge reports only his top $r < k$ preferences, this partial ranking is denoted by $\pi^* = <\pi^{-1}(1), \ldots, \pi^{-1}(r)>$ and the set of all partial rankings is denoted Ω^*. The partial ranking π^* may be viewed as a censored observation from the distribution (3.3). In that case the probability of observing π^* is the same as the probability that the full ranking is in the set $S\pi^*$ of all $\pi \in \Omega$ consistent with π^*. Thus

$$\begin{aligned} P_{\boldsymbol{\theta}}(\pi^*) &= \sum_{\pi \in S\pi^*} \exp\left\{- \sum_{j=1}^{k} \theta_j V_j(\pi)\right\} \Big/ \psi(\boldsymbol{\theta}) \\ &= \exp\left\{- \sum_{j=1}^{r} \theta_j V_j(\pi^*)\right\} \sum_{\pi \in S\pi^*} \exp\left\{- \sum_{j>r} \theta_j V_j(\pi)\right\} \Big/ \psi(\boldsymbol{\theta}) \end{aligned} \tag{3.10}$$

since $V_1, \ldots, V_r$ depend on $\pi \in S\pi^*$ only through π^*. As π takes its values in the set $S\pi^*$, the vector $(V_{r+1}, \ldots, V_{k-1})$ takes all its $(k-r)!$ possible values, thereby making the last sum in

(3.10) independent of π^* and a function of r and θ alone. Thus the induced model in the partial rankings can be simply expressed for each $\pi^* \in \Omega^*$ as

$$P_{\boldsymbol{\theta}}(\pi^*) = \exp\left\{-\sum_{j=1}^{r} \theta_j V_j(\pi^*)\right\} \Big/ \psi^*(\boldsymbol{\theta}). \tag{3.11}$$

Since (3.11) is again an exponential family, its properties follow directly from the results of sections 2 and 3.1. In particular

$$\psi^*(\boldsymbol{\theta}) = \prod_{j=1}^{r} [1 - \exp\{-(k-j+1)\theta_j\}]/\{1 - \exp(-\theta_j)\}$$

depends on $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}^* = (\theta_1, \ldots, \theta_r)$ and the maximum likelihood estimator $\hat{\theta}_j$ of θ_j is the solution to equation (3.6).

If the modal ranking π_0 is unknown, it may be estimated first as in section 2. Note that the sufficient statistics $(\bar{V}_1, \ldots, \bar{V}_r)$ contain information about the full modal ranking π_0 and not just the partial ranking π_0^*. To see this, let $e^* = \langle 1, \ldots, r \rangle$, $e_1 \in Se^*$ and $V_j^+(\pi) = V_j(\pi e_1^{-1})$. Although V_j^+ has the same distribution as V_j under (3.11),

$$\sum_{j=1}^{r} \theta_j V_j^+(\pi) \neq \sum_{j=1}^{r} \theta_j V_j(\pi), \tag{3.12}$$

and so $P_{\boldsymbol{\theta}}\{(\pi e_1^{-1})^*\} \neq P_{\boldsymbol{\theta}}(\pi^*)$ in general. On the other hand, equality does hold in (3.12) if V_j is first replaced by its indicator function $I_j = I(V_j) = I(V_j^+)$. Thus only the partial center π_0^* can be estimated from $\{\pi_i^*\}$ in the partial ranking version of model (3.8).

Rearrangement of the probabilities assigned under the model (3.8) yields a $(k-1)$ parameter extension of the basic model (2.1) for $D(\pi)$ chosen to be Cayley's distance, as we shall now discuss.

4. MODELS BASED ON CYCLIC STRUCTURE

Consideration of a ranking π as a rearrangement of items leads to natural models for sorting or encryption. For example, in the previous section, the number of discordant pairs $D_K(\pi)$ between π and e is the minimal number of adjacent transpositions needed to sort the permutation π into the proper order e. Similarly, Cayley's distance $D_C(\pi)$ between π and e may be defined as the minimal number of transpositions, not necessarily adjacent, needed to sort π. Thus $D_C(\pi) \leqslant D_K(\pi)$ for every $\pi \in \Omega$.

Cayley's distance also decomposes into a sum of $k-1$ random variables. The decomposition is based on the cyclic structure of π as follows. For any $\pi \in \Omega$, $j = 1, \ldots, k$ define the cyclic set of π at j by

$$\sigma_\pi(j) = \{\pi^i(j) \mid i = 0, 1, \ldots,\}$$

where

$$\pi^0(j) = j, \qquad \pi^2(j) = \pi(\pi(j)), \text{ etc.}$$

Let

$$X_j(\pi) = \begin{cases} 0 & \text{if } j = \max\{\sigma_\pi(j)\}, \\ 1 & \text{otherwise,} \end{cases} \qquad j = 1, \ldots, k.$$

Then

$$D_C(\pi) = \sum_{j=1}^{k-1} X_j(\pi).$$

When π is uniformly distributed on Ω, Feller (1968, p. 258) shows that the $X_j(\pi)$ are independent Bernoulli random variables with parameters $(k-j)/(k+1-j)$, $j = 1, \ldots, k$. Thus for $D(\pi) = D_C(\pi)$ in model (2.1)

$$\psi(\theta) = \prod_{j=1}^{k-1} (1 + je^{-\theta}),$$

$$E_\theta(D_C) = \sum_{j=1}^{k-1} (je^{-\theta})/(1 + je^{-\theta}),$$

and

$$\mathrm{var}_\theta(D_C) = \sum_{j=1}^{k-1} (je^{-\theta})/(1 + je^{-\theta})^2.$$

In the case where only the first $r < k$ positions are observed as $\pi^* = \langle \pi^{-1}(1), \ldots, \pi^{-1}(r) \rangle$, the truncated distance

$$D_C^*(\pi) = \sum_{j=1}^{r} X_j(\pi)$$

depends on π only through π^*. In fact, $D_C^*(\pi)$ is r minus the number of distinct cyclic sets $\sigma_\pi(j)$ whose elements are all less than or equal to r. Thus D_C^* is the Hausdorff extension of D_C (cf. Critchlow, 1985). It is also easy to show that $D_C^*(\pi)$ is the minimum possible value of $D_C(\pi)$ consistent with π^*.

The extended model (2.4) is particularly easy to use with Cayley's distance because the maximum likelihood estimator $\hat{\theta}_j$ is an elementary function of the data. The full model is

$$P_\theta(\pi) = \exp\left\{ \sum_{j=1}^{k-1} \theta_j X_j(\pi) \right\} \Big/ \prod_{j=1}^{k-1} \{1 + j \exp(-\theta_j)\}. \tag{4.1}$$

If a sample of n observations $\pi_1, \ldots, \pi_n$ is taken from (4.1), the maximum likelihood estimators of the θ_j are

$$\hat{\theta}_j = \log(k - j) - \log\{\bar{X}_j/(1 - \bar{X}_j)\}, \qquad j = 1, \ldots, k - 1,$$

where $\bar{X}_j = \sum_{i=1}^{n} X_j(\pi_i)/n$.

Note that model (4.1) assigns the same probability to any two permutations with the same cyclic sets. Thus (4.1) induces a probability function on the set of partitions $\{\sigma_j(\pi)\}$ of $\{1, \ldots, k\}$ defined by

$$P(\{\sigma_j(\pi)\}) = \prod_{j=1}^{k-1} [|\{\sigma_j(\pi)\}| - 1]!\, P_\theta(\pi). \tag{4.2}$$

Model (4.2) represents the following stagewise clustering procedure. At stage one, item number k is isolated. At stage two, item number $k-1$ is either grouped with item $k(X_{k-1} = 1)$ or starts a new cluster $(X_{k-1} = 0)$. At the rth stage item $k - r + 1$ is either placed into one of the existing clusters $(X_{k-r+1} = 1)$ or starts a new cluster $(X_{k-r+1} = 0)$. Thus a large value of θ_j in model (4.2) indicates that item j is unlike any of the items numbered above it.

The following paradigm illustrates the solution of other problems based on cyclic structure. Suppose there are k lockers, and each locker has exactly one key that will open it. One key is placed in each locker according to some scheme π where $\pi(i) = j$ if the key to the ith locker is placed in locker j $(i, j = 1, \ldots, k)$. Suppose that the cost of breaking into locker j is ϕ_j, where $0 < \phi_1 \leqslant \ldots \leqslant \phi_n$, and let $L(\pi)$ be the minimum cost of opening all lockers.

(1) If π is generated at random, what is the expected minimum cost $E[L(\pi)]$?
(2) If π is generated so that $\log P(\pi) = \theta L(\pi) + c(\theta)$, where θ is a positive constant, what is $E_\theta[L(\pi)]$?

The answer to (1) comes quickly from noticing that in searching backwards through lockers $k, k-1, \ldots, 1$, locker j need be broken if and only if j is the maximal element in $\sigma_j(\pi)$. Thus

$$L(\pi) = \sum_{j=1}^{k} \phi_j Y_j(\pi), \text{ where } Y_j(\pi) = 1 - X_j(\pi),$$

and consequently

$$E[L(\pi)] = \sum_{j=1}^{k} \phi_j/j.$$

For part (2)

$$P_\theta(\pi) = [\exp\{\theta L(\pi)\}]/k!\, M_L(\theta),$$

and

$$E_\theta(L(\pi)) = \frac{d}{d\theta} \log M_L(\theta) = \frac{d}{d\theta} \sum_{j=1}^{k} \log\{(j-1) + \exp(\phi_j\theta)\} = \sum_{j=1}^{k} \frac{\phi_j \exp(\phi_j\theta)}{(j-1) + \exp(\phi_j\theta)}. \tag{4.3}$$

Note that as $\theta \to \infty$, the distribution P_θ becomes concentrated at the identity and $E_\theta\{L(\pi)\} \to \Sigma\phi_j$, its maximum value. The distribution P_θ is the maximal entropy distribution subject to the constraint (4.3); that is, it is the distribution guaranteeing minimum expected cost (4.3) for which it is most difficult to guess the random permutation π. Such distributions may be useful for encryption.

5. DISCUSSION

In this paper, some distance based nonnull ranking models and their $(k-1)$ parameter extensions have been investigated. These extensions provide simple descriptions of the ranking process in terms of independent stages. When the modal ranking π_0 has some external validity, the parameters may be interpreted as information available to the judges at each stage. Alternatively, when the modal ranking represents the most popular selection of a population, the parameters may be interpreted as the amount of concordance within the population at each stage of the ranking process.

Our generalization of the Mallows' ϕ model to partial rankings differs from those considered by Critchlow (1985). Our approach is motivated by censoring, whereas his motivation appears to be the preservation of metric properties. In the case of Cayley's distance, our approach coincides with the Hausdorff extension developed so elegantly in Critchlow's monograph.

In addition to modeling the ranking process of a population of judges, nonnull ranking models are useful in the problem of comparing two populations of judges. These models can be used to clarify some of the issues regarding what it means for two populations of judges to be concordant. In particular two populations may be concordant at certain stages and discordant at others. Finally, the performance of procedures for comparing groups can be easily studied and compared using distance based nonnull ranking models.

6. ACKNOWLEDGEMENT

The work of M. A. Fligner was supported by the Office of Naval Research. The work of J. S. Verducci was supported in part by the U.S. Environmental Protection Agency through contract No. 68-01-6721 through the Office of Toxic Substances. The views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the

Agency. The authors thank the Graduate Record Examination Board for permission to use the primary data from the Pike Report, and Douglas Critchlow for leading us to the data. We are grateful to Ingram Olkin for his critical reading of an early draft of this paper, and to the referees for helpful comments.

REFERENCES

Cayley, A. (1849) A note on the theory of permutations. *Phil. Mag.*, **34**, 527-529.

Cohen, A. and Mallows, C. L. (1983) Assessing goodness of fit of ranking models to data. *Statistician*, **32**, 361-373.

Critchlow, D. (1985) *Metric Methods for Analyzing Partially Ranked Data.* New York: Springer Verlag.

Diaconis, P. (1982) *Group Theory in Statistics.* Harvard University Lecture Notes. To appear in the Institute of Mathematical Statistics Lecture Notes — Monograph Series.

Feigin, P., and Cohen, A. (1978) On a model of concordance between judges. *J. R. Statist. Soc.* B, **40**, 203-213.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications.* Volume 1. Third edition. New York: Wiley.

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

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

Smith, B. Babington (1950) In Discussion on Professor Ross's paper, *J. R. Statist. Soc.* B, **12**, 54.

Spearman, C. (1904) The proof and measurement of association between two things. *Amer. J. Psychol.*, **15**, 72-101.