The Annals of Statistics
1982, Vol. 10, No. 4, 1269-1276

ASYMPTOTIC THEORY FOR MEASURES OF CONCORDANCE WITH SPECIAL REFERENCE TO AVERAGE KENDALL TAU[1]

BY MAYER ALVO[2], PAUL CABILIO[3] AND PAUL D. FEIGIN

University of Ottawa, Acadia University, and Technion—Israel Institute of Technology

The problem of n rankings is considered and the asymptotic distributions of measures of concordance based on rank correlations are derived under the null model of complete randomness. The Bahadur efficiencies of the measures are computed. A matrix analysis then reveals the asymptotic distribution and superior efficiency of average Kendall tau. Some interpretation of the results is also made.

1. Introduction. Consider the situation in which n judges each rank r objects. The question of interest here is whether there exists a degree of agreement (concordance) among the n judges. The usual approach is to test the null hypothesis that the set of n rankings is a random sample from the uniform distribution on the set of $r!$ possible permutations of the numbers, $1, \cdots, r$. Although we will comment further on this null hypothesis in the concluding discussion, it forms the basis of our analysis.

The most widely used statistic in this context is Kendall's W, a statistic derived by Friedman (1937), and which is equivalent to the average of Spearman's rank correlation between each of the $\binom{n}{2}$ possible pairs of judges (see Kendall, 1970). Here we will consider a general class of concordance statistics, develop their asymptotic ($n \to \infty$) distribution theory (Section 2) and discuss their relative efficiencies (Section 3). In Section 4 we make special reference to the average Kendall tau statistic, first suggested as a preferable alternative by Ehrenberg (1952) and also discussed by Hays (1960). In fact we are able to establish its superiority over the W statistic with respect to the approximate Bahadur slope criterion and the null hypothesis mentioned above.

In Section 5 we discuss this approach to testing for agreement.

We introduce the following notation. The judges' rankings are denoted by

$$\omega_i = (\omega_i(1), \cdots, \omega_i(r))', \quad i = 1, \cdots, n,$$

and we let

$$\nu_j = (\nu_j(1), \cdots, \nu_j(r))', \quad j = 1, \cdots, k$$

be the $k = r!$ possible permutations of the numbers $1, \cdots, r$. Let

$$\mathscr{P}_r = \{\nu_1, \cdots, \nu_k\}.$$

We further denote by

$$\pi = (\pi_1, \cdots, \pi_k)'$$

the vector of probabilities according to which each judge will select a ranking, i.e.

Received April 1980; revised February and December 1981.

[1] This work was done while the first two authors were on sabbatical leave at the Technion.

[2] Work supported in part by National Research Council of Canada, Grant No. A-9068.
[3] Work supported in part by National Research Council of Canada, Grant No. A-9076.
AMS 1980 *subject classifications.* Primary 62G10, 62G20; secondary 62E15, 62E20.
Key words and phrases. n rankings, average Kendall's tau, Kendall's W, Friedman's test, Bahadur slope, weighted sum of Chi squared variables.

$$\pi_j = \text{Prob}(\omega_\iota = \nu_j). \tag{1.1}$$

The judges' rankings are assumed to be drawn independently according to the same distribution π. The null model corresponds to $\pi = \pi^0 \equiv k^{-1}\mathbf{1}$ with $\mathbf{1}$ denoting a k-vector of ones. We will let P_π denote the measure induced on the subsets of $\mathscr{P}_r$ by π according to (1.1).

2. Measures of concordance—asymptotic distributions. We describe a class of measures of concordance constructed via the following three steps:
(i) a right-invariant metric $g: \mathscr{P}_r \times \mathscr{P}_r \to [0, \infty]$ forms the basis. Diaconis and Graham (1977) consider examples of metrics g which have the right invariance property

$$g(\mu, \eta) = g(\mu\sigma, \eta\sigma), \quad \text{for all } \mu, \eta, \sigma \in \mathscr{P}_r. \tag{2.1}$$

(This definition, in which $\mu\sigma$ denotes the group multiplication of permutations, ensures that relabeling the r objects would not alter the distances between the rankings.)
(ii) By considering

$$M = \max\{g(\nu_j, \eta); \ \ j = 1, \dots, k\} \tag{2.2}$$

we now construct a rank correlation statistic,

$$\alpha(\mu, \eta) = 1 - 2g(\mu, \eta)/M, \quad \mu, \eta \in \mathscr{P}_r,$$

which has values between -1 and $+1$.
(iii) For the case of n rankings the measure of concordance is then simply defined as

$$\bar{\alpha}_n = \binom{n}{2}^{-1} \sum_{j<i} \alpha(\omega_i, \omega_j),$$

or the average of all the pairwise rank correlations.
For the measures of concordance based on Spearman's ρ, Kendall's τ and Spearman's footrule the appropriate metrics are, respectively,

$$g_S(\mu, \eta) \equiv \sum_{s=1}^r \{\mu(s) - \eta(s)\}^2 = 2\left\{\frac{r(r+1)(2r+1)}{6} - \mu'\eta\right\} \tag{2.3}$$

$$g_K(\mu, \eta) \equiv \frac{1}{2}\sum_{s<t}^r [1 - \text{sgn}\{\mu(s) - \mu(t)\}\text{sgn}\{\eta(s) - \eta(t)\}] \tag{2.4}$$

$$g_F(\mu, \eta) \equiv \sum_{s=1}^r |\mu(s) - \eta(s)|. \tag{2.5}$$

We now derive the asymptotic distribution of $\bar{\alpha}_n$, under $H_0: \pi = \pi^0$, for any measure constructed according to (i), (ii) and (iii) above. Define $k \times k$ matrices

$$G = (g(\nu_i, \nu_j)), \quad J = \mathbf{1}\,\mathbf{1}', \quad I = \text{identity}.$$

THEOREM 2.1. *Under H_0, for $\bar{\alpha}_n$ defined as in* (i), (ii), *and* (iii) *above,*

$$n(\bar{\alpha}_n - c) \to_{\mathscr{L}} X'QX - 1 + c \tag{2.6}$$

where

$$Q = J - (2/M)G, \quad Q\pi^0 = c\mathbf{1} \tag{2.7}$$

and

$$X \sim N_k(0, \Sigma_0), \quad \Sigma_0 \equiv k^{-2}(kI - J). \tag{2.8}$$

(N_k denotes the k-variate normal distribution.)

PROOF. Let $N_\ell = \sum_{\iota=1}^n 1\{\omega_\iota = \nu_\ell\}$ so that $N = (N_1, \cdots, N_k)'$ is a realization of a multinomial k-vector with parameters (n, π). Define

$$A_n = n^{-1/2}(N - n\pi^0), \quad H_n = \sum_{i,j}^n g(\omega_i, \omega_j).$$

It then follows that

$$H_n = \sum_{\ell,m}^k N_\ell N_m g(\nu_\ell, \nu_m) = N'GN = (-N'QN + n^2)M/2$$
$$= (-nA_n' QA_n - n^2 c + n^2)M/2,$$

since right-invariance ensures $Q\pi^0 = c\mathbf{1}$. Therefore

$$\bar{\alpha}_n = 1 - \binom{n}{2}^{-1} H_n/M = (A_n' QA_n - 1 + c)/(n-1) + c \tag{2.9}$$

and the required result follows from the multivariate central limit result $A_n \to_{\mathscr{L}} X$, with X as given in (2.8). □

An earlier version of Theorem 2.1 was due to Quade (1972).

Writing $q(\cdot, \cdot) = 1 - 2g(\cdot, \cdot)/M$ we see that

$$\bar{\alpha}_n = \binom{n}{2}^{-1} \sum_{j<i}^n q(\omega_i, \omega_j)$$

is a U-statistic with a kernel q of degree 2. The right invariance property, however, implies that

$$E\{q(\omega_i, \omega_j) \mid \omega_i = \nu\} = c, \quad \text{for all } \nu \in \mathscr{P}_r,$$

so that the U-statistic is degenerate. Theorem 2.1 can, therefore, also be viewed as a finite sample space ($\mathscr{P}_r$) version of Gregory's (1977) limit theorem for degenerate U-statistics.

The asymptotic form (2.6) essentially involves a quadratic form of normal deviates and an equivalent expression involving weighted sums of χ_1^2 variables can be obtained. The weights are the eigenvalues of the matrix

$$Q \textstyle\sum_0 = k^{-1}(Q - cJ) \tag{2.10}$$

(see, e.g. Johnson and Kotz, 1970). We note here that for the measure of concordance $\bar{\rho}_n$ based on Spearman's ρ it is true that $c = 0$ and that the corresponding Q matrix, Q_S say, is such that $k^{-1}(r-1)Q_S$ is idempotent with rank $r - 1$ (see (4.5)). This fact leads to the well known result that

$$(r-1)\{(n-1)\bar{\rho}_n + 1\} \to_{\mathscr{L}} \chi_{r-1}^2. \tag{2.11}$$

In Section 4 we will compute the explicit form for $\bar{\tau}_n$, the measure of concordance based on Kendall's τ.

3. Asymptotic efficiencies. In order to compare the various measures of concordance we will here consider their slopes according to the Bahadur approach to efficiency of test statistics. We let

$$T_n = n^{-1/2}\{(n-1)(\bar{\alpha}_n - c) - 1 + c\} = n^{-1/2} A_n' QA_n \tag{3.1}$$

and now determine the exact slope of $\{T_n\}$.

LEMMA 3.1 $n^{-1/2}T_n \to (\pi - \pi^0)' Q(\pi - \pi^0)$ a.s. $[P_\pi]$.

PROOF. Letting

$$V_n = n^{-1}(N_1, \cdots, N_k)' \quad \text{and} \quad A_n(\pi) = n^{1/2}(V_n - \pi)$$

we find that, from (3.1),

$$n^{-1/2}T_n = n^{-1}A_n'QA_n = n^{-1}\{A_n'(\pi)QA_n(\pi) + n(\pi - \pi^0)'Q(\pi - \pi^0) + 2\,n^{1/2}(\pi - \pi^0)'\,A_n(\pi)\} \to (\pi - \pi^0)'\,Q(\pi - \pi^0) \quad \text{a.s.} \quad [P_\pi]$$

since $n^{-1/2}A_n(\pi) \to 0$ a.s. $[P_\pi]$. □

LEMMA 3.2. *Let*

$$\Delta = \{v = (v_1, \cdots, v_k)' : v_j \geq 0, \quad \textstyle\sum_{j=1}^k v_j = 1\}$$

and

$$K(v, \pi) = \textstyle\sum_{j=1}^k v_j \log(v_j/\pi_j); \quad v, \pi \in \Delta. \tag{3.2}$$

Then

$$n^{-1}\log P_{\pi^0}(T_n \geq \sqrt{n}\, t) \to -f(t)$$

where

$$f(t) = \inf\{K(v, \pi^0) : (v - \pi^0)'\,Q(v - \pi^0) \geq t, v \in \Delta\}.$$

PROOF. We have

$$P(T_n \geq \sqrt{n}\, t) = P(n^{-1}A_n'QA_n \geq t) = P(V_n \in B_t)$$

where $B_t = \{v\colon (v - \pi^0)'Q(v - \pi^0) \geq t\}$. Applying Sanov's (1957) theorem we obtain the desired result. □

THEOREM 3.3. *The exact slope of* $\{T_n\}$ *against* π *is given by*

$$c(\pi) = 2\inf\{K(v, \pi^0)\colon (v - \pi^0)'\,Q(v - \pi^0) \geq (\pi - \pi^0)'\,Q(\pi - \pi^0)\}. \tag{3.3}$$

PROOF. The result follows from Lemmas 3.1 and 3.2 and the definition of exact slope (Bahadur, 1967) as the limit as $n \to \infty$ of $-2n^{-1}\log(\text{attained level of } T_n)$. □

In order to compare efficiencies of the various measures of concordance, one would like to evaluate (3.3) for the matrices Q_K, Q_S, Q_F corresponding to (2.3), (2.4), (2.5), but we have not been able to compute $c(\pi)$. An alternative to computing $c(\pi)$ is to calculate the approximate slope which will be a good approximation to the exact slope for local alternatives (π close to π^0). This approach is discussed in Bahadur (1967) and also in Foutz and Srivastava (1977) and can be justified directly in our multinomial situation as is shown in the following theorem.

THEOREM 3.4. *Suppose* $Q\sum_0$ *is positive semi-definite. Then the approximate slope of* $\{n^{1/2}T_n\}$ *is given by*

$$a(\pi) = (\pi - \pi^0)'\,Q(\pi - \pi^0)/\lambda_1, \tag{3.4}$$

where λ_1 *is the largest eigenvalue of* $Q\sum_0$. *Moreover,*

$$a(\pi)/c(\pi) \to 1 \text{ as } \|\pi - \pi^0\| \to 0, \tag{3.5}$$

where $\|\cdot\|$ *denotes Euclidean norm.*

PROOF. The asymptotic null distribution of $\{n^{1/2}T_n\}$ is that of $X'QX$. On applying Lemma 3.3 of Foutz and Srivastava (1977) we have that

$$\frac{1}{n}\log P(X'QX \geq A_n'QA_n) = (n^{-1}\,A_n'QA_n)\{1 + o(1)\}/(2\lambda_1). \tag{3.6}$$

Hence (3.4) follows from Lemma 3.1 and the definitions of approximate Bahadur slope as twice the limit of (3.6) as $n \to \infty$.

The result (3.5) is obtained by investigating the properties of $K(v, \pi^0)$ for v in a neighbourhood of π^0. □

Consequently, the *local* relative efficiencies of our measures of concordance can be determined from their respective approximate slopes, i.e. from equation (3.4). In the following section we will be able to establish the superiority of $\bar{\tau}_n$ over $\bar{\rho}_n$ based on these considerations.

4. Average Kendall tau. It is clear from Sections 2 and 3 that determining the asymptotic distribution and the relative efficiency of $\bar{\tau}_n$ requires an eigenanalysis of the matrix $k^{-1}Q_K$, or equivalently $Q_K \sum_0$. This analysis forms the content of this section.

In what follows, the subscripts K and S refer to the Kendall and Spearman cases respectively. The following can be determined quite readily using (2.2), (2.3), (2.7):

$$M_K = \binom{r}{2} = \frac{r(r-1)}{2}, \quad M_S = \frac{r(r+1)(r-1)}{3}, \tag{4.1}$$

$$c_K = c_S = 0, \tag{4.2}$$

$$Q_S = \left\{\frac{4}{M_S} D' D - \frac{3(r+1)}{(r-1)} J\right\}, \tag{4.3}$$

where

$$D = (\nu_1, \nu_2, \cdots, \nu_k)$$

is the $r \times k$ matrix whose columns are the $k = r!$ permutations of $1, \cdots, r$ (labeled in some order). Furthermore

$$(Q_K)_{ij} = \frac{2}{r(r-1)} \textstyle\sum_{s<t}^{r} S_i(s, t) S_j(s, t) \tag{4.4}$$

where

$$S_i(s, t) = \operatorname{sgn}\{\nu_i(s) - \nu_i(t)\}.$$

On computing DD' we may deduce further that

$$Q_S^2 = \frac{k}{(r-1)} Q_S, \tag{4.5}$$

from which (2.11) was obtained. We now state a lemma, the proof of which is available from the authors.

LEMMA 4.1. *The matrices Q_K and Q_S satisfy*

(i) $Q_K Q_S = \dfrac{2k(r+1)}{3r(r-1)} Q_S$

(4.6) (ii) $Q_K = \dfrac{2(r+1)}{3r} Q_S + A, \quad Q_S A = 0.$

(4.7) (iii) $Q_K^2 = \dfrac{4}{9} \dfrac{k(r+1)^2}{r^2(r-1)} Q_S + \dfrac{2k}{3r(r-1)} A.$

We are now able to complete the analysis of Q_K.

THEOREM 4.2. *The matrix Q_K has two distinct non-zero eigenvalues*

$$\lambda_1 = 2k(r+1)/\{3r(r-1)\}, \quad \lambda_2 = 2k/\{3r(r-1)\} \tag{4.8}$$

with corresponding principal idempotents

$$E_1 = \{(r-1)/k\}Q_S, \quad E_2 = \{3r(r-1)/2k\}A,$$

of rank $(r-1)$ *and* $\binom{r-1}{2}$ *respectively.*

PROOF. By considering the result (4.6) of Lemma 4.1 we have

$$Q_K^2 = \frac{4(r+1)^2}{9r^2}\left(\frac{k}{r-1}\right)Q_S + A^2 \tag{4.9}$$

since $Q_S^2 = \{k/(r-1)\}Q_S$. Comparing (4.9) and (4.7) of Lemma 4.1 we conclude that

$$A^2 = \{(2k)/3r(r-1)\}A.$$

We then may identify $E_2 = \{3r(r-1)/(2k)\}A$ as idempotent and by considering the traces of both sides of (4.6) we obtain

$$\operatorname{tr}(A) = k(r-2)/(3r)$$

which gives the rank of E_2 as described. □

REMARK. The above result seems to be related to that of Hajek and Sidak (1967, page 60).

Theorem 4.2 provides the key to the following two results. The first identifies the asymptotic distribution of $n\bar{\tau}_n$.

THEOREM 4.3. *Under the null model* $(P\pi^0)$, *the asymptotic distribution of* $n\bar{\tau}_n$ *is given by*

$$\frac{2}{3r(r-1)}\{(r+1)\,\chi^2_{r-1} + \chi^2_{\binom{r-1}{2}}\} - 1 \tag{4.10}$$

where the two χ^2 *variates are independent.*

PROOF. The proof follows from Theorem 2.1 and the eigenanalysis of $Q_K \sum_0 = k^{-1}Q_K$ given in Theorem 4.2. □

Critical values of the distribution (4.10) can be computed using the Wilson-Hilferty approximation described in Jensen and Solomon (1972).

Turning now to the efficiency question we have the following theorem.

THEOREM 4.4. *The approximate Bahadur slope of average Kendall tau against rho is given by*

$$a_K(\pi) = (\pi - \pi^0)'\, Q_K(\pi - \pi^0)\frac{3r(r-1)}{2(r+1)} \tag{4.11}$$

and

$$a_K(\pi) \geq a_S(\pi) \quad \textit{for all } \pi \in \Delta.$$

PROOF. The result (4.11) follows from Theorems 3.4 and 4.2. Moreover, from Lemma 4.1

$$\begin{aligned} a_K(\pi) &= \left[\{(\pi-\pi^0)'\,Q_S(\pi-\pi^0)\}\frac{2(r+1)}{3r} + (\pi-\pi^0)'\,A(\pi-\pi^0)\right]\frac{3r(r-1)}{2(r+1)} \\ &\geq (r-1)(\pi-\pi^0)'\,Q_S(\pi-\pi^0) = a_S(\pi), \end{aligned} \tag{4.12}$$

since A is positive semi-definite. □

In terms of approximate slope we see that average Kendall tau is at least as efficient as average Spearman rho for testing the null hypothesis H_0 when the probability vector π is appropriate. Bearing in mind Theorem 3.4, this superior efficiency applies also to the exact slope criterion for "local" alternatives π.

The claim that average Kendall tau is superior has appeared in the literature (Ehrenberg, 1952; Hays, 1960; Kendall, 1970) and we have here provided some justification for this claim. In addition we have identified the exact asymptotic distribution in contrast to the approximations discussed in Ehrenberg (1952) and Hays (1960).

5. Remarks. From the calculation (4.12) it is clear that the superior efficiency of $\bar{\tau}_n$ follows from the fact that it is sensitive to a wider variety of deviations from π^0 than is $\bar{\rho}_n$. This observation leads us to consider the following points.

(a). Returning to our original desire to discover "concordance," one might consider whether the additional deviations from π^0 correspond to situations which one would wish to describe as embodying agreement among the judges. In other words, in comparing tests of the null hypothesis, one should consider the particular alternatives which are relevant to the problem. This latter approach may involve non-null modeling (see, for example, Feigin and Cohen, 1978).

(b). There possibly is scope to construct other statistics which are sensitive to exactly those deviations which are to be considered as implying concordance in given problems.

(c). The likelihood ratio test (LRT) statistic is the most efficient in terms of Bahadur slope in the multinomial situation and so one may ask why not test for concordance using this statistic. There are two comments to be made here.

(i) In the light of (a) and (b) above, the fact that the LRT is sensitive to *any* deviations from π^0 is not an advantage when one is trying to discover those deviations which correspond to a particular pattern of concordance.

(ii) Technically, the LRT will be based on a multinomial sample with $k = r!$ cells. In many practical applications, this will lead to a situation in which most cells are empty. The usual χ^2_{k-1} asymptotic distribution is then not likely to be sufficiently accurate; see Fienberg (1979) on the problem of "large sparse multinomials."

(d). This last point raises the question of rates of convergence to asymptotic distributions of statistics of the form $A'_n Q A_n$. A crude calculation based on results such as those of Bhattacharya and Rao (1976, page 120) shows that the maximum deviation of the null distribution function of $A'_n Q A_n$ from its limit (as $n \to \infty$) is of the order $\nu^2 n^{-1/2}$, ν being the rank of $Q \sum_0$. For $r = 5$ for example, for $\bar{\tau}_n$ the value of ν is 10 while that for the LRT is 119; this suggests that the same accuracy may require a sample size some 20,000 times as large for the LRT χ^2_{k-1} approximation as for $\bar{\tau}_n$.

REFERENCES

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BHATTACHARYA, R. N. and RAO, R. RANGA (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

DIACONIS, P. and GRAHAM, R. L. (1977). Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B.* **39** 262–268.

EHRENBERG, A. S. C. (1952). On sampling from a population of rankers. *Biometrika* **39** 82–87.

FEIGIN, P. D. and COHEN, A. (1978). On a model for concordance between judges. *J. Roy. Statist. Soc. Ser. B.* **40** 203–213.

FIENBERG, S. E. (1979). The use of chi-squared statistics for categorical data problems. *J. Roy. Statist. Soc. Ser. B.* **41** 54–64.

FOUTZ, R. V., and SRIVASTAVA, R. C. (1977). The performance of the likelihood ratio test when the model is incorrect. *Ann. Statist.* **5** 1183–1194.

GREGORY, G. C. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

HÁJEK, J. and SIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

HAYS, W. L. (1960). A note on average tau as a measure of concordance. *J. Amer. Statist. Assoc.* **55** 331–341.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

JENSEN, D. R. and SOLOMON, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *J. Amer. Statist. Assoc.* **67** 898–902.

JOHNSON N. L. and KOTZ S. (1970). *Continuous Univariate Distributions 2.* Houghton Mifflin, Boston.

KENDALL, M. G. (1970). *Rank Correlation Methods.* Hafner, New York.

QUADE, DANA (1972). Technical report. Mathematical Centre, University of Amsterdam.

SANOV, C. (1957). On the probability of large deviations of random variables. *Transl. Math. Statist. Probab.* **1** 213–244.

MAYER ALVO
DEPT. OF MATHEMATICS
UNIVERSITY OF OTTAWA
OTTAWA, ONTARIO
CANADA

PAUL CABILIO
DEPT. OF MATHEMATICS
ACADIA UNIVERSITY
WOLFVILLE, NOVA SCOTIA
CANADA

PAUL D. FEIGIN
FACULTY OF INDUSTRIAL
ENGINEERING AND MANAGEMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY,
HAIFA, ISRAEL.