

SPEARMAN'S FOOTRULE AS A MEASURE OF DISARRAY

ACCESSION FOR	
NTIS	White Section ☒
DOJ	Bluff Section ☐
UNANNOUNCED	☐
JUSTIFICATION.....	
BY.....	
DISTRIBUTION.....	

A

AD-A030 697

by

Persi Diaconis and R. L. Graham

Technical Report No. 237

September 2, 1976

Prepared under Contract N00014-76-C-0475
(NR-042-267)

Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

D D C
RECEIVED
OCT 13 1976
REGISTRATION
D

Also issued as Technical Report No. 88 under National Science Foundation
Grant MPS74-21416.

1. Introduction.

Widely used nonparametric measures of association such as Kendall's τ and Spearman's ρ lead to natural metrics on the set S_n of permutations of n letters. Let π and σ be elements of S_n . Four metrics which we consider are

$$(1-1) \quad D(\pi, \sigma) = \sum_{i=1}^n |\pi(i) - \sigma(i)|$$

$$(1-2) \quad S(\pi, \sigma) = \sum_{i=1}^n (\pi(i) - \sigma(i))^2$$

$$(1-3) \quad T(\pi, \sigma) = \text{the minimum number of transpositions required to bring } (\pi(1), \dots, \pi(n)) \text{ into the order } (\sigma(1), \dots, \sigma(n)).$$

$$(1-4) \quad I(\pi, \sigma) = \text{the minimum number of pairwise adjacent transpositions required to bring } (\pi^{-1}(1), \dots, \pi^{-1}(n)) \text{ into the order } (\sigma^{-1}(1), \dots, \sigma^{-1}(n)). \text{ Here } \pi^{-1} \text{ and } \sigma^{-1} \text{ are the permutations inverse to } \pi \text{ and } \sigma.$$

In section 2 we derive the mean, variance and limiting normality of $D(\pi, \sigma)$ when π and σ are chosen independently and uniformly from S_n . Similar, known, results for S , T and I are summarized in Table 1. In section 3 we prove that $I+T \leq D \leq 2I$. The proof and subsequent discussion of boundary cases depends on combinatorial arguments of independent interest. In section 4 we discuss the statistical implications of section 2 and 3 along with some open problems.

2. Properties of the 4 Metrics.

If ρ denotes one of the functions D , S , T or I , then it is easy to verify that ρ is a metric on S_n ; that is:

$$\rho(\pi, \sigma) \geq 0 \quad \text{and} \quad \rho(\pi, \sigma) = 0 \quad \text{if and only if} \quad \pi = \sigma,$$

$$\rho(\pi, \sigma) = \rho(\sigma, \pi) \quad \text{and} \quad \rho(\pi, \sigma) \leq \rho(\pi, \eta) + \rho(\eta, \sigma).$$

Further, all four metrics are right invariant in the sense that

$$\rho(\pi, \sigma) = \rho(\pi\eta, \sigma\eta) \quad \text{for} \quad \eta \in S_n. \quad \text{In particular} \quad \rho(1, \pi) = \rho(1, \pi^{-1}),$$

where we write 1 for the identity permutation in S_n . We shall usually abbreviate $\rho(1, \pi)$ by $\rho(\pi)$. We consider more specific properties below.

$D(\pi, \sigma)$: The sum of the absolute values of the difference in the ranks is the metric associated with Spearman's footrule. Kendall (1970) discusses the footrule as a measure of association and dismisses it because of a lack of available statistical properties. Toward this end we prove

Theorem 1. Let π and σ be permutations chosen independently and uniformly in S_n . Then, as n goes to infinity

$$(2-1) \quad E\{D(\pi, \sigma)\} = \frac{n^2}{3} + o(n)$$

$$(2-2) \quad \text{Var}\{D(\pi, \sigma)\} = \frac{2}{45} n^3 + o(n^2)$$

$$(2-3) \quad \left| P\left\{ \frac{D(\pi, \sigma) - \frac{n^2}{3}}{\sqrt{\frac{2}{45} n^3}} \leq x \right\} - \Phi(x) \right| \rightarrow 0$$

$$\text{where} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx.$$

Proof. Using right invariance, the distribution of $D(\pi, \sigma)$ is the same as the distribution of $D(1, \pi) = D(\pi)$. Thus we may restrict our attention to $D(\pi)$.

We first compute

$$E(D(\pi)) = \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n |i-j| = \frac{n^2}{3} + O(n).$$

This proves (2-1). The proof of (2-2) is a lengthy computation and we only sketch the details. Consider

$$\begin{aligned} (2-4) \quad \text{var}\{D(\pi)\} &= \sum_{i=1}^n \text{var}|i-\pi(i)| + 2 \sum_{i < j} \text{cov}(|i-\pi(i)|, |j-\pi(j)|) \\ &= \frac{n^3}{20} + O(n^2) + 2 \sum_{i < j} \text{cov}(|i-\pi(i)|, |j-\pi(j)|). \end{aligned}$$

To simplify the evaluation of the covariance sum note that for any bounded functions f and g ,

$$\begin{aligned} E\{f(\pi(i))g(\pi(j))\} &= \frac{1}{n(n-1)} \cdot \sum_{\ell \neq k} f(\ell)g(k) \\ &= \frac{n}{n-1} E\{f(\pi(1))\}E\{g(\pi(1))\} - \frac{1}{n-1} E\{f(\pi(1))g(\pi(1))\} \\ &= E\{f(\pi(1))\}E\{g(\pi(1))\} - \frac{1}{n-1} \text{cov}(f(\pi(1)), g(\pi(1))). \end{aligned}$$

In particular,

$$(2-5) \quad \text{cov}(|i-\pi(i)|, |j-\pi(j)|) = -\frac{1}{n-1} \text{cov}(|i-\pi(1)|, |j-\pi(1)|).$$

Consider

$$\begin{aligned}
 (2-6) \quad \text{cov}(|1-\pi(1)|, |j-\pi(1)|) &= \frac{1}{n} \sum_{\ell=1}^n |1-\ell| |j-\ell| \\
 &- \frac{1}{4n^2} \{1(1-1) + n(n-1-1)\} \{j(j-1) + (n-j)(n-j-1)\} \\
 &= f_1(1, j) + f_2(1, j) .
 \end{aligned}$$

Summing $f_1(1, j)$ in $i < j$ for fixed j and then in j yields

$$(2-7) \quad \sum_{i < j} f_1(1, j) = \frac{7n^4}{120} + o(n^3) .$$

Similarly

$$(2-8) \quad \sum_{i < j} f_2(1, j) = -\frac{n^4}{18} + o(n^3) .$$

(2-8) and (2-7) along with (2-6) and (2-5) show that

$$\sum_{i < j} \text{cov}(|1-\pi(i)|, |j-\pi(j)|) = -\frac{n^3}{360} + o(n^2) .$$

Using this in (2-4) completes the proof of (2-2).

To prove (2-3) we use Hoeffding's (1951) combinatorial central limit theorem. This states that if $\{a_{ij}^n\}$ $i, j=1, \dots, n$ is a sequence of square array's and if $W_n = \sum_{i=1}^n a_{i\pi(i)}^n$ where π is a random permutation in S_n then, subject to growth conditions on a_{ij}^n ,

$$\lim_{n \rightarrow \infty} \left| P_n \left(\frac{W_n - E(W_n)}{\sqrt{\text{var } W_n}} \leq x \right) - \Phi(x) \right| = 0$$

Using (2-2) we readily verify the sufficient condition given in equation (12) of Hoeffding (1951) for the array

$$a_{ij}^n = |i-j| \quad i, j=1, \dots, n.$$

This completes the proof of theorem 1. $\square$

$S(\pi, \sigma)$: The sum of squares of the difference in the ranks is the metric associated with Spearman's coefficient of association ρ . All of the results listed in table 1 may be found in Kendall (1970).

$T(\pi, \sigma)$: This seemingly natural metric has not received much attention in the statistical literature. Note first that if ψ and η are in S_n , we have

$$T(\pi, \sigma) = T(\psi\pi\eta, \psi\sigma\eta) = T(1, \pi^{-1}\sigma) = T(\pi^{-1}\sigma).$$

A result due to Caley (1849) states that $T(\pi^{-1}\sigma) = n - C(\pi^{-1}\sigma)$ where $C(\eta)$ is the number of cycles in η . This gives an easy way to compute $T(\pi, \sigma)$. Feller (1968), pg. 256, gives a representation of the number of cycles in a random permutation as a sum of independent random variables. This representation and standard theorems from probability theory easily imply the limiting normality of T as well as the results listed in table 1.

$I(\pi, \sigma)$: This metric arises from Kendall's measure of association τ . It is thoroughly discussed in Kendall (1970). A simple approach to the analysis

of the stochastic properties of this metric is as follows: A permutation η is said to have an inversion at (k, ℓ) if $k < \ell$ and $\eta(k) > \eta(\ell)$. A standard result (see for example Knuth (1973) Sec. 5.1.1) is that $I(\pi, \sigma) = I(1, \pi\sigma^{-1})$ = the number of inversions in $\pi\sigma^{-1}$. Feller (1968), pg. 256 gives a representation for the number of inversions in a random permutation as a sum of independent random variables. From this the limiting normality as well as the results listed in Table 1 follow easily.

The following table summarizes the means and variances of these metrics. Only the leading term of the mean and variance is indicated, the notation $[x]$ is used for the integral part of x .

Table 1

	Max	Mean	Variance
D	$[\frac{n^2}{2}]$	$\frac{n^2}{3}$	$\frac{2}{45} n^3$
S	$\frac{n^3-n}{3}$	$\frac{n^3}{6}$	$\frac{n^5}{36}$
T	$n-1$	$n-\log n$	$\log n$
I	$\frac{n^2-n}{2}$	$\frac{n^2}{4}$	$\frac{n^3}{36}$

3. Some Inequalities Involving $D(\pi, \sigma)$.

There are important relations between the various metrics which generally take the form of inequalities which must be satisfied. For example, Kendall (1970) discusses the Durbin-Stuart inequality, which

in our notation is:

$$S(\pi, \sigma) \geq \frac{4}{3} I(\pi, \sigma) \left\{ 1 + \frac{I(\pi, \sigma)}{n} \right\} .$$

We now prove:

Theorem 2.

$$(3-1) \quad I(\pi, \sigma) + T(\pi, \sigma) \leq D(\pi, \sigma) \leq 2I(\pi, \sigma) .$$

Proof. Since all the quantities in (3-1) are right invariant it is sufficient to show that

$$(3-2) \quad I(\pi) + T(\pi) \leq D(\pi) \leq 2I(\pi) .$$

We prove the right hand inequality first.

$I(\pi) = I(\pi^{-1})$ is the smallest number of pairwise adjacent transpositions required to bring π to the identity. Let $x_i, 1 \leq i \leq I(\pi)$, be a sequence of integers which index's a sequence of transpositions which transform 1 to π_1 to π_2 to ... to π . The i^{th} transposition transforms π_i to π_{i+1} by interchanging $\pi_i(x_i)$ and $\pi_i(x_i+1)$. We may assume that the sequence x_i is chosen so that $\pi_i(x_i) < \pi_i(x_i+1)$. Consider the difference

$$\begin{aligned} \Delta_{i+1} &= D(\pi_{i+1}) - D(\pi_i) \\ &= |x_i - \pi_i(x_i+1)| + |x_i+1 - \pi_i(x_i)| - |x_i - \pi_i(x_i)| - |x_i+1 - \pi_i(x_i+1)| . \end{aligned}$$

There are three possibilities

Case 1. $\pi_i(x_i+1) \leq x_i$. Then $\pi_i(x_i) < x_i$ and $\Delta_{i+1} = 0$.

Case 2. $\pi_1(x_1) \geq x_1 + 1$. Then $\Delta_{1+1} = 0$.

Case 3. $\pi_1(x_1 + 1) \geq x_1 + 1$, $\pi_1(x_1) \leq x_1$. Then $\Delta_{1+1} = 2$.

$$\text{Thus } D(\pi) = \sum_{i=1}^{I(\pi)} \Delta_i \leq 2I(\pi) \quad \text{as desired.}$$

To prove the left hand side of the inequality (3-2) we need some more notation.

Denote the inversion $\pi(k) > \pi(\ell)$ with $k < \ell$ by $[k; \ell]$. Let us call $[k; \ell]$ a Type I inversion if $\pi(k) \geq \ell$ and a Type II inversion if $\pi(k) \leq \ell$. Thus, every inversion of π is either a Type I or Type II inversion and some inversions may be of both types.

For a fixed k , if $[k; y]$ is a Type I inversion then we must have

$$k < y \leq \pi(k).$$

Therefore, denoting the number of elements in a finite set A by $|A|$,

$$(3-3) \quad |\{y: [k; y] \text{ is a Type I inversion}\}| \leq \pi(k) - k.$$

Similarly, if $[x; \ell]$ is a Type II inversion then

$$\pi(\ell) < \pi(x) \leq \ell$$

and so,

$$(3-4) \quad |\{x: [x; \ell] \text{ is a Type II inversion}\}| \leq \ell - \pi(\ell).$$

Thus, it follows that

$$\begin{aligned}
D(\pi) &= \sum_k |k - \pi(k)| \\
&= \sum_{\pi(k) \geq k} (\pi(k) - k) + \sum_{\pi(\ell) \leq \ell} (\ell - \pi(\ell)) \\
(3-5) \quad &\geq \# \text{ of Type I inversions} + \# \text{ of Type II inversions} \\
&\geq I(\pi) .
\end{aligned}$$

(3-5) is clearly a step in the right direction toward proving (3-2). In order to actually prove (3-2) we have to examine (3-5) more carefully. In particular we must see how much was given away at each of the two inequalities in (3-5).

In the first place, suppose for some integer y ,

$$(3-6) \quad k < y \leq \pi(k) < \pi(y) .$$

Such a y is counted by the sum $\sum_{\pi(k) \geq k} (\pi(k) - k)$ but is not counted in the sum of the numbers of Type I and Type II inversions since $[k; y]$ is not an inversion.

Similarly, if x satisfies

$$(3-7) \quad \pi(x) < \pi(\ell) \leq x < \ell$$

then x is counted by the sum $\sum_{\pi(\ell) \leq \ell} (\ell - \pi(\ell))$ but $[x; \ell]$ is not an inversion. Thus, for each such x and y we get a contribution of 1 to $D(\pi) - I(\pi)$.

The other place in (3-5) where we have a potential gain of $D(\pi)$ over $I(\pi)$ is the last inequality. Here, we gain 1 for each inversion

of π which is both Type I and Type II, i.e., for each x having

$$(3-8) \quad x < \pi(x), \quad \pi^2(x) < \pi(x).$$

Let us split (3-6) and (3-7) into two pieces each:

$$(3-6') \quad k < \pi(k) < \pi^2(k) \quad (\text{with } y = \pi(k))$$

$$(3-6'') \quad k < y < \pi(k) < \pi(y)$$

$$(3-7') \quad \pi^2(l) < \pi(l) < l \quad (\text{with } x = \pi(l))$$

$$(3-7'') \quad \pi(x) < \pi(l) < x < l.$$

Let us now restrict our attention to a cycle $C = (c_1, c_2, \dots, c_m)$ of π of length $|C| = m$. Thus, $\pi(c_k) = c_{k+1}$ for $1 \leq k < m$, and $\pi(c_m) = c_1$. We call an element or pair of elements of C satisfying (3-6'), (3-6''), (3-7'), (3-7'') or (3-8) a critical configuration of C .

We now show that C has at least $m-1$ critical configurations. We proceed to induction on m . For $m=1$ there is nothing to prove. For $m=2$ it is immediate since in this case (3-8) clearly holds for the smaller element of C . Assume for some $m > 2$ that the claim is valid for all values $< m$ and suppose C is an m -cycle which has at most $m-2$ critical configurations. There are several cases:

Case 1. Suppose $x \in C$ satisfies (3-6'), i.e., $x < \pi(x) < \pi^2(x)$. Form the $(m-1)$ -cycle C' by removing x from C . It is not hard to check that we have reduced the total number of critical configurations by at least 1 so that C' has at most $m-3$ critical configurations, which is a contradiction to the induction hypothesis.

Case 2. Similarly, if $x \in C$ satisfies (3-7') then by deleting it from C , we form an $(m-1)$ -cycle C' having at most $m-3$ critical configurations, which is impossible.

Case 3. Suppose there exists $x \in C$ satisfying

$$x < \pi^2(x) < \pi(x) < \pi^3(x) .$$

Thus, C has a critical configuration satisfying (3-6'') with $k=x$ and $y=\pi^2(x)$. Form the $(m-2)$ -cycle C' by deleting $\pi(x)$ and $\pi^2(x)$ from C . It is now not difficult to see that the total number of critical configurations remaining is at most $m-4$ (note that x also satisfies (3-8)), which contradicts the induction hypothesis.

Case 4. A similar argument applies if some $y \in C$ satisfies

$$y > \pi^2(y) > \pi(y) > \pi^3(y) .$$

Therefore, we may assume that cases 1 through 4 do not occur in C . However, if c denotes the least element of C then this assumption forces $\pi^{k+2}(c)$ always to be strictly between $\pi^k(c)$ and $\pi^{k+1}(c)$ for any $k \geq 0$. This is clearly impossible (since C is a cycle) and completes the induction.

Thus, each m -cycle of π contributes at least $m-1$ to the difference $D(\pi) - I(\pi)$. Therefore

$$D(\pi) - I(\pi) \geq \sum_{\substack{C=\text{cycle} \\ \text{of } \pi}} (|C|-1) = \sum_C |C| - C(\pi) = n - C(\pi) .$$

This proves Theorem 2. $\square$

As a numerical example, Table 2 lists the values of the 4 metrics when $n=4$.

Table 2.

π	cycles	$T(\pi)$	$I(\pi)$	$D(\pi)$	$S(\pi)$
(1234)	(1)(2)(3)(4)	0	0	0	0
(1243)	(1)(2)(34)	1	1	2	2
(1324)	(1)(23)(4)	1	1	2	2
(1342)	(1)(234)	2	2	4	6
(1423)	(1)(243)	2	2	4	6
(1432)	(1)(24)(3)	1	3	4	8
(2134)	(12)(3)(4)	1	1	2	2
(2143)	(12)(34)	2	2	4	4
(2314)	(123)(4)	2	2	4	6
(2341)	(1234)	3	3	6	12
(2413)	(1243)	3	3	6	10
(2431)	(124)(3)	2	4	6	14
(3124)	(132)(4)	2	2	4	6
(3142)	(1342)	3	3	6	10
(3214)	(13)(2)(4)	1	3	4	8
(3241)	(134)(2)	2	4	6	14
(3412)	(13)(24)	2	4	8	16
(3421)	(1324)	3	5	8	18
(4123)	(1432)	3	3	6	12
(4132)	(142)(3)	2	4	6	14
(4213)	(143)(2)	2	4	6	14
(4231)	(14)(2)(3)	1	5	6	18
(4312)	(1423)	3	5	8	18
(4321)	(14)(23)	2	6	8	20

- Remarks. 1. An immediate deduction from Theorem 1 is the bound $C(\pi) + I(\pi) \geq n$.
2. We can easily characterize the permutations $\pi \in S_n$ where $D(\pi)$ takes on its maximum value. For n even, $D(\pi) \leq \frac{n^2}{2}$ with equality if and only if $\pi(i) > \frac{n}{2}$ for $i=1,2,\dots,\frac{n}{2}$. Thus $D(\pi)$ takes on its maximum value for $((\frac{n}{2})!)^2$ choices of π . A similar bound holds for odd n . This skewness partially explains why the mean of $D(\pi)$ is larger than $\frac{n^2}{4}$.
3. It is natural to consider when equality is attained in theorem 2. The example in table 2 shows equality of the lower bound in 23 out of 24 cases. For larger n consideration of the mean values shows that equality can occur for at most $o(n!)$ elements of S_n . We can characterize the elements $\pi \in S_n$ for which equality occurs in the upper bound. Let us say that π has a 3-inversion if there exist $i < j < k$ with $\pi(i) > \pi(j) > \pi(k)$. It can be shown that $D(\pi) = 2I(\pi)$ if and only if π has no 3-inversions. Knuth (1973), Sec. 5.1.4, notes that the number of permutations in S_n with no 3-inversions is exactly the well known Catalan number $\frac{1}{2n+1} \binom{2n+1}{n}$.
4. The number of permutations π satisfying $I(\pi) = T(\pi)$ is easily seen to be the Fibonacci number F_{2n-1} defined by $F_0 = 0$, $F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$ for $n \geq 0$. This shows that simultaneous equality of both bounds in Theorem 2 holds exponentially often.

4. Interpretation of Results and Some Open Problems.

Statisticians most often normalize metrics so that they have the properties of a correlation coefficient. The translation is straightforward, if ρ is a metric on S_n and its maximum value is m , define a measure of association by: $R(\pi, \sigma) = 1 - \frac{2\rho(\pi, \sigma)}{m}$. We note in this connection the confusion on pgs. 32-33 in Kendall (1970) concerning Spearman's footrule. It is clear that the proper choice of m for the footrule is $\lfloor \frac{n^2}{2} \rfloor$. Knuth (1973) works with I and T directly as measures of disarray.

An example from Kendall (1970), will help clarify a discussion of invariance. Consider a class of 10 students ranked in mathematics and in music.

Pupil	a	b	c	d	e	f	g	h	i	j
Mathematics (π)	7	4	3	10	6	2	9	8	1	5
Music (σ)	5	7	3	10	1	9	6	2	8	4

A typical use of the metrics we suggest would be to compute the distance between the permutations π and σ as a measure of association. Right invariance here refers to invariance under changes in the labels a through j , of the students. Right invariance is compelling in so many situations that we have forced all of the metrics to be right invariant. This accounts for the at first unnatural appearance of the definition of $I(\pi, \sigma)$. Left invariance corresponds to relabeling the ranks; as an example, it could be that 10 means the worst ranking and 1 the best, instead of the opposite relation. Among the metrics we consider, T is left invariant as well as right invariant.

The results in Table 1 suggest that T is unsuitable for general use, having very small variance about a mean very close to its maximum value. Of the three remaining metrics S has the largest range with corresponding greater variability. S has the natural interpretation of the euclidian distance between the vectors π and σ regarded as points on the surface of a sphere in n dimensional space. I and D seem roughly similar, D seems somewhat easier to interpret directly while I has the advantage of having its distribution tabulated for small sample sizes (cf. Kendall [1970]). The inequalities in Section 3 also suggest that the difference between I and D is not very great.

The quantities I and T arise in the analysis of sorting algorithms. In particular Knuth (1973, pg. 141) considers the quantity $B(\pi)$ = the number of right to left maxima in π . Knuth (1973, pg. 157) shows that, in our notation $2B(\pi) \leq I(\pi) + T(\pi) \leq \lfloor \frac{n^2}{2} \rfloor$.

Use of theorem 2 actually gives the stronger inequality $2B(\pi) \leq D(\pi)$. The quantity $D(\pi)$ also appears in Problem 5.2.1-7 of Knuth (1973) in connection with another sorting algorithm.

We conclude by listing some open problems.

1. Find inequalities relating $D(1, \pi)$ to $S(1, \pi)$. There is the obvious bound $D(1, \pi) \leq S(1, \pi)$. Note that even this weak bound holds exponentially often. Indeed, it is not hard to see that $D(\pi) = S(\pi)$ for exactly F_{n+1} permutations $\pi \in S_n$ (cf. Remark 4).
2. Find reasonable 2-sided invariant metrics.
3. Characterize the permutations in S_n where equality of the lower bound of Theorem 2 holds.

Acknowledgement. It is a pleasure to thank Professor Herbert Solomon for first suggesting the use of metrics rather than correlations. Professor Donald Knuth provided valuable references and suggestions.

References

- Caley, A. (1849), *Philos. Mag.*, 34, 527-529.
- Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd ed. Wiley, New York.
- Hoeffding, W. (1951), *Ann. Math. Stat.*, 22, 558-566.
- Kendall, M. (1970), *Rank Correlation Methods*, 4th ed., Griffin, London.
- Knuth, D. (1973), *The Art of Computer Programming*, Vol. 2, Addison-Wesley, Reading, Massachusetts.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 237	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Spearman's Footrule as a Measure of Disarray.	5. TYPE OF REPORT & PERIOD COVERED Technical Report.	
7. AUTHOR(s) Persi/Diaconis and R. L./Graham	8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-267)	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217	12. REPORT DATE 2 Sep 1976	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 19p. (14) TR-237	13. NUMBER OF PAGES 16	
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution Unlimited		15. SECURITY CLASS. (of this report) Unclassified
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Also issued as Technical Report No. 88 under National Science Foundation Grant MPS74-21416		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) non-parametric measures of correlation; permutations; combinatorics		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) → Spearman's rho and Kendall's tau are related to two less widely used measures of association: Spearman's footrule and the minimum number of transpositions. The means, variances, and limiting normality of all four metrics are derived. ←		

DD FORM 1473
1 JAN 73EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

332 580

4B