EDITOR-IN-CHIEF

SAMUEL KOTZ
The George Washington University

EXECUTIVE EDITOR

CAMPBELL B. READ
Southern Methodist University

ASSOCIATE EDITOR

DAVID L. BANKS
National Institute of Standards and Technology

EDITORIAL BOARD

NORMAN L. JOHNSON, CHAIR
University of North Carolina at Chapel Hill

STEPHEN E. FIENBERG
Carnegie Mellon University

NANCY REID
University of Toronto

D. MICHAEL TITTERINGTON
University of Glasgow

GEOFFREY S. WATSON (deceased)
Princeton University
CIRCULAR DATA, RAO'S SPACING TEST OF

Rao's spacing test is a useful and powerful statistic for testing uniformity of circular ("directional") data. As with other circular statistics, the test is applicable for analysis of directional data, time series, similarity judgments involving musical pitch or color [5], phase relations in studies of movement [2], and spatial trends in geographical research [4].

The statistic was first conceived in the doctoral dissertation of J. S. Rao [6] and is further described in Rao [7, 8] and Batschelet [1]. Rao's spacing test is based on the idea that if the underlying distribution is uniform, n successive observations should be approximately evenly spaced, about 360°/n apart. Large deviations from this distribution, resulting from unusually large spaces or unusually short spaces between observations, are evidence for directionality. The test is related to the general class of linear statistical tests based on successive order statistics and spacings. A similar test for higher dimensions was independently described by Foutz [3], who apparently was unaware of Rao's previous work, and did not know the exact distribution.

Rao's test statistic $U$ is defined as follows. If $(f_1, f_2, \ldots, f_n)$ denote the successive observations, either clockwise or counterclockwise, then

$$U = \frac{1}{2} \sum_{i=1}^{n} |T_i - \lambda|,$$

where $\lambda = 360°/n$ and

$$T_i = f_{i+1} - f_i, \quad 1 \leq i \leq n - 1,$$

$$T_n = (360°/f_n) + f_1.$$

Because the sum of the positive deviations must equal the sum of the negative ones, a simpler computational form eliminates absolute values, so that

$$U = \sum_{i=1}^{n} (T_i - \lambda)^+, $$

summed across positive deviations only. The density function of $U$ is known [8] to be:

$$f(u) = \frac{1}{(n-1)!} \times \sum_{j=1}^{n-1} \binom{n}{j} \left( \frac{u}{2\pi} \right)^{n-j-1} \frac{g_j(nu)}{(n-j-1)!n^{j-1}},$$

$$0 < u < 2\pi \left( 1 - \frac{1}{n} \right),$$

where

S. D. BROWN
CIRCULAR DATA, RAO'S SPACING TEST OF

\[ g_j(x) = \frac{1}{(j - 1)!} \left( \frac{x}{2\pi} - k \right)^{j-1} \]

An expanded set of critical values of the statistic \( U \) has been tabulated and published [9].

Example. Suppose one wishes to know whether birth times at a hospital are uniformly distributed throughout the day. The alternative hypothesis is that there is a time (or times) when births are more frequent. Table 1 displays hypothetical data for delivery times collected across several days. These time series data form a continuous circular distribution, with times of day converted to angles around a circle, such that 12 midnight = 0°, 6 A.M. = 90°, and each minute corresponds to 360°/(24 h × 60 min) = 0.25°. The distribution around the circle is shown graphically in Fig. 1.

We find that \( \sum_{i=1}^{n} |T_i - \lambda| = 354 \), and so

\[ U = \frac{1}{2} \sum_{i=1}^{n} |T_i - \lambda| = 354/2 = 177. \]

From the tabulated values for \( U(15) \) [9], we have \( p < 0.02 \), so we reject the hypothesis of uniformity in favor of a hypothesis of modality or multimodality. Close inspection of the data reveals bimodality, with peaks near 12:40 A.M. and 6:38 A.M.

[Jammalamadaka S. Rao currently publishes under the name S. Rao Jammalamadaka.]

Table 1 Calculation of the Test Statistic \( U \) for Sample Data Given in Text

| \( i \) | Time   | \( f \) | \( T_i \) | \( |T_i - \lambda| \) |
|-------|--------|--------|----------|-----------------|
| 1     | 12:20 A.M. | 5      | 5        | 19              |
| 2     | 12:40 A.M. | 10     | 0        | 24              |
| 3     | 12:40 A.M. | 10     | 2        | 22              |
| 4     | 12:48 A.M. | 12     | 5        | 19              |
| 5     | 1:08 A.M.  | 17     | 68       | 44              |
| 6     | 5:40 A.M.  | 85     | 5        | 19              |
| 7     | 6:00 A.M.  | 90     | 9        | 15              |
| 8     | 6:36 A.M.  | 99     | 1        | 23              |
| 9     | 6:40 A.M.  | 100    | 10       | 14              |
| 10    | 7:20 A.M.  | 110    | 43       | 19              |
| 11    | 10:12 A.M. | 153    | 80       | 56              |
| 12    | 3:32 P.M.  | 233    | 2        | 22              |
| 13    | 3:40 P.M.  | 235    | 61       | 37              |
| 14    | 7:44 P.M.  | 296    | 35       | 11              |
| 15    | 10:04 P.M. | 331    | 34       | 10              |

\( n = 15 \quad \sum = 354 \)

Arc length in this example, \( \lambda = 360^\circ/n = 24^\circ \). \( U = \frac{1}{2} \sum_{i=1}^{n} |T_i - \lambda| = 354/2 = 177 \). From tabulated values, \( U(15) \) yields \( p < 0.02 \), so we reject the hypothesis of uniformity.

References

COCHRAN’S THEOREM

Cochran’s theorem and its extensions specify conditions under which a set of random variables (matrices) forms a family of independent chi-square variables (Wishart matrices). Such criteria are used in regression analysis, experimental design*, and analysis of variance*. For example, an illustration involving a test statistic for the general linear model* is given in Graybill and Marsaglia [12], and applications to the two-way layout and the Latin square* design are discussed in Scheffé [33].

Cochran [7] obtained the following fundamental result in connection with an analysis of covariance* problem: Let \( X \) be a vector of \( k \) independent standard normal variables, and \( A_i \) be a symmetric matrix of rank \( r_i \); for \( i = 1, \ldots, k \); then, whenever the sum of squares \( X'X \) can be partitioned into the sum of \( k \) quadratic forms \( Q_i = X'A_iX \), that is, \( X'X = \sum_{i=1}^{k} X'A_iX \), a necessary and sufficient condition for the \( Q_i \)’s to be independently distributed as chi-square variables is that \( \sum_{i=1}^{k} r_i = p \). Previously, Fisher [11] had considered a related problem, showing that for \( Q_1 \sim \chi^2_p \), one has \( X'X - Q_1 \sim \chi^2_{p-r} \) independently of \( Q_1 \).

Using a matrix-theoretic approach, James [14] and Lancaster [20] proved the more general result: Let

\[
\sum_{i=1}^{k} A_i = I_p, \tag{1}
\]

where \( A_i \) is a \( p \times p \) symmetric matrix of rank \( r_i \); for \( i = 1, \ldots, k \), and \( I_p \) denotes the identity matrix of order \( p \); then the following conditions are equivalent:

1. \( A_iA_j = 0 \) (the null matrix) for \( i \neq j \),
2. \( A_i^2 = A_i \) for \( i = 1, \ldots, k \), and
3. \( \sum_{i=1}^{k} r_i = p \).

In terms of the quadratic forms \( Q_i \) defined above, condition (i) states that the \( Q_i \)’s are mutually independent: under (i), the joint moment-generating function of \( Q_1, \ldots, Q_k \) can be factorized and (i) can then be obtained from the rarely used necessity part of the Craig–Sakamoto theorem*. Moreover, condition (ii) states that each \( Q_i \) follows a chi-square distribution with \( r_i \) degrees of freedom; refs. [8, 32, 5, 27, 24, 25, 29] give sets of necessary and sufficient conditions under which a single quadratic form in a normal vector is distributed as a (possibly noncentral) chi-square variable; see also GENERAL LINEAR MODEL, Theorem 2. Thus, the original version of Cochran’s theorem may be formulated as follows: Given (1), one has (i) \iff (ii).

Several generalizations have been considered. For example, on replacing (1) with

\[
\sum_{i=1}^{k} A_i = A, \quad \text{where} \quad A = A^2, \tag{1'}
\]

it can be shown that any two of the statements (1’), (i), and (ii) imply all four and that (1’) and (iii) imply (i) and (ii) [16, 30]. Extensions of Cochran’s theorem were obtained by Madow [23] under (1) for the noncentral case, that is, for quadratic forms in normally distributed vectors with mean \( \mu \) and covariance matrix \( \Sigma = I_p \); by Ogawa [28] under (1’) for \( \mu = 0 \) and \( \Sigma \) positive definite (\( \Sigma > 0 \)); by Ogasawara and Takahashi [27] for \( \mu = 0 \) and \( \Sigma \) nonnegative definite (\( \Sigma \succeq 0 \)) and for any \( \mu \) and \( \Sigma > 0 \) (see also refs. [12, 3, 21, 34, 13] for the latter case); and by Styfan [35] and Tan [36] for any...