The Annals of Probability
1987, Vol. 15, No. 1, 416–422

INDEPENDENT SUBSETS OF CORRELATION AND OTHER MATRICES

BY TIMOTHY C. BROWN[1]

University of Melbourne

It is known that the set of correlation coefficients formed from k independent normal samples exhibits pairwise independence of its members (Geisser and Mantel (1962)). Here it is shown that *many* much larger subsets of the matrix are *fully independent*. The main result characterises such subsets in a simple way. Because the results are framed in abstract terms, they also apply to rank correlation coefficients and χ^2 statistics.

Suppose $X_1, \dots, X_k$ are independent uniform random elements of a metric space (S, d). By *uniform* we mean that the distribution of X_i, $i = 1, \dots, k$, attributes equal probability to balls of equal radius. The existence of a uniform distribution is a consequence, in most cases, of an additional group of transformations on S (when the uniform distribution is normalised Haar measure), but it is specified uniquely by the above requirement (Christensen (1970)).

This note is concerned with the random variables $d(X_i, X_j)$, $1 \le i < j \le k$, each of which will henceforth be shortened to *ij*. Similarly, the random element X_i will be abbreviated to i. The motivation for studying these pairwise distances comes from the fact that they are one-to-one functions of statistics commonly used to test certain null hypotheses. Hence, the dependence or independence of subsets of the statistics is equivalent to the dependence or independence of the corresponding distances.

For example, suppose $\mathbf{A}_1, \dots, \mathbf{A}_k$ are normal samples each with n (≥ 3) independent and identically distributed observations and that X_i is the n random vector

$$X_i = \frac{\mathbf{A}_i - \overline{A}_i \mathbf{1}}{\|\mathbf{A}_i - \overline{A}_i \mathbf{1}\|},$$

where $\overline{A}_i$ is the sample mean of $\mathbf{A}_i$ and $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^n$. Then, $X_1, \dots, X_k$ lie on the sphere, S, in $n - 1$ dimensions,

$$S = \left\{ \mathbf{x} \in \mathbb{R}^n \colon \sum_{i=1}^{n} x_i^2 = 1,\ \sum_{i=1}^{n} x_i = 0 \right\}.$$

Moreover, the product moment correlation coefficient is

$$r(\mathbf{A}_i, \mathbf{A}_j) = X_i \cdot X_j,$$

which is in one-to-one correspondence with the usual distance $d(X_i, X_j)$ on S.

Received January 1985; revised July 1985.

[1]Part of this work was done while the author was visiting the Open University, U.K.

AMS 1980 *subject classifications.* Primary 62J15; secondary 60E99, 60G55.

Key words and phrases. Correlation matrix, partial independence, χ^2 statistics, distances on metric space, uniform distribution.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

Finally, under normality $X_1, \dots, X_k$ have uniform distributions over S (Fisher (1915)) and the null hypothesis of no correlations among the **A**'s implies the independence of $X_1, \dots, X_k$.

Other statistical examples include rank correlation coefficients, in which

$$S = \{\mathbf{i} \in \mathbb{N}^n \colon \mathbf{i} \text{ is a permutation of } (1, 2, \dots, n)\}.$$

Spearman's rank correlation coefficient is then a one-to-one function of Euclidean distance on S and Kendall's tau is a one-to-one function of the pairwise-interchange metric on S. If, for some natural number N,

$$S = \left\{\mathbf{i} \in \mathbb{N}^n \colon i_j \in \{0,1\} \text{ and } \sum_{j=1}^{n} i_j = N\right\}$$

then, by thinking of i_j as the presence or absence of an attribute among n people (in which a total of N have the attribute), the random elements $X_1, \dots, X_k$ on S record the presence or absence of k attributes among each of the n people. Any two of these $X_1, \dots, X_k$ generate a 2×2 contingency table whose margins are both fixed at N and $n - N$. A common distribution used to test the null hypothesis of no association between the attributes is that of independence and uniformity of $X_1, \dots, X_k$—this distribution corresponds to multinomial allocation into the 2^p cells of the table, conditional on the margins for each attribute being N and $n - N$. Further, Pearson's χ^2 in such tables is a one-to-one function of the Euclidean metric on S. [The more realistic case of nonbinary attributes and differing margins will be considered later.]

It is known that product moment correlation coefficients are pairwise independent (Geisser and Mantel (1962)) and it has been pointed out that this pairwise independence extends to the above framework (Silverman (1978), Brown and Eagleson (1984), Brown, Cartwright and Eagleson (1986)). This is at first sight surprising since 12 and 13 share 1 in common. What is perhaps even more surprising is that there are *many* sets of the distances which are fully independent. The theorem below characterises these precisely in graph theoretic terms. Let $K = \binom{k}{2}$. A subset $\{i_1 j_1, \dots, i_p j_p\}$ $(p \le K)$ of distances may be visualised by a graph whose vertices are $\{i_1, j_1, \dots, i_p, j_p\}$ and edges are $i_1 j_1, \dots, i_p j_p$. We then have

THEOREM 1. *Suppose $X_1, \dots, X_k$ are independent uniform random elements of a separable metric space, (S, d), with more than one element. A subset of the distances formed from $X_1, \dots, X_k$ is independent if, and only if, the corresponding graph has no cycles.*

REMARK 1. The separability is only used in the proof that an independent subset has a graph with no cycles. The *existence* of a uniform distribution implies precompactness and thus there are virtually no cases in which the theorem has content but the metric space is not compact.

REMARK 2. In the trivial case where the metric space has one element, the theorem is obviously false since all distances are deterministically zero and hence fully independent.

PROOF. Assume first that the corresponding graph has no cycles. Then, by a theorem of graph theory (e.g., Street and Wallis (1977), Lemma 9, page 389), the graph has at least one vertex with valency 1. Without loss of generality, we may label this vertex 1. Suppose that this is connected to vertex 2 alone and label the other edges in the subset $i_2 j_2, \dots, i_p j_p$ $(p \le K)$. For arbitrary $r_1, \dots, r_p \ge 0$ we seek

$$\begin{aligned} &P(12 \le r_1, i_2 j_2 \le r_2, \dots, i_p j_p \le r_p) \\ &\quad = E\big(P(12 \le r_1, i_2 j_2 \le r_2, \dots, i_p j_p \le r_p | 2, i_2, j_2, \dots, i_p, j_p)\big) \\ &\quad = E\big(I\left[i_2 j_2 \le r_2, \dots, i_p j_p \le r_p\right] P(12 \le r_1 | 2)\big) \end{aligned}$$

using the independence of $1, 2, \dots, k$ and standard properties of conditional expectation. But, for all x in S,

$$\begin{aligned} P(12 \le r_1 | 2 = x) &= P(1 \in \text{closed ball, centre } x\text{, radius } r_1) \\ &= P(12 \le r_1) \end{aligned} \tag{1}$$

since the second probability is the same for all x, by assumption. Combining this with the last equation gives

$$\begin{aligned} &P(12 \le r_1, i_2 j_2 \le r_2, \dots, i_p j_p \le r_p) \\ &\quad = P(12 \le r_1) P(i_2 j_2 \le r_2, \dots, i_p j_p \le r_p). \end{aligned}$$

Since the subgraph with edges $i_2 j_2, \dots, i_p j_p$ cannot have cycles when its parent does not have them, we may apply the same argument to remove one more marginal probability from the joint probability. Applying the same argument a further $p - 3$ times yields the required factorisation and independence is established.

Suppose on the contrary that the corresponding graph does have a cycle. Without loss of generality, we may suppose that the edges in the cycle are $12, 23, \dots, p1$ $(2 \le p \le k)$. An $r > 0$ will be found such that

$$P(23 \le r) P(34 \le r) \cdots P(p1 \le r) P(12 > (p-1)r) > 0. \tag{2}$$

This suffices to show that $12, 23, \dots, p1$ are dependent because by the triangle inequality $23 \le r, 34 \le r, \dots, p1 \le r \Rightarrow 12 \le (p-1)r$ and so

$$P(23 \le r, 34 \le r, \dots, p1 \le r, 12 > (p-1)r) = 0.$$

We first show that $P(12 > 0) > 0$. If not, then $P(12 = 0) = 1$. Choosing any x in S, we then have $P(1 = x) = 1$, by applying equation (1) with $r_1 = 0$. Since S has more than one element a contradiction is obtained.

Since

$$0 < P(12 > 0) = \lim_{R \to 0} P(12 > R)$$

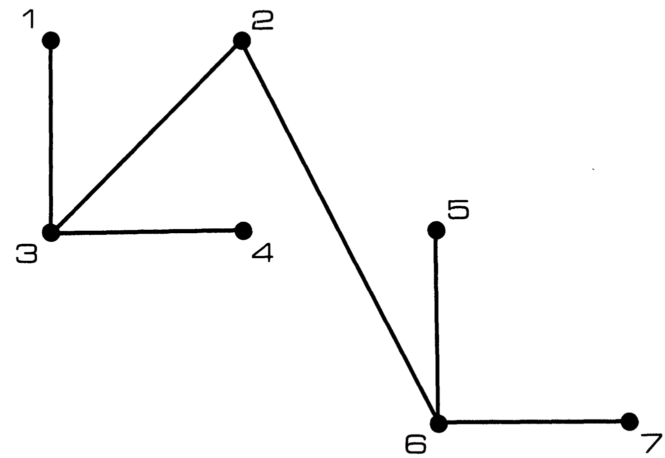

FIG. 1.

there must exist $R > 0$ such that $P(12 > R) > 0$. We set $r = R/(p - 1)$ for such R, so that the last factor of the left of inequality (2) *is* positive. To see that the others are also positive, note that, by separability, S may be covered by a countable number of closed balls, $B_1, B_2, \ldots$ of radius r. Then, if $P(1 \in B_1) = P(1 \in B_2) = \cdots = 0$, we would have

$$P(1 \in S) \leq \sum_{n=1}^{\infty} P(1 \in B_n) = 0.$$

Thus, $P(1 \in B_1) > 0$ and this is enough by equation (1) with $r_1 = r$ and the fact that $1, 2, \ldots, k$ are independent and indentically distributed. □

EXAMPLES. (a) It is perhaps not immediately apparent, without the theorem, that $\{13, 23, 34, 26, 56, 67\}$ is an independent set of random variables. However it is trivial to confirm this from the corresponding graph drawn in Figure 1.

(b) The set $\{12, 23, 25, 45, 56, 36\}$ is dependent as may be seen from Figure 2.

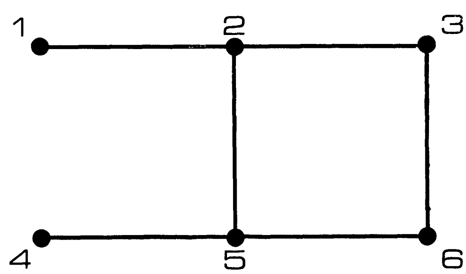

FIG. 2.

Corollary 2 below expresses the theorem without using the language of graphs. A direct proof of the corollary is possible, but is lengthier than that obtained by showing the corollary's condition for independence is equivalent to that in Theorem 1.

COROLLARY 2. *With the same setup as Theorem* 1, *a subset I of the distances is independent if, and only if, for each subset* $J = \{i_1 j_1, \dots, i_p j_p\}$ *of I, the set* $\{i_1, j_1, \dots, i_p, j_p\}$ *has at least* $p + 1$ *members.*

PROOF. We show that the graph corresponding to I has a cycle if, and only if, there exists a subset J of I with $\{i_1, j_1, \dots, i_p, j_p\}$, having at most p members. A cycle $J = \{i_1 j_1, i_2 j_2, \dots, i_p j_p\}$ has $j_p = i_1$, $j_1 = i_2, \dots, j_{p-1} = i_p$ so $\{i_1, j_1, \dots, i_p, j_p\} = \{i_1, \dots, i_p\}$, a set of cardinality p. On the other hand, suppose J is a subset of smallest size with $\{i_1, \dots, j_p\}$ having at most p members. If the graph corresponding to J had a vertex of valency 1, we could remove that vertex and obtain a smaller subset still having the same property. Hence, every vertex has valency at least 2 and the previously used combinatorial result gives the existence of a cycle. (A little thought shows that J actually is a cycle.) □

As stated previously, Theorem 1 only applies to Pearson's χ^2 in the marginal tables of a 2^p contingency table with all margins fixed identically. We can however generalise the setup to obtain an independence condition for subsets of similar χ^2 statistics in general, nonbinary, p-way tables with differing but fixed margins. Suppose now that X_i takes values in a closed subspace S_i of the space S. We suppose that S is a homogeneous space under a group G (for definitions see Federer (1969)). We assume moreover that each S_i is a homogeneous subspace, so that for each y and z in S_i there exists g in G for which $gy = z$. The function d is now only required to be symmetric and is thus not necessarily a metric. However, it is required that the distribution of X_i be invariant under G and that d be invariant, so that for all g in G,

$$gX_i =_d X_i, \qquad i = 1, \dots, k, \tag{3}$$

and, for all x, y in S,

$$d(gx, gy) = d(x, y). \tag{4}$$

Then we may adapt an argument of Silverman (1978) to prove

THEOREM 3. *In the above setup, a subset of* $\{d(X_i, X_j), 1 \le i < j \le k\}$ *is independent if the corresponding graph has no cycles.*

PROOF. Unexplained notation is the same as in the corresponding part of Theorem 1. Assume again that 1 is a vertex with valency one and that its connection is to 2. Fix $x \in S_2$.

The mapping

$$\phi\colon G \to S_2, \qquad \phi(g) = gx$$

is continuous and onto S_2 by assumption. If S_2 is countable, say $S_2 = \{y_1, y_2, \dots\}$, then we may choose g_i such that $g_i y_i = x$ and the mapping $g^*\colon y_i \mapsto g_i$ is automatically measurable. If S_2 is uncountable, we may use Theorem 2.7 of Parthasarathy (1967) to construct a subset C of G such that ϕ restricted to C is a homeomorphism, with inverse τ (say). The mapping $\tau\colon S_2 \to C$ is then measurable because it is continuous. In this case let $g^*(y) = \{\tau(y)\}^{-1}$ so that $\{g^*(y)\}y = x$. Set G^* to be the random element of G equal to $g^*(2)$. We then have

$$G^*2 = x.$$

Since S is separable and $1, 3, \dots, k$ are independent of G^*, $G^*1, G^*3, \dots, G^*k$ are also random elements of $S_1, S_3, \dots, S_k$ (see Brown, Cartwright and Eagleson (1985)). Moreover, conditioning on G^* and using equation (3), it can be seen that the joint distribution of $G^*1, G^*3, \dots, G^*k$ is the same as that of $1, 3, \dots, k$. But

$$d(i, j) = d(G^*i, G^*j)$$

by equation (4). Hence

$$\begin{aligned} &P\big(12 \le r_1, i_2 j_2 \le r_2, \dots, i_p j_p \le r_p\big) \\ &\quad = P\big(d(G^*1, x) \le r_1, d(G^*i_2, G^*j_2) \le r_2, \dots, d(G^*i_p, G^*j_p) \le r_p\big), \end{aligned}$$

which factors as in Theorem 1 since $1 \notin \{i_2, j_2, \dots, i_p, j_p\}$ and thus G^*1 is independent of $G^*i_2, \dots, G^*j_p$. The argument concludes as in Theorem 1. □

To apply this to χ^2 statistics in a general table, suppose that n individuals are classified in k ways. By possibly extending the classifications with null classifications, we may assume that each classification has J levels $1, 2, \dots, J$. Thus

$$S = \big\{\mathbf{i} \in \mathbb{N}^n\colon i_j \in \{1, \dots, J\}\big\}$$

where $\mathbf{i}$ gives the results for the individuals of a particular classification. If the marginal number for each level of a particular classification, say 1, is fixed, then the observations for this classification will lie in a subspace, S_1, consisting of those $\mathbf{i}$ for which the number $n_j^{(1)}$ of $i_1, \dots, i_n$ equal to j ($j = 1, \dots, J$) are fixed. As before, a common distribution for the null hypothesis of no association between attributes is that the k random classifications $X_1, \dots, X_k$ are independent and uniform on $S_1, \dots, S_k$. Then χ^2 statistics are then, for $1 \le i < j \le k$,

$$\chi^2(X_i, X_j) = \sum_{l, m=1}^{J} n\big(N_{lm} - n_l^{(i)} n_m^{(j)}/n\big)^2 / \big(n_l^{(i)} n_m^{(j)}\big),$$

where N_{lm} is the number of individuals with classification l in X_i and classification m in X_j. The spaces $S, S_1, \dots, S_k$ are homogeneous under the group G of permutations. Furthermore χ^2 is invariant under this group, as are the distributions of $X_1, \dots, X_k$. Thus, Theorem 2 applies with d substituted by χ^2, making possible easy analysis of many subsets of the pairwise χ^2 statistics.

REFERENCES

BROWN, T. C. and EAGLESON, G. K. (1984). A useful property of some symmetric statistics. *Amer. Statist.* **38** 63–65.

BROWN, T. C., CARTWRIGHT, D. I. and EAGLESON, G. K. (1985). Characterizations of invariant distributions. *Math. Proc. Cambridge Philos. Soc.* **97** 349–355.

BROWN, T. C., CARTWRIGHT, D. I. and EAGLESON, G. K. (1986). Correlations and characterizations of the uniform distribution. *Austral. J. Statist.* **28** 89–96.

CHRISTENSEN, J. P. R. (1970). On some measures analogous to Haar measures. *Math. Scand.* **26** 103–106.

FEDERER, H. (1969). *Geometric Measure Theory*. Springer, Berlin.

FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in a sample of values from an indefinitely large population. *Biometrika* **10** 507–521.

GEISSER, S. and MANTEL, N. (1962). Pairwise independence of jointly dependent variables. *Ann. Math. Statist.* **33** 290–291.

PARTHASARATHY, K. (1967). *Probability Measures on Metric Spaces*. Academic, New York.

SILVERMAN, B. W. (1978). Distances on circles, toruses and spheres. *J. Appl. Probab.* **15** 136–143.

STREET, A. P. and WALLIS, W. D. (1977). *Combinatorial Theory: An Introduction*. Charles Babbage Research Centre, Manitoba.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF MELBOURNE
PARKVILLE, 3052
AUSTRALIA