PAIRWISE STATISTICAL INDEPENDENCE

By H. O. Lancaster

University of Sydney

1. Introduction. It has long been known that pairwise stochastic independence is not sufficient for stochastic independence of sets of more than two random variables. In Section 2, the well known example from Bernstein's text-book is generalized by giving a solution, for every n greater than three, to the problem of how to define a measure on n points in such a way as to yield a set of $(n - 1)$ pairwise stochastically independent, random variables. The examples yield, furthermore, because of the symmetries of the measure, readily computable generalized coefficients of correlation in the sense of Lancaster (1960). The problem, specialized by prescribing equal measures to the n points, has a solution only for those values of n for which a Hadamard matrix of elements ± 1 exists.

As a further generalization of the results of Section 2, it is shown that on any atom-free measure space, a constant function and a set of pairwise independent random variables can be found, which together form a complete orthonormal set of functions on the measure space, and the members of which are products of a set of completely independent random variables taken 1, 2, $\cdots$ at a time.

2. Pairwise independence on n points. Not more than $\log_2 n$ mutually independent random variables can be defined on n points, for the condition for complete independence of k random variables requires the product measure to be non-zero at not less than 2^k points; this is evident from equation (3) of page 10 of Kolmogorov (1933) and occurs in the statement of Theorem 1 of Bell (1961). It is now shown that on any space of more than three points, a probability measure can be defined in such a way that there is a set of $(n - 1)$ pairwise independent random variables.

Theorem 2.1. *For any probability measure on a space of n distinct points, a set of at most $(n - 1)$ pairwise independent random variables can be defined. A maximal set can be obtained only if each random variable takes precisely two distinct values with positive measure. A maximal set can be obtained for each value of $n > 3$. If the measure, n^{-1}, is assigned to each point of the space the solution is equivalent to determining a Hadamard matrix of size n.*

Proof. Since the space contains only a finite number of points, it can be assumed that each of the random variables X_j possesses finite moments of all orders and hence that each X_j is standardized to have zero mean and unit variance. The values of the jth variable can be written as the elements of a column vector $\mathbf{x}_j$ so that on the ith point of the space X_j takes the value x_{ij}. It is convenient to write the constant vector as $\mathbf{x}_n$ with each element $x_{in} = 1$. The number of pairwise independent random variables is not greater than

Received 30 December 1963; revised 12 February 1965.

$(n-1)$; for the random variables and the constant function are mutually orthogonal with respect to the measure, $\{p_i\}$ say, (since if $k \neq k'$,

$$\sum_{i=1}^{n} x_{ik}\, x_{ik'}\, p_i = E\,(X_k X_{k'}) = EX_k\, EX_{k'} = 0, \tag{2.1}$$

by the pairwise independence conditions) and there cannot be a set of more than n mutually orthogonal functions on a measure space of n points. The problem is now to determine the conditions under which a maximal set of $(n-1)$ pairwise independent random variables is possible. Such a maximal set of $(n-1)$ random variables cannot include any member which can assume more than two distinct values, for in this case it would be possible to define a second non-constant function, ξ say, on the range of the random variable, taken without loss of generality to be X_1. For example, X_1 itself is the first orthogonal polynomial and ξ could be taken to be the second orthogonal polynomial. Now ξ is orthogonal to X_1 and to the constant function by definition. ξ is also orthogonal to $X_2, X_3, \cdots, X_{n-1}$ by the independence condition. This would then constitute a contradiction since a function orthogonal to n mutually orthogonal functions on a space of n points is zero.

From the mutual orthogonality of the variables on the measure space, it follows that the matrix $\mathbf{PX}$ is orthogonal, where $\mathbf{P}$ is the diagonal matrix with diagonal elements, $p_i^{\frac{1}{2}}$. Consequently

$$(x_{i1}^2 + x_{i2}^2 \cdots + x_{i,n-1}^2 + 1)p_i = 1, \text{ for each } i. \tag{2.2}$$

Multiplication by x_{i1} on both sides and summation with respect to the index i yields

$$EX_1^3 + \sum_2^{n-1} EX_1X_k^2 + EX_1 = sA + (n-s)a, \tag{2.3}$$

where X_1 takes the value A, s times and the value a, $(n-s)$ times. But only the first of the expressions on the left is possibly non-zero. The second and third expressions vanish because they contain terms of the form $EX_1EX_k^2$ and EX_1, respectively. Therefore

$$\mu_3^{(1)} = sA + (n-s)a, \quad \text{where} \quad \mu_3^{(1)} = EX_1^3. \tag{2.4}$$

Let $P\{X_1 = A\}$ be p and $P\{X_1 = a\}$ be q; $p + q = 1$. Then a simple calculation gives

$$A = (q/p)^{\frac{1}{2}}, \qquad a = -1/A; \qquad \mu_3^{(1)} = (q-p)/(pq)^{\frac{1}{2}}, \tag{2.5}$$

and so from (2.4) and (2.5),

$$p = (s-1)/(n-2). \tag{2.6}$$

For $n > 3$, there always exists a solution in which the required measure and random variables are given, for $1 \leqq i \leqq n$ and $1 \leqq j \leqq n-1$, by

$$\begin{aligned} &p_i = (n-3)(n-2)^{-2}, \quad (i \leqq n-1), \qquad p_n = (n-2)^{-2}; \\ &x_{ij} = (n-3)^{\frac{1}{2}} \text{ if } i = j \text{ or } i = n, \\ &x_{ij} = -(n-3)^{-\frac{1}{2}} \text{ if } i \neq j \text{ and } i \neq n, \\ &x_{in} = 1. \end{aligned} \tag{2.7}$$

The final statement of the theorem follows after substituting $p_i = n^{-1}$: Summing as before, one obtains $\mu_3^{(1)} = 0$ since the right hand side of (2.3) is now nEX_1 and hence equal to zero. But since X_1 takes only two distinct values, (2.5) ensures that $p = \frac{1}{2}$ and X_1 takes values, ± 1. A comparison of this proposition with the orthogonality of **PX** shows that **X** is a matrix with elements, ± 1. **X** is thus of the required Hadamard form.

COROLLARY. *An orthogonal matrix,* **PX**, *with rational elements can be determined if* $(n - 3)$ *is a perfect square.*

EXAMPLES.

(i) $n = 4$ is the example of Bernstein (1945).

(ii)

$$5\mathbf{PX} = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 & -1 & 2 \\ -1 & 4 & -1 & -1 & -1 & -1 & 2 \\ -1 & -1 & 4 & -1 & -1 & -1 & 2 \\ -1 & -1 & -1 & 4 & -1 & -1 & 2 \\ -1 & -1 & -1 & -1 & 4 & -1 & 2 \\ -1 & -1 & -1 & -1 & -1 & 4 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{bmatrix}, \qquad \text{when} \quad n = 7.$$

(iii) Generalized coefficients of correlation have been defined in Lancaster (1960) as the expectations of the product of functions, orthonormal on the marginal distributions. In the multivariate distribution of $X_1, X_2, \cdots, X_{n-1}$, constructed in Theorem 2, these coefficients are relatively easy to compute because of the symmetry. The only possible coefficients are of the form,

$$\begin{aligned} \rho_{12} &= E(X_1X_2) = 0 \\ \rho_{123} &= E(X_1X_2X_3) = (n-3)^{-\frac{1}{2}} \\ \rho_{1234} &= E(X_1X_2X_3X_4) = (n-5)/(n-3), \cdots . \end{aligned} \tag{2.8}$$

These coefficients may be thought of as being of the first, second, third order. The coefficient of the $(n - 2)$th order is given by

$$\rho_{12\cdots(n-1)} = \{(n-3)^{\frac{1}{2}(n-1)} + (-1)^n(n-1)(n-3)^{-\frac{1}{2}(n-5)}\}(n-2)^{-2}. \tag{2.9}$$

When $n = 4$, this coefficient is unity and when $n = 7$, it is 2.5, an example of the generalized coefficients exceeding unity.

(iv) Let $X_1, X_2, \cdots, X_n$, be a set of n random variables, each rectangularly distributed in the unit interval, whose joint density function is given by

$$1 + \sum_{i,j} a_{ij} \operatorname{Sin} 2\pi x_i \operatorname{Sin} 2\pi x_j + \sum_{i,j,k} a_{ijk} \operatorname{Sin} 2\pi x_i \operatorname{Sin} 2\pi x_j \operatorname{Sin} 2\pi x_k \cdots, \tag{2.10}$$

for $0 \leqq x_i \leqq 1$ and $\sum |a| \leqq 1$ where all summations are over distinct indices. By appropriate choice of the $a_{ij}, a_{ijk}, a_{ijkl}, \cdots$ any required combination of interactions in the sense of Lancaster (1960) can be obtained. For example, $a_{ij} = 0$, yields pairwise independent random variables.

(v) Let the rows, columns and letters of the latin alphabet of a latin square be numbered 0, 1, 2, $\cdots$, $(n - 1)$. If the $(z + 1)$th letter occurs at the inter-

section of the $(x + 1)$th row and $(y + 1)$th column, we say that $\{x, y, z\}$ occurs in the description of the latin square. The latin square can be described by n^2 such triplets. A three dimensional distribution in variables, X, Y and Z, can now be defined by setting $P\{X = x, Y = y, Z = z\}$ equal to n^{-2}, if the $\{x, y, z\}$ occurs in the description of the latin square and equal to zero otherwise. Since each pair $\{x, y\}$ occurs precisely once with non-zero probability it is easily verified that the variables are independent in pairs. Obviously, they are not independent in a set of three, for the values of Z are completely determined by the values taken by X and Y; alternatively, with independence and the same marginal distributions, $P(\{X = x, Y = y, Z = z\}) = n^{-3}$ for all permissible values of x, y and z. This example can be generalized further by considering product measures of the same form as that given for $\{X, Y, Z\}$. In other words, the vector sets $\{X_i, Y_i, Z_i\}$ are to be mutually independent. Linear forms, U_k, V_k and W_k in $\{X_i\}$, $\{Y_i\}$ and $\{Z_i\}$,

$$U_k = n^{-1}X_1 + n^{-2}X_2 + \cdots + n^{-k}X_k, \tag{2.11}$$

and similarly V_k and W_k are then defined and it is easily shown that U_k, V_k and W_k are pairwise independent but not independent as a triple. If $k \to \infty$, then $U_k \to_L U$, $V_k \to_L V$ and $W_k \to_L W$, where U, V and W are rectangular variables, which are pairwise independent but not independent as a triple.

3. Products of random variables as complete sets. Walsh (1923) defined a set of orthonormal functions on the unit interval, which have been applied by Kaczmarz (1929) and others to number theory and probability. The Walsh functions can be exhibited as products of Rademacher functions as follows. Let any integer i be expanded in the binary system by means of the equation

$$2i = 2^{i_1} + 2^{i_2} + \cdots + 2^{i_s}, \; i_1 < i_2 < \cdots < i_s. \tag{3.1}$$

Then the Walsh function of order i may be written as a product of the Rademacher functions of orders $i_1, i_2, \cdots, i_s$. The Rademacher functions are defined on the unit interval as

$$r_k(t) = +1, \text{ if the } k\text{th digit of the binary expansion of } t \text{ is zero,} \tag{3.2}$$

and

$$= -1, \text{ if the } k\text{th digit is unity.}$$

Otherwise, $r_k(t)$ can be defined as sign (Sin $2^k\pi t$). Kac (1959) and Alexits (1961) may be consulted for the many interesting properties of the Rademacher and Walsh functions. The following theorem characterizes the two sets of functions by pairwise independence of random variables on a non-atomic probability measure space.

THEOREM 3.2. *If a set of pairwise stochastically independent random variables, $\{X_i\}$, together with the constant function, forms a complete orthonormal set on a measurable space, Ω, on which is defined a non-atomic probability measure, then each X_i takes only two distinct values with non-zero probability. If further the set $\{X_i\}$ can be represented, perhaps after renumbering, in the form*

(3.3) $X_i = Y_{i_1}Y_{i_2}\cdots Y_{i_s}$, *where* $2i = 2^{i_1} + 2^{i_2} + \cdots + 2^{i_s}$,

with $i_1 < i_2 < \cdots < i_s$, *and where* $\{Y_i\}$ *is a set of mutually independent random variables, then the* X_i *are the Walsh functions and the* Y_i *are the Rademacher functions, defined on the same measure space, on which*

(3.4) $$Y = \sum_j (1 - Y_j)/2^{j+1}$$

is a random variable rectangularly distributed on the unit interval.

PROOF. As in the proof of Theorem 2.1, each X_i can take no more than, and hence precisely, two distinct values with positive probability. For otherwise, if X_1, say, took more than two values we could define ξ, a function of X_1 orthogonal to X_1, the constant function and every other member of a complete set of orthonormal functions, which would be a contradiction. X_1 furthermore would be degenerate if it did not take two distinct values. Each X_j can be therefore considered to be in standardized form with zero mean and unit variance. Each Y_j also is an X_k. Now it is given that some X_i is of the form Y_1Y_2, where Y_1 and Y_2 are independent random variables. Suppose that Y_1 takes its two values on complementary sets, A and A', and that Y_2 takes its two values on complementary sets, B and B'. Then X_i assumes a constant value on each of the four sets, AB, AB', $A'B$ and $A'B'$ but only two of these values are distinct. So that X_i is constant on $AB \cup A'B'$ and on $AB' \cup A'B$. This is only possible if the two values of Y_1 and Y_2 are ± 1. The reasoning is general and so every Y_j takes only two distinct values, namely ± 1. Since each X_i has been standardized, so is each Y_j and so Y_j takes each of its two values with probability, $\frac{1}{2}$. It then follows from the mutual independence of any finite subset of k variables of $\{Y_i\}$, that the values ± 1, assigned to each, partition the space into 2^k sets, each corresponding to a probability measure, 2^{-k}. The proof is complete.

A similar theorem can be proved for measures positive at each of n points, where n is a power of 2, but is of little interest. The procedure used in the theorem is only possible if the measure is non-atomic, since it would not have been possible to find k independent random variables if the measure of the largest atom was greater than 2^{-k}.

REFERENCES

ALEXITS, G. (1961). *Convergence Problem of Orthogonal Series*. Pergamon Press, England.

BELL, C. B. (1961). Maximal independent stochastic processes. *Ann. Math. Statist.* **32** 704–708.

BERNSTEIN, S. (1945). *Theory of Probability* (3rd edition). GTTI, Moscow.

KAC, M. (1959). *Statistical Independence in Probability Analysis and Number Theory*. Math. Assoc. America, Rahway, N.J.

KACZMARZ, S. (1929), Über ein Orthogonalsystem. *C.R. du 1e Cong. des Math. des Payes Slaves*, 189–192.

KOLMOGOROV, A. N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik II, **4** Springer, Berlin; translated as *Foundations of the Theory of Probability*. Chelsea, New York, 1950.

LANCASTER, H. O. (1960). On tests of independence in several dimensions. *J. Austral. Math. Soc.* **1** 241–254; Correction. *J. Austral. Math. Soc.* **1** 496.

WALSH, J. L. (1923). A closed set of normal orthogonal functions. *Amer. J. Math.* **45** 5–24.