ON STATISTICAL INDEPENDENCE AND ZERO CORRELATION IN SEVERAL DIMENSIONS.

H. O. LANCASTER

(received 3rd May, 1960; revised 30th June, 1960)

Bivariate distributions, subject to a condition of ϕ^2 boundedness to be defined later, can be written in a canonical form. Sarmanov [4] used such a form to deduce that two random variables are independent if and only if the maximal correlation of any square summable function, $\xi(x_1)$, of the first variable with any square summable function, $\eta(x_2)$, of the second variable is zero. This is equivalent to the condition that the canonical correlations are all zero. The theorem of Sarmanov [4] was proved without any restriction in Lancaster [2] and the proof is now extended to an arbitrary number of dimensions.

Notational. Let $x_1, x_2, \cdots, x_n$ be a set of n random variables with joint probability function, $P(A)$, and product probability function, $P^*(A)$, defined by

$$(1) \qquad P(A) \equiv P(A_{\{j\}}^{\{i_j\}}) = P(x_1 \in A_1^{(i_1)}, x_2 \in A_2^{(i_2)}, \cdots, x_n \in A_n^{(i_n)})$$

$$(2) \qquad P^*(A) \equiv P^*(A_{\{j\}}^{\{i_j\}}) = \prod_j P(x_j \in A_j^{(i_j)}).$$

The affixes $\{j\}$ and $\{i_j\}$ are to be interpreted as $12\cdots n$ and $i_1 i_2 \cdots i_n$. We shall write sums and integrals using the integral sign and dP or dP^* as the case may be. Real orthonormalised functions can be defined on the distribution of each x_j and will be denoted by $x_j^{(i_j)}$, $i_j \neq 0$; $x_j^{(0)} = 1$. The expectation of a product $\prod_j x_j^{(i_j)}$ on the joint distribution such that at least two superscripts i_j are not zero will be called a generalised coefficient of correlation and written $\rho^{\{i_j\}}$. These coefficients may not be less than unity in absolute value. If precisely two of the i_j are non-zero for any such coefficient, it becomes an ordinary coefficient of correlation. In the enunciation of the theorems below a mention is made of complete sets. This will only be appropriate for marginal distributions with an infinite number of points of increase. It is to be understood if there are only a finite number of such points that the complete set is to be replaced by an orthonormal set forming a basis, namely one consisting of unity and $n_j - 1$ orthonormal functions and the notation is to be taken as modified accordingly. It is convenient to give two theorems, the first of which is a special case of the second. Two lemmas are first proved.

LEMMA 1. *If the random variables, $x_1, x_2, \cdots, x_n$, do not form a mutually independent set, then step functions, $X_1, X_2, \cdots, X_n$, everywhere finite and normalised to have zero mean and unit variance on the appropriate marginal distributions, can be found such that at least one of the products of the X_i taken two or more at a time has non-zero expectation.*

PROOF. Let normalised step functions having only two values be defined as follows on each of the marginal distributions of the form,

$$(3) \qquad \begin{aligned} X(x) &= -\sqrt{\{P(x \in A^{(2)})/P(x \in A^{(1)})\}} \quad \text{for } x \in A^{(1)} \\ &= +\sqrt{\{P(x \in A^{(1)})/P(x \in A^{(2)})\}} \quad \text{for } x \in A^{(2)} \end{aligned}$$

where $P(A^{(1)}) + P(A^{(2)}) = 1$, $A^{(1)}$ and $A^{(2)}$ are mutually exclusive sets, neither $P(A^{(i)})$ being null. X is normalised on the distribution of x as well as on its own two-point distribution.

Let it now be supposed that the x_j do not form a mutually independent set. Then there exists at least one division of each of the marginal distributions into two proper mutually exclusive sets so that the $P(A) \equiv P(A^{\{i_j\}}_{\{j\}})$ of (1) is not equal to $P^*(A)$ of (2) for some set $\{i_j\}$. The superscripts i_j take values, 1 and 2. Now let us define a set of variables $X_j \equiv X_j(x_j)$ corresponding to this partition and consider the joint distribution of the $\{X_i\}$. These new variables do not form a mutually independent set and the probability distribution is defined on 2^n points. The product set of functions $\{1, X_1\} \times \{1, X_2\} \times \cdots \times \{1, X_n\}$ are all mutually orthogonal and indeed orthonormal on P^* and form an orthonormal basis for finite functions on the partition of P^* into 2^n sets of positive probability. In particular, the ratio, $Q \equiv Q(\{X_j\})$, of the measures assigned to the 2^n "quadrants" by P and P^* by making the transformation, $\{x_j\} \to \{X_j\}$, is finite and can be expressed as a linear form in the elements of the orthonormal basis. The coefficients of the elements in this linear form will be equal to $\int QX_1^{i_1}X_2^{i_2} \cdots X_n^{i_n} dP^*$, which is equal to $\int X_1^{i_1}X_2^{i_2} \cdots X_n^{i_n} dP$, since on P and P^*, Q is a step function constant over any given "quadrant" and so is the product, $X_1^{i_1}X_2^{i_2} \cdots X_n^{i_n} i_j = 0$ or 1. The coefficient of the constant term is unity and all the other coefficients cannot vanish since if they did the P-measure assigned to each quadrant would be P^*-measure. This would constitute a contradiction since the hypothesis given assumes that the probability function $P(A)$ does not factorise for this particular partition. The coefficient of $X_1, X_2, \cdots X_n$ taken singly in the linear form are all zero.

https://doi.org/10.1017/S1446788700026288 Published online by Cambridge University Press

LEMMA 2. *If $x_1, x_2, \cdots, x_k$ form a mutually independent set, then a product of orthonormal functions, $X_1 X_2 \cdots X_k$, can be approximated in quadratic mean arbitrarily closely by expressions of the form, $S_1 S_2 \cdots S_k$, where*

$$S_j = \sum_{i=1}^{N_j} a_j^{(i)} x^{(i)}, \qquad j = 1, 2, \cdots, k \tag{4}$$

The proof is by induction. Given that the lemma is true for $(k-1)$, $S_1, S_2, \cdots, S_k$ can be chosen such that, for arbitrary ε, $E(X_1 \cdots X_{k-1} - S_1 \cdots S_{k-1})_2 \leqq \varepsilon^2/4$ and also $E(X_k - S_k)^2 \leqq \varepsilon^2/4$. Then, since for real numbers, $(a+b)^2 \leqq 2(a^2+b^2)$,

$$(X_1 X_2 \cdots X_k - S_1 S_2 \cdots S_k)^2 \leqq 2[\{X_1 \cdots X_{k-1}(X_k - S_k)\}^2 + \{S_k(X_1 X_2 \cdots X_{k-1} - S_1 S_2 \cdots S_{k-1})\}^2]. \tag{5}$$

Taking expectations, we find that the expectation on the left of (5) is not greater than ε^2, for the independence of the k variables enables the expectations of the two expressions on the right to be evaluated as products of expectations. Further $E(S_k^2) \leqq E(X_k^2) = 1$.

THEOREM 1. *Let $\{x_1^{(i_1)}\}$ and $\{x_2^{(i_2)}\}$ be complete orthonormal sets on the marginal distributions of two random variables, which have a joint probability measure, $P \equiv P(x_1, x_2)$, and let $x_1^{(0)} = x_2^{(0)} = 1$. Then a necessary and sufficient condition for the independence of x_1 and x_2 is that every coefficient of correlation, $\rho_{i_1 i_2} = \int x_1^{(i_1)} x_2^{(i_2)} dP$, should be zero for every $i_1 > 0$, $i_2 > 0$.*

PROOF. The *necessity* is obvious.

$$\int x_1^{(i_1)} x_2^{(i_2)} dP(x_1, x_2) = \int x_1^{(i_1)} dP(x_1) \int x_2^{(i_2)} dP(x_2) = 0. \tag{6}$$

Sufficiency. By Lemma 1, we have at least one pair of orthonormalised step functions, X_1 and X_2, such that

$$0 \neq \rho = \int X_1 X_2 dP. \tag{7}$$

By taking finite numbers of terms in the expansions, S_1 and S_2, the expectations of $(X_1 - S_1)^2$ and $(X_2 - S_2)^2$ can both be made less than ε^2 on their respective marginal distributions and hence with respect to the P-measure. The identity holds,

$$X_1 X_2 = X_1(X_2 - S_2) + X_2(X_1 - S_1) - (X_1 - S_1)(X_2 - S_2) - S_1 S_2. \tag{8}$$

Integrating with respect to P-measure, it follows that

$$\rho \leqq 2\varepsilon + \varepsilon^2, \tag{9}$$

since by an application of the Schwarz inequality the first two terms on the right are each less than ε and the third is less than ε^2. The fourth term is identically zero. ε can be made arbitrarily small and so (9) is a contradiction. x_1 is independent of x_2 if all the correlations vanish.

https://doi.org/10.1017/S1446788700026288 Published online by Cambridge University Press

THEOREM 2. *The random variables* $\{x_i\}$ *are mutually independent if and only if the generalised coefficients of correlation,* $\rho^{\{i_j\}}$, *corresponding to complete sets of orthonormal functions, are all zero.*

PROOF. The *necessity* is evident as in Theorem 1. For the *sufficiency*, it is impossible that any product of the form, $X_i X_j$, $i \neq j$ should have a non-zero expectation by Theorem 1. Let us now suppose the theorem true for sets of k variables and prove it true for a set of $k+1$ variables, $2 \leqq k < n$. Without loss of generality, we take the variables to be $x_1, x_2, \cdots, x_k$. Let us suppose that $x_1, x_2, \cdots, x_{k+1}$ do not form a mutually independent set. By Lemma 2 at least one product of the form $X_{j_1} X_{j_2} \cdots$ can be found which has a non-zero expectation. This can only be $X_1 X_2 \cdots X_{k+1}$ since any such product with less than $k+1$ factors would have been zero. But now as before we can consider the identity

$$\begin{aligned}(10) \qquad &(X_1 X_2 \cdots X_k) X_{k+1} \\ &= X_{k+1}(X_1 X_2 \cdots X_k - S_1 S_2 \cdots S_k) + X_1 X_2 \cdots X_k (X_{k+1} - S_{k+1}) \\ &\quad - (X_{k+1} - S_{k+1})(X_1 X_2 \cdots X_k - S_1 S_2 \cdots S_k) - S_1 S_2 \cdots S_{k+1}\end{aligned}$$

and there follows

$$(11) \qquad 0 \neq E(X_1 X_2 \cdots X_{k+1}) < 2\varepsilon + \varepsilon^2$$

and a contradiction has been reached. The vanishing of the generalised coefficients of correlation ensure the independence of the marginal variables taken in pairs, then in threes, fours, $\cdots$, and finally in a set of n. They form a completely independent set.

COROLLARY. *A necessary and sufficient condition for independence of two variables,* x *and* y, *is that the maximum correlation between any two functions* $\xi(x)$ *and* $\eta(y)$ *should be zero.* (Sarmanov [4]).

PROOF. The necessity is obvious. The sufficiency follows as in Theorem 1 by noting that $E\{\xi(x)\eta(y)\} = 0$ forces all the $\rho_{ij} = E(x_1^{(i)} x_2^{(j)})$ to be zero since otherwise $\xi(x)$ and $\eta(y)$ would not have maximal correlation.

EXAMPLE. Let $n = 2$ and let the marginal distributions of the variables, x and y, be uniform in the interval, -1 to $+1$. Define complete orthonormal sets of functions of the form $x^{(i)} = P_i(x)\sqrt{(2i+1)}$, where $P_i(x)$ is the Legendre polynomial of degree, i. Let

$$(12) \quad f(x, y) = \tfrac{1}{4}(1 + a_{11} x^{(1)} y^{(1)} + a_{21} x^{(2)} y^{(1)} + a_{12} x^{(1)} y^{(2)} + a_{31} x^{(3)} y^{(1)} + \cdots),$$

where

$$(13) \qquad 2^{i+j-1}(i+j-1)|a_{ij}| \leqq |[\max. x^{(i)} \max. y^{(j)}]|^{-1}$$

Then $f(x, y)$ is never negative over the ranges of its marginal variables and is thus a probability density function with the required marginal densities.

Further $E(x^{(i)}y^{(j)}) = a_{ij}$ and x and y are independent only if all the a_{ij} vanish. Theorem 1 evidently requires that each of the members of a complete set on the first distribution is uncorrelated with each of the members of a complete set on the second distribution. This example can be extended to show that the orthonormal sets chosen must be complete in the enunciation of Theorem 2.

Much has been written on the relation between independence and zero correlation but it appears that these theorems are more general than any previously proved. They are more general than that of Sarmanov [4] because they hold for several dimensions and avoid making the bounded ϕ^2 restriction namely that dP/dP^*, the Radon-Nikodym derivative of P with regard to P^*, is square summable on P^*. The probability function in multivariate ϕ^2 bounded distributions can be written as a product of the marginal probabilities by a series in products of orthonormal functions, in which the coefficient of each product is simply its expectation. Such expansions are given in Lancaster [1] and [3]. Dependencies in multivariate distributions are classified in Lancaster [3] by the vanishing of various classes of the generalised correlation coefficients.

It is evident from these papers that the ordinary Pearson χ^2 chooses the indicator variables on the marginal distributions as the set of functions to be normalised. However, the method can be generalised to use, for example, the Hermite-Chebyshev polynomials in the joint normal case. The χ^2 in either of these cases tests whether the sum of the squares of the generalised coefficients of correlation can be all considered to be zero.

References

[1] Lancaster, H. O., The structure of bivariate distributions, Ann. Math. Statist. 29 (1958), 719–736.

[2] Lancaster, H. O., Zero correlation and independence, Aust. J. Statist. 1 (1959), 53–56.

[3] Lancaster, H. O., On tests of independence in several dimensions, J. Aust. Math. Soc. 1 (1960), 241–254.

[4] Sarmanov, O. V., Maximum correlation coefficient (non-symmetrical case), Dokl. Akad. Nauk S.S.S.R. 121 (1958), 52–55.

Department of Mathematical Statistics,
University of Sydney.

CORRECTIONS

to H. O. Lancaster: "On tests of independence in several dimensions". This Journ. 1 (1960), 241–254.

At the top of page 244, equation (3) should read

$$\phi_{xy}^2 + \phi_{xz}^2 + \phi_{yz}^2 + \phi_{xyz} = \phi^2 \tag{3}$$

The statement in the lines following (3) that "each ρ is less in absolute value than unity" is not universally true.

In the statement of Theorem 8, an integral sign has been omitted after "expression".

In (18) for $\pm$ read $-$.

On page 253, in lines 14 and 15 in place of "$N^{\frac{1}{2}}$ times square roots of the ϕ_{xy}^2, ϕ_{xz}^2 and ϕ_{yz}^2", read "$N\phi_{xy}^2$, $N\phi_{xz}^2$ and $N\phi_{yz}^2$".