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.

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