CORRELATION AND COMPLETE DEPENDENCE OF RANDOM VARIABLES

By H. O. Lancaster

University of Sydney

1. Introduction. In the bivariate joint normal distribution, zero correlation implies independence and unit correlation implies that one variable is a linear function of the other. The coefficient of correlation in such a joint normal distribution is, moreover, a parameter determining the distribution once the means and variances have been given. Between the limits of zero and unity, an interpretation can be given to the absolute value of the coefficient of correlation. Without the hypothesis of joint normality it can no longer be assumed that zero correlation implies independence; in fact, in the general case, a necessary and sufficient condition for independence is that every standardized function of the first variable should have zero correlation with any standardized function of the second variable. Necessary and sufficient conditions are now given for the complete mutual dependence of random variables. Examples are given to show that the results cannot be improved and that certain measures of dependence, valid for the normal distribution have little interest in the more general case.

2. Complete mutual dependence of random variables. A random variable Y will be said to be completely dependent on a random variable X if Y takes only one value for each value of X with probability one. More formally if there exists a single valued function, $Y \equiv f(X)$, so that the set $\{(x, f(x)), \text{all } x\}$ is measurable and has probability one, then Y will be said to be completely dependent on X. If also X is completely dependent on Y, the two variables will be said to be mutually completely dependent.

Examples. Let X be rectangularly distributed in the open interval $(0, 1)$ in each of the following examples

(2.i) Let $Y = -1$ when $X \leqq \frac{1}{2}$, $Y = +1$ when $X > \frac{1}{2}$. Then Y is completely dependent on X but X is not completely dependent on Y.

In the following examples, X and Y are mutually completely dependent.

(2.ii) Let $Y = aX + b$, where $a \neq 0$ and a and b are constants.

(2.iii) Let Y be any strictly monotonic function of X.

(2.iv) Let X and Y be related by any bi-unique transformation,

$$Y = f(X), \qquad X = f^{-1}(Y), \tag{2.1}$$

true with probability one.

Let now X and Y be random variables with a general bivariate distribution function, $F(x, y)$ and marginal distribution functions, $G(x)$ and $H(y)$. E is used as the operator for taking mathematical expectations or integrating with respect to the probability measure, $F(x, y)$.

Received February 11, 1963; revised June 16, 1963.

The main theorem will be established with the aid of the theory of orthonormal functions, so that it will be true only with probability one. A random variable with finite variance (or square summable function) is said to be normalized if it is centred and scaled to have zero mean and unit variance. Normalized random variables cannot be degenerate. Correlation between two random variables can be defined if and only if each has finite variance. Two random variables, $\zeta_1(X, Y)$ and $\zeta_2(X, Y)$, will be said to be equivalent in the mean square sense, or more briefly, substantially equivalent if

$$E(\zeta_1 - \zeta_2)^2 \equiv \int \{\zeta_1(x, y) - \zeta_2(x, y)\}^2 \, dF(x, y) = 0. \tag{2.2}$$

In other words, ζ_1 and ζ_2 are equivalent if $\zeta_1 = \zeta_2$ almost everywhere with respect to the probability measure, $F(x, y)$.

LEMMA 1. *If two normalized functions are equivalent, they have unit correlation and conversely.*

PROOF.

$$E(\zeta_1 - \zeta_2)^2 = E\zeta_1^2 - 2E\zeta_1\zeta_2 + E\zeta_2^2 = 2 - 2 \operatorname{corr}(\zeta_1, \zeta_2). \tag{2.3}$$

LEMMA 2. *The normalized functions having unit correlation with a given normalized function form an equivalence class.*

PROOF. Only the transitivity needs proof. Let ζ_1 and ζ_2 have unit correlation with ζ. Then

$$E(\zeta_1 - \zeta_2)^2 = E(\zeta_1 - \zeta)^2 + E(\zeta_2 - \zeta)^2 - 2E(\zeta_1 - \zeta)(\zeta_2 - \zeta) = 0, \tag{2.4}$$

for the first two integrals vanish by hypothesis and the third by an application of the Bunyakovsky-Schwarz inequality.

THEOREM 1. *Let $\{x^{(i)}\}$, $x^{(0)} = 1$, and $\{y^{(i)}\}$, $y^{(0)} = 1$, be complete sets of orthonormal functions defined on the marginal distributions, $G(x)$, and $H(y)$, of the random variables, X and Y, which have a joint distribution function, $F(x, y)$. Then for the complete mutual dependence of X and Y, either of the following conditions is necessary and sufficient:*

(i) *the matrix of correlations, $\mathbf{A}$, defined by*

$$a_{ij} = \operatorname{corr}(x^{(i)}, y^{(j)}) = \int x^{(i)} y^{(j)} \, dF(x, y), \qquad i > 0, j > 0 \tag{2.5}$$

is orthogonal,

(ii) *every normalized function of X (or Y) has a correlation of $+1$ with a substantially unique normalized function of Y (or X).*

PROOF. (i) implies (ii) for the elements of $\{y^{(i)}\}$ can be written as the elements of a vector, $\mathbf{y}$. Defining a new set by

$$\mathbf{y}^* = \mathbf{A}\mathbf{y}, \tag{2.6}$$

the matrix of correlations between the elements of $\{x^{(i)}\}$ and $\{y^{*(j)}\}$ is the unit matrix, for

$$E\mathbf{x}\mathbf{y}^{*T} = E\mathbf{x}\mathbf{y}^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \mathbf{1}, \tag{2.7}$$

where we have used the convention that $E(z_{ij}) = (Ez_{ij})$. Now let ξ be a normalized function of X. It follows from the usual theory of orthonormal functions that

$$\xi = \sum_i b_i x^{(i)}, \qquad \sum b_i^2 = 1, \qquad b_0 = 0. \tag{2.8}$$

If $\eta = \sum_i b_i y^{*(i)}$, then the correlation of ξ and η is given by

$$E(\xi\eta) = E\mathbf{b}^T\mathbf{x}\mathbf{y}^{*T}\mathbf{b} = \mathbf{b}^T\mathbf{1}\mathbf{b} = 1, \tag{2.9}$$

and by Lemma 2, η is substantially unique.

(ii) implies (i). For let $\{x^{(i)}\}$ be a complete orthonormal set on $G(x)$. Then it is given that there is a set of functions, $\{y^{*(i)}\}$ say, such that

$$Ex^{(i)}y^{*(i)} = 1, \qquad i = 1, 2, \cdots \tag{2.10}$$

where each of the $y^{*(i)}$ has unit variance. Then $\{y^{*(i)}\}$ is a set orthonormal on the marginal distribution, since it is orthonormal on the bivariate distribution, for

$$Ey^{*(i)}y^{*(k)} = Ey^{*(i)}(y^{*(k)} - x^{(k)}) + E(y^{*(i)} - x^{(i)})x^{(k)} + Ex^{(i)}x^{(k)}. \tag{2.11}$$

The first two integrals vanish because of the Bunyakovsky-Schwarz inequality and the third because $\{x^{(i)}\}$ is orthonormal. Now $\{y^{*(i)}\}$ may be not complete on $H(y)$. If not, let there be adjoined the orthogonal complement of $\{y^{*(i)}\}$ and call it $\{z^{(i)}\}$. But by hypothesis orthonormal functions in X, $\{w^{(i)}\}$ say, can be found such that $Ew^{(i)}z^{(i)} = 1$. By the argument above, each $w^{(i)}$ is a square summable function of X orthogonal to each $x^{(j)}$. This is a contradiction unless $\{w^{(i)}\}$, and hence $\{z^{(i)}\}$, is empty. $\{y^{*(i)}\}$ is therefore complete with respect to $H(y)$ and is consequently an orthogonal transform of any other complete orthonormal set, $\{y^{(i)}\}$. It follows that

$$\mathbf{y} = \mathbf{A}^T\mathbf{y}^*, \quad \mathbf{A} \text{ orthogonal, and } \quad \mathbf{A} = E\mathbf{x}\mathbf{y}^T. \tag{2.12}$$

The theorem will now be proved for condition (ii).

NECESSITY. Suppose that the variables are completely mutually dependent and let $\xi = \xi(X)$ be an arbitrary normalized function. For any point of increase of $H(Y)$ define $\eta = \eta(Y) = \xi(X)$. Since Y determines a unique X a.e., this rule defines η uniquely as a function of Y and obviously the correlation of ξ and η is unity.

SUFFICIENCY. The method of proof is to make increasingly fine partitions of the spaces of the marginal variables. For $k = 1, 2, 3, \cdots$, let the space of the distribution of X be partitioned into n_k sets, $A_k^{(i)}$, such that $P(X \varepsilon A_k^{(i)})$ is not zero for any i. Then this partition also determines a partition of the space of the distribution of Y into n_k sets, $B_k^{(i)}$, and $P(Y \varepsilon B_k^{(i)})$ is not zero for any i. For suppose that $P(X \varepsilon A_k^{(i)}) = p \neq 0$. Then a normalized variable $\xi^{(i)}$ may be defined to be $-\sqrt{\{p/(1-p)\}}$ when X is not in $A_k^{(i)}$ and to be $\sqrt{\{(1-p)/p\}}$ when X is in $A_k^{(i)}$. By hypothesis there is a normalized function of Y, $\eta^{(i)}$, such that $E\eta^{(i)}\xi^{(i)} = 1$. $\eta^{(i)}$ is therefore also a step function with two distinct values and partitions the space of the distribution of Y into $B_k^{(i)}$ and its complement. Further, the intersection of $A_k^{(i)}$ and the complement of $B_k^{(i)}$ is empty, and similarly with $B_k^{(i)}$ and the complement of $A_k^{(i)}$. The bivariate distribution is thus partitioned into two disjunct pieces. It is readily seen that the intersection of $A_k^{(i)}$ and $B_k^{(j)}$ is empty if $i \neq j$ since $B_k^{(j)}$ lies in the complement of $B_k^{(i)}$. The space of Y is broken into n_k disjunct pieces. If possible, let the space of the distribution of Y be subpartitioned into sets, $\{B_{k+1}^{(i)}\}$, each of which is associated with positive probability. This in turn leads to a refinement of the partition of the space of the distribution of X into the sets, $\{A_{k+1}^{(i)}\}$, say, and the bivariate distribution is broken down into n_{k+1} disjunct parts. The notation has implied that X and Y are numerical variables but this is not necessary for the argument. The partitioning of the marginal probability spaces can proceed to any assigned degree of refinement or a stage is reached when no further partition is possible, as in the discrete distributions; in either case we have thus set up a bi-unique transformation,

$$X \,\varepsilon\, A_k^{(i)} \leftrightarrow Y \,\varepsilon\, B_k^{(i)}, \quad \text{or} \tag{2.13}$$

$$X = f(Y), \qquad Y = f^{-1}(X). \tag{2.14}$$

3. Correlations in arbitrary distributions. We now construct some bivariate distributions to show that generalizations from the normal distribution to other distributions may not be useful. For this purpose, we will use mixtures of bivariate distributions and hence we need to be able to specify complete orthonormal sets on mixtures of marginal distributions. To simplify the discussion, we choose two distribution functions, $G(x)$ and $J(x)$, such that there is no set of points, to which positive measure is assigned by both distributions: we can say that the two spaces are disjunct. A mixture can be obtained by taking a linear combination,

$$K(x) = \beta_1 G(x) + \beta_2 J(x), \qquad 0 < \beta_j < 1, \beta_1 + \beta_2 = 1. \tag{3.1}$$

LEMMA 3. *If* $\{u^{(i)}\}$ *and* $\{v^{(i)}\}$ *are complete sets on* $G(x)$ *and* $J(x)$, *respectively and* $u^{(0)} = v^{(0)} = 1$, *then a complete set on* $K(x)$ *can be obtained by means of the following*:

$$\begin{aligned}
k^{(0)}(x) &= 1 \quad \textit{for both spaces},\\
k^{(1)}(x) &= -\,(\beta_2/\beta_1)^{\frac{1}{2}} \quad \textit{for the space of } G(x),\\
&= +\,(\beta_1/\beta_2)^{\frac{1}{2}} \quad \textit{for the space of } J(x),\\
k^{(2i)}(x) &= \beta_1^{-\frac{1}{2}}\, u^{(i)},\ i = 1, 2, \cdots \quad \textit{for the space of } G(x),\\
&= 0 \quad \textit{otherwise}.\\
k^{(2i+1)}(x) &= \beta_2^{-\frac{1}{2}}\, v^{(i)},\ i = 1, 2, \cdots \quad \textit{for the space of } J(x),\\
&= 0 \quad \textit{otherwise}.
\end{aligned} \tag{3.2}$$

If either $G(x)$ or $J(x)$ has only a finite number of points of increase, some notational changes will have to be made in (3.2).

A bivariate distribution is said to be ϕ^2-bounded with respect to the marginal distributions or simply ϕ^2-bounded if

$$\phi^2 + 1 = \int \Omega^2(x, y)\, dG(x)\, dH(y) < \infty, \tag{3.3}$$

where $\Omega(x, y)$ is the Radon-Nikodym derivative of $F(x, y)$ with respect to $G(x)H(y)$; $\Omega(x, y) \equiv dF(x, y)/\{dG(x)\, dH(y)\}$, Lancaster (1958). A finite value of ϕ^2 can be given an interpretation,

$$\phi^2 = \sum_{i,j} a_{ij}^2, \tag{3.4}$$

where $a_{ij} = Ex^{(i)}y^{(j)}$; $i > 0, j > 0$, and $\{x^{(i)}\}$ and $\{y^{(i)}\}$ are complete sets on the marginal distributions. Of the following examples, only the first is ϕ^2-bounded.

Example 3.i. Let the random variables, X and Y, have a bivariate distribution uniform on the sides of a square with corners at $(\pm 1, \pm 1)$. The marginal distribution of X can be expressed in the form (3.1) with $\beta_1 = \beta_2 = \frac{1}{2}$, $G(x)$ being absolutely continuous with the density function, $G'(x) = \frac{1}{2}$, on the open interval $(-1, 1)$ and $J(x)$ having saltuses of $\frac{1}{2}$ at ± 1. The complete set is $k^{(0)}(x) = 1$, $k^{(1)}(x) = -1$ for $-1 < x < 1$ $k^{(1)}(x) = +1$ for $x = \pm 1$ and $k^{(1)}(x) = 0$ otherwise. Corresponding to the $k^{(2i)}(x)$, is $\sqrt{2}u^{(i)}$ where $u^{(i)}$ are the normalized Legendre polynomials. There is only one $k^{(2i+1)}(x)$, namely $k^{(3)}(x) = \pm 1$ according as $x = \pm 1$ and $k^{(3)}(x) = 0$ elsewhere. All other $k^{(2i+1)}(x)$ are not defined, so that the notation can be changed so that $k^{(2i)}(x)$ becomes $k^{(i+2)}(x)$ for $i > 1$. There is but a single non-zero correlation namely one of -1 between variables of the form $k^{(1)}(x)$. $\phi^2 = 1$.

Example 3.ii. A mixture of a uniform distribution along the straight line from $(-1, -1)$ to the origin with a second uniform distribution over the interior of a square with corners, $(0, 0)$, $(0, 1)$, $(1, 1)$ and $(1, 0)$. In this case, the orthonormal set is given by (3.2) with $k^{(2i)}(x)$ and $k^{(2i+1)}(x)$ orthonormal polynomials on the two rectangular distributions on the intervals $(-1, 0)$ and $(0, 1)$ respectively. The correlation between $k^{(1)}(X)$ and $k^{(1)}(Y)$ is $+1$, between $k^{(2i)}(X)$ and $k^{(2i)}(Y)$ is $+1$, and between $k^{(2i+1)}(X)$ and $k^{(2i+1)}(Y)$ is zero, $i = 1, 2, 3, \cdots$. All correlations between unlike functions of X and Y are zero. ϕ^2 is infinite. There is an infinity of maximal correlations of unity. It may be noted that for every positive value of β_1, there is an infinity of unit correlations and ϕ^2 is infinite. A measure of dependence should give some weight to the fact that there is a probability of β_1 of knowing the exact value of Y after sampling X. If β_1 is very small, knowledge of X will yield little information on Y.

Example 3.iii. A mixture of two distributions, the first a uniform (singular) distribution along the diagonal and the second a uniform distribution over the interior of a square with corners at $(\pm 1, \pm 1)$. In this case, the complete orthonormal sets are the standardized Legendre polynomials. The correlation between

$P^{(i)}(X)$ and $P^{(j)}(Y)$ is $\beta_1 \delta_{ij}$. ϕ^2 is unbounded. It is easily verified that the maximal correlation is β_1 and that any normalized function $g(X)$ has a correlation of β_1 with $g(Y)$.

EXAMPLE 3.iv. A mixture of a uniform (singular) distribution along the line, joining $(-1, -1)$ and $(+1, +1)$, excluding the end points and a distribution of four weights $\frac{1}{4}(1 + \rho XY)$, $|\rho| < 1$ at the four points, $(\pm 1, \pm 1)$. The orthonormal set is given as in (3.2), $k^{(2i)}(X)$ being the standardized Legendre polynomials $k^{(3)}(X)$ taking values $\pm \beta_2^{-\frac{1}{2}}$, all the other odd functions not being defined. The correlation between $k^{(1)}(X)$ and $k^{(1)}(Y)$ is unity. The correlation of the pairs of the even series is also unity. The correlation of $k^{(3)}(X)$ with $k^{(3)}(Y)$ is ρ. In this example, there is an infinity of unit correlations and a single correlation not unity. This last is thus sufficient to ensure that X and Y are not mutually completely dependent. The distribution shows that Linfoot's (1957) information measure of correlation can be unity without complete mutual dependence.

Certain generalizations, useful in normal correlation, do not carry over to such general distributions.

(1) ϕ^2. In normal theory, $\phi^2 = 0$ implies independence; and infinite ϕ^2 implies mutual complete dependence. Examples 3.ii, 3.iii and 3.iv show that ϕ^2 can be infinite without the variables being mutually completely dependent.

(2) *The maximal correlations.* In normal theory, the maximal correlation is $|\rho|$; if the maximal correlation is unity the marginal (normal) variables are mutually completely dependent. Examples 3.ii and 3.iv show that this does not suffice to give mutual complete dependence in the general case. Gebelein (1942 and 1952) found, however, that it was sufficient in continuous distributions for X to have unit correlation with Y. This can be deduced as a corollary from Lemma 1.

(3) *Averaged correlations.* In the contingency tables various authors have suggested averaged values of the squares of the coefficients of correlation, such as $\phi^2/\sqrt{\{(r-1)(c-1)\}}$ or $\phi^2/(r-1)$, where r is the number of rows and c is the number of columns. It might be thought that some average value of the correlations or their squares might be defined for the general case. Example 3.iv shows that this hope is illusory; for in this example, we would have a limiting ratio of unity without complete mutual dependence. The examples also show that there is no necessary relation between ϕ^2 and the maximal correlation except that the latter is bounded by ϕ^2. Reviews of various measures of association have been given by Gebelein (1941), Rényi (1959) and Höffding (1942).

Condition (ii) in Theorem 1 of this paper can be regarded as an extension of the main theorem of Lancaster (1963).

REFERENCES

[1] GEBELEIN, H. (1941). Das statistische Problem der Korrelation als Variations—und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.* **21** 364–379.

[2] GEBELEIN, H. (1942). Bermerkung über ein von W. Höffding vorgeschlagenes, massstabsinvariantes Korrelationsmasz. *Z. Angew. Math. Mech.* **22** 171–173.

[3] GEBELEIN, H. (1952). Maximalkorrelation und Korrelationsspektrum. *Z. Angew. Math. Mech.* **32** 9–19.

[4] HÖFFDING, W. (1942). Stochastische Abhängigkeit und funktionaler Zusammenhang. *Skand. Aktuarietidskr.* **25** 200–227.

[5] LINFOOT, E. H. (1957). An informational measure of association. *Information & Control* **1** 85–89.

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

[7] LANCASTER, H. O. (1963). Correlations and canonical forms of bivariate distributions. *Ann. Math. Statist.* **34** 532–538.

[8] RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10** 441–451.