CORRELATIONS AND CANONICAL FORMS OF BIVARIATE DISTRIBUTIONS

By H. O. Lancaster

University of Sydney

1. Introduction. This note gives an improved method of deriving the expansion for the bivariate measure in terms of the two marginal measures and the canonical correlations and functions. Although such an expansion will not be available for bivariate measures, which are not ϕ^2-bounded with respect to the product measure, the notion of canonical variable may still be applicable and an example is given. A theorem is proved that a bivariate measure is completely specified by its marginal measures and the correlations between the members of complete sets of orthonormal functions defined on the margins. This generalizes a theorem that independence between two variables is determined by the vanishing of every correlation between the members of complete orthonormal sets on the marginal distributions.

2. Canonical forms of bivariate distributions. The canonical variables, $\xi^{(i)}$ and $\eta^{(i)}$, and correlations, ρ_i, have been defined in Lancaster [1]. The same notation is followed here except that we write $\{\xi^{*(i)}\}$ and $\{\eta^{*(i)}\}$ for the sets complementary to the canonical sets such that the direct sums, $\{\xi^{(i)}\} \dotplus \{\xi^{*(i)}\}$ and $\{\eta^{(i)}\} \dotplus \{\eta^{*(i)}\}$, are complete on the respective marginal distributions. Throughout we shall use $\{x^{(i)}\}$ and $\{y^{(i)}\}$ as complete sets on the marginal distributions. $F(x, y)$ is taken to be a general bivariate distribution function. In this section, we assume that F is ϕ^2-bounded with respect to the product of the marginal distributions, $G(x)$ and $H(y)$. In other words, we assume that

$$\begin{aligned} \phi^2 + 1 &= \int \{dF/(dG\,dH)\}^2 dG\,dH \\ &\equiv \int \Omega^2\, dG\,dH \equiv \int (dF)^2/(dG\,dH), \end{aligned} \tag{1}$$

is finite. $\Omega \equiv \Omega(x, y)$ is the Radon-Nikodym derivative of $F(x, y)$ with respect to $G(x)H(y)$. We recall Theorem 2 of the previous paper and extend it slightly as the following:

Theorem A. *The canonical variables obey a second set of orthogonal conditions,*

$$\int \xi^{(i)} \eta^{(j)}\, dF = 0 \quad \text{if} \quad i \neq j, \tag{2}$$

and $\xi^{(i)}$ is also orthogonal to every square-summable function of Y, orthogonal to the canonical variables; similarly $\eta^{(i)}$ is orthogonal to every square-summable function of X, which is orthogonal to $\xi^{(i)}$.

Proof. The proof given in the previous paper is applicable.

Received May 29, 1962.

THEOREM B. *If $F(x, y)$ is a ϕ^2-bounded bivariate distribution with marginal distribution, $G(x)$ and $H(y)$, then complete sets of orthonormal functions can be defined on the marginal distributions such that each member of a set of canonical variables appears as a member of the complete set of orthonormal functions. The element of frequency can be expressed in terms of the marginal distributions,*

$$dF(x, y) = \left\{1 + \sum_1^\infty \rho_i x^{(i)} y^{(i)}\right\} dG(x)\, dH(y), \quad \text{a.e.,} \tag{3}$$

and

$$\phi^2 = \sum_{i=1}^{\infty} \rho_i^2 . \tag{4}$$

PROOF. We give the proof for the case when $G(x)$ and $H(y)$ possess infinitely many points of increase. The proof is easily adapted to the case when one of the marginal distributions has only finitely many jumps. For the case where both have only finitely many jumps, the proof has already been given in that paper, where reference has been made to other proofs.

The general theory of orthonormal functions can be invoked to show that the set of the products, $x^{(i)}y^{(j)}$, namely $\{x^{(i)}\} \times \{y^{(i)}\}$, is complete on the product distribution. Ω is square summable by definition so that we can minimize

$$\int (\Omega - S_{mn})^2\, dG\, dH = q, \tag{5}$$

where

$$\begin{aligned} S_{mn} &= 1 + \sum_{i,j}^{m,n} \lambda_{ij}\, x^{(i)}\, y^{(j)} \\ &\equiv 1 + \sum_{i=1}^{m_1} \sum_{j=1}^{m_1} a_{ij}\, \xi^{(i)}\, \eta^{(j)} + \sum_{i=1}^{m_1} \sum_{j=1}^{n_2} b_{ij}\, \xi^{(i)}\, \eta^{*(j)} \\ &\quad + \sum_{i=1}^{m_2} \sum_{j=1}^{m_1} c_{ij}\, \xi^{*(i)}\, \eta^{(j)} + \sum_{i=1}^{m_2} \sum_{j=1}^{n_2} d_{ij}\, \xi^{*(i)}\, \eta^{*(j)}, m = m_1 + m_2, n = m_1 + n_2, \end{aligned}$$

by choice of the coefficients. After Theorem A, the solution is

$$S_{mn} = 1 + \sum_i \rho_i \xi^{(i)} \eta^{(i)} + \sum_i \sum_j \gamma_{ij} \xi^{*(i)} \eta^{*(j)} \tag{6}$$

where

$$\gamma_{ij} = \int \xi^{*(i)}\, \eta^{*(j)}\, dF. \tag{7}$$

We now prove that every γ_{ij} is zero. By Bessel's inequality,

$$\sum_1^{m_1} \rho_i^2 + \sum_1^{m_2} \sum_1^{n_2} \gamma_{ij}^2 \leqq \phi^2. \tag{8}$$

As $i \rightarrow \infty$, ρ_i must either be zero or approach zero as a limit. Suppose now that γ_{uv} is not zero for some pair, $\xi^{*(u)}$ and $\eta^{*(v)}$. As $m_1 \rightarrow \infty$, there is a ρ_i, $i < m_1$, such that $\rho_i < |\gamma_{uv}|$. This is a contradiction for the pair, $\xi^{*(u)}$ and $\eta^{*(v)}$ obey all the necessary conditions to be a canonical pair. γ_{uv} is thus zero for all u and v. The minimization process, therefore, yields (3).

3. Bivariate distributions characterized by correlation.

THEOREM 1. *A bivariate distribution is completely characterized* a.e. *by its marginal distributions and the matrix of correlations of any pair of complete sets of orthonormal functions on the marginal distributions.*

PROOF. Suppose that $F_1(X, Y)$ and $F_2(X, Y)$ are two distinct bivariate distributions and let $\{x^{(i)}\}$ and $\{y^{(i)}\}$ be complete sets on the common marginal distributions, $G(x)$ and $H(y)$. Suppose that the correlations are

$$\rho_{ij}^{(k)} = \int x^{(i)} y^{(j)} \, dF_k(x, y), \qquad i = 1, 2 \cdots ; j = 1, 2 \cdots ; \; k = 1, 2. \tag{9}$$

If F_1 and F_2 are not identical then there is a pair of sets, G^* and H^* say, such that the measures, assigned to the intersection of the sets, $P(X \varepsilon G^*, Y \varepsilon H^*)$ by F_1 and F_2 are not identical and where neither $P(X \varepsilon G^*)$ nor $P(Y \varepsilon H^*)$ is either 0 or 1. Let us form a bivariate distribution of an orthonormal step function taking two values, constant over G^* and its complement and an orthonormal step function taking two values constant over the set H^* and its complement respectively. These two step functions, ξ and η say, will have a correlation depending on $P(X \varepsilon G^*, Y \varepsilon H^*)$ and this correlation will be different for F_1 and F_2. The two values may be written ρ_1 and ρ_2. We have to show that this leads to a contradiction. Let us approximate to ξ and η by series.

$$S_m = \sum_1^m a_i x^{(i)}, \qquad S_n' = \sum_1^n b_j y^{(j)} \tag{10}$$

so that $\int (\xi - S_m)^2 \, dG$ and $\int (\eta - S_n')^2 \, dH$ are each less than ϵ^2. By the Schwarz inequality, for $i = 1, 2$,

$$\begin{aligned} \left| \int \xi(\eta - S_n') \, dF_i \right| &\leqq \left[\int \xi^2 \, dF_i \int (\eta - S_n')^2 \, dF_i \right]^{\frac{1}{2}} \\ &= \left[\int \xi^2 \, dG \int (\eta - S_n')^2 \, dH \right]^{\frac{1}{2}} < \epsilon. \end{aligned} \tag{11}$$

Similarly $|\int S_n'(\xi - S_m) \, dF_i| < \epsilon$, since $\int S_n'^2 \, dH \leqq 1$. We now have

$$\begin{aligned} \rho_1 - \rho_2 &= \int \xi\eta \, dF_1 - \int \xi\eta \, dF_2 \\ &= \int (\xi\eta - S_n' S_m) \, dF_1 - \int (\xi\eta - S_n' S_m) \, dF_2 \\ &= \int [\xi(\eta - S_n') + S_n' (\xi - S_m)] \, dF_1 \\ &\quad - \int [\xi(\eta - S_n') + S_n' (\xi - S_m) \, dF_2. \end{aligned} \tag{12}$$

Taking absolute values after breaking the two integrals up into their two parts,

$$|\rho_1 - \rho_2| < 4\epsilon. \tag{13}$$

Since ϵ is arbitrary a contradiction has been reached if $\rho_1 \neq \rho_2$. F_1 and F_2 are therefore functions identical a.e.

COROLLARY. *A necessary and sufficient condition for independence of the marginal variables of a bivariate statistical distribution is that $\rho_{ij} = 0$, for $i > 0$ and $j > 0$, where the ρ_{ij} are defined by* (9).

PROOF. Let $F_1(X, Y)$ be the bivariate distribution and let $G(X)$ and $H(Y)$ by the marginal distributions. Let the product distribution, F_2, be defined so that $F_2(X, Y) = G(X)H(Y)$. Then F_1 and F_2 have the same marginal distributions and further the same correlation matrix; for the matrix of correlations of F_1 consists of zeroes by hypothesis and the matrix of correlations of F_2 is easily verified to be the null matrix. F_1 and F_2 are therefore identical a.e. and so $F_1(X, Y) \equiv F_2(X, Y) \equiv G(X)H(Y)$, which shows that the two random variables are mutually independent, since the joint distribution function factorizes.

4. Canonical correlations in ϕ^2-unbounded distributions. The canonical form of bivariate distribution has always been obtained by assuming that the bivariate distribution is ϕ^2-bounded with respect to the product of the marginal distributions. However, in Section 3, we have not assumed ϕ^2-boundedness and so we might ask whether we can simplify the matrix, $\mathbf{R}$, in the ϕ^2-unbounded case. Let us consider arbitrary orthonormal linear forms, $\mathbf{a}^T\mathbf{x}$ and $\mathbf{b}^T\mathbf{y}$; the integrals of the squares of these forms are unities and so the integral of $\mathbf{a}^T\mathbf{x}\mathbf{y}^T\mathbf{b}$ exists and is bounded by unity in absolute value. Let $\mathbf{a}$ and $\mathbf{b}$ take all permissible values, then since $|E\mathbf{a}^T\mathbf{x}\mathbf{y}^T\mathbf{b}|$ is bounded it will have upper limiting value. Let the linear orthonormal forms, $\mathbf{a}^T\mathbf{x}$ and $\mathbf{b}^T\mathbf{y}$, for this maximum be taken as the first pair of variables in new orthonormal sets. The correlation matrix in the new set will have no other non-zero term in the first row or column than the leading term. A second pair can now be found orthogonal to the first pair and the correlation matrix will have two non-zeros along the diagonal but every other correlation in the first row or column will be zero. $\mathbf{R}$ can be diagonalized by a repetition of this process. However, in a ϕ^2-unbounded distribution, there may remain pairs of functions, which have positive correlation and do not occur anywhere in the series.

EXAMPLE. Let $\psi_k^{(i)}(x)$ be defined so that $\psi_k^{(i)}(x)$ is zero outside the open interval $(k - 1, k)$ and in the interval is a polynomial of precise degree i and further so that

$$\int_{k-1}^{k} \psi_k^{(i)}(x)\psi_k^{(j)}(x)\, dx = \delta_{ij}. \tag{14}$$

Now define the kth bivariate distribution, $Q_k(x, y)$, with a probability measure of β_k uniformly distributed along the diagonal and $(1 - \beta_k)$ uniformly distributed over the square. It is easily verified that

$$\int \psi_k^{(i)}(x)\psi_k^{(j)}(y)\, dQ_k(x, y) = \beta_k\, \delta_{ij}. \tag{15}$$

Now two of these distributions can be mixed

$$Q(x, y) = a_1Q_1(x\ y) + a_2Q_2(x, y), \qquad a_1 + a_2 = 1, \quad 1 > a_1, a_2 > 0 \tag{16}$$

and orthonormal sets defined of the form

$$\begin{aligned} x^{(0)} &= 1, \quad X \,\varepsilon\, (0, 1) \text{ or } (1, 2) \\ x^{(1)} &= -a_2^{\frac{1}{2}}a_1^{-\frac{1}{2}} \text{ if } X \,\varepsilon\, (0, 1) \\ x^{(1)} &= a_1^{\frac{1}{2}}a_2^{-\frac{1}{2}} \text{ if } X \,\varepsilon\, (1, 2) \\ x^{(2i)} &= a_1^{-\frac{1}{2}}\psi_1^{(i)}(x), \quad i > 0 \\ x^{(2i+1)} &= a_2^{-\frac{1}{2}}\psi_2^{(i)}(x), \quad i > 0 \end{aligned} \tag{17}$$

$\{x^{(i)}\}$ is complete. $\{y^{(i)}\}$ can be formed in a similar manner. $x^{(1)}$ and $y^{(1)}$ are the first pair of canonical variables with a canonical correlation of unity. If $\beta_1 > \beta_2$, $x^{(2i)}$ is the $(i + 1)$th canonical variable and β_1 is the $(i + 1)$th canonical correlation for $i > 0$. The canonical set is not complete since members of the set $\{\psi_2^{(i)}(x)\}$ never appear in it.

This procedure could be varied by arranging the distributions to be mixed in a checkerboard pattern. Let us now give a bivariate distribution which is not ϕ^2-bounded but nevertheless has a canonical set of variables and distinct canonical correlations.

Let a probability measure be defined by

$$\mu(X = i, Y = j) = p_{ij}, \qquad \begin{array}{l} i = 1, 2, 3 \cdots \\ j = 1, 2, 3 \cdots \end{array} \tag{18}$$

so that $p_{ij} = p_{ji}$ and let

$$\sum_i p_{ij} = \sum_i p_{ji} = w_j \tag{19}$$

$$\sum_{j=1}^{k} w_j = W_k. \tag{20}$$

An infinite set of orthogonal functions can now be defined by

$$\begin{aligned} x^{(i)} &= w_{i+1}^{\frac{1}{2}}/(W_iW_{i+1})^{\frac{1}{2}}, && X = 1, 2 \cdots i \\ &= -W_i^{\frac{1}{2}}/(w_{i+1}W_{i+1})^{\frac{1}{2}}, && X = i + 1 \\ &= 0 \quad \text{for } X > i + 1. \end{aligned} \tag{21}$$

A similar set of orthonormal functions can be defined on the other margin.

Let now a set of positive numbers, the required canonical correlations, be given so that $1 > a_1 \geqq a_2 \geqq \cdots \geqq 0$. It is required to find a symmetric bivariate distribution with these canonical correlations, $\{a_i\}$, and with marginal frequencies, $\{w_i\}$, with the obvious limitations $w_i > 0$ and $\sum w_i = 1$ and a further requirement, $W_2 = w_1 + w_2 > a_1$. A probability with a single canonical correlation of a_1 is now obtained by setting

$$(22)\qquad \begin{aligned} p_{11}^{(1)} &= w_1 w_1 + \delta_1 \\ p_{12}^{(1)} &= p_{21}^{(1)} = w_1 w_2 - \delta_1 \\ p_{22}^{(1)} &= w_2 w_2 + \delta_1 \\ p_{ij}^{(1)} &= w_i w_j, \quad \text{for } i > 2 \text{ or } j > 2. \end{aligned}$$

This gives a probability distribution if $\delta_1 \leqq w_1 w_2$. In this distribution, the correlation of $x^{(1)}$ and $y^{(1)}$ is contributed to only by the four cell probabilities given in (22) and is $\delta_1 W_2/(w_1 w_2)$. It can take any value between 0 and $W_2 < 1$ and consequently the value, a_1. The probabilities, $\{p_{ij}^{(1)}\}$, can be modified without altering the first correlation if the amounts added in the four leading cells are proportional to w_1w_1, w_1w_2, w_2w_1 and w_2w_2 respectively. We define

$$(23)\qquad \begin{aligned} p_{ij}^{(2)} &= p_{ij}^{(1)} + \delta_2 w_i w_j / W_2^2, & i &= 1, 2; j = 1, 2 \\ p_{3j}^{(2)} &= p_{j3}^{(2)} = w_j w_3 - \delta_2 w_j / W_2, & j &= 1, 2 \\ p_{33}^{(2)} &= w_3^2 + \delta_2 & & \\ p_{ij}^{(2)} &= w_i w_j = p_{ij}^{(1)} \quad \text{for } i > 3 \text{ or } j > 3. & & \end{aligned}$$

Once again the correlation between $x^{(2)}$ and $y^{(2)}$ is $\delta_2 W_3/(w_3 W_2)$ and this can take any value between zero and W_3. But $a_2 \leqq a_1 < W_2 < W_3$, δ_2 can be chosen to give a second correlation of a_2. The correlation between $x^{(1)}$ and $y^{(1)}$ remains a_1 since the probabilities added in the cells of the leading 2×2 submatrix are proportional to the product of the marginal distribution. $\{p_{ij}^{(3)}\}$ can be defined in a similar manner and the process can be continued. p_{ij} is defined as the limit of $p_{ij}^{(n)}$ as $n \to \infty$. It is easily verified that $x^{(i)}$ is uncorrelated with $y^{(j)}$ for $i \neq j$. The correlations defined are thus canonical. No restrictions have been put on the a_i except $a_1 < 1$ and so Σa_i^2 may be infinite. A ϕ^2-unbounded distribution with an arbitrary set of canonical correlations has therefore been constructed. In this artificial probability distribution the expansion holds

$$(24)\qquad p_{ij} = w_i w_j (1 + a_1 x^{(1)} y^{(1)} + a_2 x^{(2)} y^{(2)} \cdots)$$

for the successive terms are the δ's and the series is convergent. Suppose $i \geqq j$ for definiteness. Then the terms in (24) with superscripts less than $(i - 1)$ are each zero. The sum of the terms with superscript greater than $(i - 1)$ is convergent, for

$$(25)\qquad \left|\sum_i^\infty a_k x^{(k)} y^{(k)}\right| \leqq \sum_i^\infty |x^{(k)} y^{(k)}| \leqq \sum_i^\infty w_{k+1} W_k^{-1} W_{k+1}^{-1} < W_i^{-2} \sum_i^\infty w_k < W_i^{-2}.$$

The term with superscript $(i - 1)$ is $a_{i-1}W_{i-1}w_i^{-1}W_i^{-1}$, when $i = j$ and this will be unbounded for an infinity of values of i. If $i > j$ the corresponding term is $-a_{i-1}/W_i$, which is always finite.

REFERENCES

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

[2] LANCASTER, H. O. (1960). On statistical independence and zero correlation in several dimensions. *J. Austral. Math. Soc.* **1** 492–496.