JOURNAL OF MULTIVARIATE ANALYSIS 6, 526–537 (1976)

The Canonical Decomposition of Bivariate Distributions

P. L. CHESSON*

Department of Statistics, University of Adelaide, South Australia

Communicated by E. J. Hannan

The ordinary notion of a bivariate distribution has a natural generalisation. For this generalisation it is shown that a bivariate distribution can be characterised by a Hilbert space $\mathscr{H}$ and a family $\mathscr{M}_\rho$, $0 \leqslant \rho \leqslant 1$, of subspaces of $\mathscr{H}$. $\mathscr{H}$ specifies the marginal distributions whilst $\mathscr{M}_\rho$ is a summary of the dependence structure. This characterisation extends existing ideas on canonical correlation.

1. INTRODUCTION

The general theory of canonical correlation in bivariate distributions can be said to have originated with the paper of Lancaster [5] although previous workers considered the special case of a finite probability space. An extensive bibliography can be found in Lancaster [6]. Lancaster was chiefly concerned with finding expansions of the Radon–Nikodym derivative of a bivariate distribution relative to the product of its margins. However in more general situations than Lancaster considered such expansions may not exist.

The purpose of this paper is to generalise the notion of a bivariate distribution and to obtain decompositions of such distributions which are general analogues of Lancaster's results.

It will be shown that any bivariate distribution can be characterised in a natural way by a Hilbert space $\mathscr{H}$ and a family $\{\mathscr{M}_\rho, \rho \in [0, 1]\}$ of subspaces of $\mathscr{H}$. $\mathscr{H}$ depends only on the two marginal distributions and specifies them completely whilst $\mathscr{M}_\rho$ summarises the dependence structure of the bivariate distribution.

The main tool in the analysis is the spectral theorem for self-adjoint operators on real Hilbert spaces. Most discussions of spectral theory deal exclusively with

Received April 13, 1976.

AMS 1970 subject classifications: Primary 60E05, 62H05.

Key words and phrases: Bivariate distribution, Canonical correlation, Spectral theorem.

* This work was supported by a C.S.I.R.O. Postgraduate Studentship.

complex Hilbert spaces. However all of the results needed here are easily proved from their complex counterparts.

Hannan [4] was the first to apply general Hilbert space operator theory to canonical correlation. The approach given here is closely related to Hannan's but the use of the polar decomposition of a bounded linear operator is deliberately avoided because of the asymmetry inherent in that approach; however the polar decomposition does provide a useful alternative method for deriving the results given in this paper.

By a subspace of a Hilbert space $\mathscr{H}$ we will always mean a subset which is a Hilbert space with the same inner product. If $\mathscr{M}$ and $\mathscr{N}$ are two subspaces of $\mathscr{H}$, $\mathscr{M} \oplus \mathscr{N}$ is the orthogonal sum of $\mathscr{M}$ and $\mathscr{N}$, and $\mathscr{M} \ominus \mathscr{N}$ is the orthogonal complement of $\mathscr{N}$ in $\mathscr{M}$.

2. The Notion of a Bivariate Distribution

Let $\mathscr{F}$ and $\mathscr{G}$ be σ-fields of subsets of some set Ω. Define $\mathscr{F} \vee \mathscr{G}$ as the σ-field generated by the set $\mathscr{S} = \{A \cap B \mid A \in \mathscr{F}, B \in \mathscr{G}\}$.

Definition. A probability measure P is a bivariate distribution for $\mathscr{F}$ and $\mathscr{G}$ if P is a probability measure on $\mathscr{F} \vee \mathscr{G}$.

To see the analogy with the usual concept let X and Y be random variables on some probability space $(\Omega, \mathscr{A}, Q)$. The bivariate distribution of X and Y is generally thought of as the measure P' on $\mathscr{B}^2$, the Borel sets in R^2, defined by $P'(B) = Q((X, Y) \in B)$, for $B \in \mathscr{B}^2$. In terms of the definition above P' is a bivariate distribution for the σ-fields $\{B \times R^1 \mid B \in \mathscr{B}\}$ and $\{R^1 \times B \mid B \in \mathscr{B}\}$ where $\mathscr{B}$ is the Borel sets in R^1. More naturally we could consider the corresponding measure in the original space, that is, P defined as the restriction of Q to the sets $\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in B\}$ for $B \in \mathscr{B}^2$. Clearly if $\mathscr{F}$ and $\mathscr{G}$ are the σ-subfields of $\mathscr{A}$ generated by X and Y respectively then P is a bivariate distribution for $\mathscr{F}$ and $\mathscr{G}$. We will sometimes refer to a bivariate distribution arising in this way as a bivariate distribution for X and Y.

The general definition allows one to consider the joint distributions of pairs of random variables taking values in arbitrary measurable spaces, and includes the case where these random variables are stochastic processes such as considered by Hannan [4].

Given a bivariate distribution P for the σ-fields $\mathscr{F}$ and $\mathscr{G}$ of subsets of Ω, let F and G be the restrictions of P to $\mathscr{F}$ and $\mathscr{G}$, respectively. The probability spaces $(\Omega, \mathscr{F}, F)$ and $(\Omega, \mathscr{G}, G)$ will be referred to as the margin spaces. Define $\mathscr{H}$ to be the set of ordered pairs, (f, g), of real valued functions on Ω such that f is $\mathscr{F}$-measurable, g is $\mathscr{G}$-measurable and $E[f^2 + g^2] < \infty$.

$\mathscr{H}$ is a real Hilbert space under the inner product

$$\langle (f_1, g_1), (f_2, g_2) \rangle = \tfrac{1}{2}E[f_1 f_2 + g_1 g_2]$$

since $\mathscr{H}$ is essentially the direct sum of the spaces $\mathscr{L}^2(\Omega, \mathscr{F}, F)$ and $\mathscr{L}^2(\Omega, \mathscr{G}, G)$. Write $\mathscr{H} = \mathscr{L}^2(\Omega, \mathscr{G}, G) \dotplus \mathscr{L}^2(\Omega, \mathscr{F}, F)$.

Defined in this way $\mathscr{H}$ determines the margin spaces but tells us nothing about the dependence between them. To investigate this dependence note that $\mathscr{S}$ is a semiring generating $\mathscr{F} \vee \mathscr{G}$. Thus P is determined on $\mathscr{F} \vee \mathscr{G}$ by $E[fg]$ for $(f, g) \in \mathscr{H}$.

Lancaster [5, 6] was concerned with the space $(R^2, \mathscr{B}^2, P')$ induced by two random variables X and Y as discussed above. Let $F'(B) = Q(X \in B)$, $G'(B) = Q(Y \in B)$ for $B \in \mathscr{B}$, then Lancaster's result can be stated as follows:

If P' has a square summable Radon–Nikodym derivative, p, with respect to $F' \times G'$ then p has the representation

$$p(x, y) = \sum_{n=0}^{l} \rho_n \xi_n(x)\, \eta_n(y) \tag{1}$$

where the series is mean square convergent with respect to $F' \times G'$, $\{\xi_n\}$ and $\{\eta_n\}$ are orthonormal sequences of real valued functions relative to F' and G' respectively, $\rho_n \geqslant \rho_{n+1} > 0$, $\sum_0^l \rho_n^2 < \infty$ and $l \in \{0, 1, \ldots, \infty\}$.

In expression (1) the (ξ_n, η_n) are unique up to a change of sign when the ρ_n are disitnct. If, however, $\rho_k = \rho_{k+1} = \cdots = \rho_m$ then uniqueness holds only up to an equivalence. Specifically $\{(\xi_n, \eta_n)\}_{n=k}^m$ can be replaced by any set $\{\sum_{r=k}^m h_{nr}(\xi_r, \eta_r)\}_{n=k}^m$ where (h_{nr}), $n, r = k, \ldots, m$ is an orthogonal matrix. Later we shall find a convenient way to summarise this uniqueness.

Lancaster referred to the condition that p is square summable relative to $F' \times G'$ as "φ^2-boundedness."

In the general setting p may not even exist but from (1) we can easily obtain an expression (2) which does not involve p and so is more suitable for generalisation:

$$E[f(X)\, g(Y)] = \sum_{n=0}^{l} \rho_n E[f(X)\, \xi_n(X)]\, E[g(Y)\, \eta_n(Y)] \tag{2}$$

for all f and g such that $E[f^2(X) + g^2(Y)] < \infty$.

Note that the requirement $\sum_{n=0}^{l} \rho_n^2 < \infty$ in (2) implies the existence and square summability of p and (1) follows from this. Thus (1) and (2) are equivalent.

To see how a family of subspaces summarises the dependence structure define $\mathscr{M}_\rho$ to be the subspace generated by the set $\{(\xi_n(X), \eta_n(Y)) \mid \rho_n \leqslant \rho\}$ where $\{(\xi_n(X), \eta_n(Y))\}_{n=0}^{l}$ is any particular sequence for which (2) (or (1)) holds. Expression (2) is true when and only when $\{(\xi_n(X), \eta_n(Y))\}_{n=0}^{l}$ is some orthonormal basis of $\mathscr{M}_1$ with $(\xi_n(X), \eta_n(Y)) \in \mathscr{M}_{\rho_n} \ominus \mathscr{M}_\rho$ for $\rho_n > \rho$. The sequence $\{\rho_n\}$ is determined by the jumps in $\mathscr{M}_\rho$.

3. Main Results

Let P be a bivariate distribution for the σ-fields $\mathscr{F}$ and $\mathscr{G}$ of subsets of some set Ω. Using the definitions of $\mathscr{H}$, F and G given in Section 2 we have the following theorem.

Theorem. *There exists a unique family of subspaces $\mathscr{M}_\rho$ of $\mathscr{H}$, $0 \leqslant \rho \leqslant 1$, such that*

(i) $\bigcap_{\rho>\rho'} \mathscr{M}_\rho = \mathscr{M}_{\rho'}$, $0 \leqslant \rho' < 1$, *and* $\mathscr{M}_0 = \{0\}$.

(ii) *If* $\{(\xi_t^\rho, \eta_t^\rho)\}_{t\in T_\rho}$ *is an orthonormal basis for* $\mathscr{M}_\rho$ *then* $\{\xi_t^\rho\}_{t\in T_\rho}$ *and* $\{\eta_t^\rho\}_{t\in T_\rho}$ *are orthonormal families of functions on the margin spaces* $(\Omega, \mathscr{F}, F)$ *and* $(\Omega, \mathscr{G}, G)$ *respectively.*

(iii) *For* $(f, g) \in \mathscr{H}$ *and* $\{(\xi_t^\rho, \eta_t^\rho)\}_{t\in T_\rho}$ *as in* (ii)

$$E[fg] = \int_{(0,1]} \rho \, dQ(\rho)$$

where

$$Q(\rho) = \sum_{t\in T_\rho} \left(\int f\xi_t^\rho \, dF\right)\left(\int g\eta_t^\rho \, dG\right)$$

$$= \sum_{t\in T_\rho} E[f\xi_t^\rho] \, E[g\eta_t^\rho].$$

Proof. Define the linear operator $\mathbf{A}$ on $\mathscr{H}$ by $\mathbf{A}(f, g) = (E[g \mid \mathscr{F}], E[f \mid \mathscr{G}])$. To check the boundedness of $\mathbf{A}$ note that

$$E\{E[g \mid \mathscr{F}]\}^2 + E\{E[f \mid \mathscr{G}]\}^2 \leqslant E[f^2 + g^2] = 2 \mid (f, g)\mid^2.$$

This means $\mid \mathbf{A}(f, g)\mid^2 \leqslant \mid(f, g)\mid^2$ so that $\mid \mathbf{A} \mid \leqslant 1$ and from the fact that $\mathbf{A}(1, 1) = (1, 1)$ we see that $\mid \mathbf{A} \mid = 1$. Finally the symmetry of the expression $\langle \mathbf{A}(f_1, g_1), (f_2, g_2)\rangle = \frac{1}{2}E[f_2 g_1 + f_1 g_2]$ indicates that $\mathbf{A}$ is self-adjoint.

Now $\langle \mathbf{A}(f, g), (f, g)\rangle = E[fg]$ and it is the spectral decomposition of $\mathbf{A}$ that will be used to give the desired decomposition of $\mathscr{H}$.

From the above properties of $\mathbf{A}$ there is a spectral measure $\mathbf{P}$ on $\mathscr{B}$ such that $\mathbf{A} = \int_{[-1,1]} \rho \, d\mathbf{P}(\rho)$ and so

$$E[fg] = \int_{[-1,1]} \rho \, d\langle \mathbf{P}(\rho)(f, g), (f, g)\rangle. \tag{3}$$

Consider the self-adjoint isometry $\mathbf{L}$ defined by $\mathbf{L}(f, g) = (f, -g)$. Now $\mathbf{LAL} = -\mathbf{A}$ so that $-\mathbf{A} = \int_{[-1,1]} \rho \, d\mathbf{LP}(\rho)\mathbf{L} = \int_{[-1,1]} \rho \, d\mathbf{P}'(\rho)$ where $\mathbf{P}'(B) = \mathbf{P}(-B)$ for $B \in \mathscr{B}$. Since both $\mathbf{LP}(\cdot)\mathbf{L}$ and $\mathbf{P}'(\cdot)$ are spectral measures it follows from the spectral theorem that they are equal. Hence (3) can be written as

$$E[fg] = \int_{(0,1]} \rho \, d\langle(\mathbf{P}(\rho) - \mathbf{LP}(\rho)\mathbf{L})(f, g), (f, g)\rangle$$

and the integral can be interpreted as a Lebesgue–Stieltjes integral with respect to the function of bounded variation

$$Q(\rho) = \langle(\mathbf{P}(0, \rho] - \mathbf{LP}(0, \rho]\mathbf{L})(f, g), (f, g)\rangle.$$

Define $\mathscr{M}_\rho = \mathbf{P}(0, \rho]\mathscr{H}$. $\mathbf{P}(0, \rho]$ can be written

$$\sum_{t \in T_\rho} \langle(f, g), (\xi_t^\rho, \eta_t^\rho)\rangle(\xi_t^\rho, \eta_t^\rho) = \sum_{t \in T_\rho} \tfrac{1}{2}E[f\xi_t^\rho + g\eta_t^\rho](\xi_t^\rho, \eta_t^\rho)$$

where $\{(\xi_t^\rho, \eta_t^\rho)\}_{t \in T_\rho}$ is a basis for $\mathscr{M}_\rho$.

$\mathbf{LP}(0, \rho]\,\mathbf{L}(f, g)$ is then $\sum_{t \in T_\rho} \tfrac{1}{2}E[f\xi_t^\rho - g\eta_t^\rho](\xi_t^\rho, -\eta_t^\rho)$ and

$$(\mathbf{P}(0, \rho] - \mathbf{LP}(0, \rho]\mathbf{L})(f, g) = \sum_{t \in T_\rho} (E[g\eta_t^\rho]\xi_t^\rho, E[f\xi_t^\rho]\eta_t^\rho)$$

from which we conclude $Q(\rho) = \sum_{t \in T_\rho} E[f\xi_t^\rho]\, E[g\eta_t^\rho]$. This proves (iii).

$\mathscr{M}_\rho$ clearly has property (i); to verify (ii) suppose that (f_1, g_1), (f_2, g_2) belong to $\mathscr{M}_1$ and are orthogonal, that is $E[f_1f_2 + g_1g_2] = 0$. $(f_1, -g_1)$ is an element of $\mathbf{L}\mathscr{M}_1 = \mathbf{P}[-1, 0)\mathscr{H}$ and so is orthogonal to both (f_1, g_1) and (f_2, g_2) giving the additional equations $E[f_1^2] - E[g_1^2] = 0$, $E[f_1f_2] - E[g_1g_2] = 0$. Hence $E[f_1^2] = E[g_1^2]$ and $E[f_1f_2] = E[g_1g_2] = 0$ from which (ii) follows.

To see that $\mathscr{M}_\rho$ is unique assume $\mathscr{M}_\rho'$ is another family with properties (i) and (iii) and let $\mathbf{P}_\rho'$ be the projection on $\mathscr{M}_\rho'$. Define $\mathbf{A}_1 = \int_{(0,1]} \rho \, d\mathbf{P}_\rho'$ and $\mathbf{A}_2 = \mathbf{A}_1 - \mathbf{LA}_1\mathbf{L}$. With this construction $\langle\mathbf{A}_2(f, g), (f, g)\rangle = E[fg] = \langle\mathbf{A}(f, g), (f, g)\rangle$ and since both operators are self-adjoint they must be equal. Hence $\mathscr{M}_\rho' = \mathscr{M}_\rho$.

The origin of $\mathscr{M}_\rho$ from the spectral decomposition of $\mathbf{A}$ immediately yields a number of interesting properties.

(iv) *If $(f, g) \in \mathscr{M}_\beta \ominus \mathscr{M}_\alpha$ and $E[f^2] = 1$ $(=E[g^2])$ then $\alpha < E[fg] \leqslant \beta$.*

(v) *$\mathscr{H}$ can be decomposed as $\mathscr{H} = \mathscr{M}_1 \oplus \mathscr{O} \oplus \mathscr{M}_{-1}$ where, for $(f, g) \neq 0$, $E[fg]$ is positive, zero, or negative according to whether (f, g) belongs to $\mathscr{M}_1$, $\mathscr{O}$ or $\mathscr{M}_{-1}$ respectively. Furthermore $\mathscr{O}$ is the unique subspace with the property $E[fg'] = E[f'g] = 0$ for all $(f, g) \in \mathscr{O}$ and $(f', g') \in \mathscr{H}$.*

(vi) *If* $(f, g) \in \mathscr{H}$ *is written as the sum* $\sum_{i=1}^{4} (f_i, g_i)$, $(f_1, g_1) \in \mathscr{M}_1 \ominus \mathscr{M}_\rho$, $(f_2, g_2) \in \mathscr{M}_\rho$, $(f_3, g_3) \in \mathscr{O}$, $(f_4, g_4) \in \mathscr{M}_{-1}$ *then* $E[fg] = \sum_{i=1}^{4} E[f_i g_i]$.

(vii) *The properties* $E[f_1 g_1] > \rho\,|(f_1 g_1)|^2$ *for* $0 \neq (f_1, g_1) \in \mathscr{M}_1 \ominus \mathscr{M}_\rho$, $E[f_2 g_2] \leqslant \rho\,|(f_2, g_2)|^2$ *for* $(f_2, g_2) \in (\mathscr{M}_1 \ominus \mathscr{M}_\rho)^\perp$, *and* $E[(f_1 + f_2)(g_1 + g_2)] = E[f_1 g_1 + f_2 g_2]$ *characterise* $\mathscr{M}_1 \ominus \mathscr{M}_\rho$.

Proof. (iv) If $(f, g) \in \mathscr{M}_\beta \ominus \mathscr{M}_\alpha$ with $|(f, g)|^2 = 1$, or equivalently $E[f^2] = E[g^2] = 1$, the monotonic function $\langle \mathbf{P}(\rho)(f, g), (f, g)\rangle$ has the value 1 for $\rho \geqslant \beta$ and 0 for $\rho \leqslant \alpha$. Hence

$$\alpha < \int \rho \, d\langle \mathbf{P}(\rho)(f, g), (f, g)\rangle \leqslant \beta$$

and so by (3) $\alpha < E[fg] \leqslant \beta$.

(v) Define $\mathscr{M}_{-1}$ as $\mathbf{L}\mathscr{M}_1$ and $\mathscr{O}$ as $\mathscr{H} \ominus (\mathscr{M}_1 \oplus \mathscr{M}_{-1})$. Clearly $E[fg] > 0$ for $0 \neq (f, g) \in \mathscr{M}_1$, and $E[fg] < 0$ for $0 \neq (f, g) \in \mathscr{M}_{-1}$. $\mathscr{O}$ is the null space of $\mathbf{A}$ so that if $(f, g) \in \mathscr{O}$, $E[fg] = \langle \mathbf{A}(f, g), (f, g)\rangle = 0$. Also $\langle \mathbf{A}(f, g), (f', g')\rangle = 0$ for any $(f', g') \in \mathscr{H}$ yielding $E[fg'] + E[f'g] = 0$. Since this is also true for $(f', -g')$ it follows that $E[fg'] = E[f'g] = 0$.

On the other hand if (f, g) has the property $E[fg'] = E[f'g] = 0$ for all $(f', g') \in \mathscr{H}$ then clearly $\mathbf{A}(f, g) = 0$ or $(f, g) \in \mathscr{O}$.

(vi) The four subspaces $\mathscr{M}_1 \ominus \mathscr{M}_\rho$, $\mathscr{M}_\rho$, $\mathscr{O}$ and $\mathscr{M}_{-1}$ are mutually orthogonal invariant subspaces of $\mathbf{A}$.

(vii) Let $\mathscr{L}$ be another subspace with these properties. The equation $E[(f_1 + f_2)(g_1 + g_2)] = E[f_1 g_1 + f_2 g_2]$ for $(f_1, g_1) \in \mathscr{L}$, $(f_2, g_2) \in \mathscr{L}^\perp$, entails that $E[f_2 g_1 + f_1 g_2] = 0$ or $\langle \mathbf{A}(f_1, g_1), (f_2, g_2)\rangle = 0$ and hence that $\mathbf{A}(f_1, g_1) \in \mathscr{L}$ for $(f_1, g_1) \in \mathscr{L}$. Thus $\mathscr{L}$ is an invariant subspace of $\mathbf{A}$ and so

$$\mathscr{L} = \mathscr{L} \cap (\mathscr{M}_1 \ominus \mathscr{M}_\rho) \oplus \mathscr{L} \cap (\mathscr{M}_1 \ominus \mathscr{M}_\rho)^\perp,$$

$$\mathscr{L}^\perp = \mathscr{L}^\perp \cap (\mathscr{M}_1 \ominus \mathscr{M}_\rho) \oplus \mathscr{L}^\perp \cap (\mathscr{M}_1 \ominus \mathscr{M}_\rho)^\perp.$$

Considering the remaining common properties of $\mathscr{L}$ and $\mathscr{M}_1 \ominus \mathscr{M}_\rho$ it is clear that $\mathscr{L} = \mathscr{M}_1 \ominus \mathscr{M}_\rho$.

4. The φ^2-bounded Case

In this section an analogous condition to Lancaster's φ^2-boundedness is investigated.

We need to relate the bivariate distribution P, for $\mathscr{F}$ and $\mathscr{G}$, to a measure which is in some sense the product of F and G. There are two approaches that could be taken to do this:

A. Attempt to construct a measure $F \cdot G$, say, on $\mathscr{F} \vee \mathscr{G}$ such that $\mathscr{F}$ and $\mathscr{G}$ are independent under $F \cdot G$, and the restrictions of $F \cdot G$ to $\mathscr{F}$ and $\mathscr{G}$ are, respectively, F and G.

B. Redefine P on $\mathscr{F} \times \mathscr{G}$, a σ-field for which we know a suitable product measure exists.

Since approach A is more natural and more informative it will be adopted here. On the other hand approach B is not devoid of interest and so a brief outline of that approach will be given at the end of the section.

We want to define $F \cdot G(A \cap B)$ as $F(A)\, G(B)$ for $A \in \mathscr{F}$, $B \in \mathscr{G}$. In order for $F \cdot G$ to be well defined we need to impose the condition:

I. $A \cap B = \varnothing$ implies $F(A)\, G(B) = 0$ for all $A \in \mathscr{F}$, $B \in \mathscr{G}$.

However it has only been possible to prove finite additivity of $F \cdot G$ without further restrictions. To prove countable additivity it seems necessary to require:

II. $A \cap B = \varnothing$ implies $A = \varnothing$ or $B = \varnothing$ for all $A \in \mathscr{F}$, $B \in \mathscr{G}$. Renyi [7, p. 115] gives the following lemma which we prove here by a simpler method.

LEMMA. *Let condition II hold and* $F \cdot G\ (A \cap B) = F(A)\, G(B)$, $A \in \mathscr{F}$. $B \in \mathscr{G}$. *$F \cdot G$ is countably additive on $\mathscr{S}$ and so has a unique extension to a probability measure on $\mathscr{F} \vee \mathscr{G}$.*

Proof. Let $I\colon \Omega \to \Omega^2$ be defined by $I(\omega) = (\omega, \omega)$. The inverse image, I^{-1}, maps the semiring $\mathscr{T} = \{A \times B \mid A \in \mathscr{F}, B \in \mathscr{G}\}$ onto $\mathscr{S}$. Condition II entails that $I^{-1}(C) = \varnothing$ implies $C = \varnothing$ for $C \in \mathscr{T}$. This fact will be used to infer the countable additivity of $F \cdot G$ on $\mathscr{S}$ from that of $F \times G$ on $\mathscr{T}$.

Suppose $A \cap B = \sum_{n=0}^{\infty} A_n \cap B_n$; A, $A_n \in \mathscr{F}$; B, $B_n \in \mathscr{G}$; and the union is disjoint. From condition II it follows that $\sum_{n=0}^{\infty} A_n \times B_n$ is a disjoint union. Now

$$\varnothing = A \cap B - \bigcup_{n=0}^{\infty} A_n \cap B_n$$

$$= I^{-1}\left(A \times B - \bigcup_{n=0}^{\infty} A_n \times B_n\right)$$

$$= I^{-1}\left(\bigcap_{m=0}^{\infty} A \times B - \bigcup_{n=0}^{m} A_n \times B_n\right).$$

Using the semiring properties of $\mathscr{T}$ this can be written

$$I^{-1}\left(\bigcap_{m=0}^{\infty}\bigcup_{n=0}^{r_m} A_{nm}\times B_{nm}\right),\qquad \text{for}\quad A_{nm}\times B_{nm}\in\mathscr{T},$$

$$=I^{-1}\left(\bigcup_{f}\bigcap_{m=0}^{\infty} A_{f(m)m}\times B_{f(m)m}\right)$$

$$=I^{-1}\left(\bigcup_{f}\left(\bigcap_{m=0}^{\infty} A_{f(m)m}\right)\times\left(\bigcap_{m=0}^{\infty} B_{f(m)m}\right)\right)$$

where the union is over all functions $f:\{0,1,...\}\to\{0,1,...\}$ satisfying $f(m)\leqslant r_m$. Hence

$$\varnothing=\bigcup_{f} I^{-1}\left(\bigcap_{m=0}^{\infty} A_{f(m)m}\times\bigcap_{m=0}^{\infty} B_{f(m)m}\right)$$

and so

$$\bigcap_{m=0}^{\infty} A_{f(m)m}\times\bigcap_{m=0}^{\infty} B_{f(m)m}=\varnothing$$

which means

$$A\times B-\bigcup_{n=0}^{\infty} A_n\times B_n=\varnothing.$$

It is easy to see that $\bigcup_{n=0}^{\infty} A_n\times B_n-A\times B=\varnothing$ giving $A\times B=\bigcup_{n=0}^{\infty} A_n\times B_n$. Finally

$$F\cdot G(A\cap B)=F\times G(A\times B)$$

$$=\sum_{n=0}^{\infty} F\times G(A_n\times B_n)=\sum_{n=0}^{\infty} F\cdot G(A_n\cap B_n)$$

Hence $F\cdot G$ is countably additive on $\mathscr{S}$.

The extension of $F\cdot G$ to $\mathscr{F}\vee\mathscr{G}$ will be referred to as $F\cdot G$.

We now impose a φ^2-boundedness condition.

III. Suppose that P is absolutely continuous with respect to $F\cdot G$ and the Radon–Nikodym derivative, p, is square summable relative to $F\cdot G$.

Let $\{\varphi_s\}_{s\in S}$ be an orthonormal basis for $\mathscr{L}^2(\Omega,\mathscr{F},F)$ and $\{\psi_t\}_{t\in T}$ an orthonormal basis for $\mathscr{L}^2(\Omega,\mathscr{G},G)$. $\{\varphi_s\psi_t\}_{(s,t)\in S\times T}$ is an orthonormal set (indeed a basis) for $\mathscr{L}^2(\Omega,\mathscr{F}\vee\mathscr{G},F\cdot G)$. Condition III entails that $p\in\mathscr{L}^2(\Omega,\mathscr{F}\vee\mathscr{G},F\cdot G)$ so that $c_{st}=\int p\varphi_s\psi_t\,dF\cdot G$ satisfies $\sum_{s,t}c_{st}^2<\infty$.

Now

$$\{x_u\}_{u\in U} = \{2^{1/2}(\varphi_s , 0)\}_{s\in S} \cup \{2^{1/2}(0, \psi_t)\}_{t\in T}$$

is an orthonormal basis for $\mathscr{H}$.

$$c_{st} = E[\varphi_s\psi_t] = 2\langle \mathbf{A}(\varphi_s , 0), (0, \psi_t)\rangle = 2\langle \mathbf{A}(0, \psi_t), (\varphi_s , 0)\rangle$$

and $\langle \mathbf{A}(\varphi_t , 0), (\varphi_{t'} , 0)\rangle = \langle \mathbf{A}(0, \psi_s), (0, \psi_{s'})\rangle = 0$ so that $\sum_{u,v} |\langle \mathbf{A}x_u , x_v\rangle|^2 = 8 \sum_{s,t} c_{st}^2 = <\infty$. Hence $\mathbf{A}$ is a Hilbert–Schmidt operator (see Dunford and Schwartz [1]). Thus the only element of the continuous spectrum of $\mathbf{A}$ is the point 0. The discrete spectrum consists of a set $\{\rho_n\}_{n=0}^{\infty} \cup \{-\rho_n\}_{n=0}^{\infty}$, $\rho_n \geqslant 0$ and $\rho_n \downarrow 0$. $\mathscr{M}_\rho$ is the Hilbert space generated by the eigenvectors of $\mathbf{A}$ corresponding to positive eigenvalues $\leqslant \rho$. Finally, if $l + 1 = \dim \mathscr{M}_1$, 3(iii) reduces to

$$E[fg] = \sum_{n=0}^{l} \rho_n E[f\xi_n]\, E[g\eta_n]$$

for all $(f, g) \in \mathscr{H}$ and $\{(\xi_n , \eta_n)\}_{n=0}^{l}$ is a basis for $\mathscr{M}_1$ such that $\rho_n > \rho$ implies

$$(\xi_n , \eta_n) \in \mathscr{M}_{\rho_n} \ominus \mathscr{M}_\rho .$$

To proceed by approach B note that $I^{-1}(\mathscr{F} \times \mathscr{G}) = \mathscr{F} \vee \mathscr{G}$ and so define $P'(C)$ as $P(I^{-1}(C))$ for $C \in \mathscr{F} \times \mathscr{G}$. Since $\mathscr{F} \times \mathscr{G} = \mathscr{F}' \vee \mathscr{G}'$, where $\mathscr{F}' = \{A \times \Omega \mid A \in \mathscr{F}\}$, $\mathscr{G}' = \{\Omega \times B \mid B \in \mathscr{G}\}$, P' is a bivariate distribution for $\mathscr{F}'$ and $\mathscr{G}'$ and the derivation can proceed as a special case of approach A by passing the lemma since $F \times G$ replaces $F \cdot G$. At the final stage it is necessary to relate the $\mathscr{F}'$ measurable functions on $\Omega \times \Omega$ to the $\mathscr{F}$ measurable functions on Ω, etc, to obtain a decomposition on the original space.

It is interesting to note that the absolute continuity of P' relative to $F \times G$ implies condition I. Further if P' is absolutely continuous relative to $F \times G$ for every bivariate distribution P, for $\mathscr{F}$ and $\mathscr{G}$, then condition II holds.

5. Examples of φ^2-unbounded Bivariate Distributions

(i) *Mixtures of bivariate normals.* Let P_ρ be the standarised bivariate normal measure on $(R^2, \mathscr{B}^2)$ with correlation $\rho \geqslant 0$, and Φ the standard normal measure on $(R, \mathscr{B})$.

Let X and Y be the functions from R^2 to R defined by $X(x, y) = x$ and $Y(x, y) = y$. It is well known that P_ρ has the decomposition

$$E[f(X)\, g(Y)] = \sum_{n=0}^{\infty} \rho^n E[H_n(X) f(X)]\, E[H_n(Y)\, g(Y)]$$

r $(f(X), g(Y)) \in \mathscr{L}^2(R, \mathscr{B}, \Phi) \dotplus \mathscr{L}^2(R, \mathscr{B}, \Phi) = \mathscr{H}$. H_n is the nth Hermite polynomial normalised to yield $\int H_n^2(x)\, d\Phi(x) = 1$.

Let Q be a probability measure on the Borel sets in $[0, 1]$ then $P = \int P_\rho\, dQ(\rho)$ has the expansion

$$E[f(X)g(Y)] = \sum_{n=0}^{\infty} \rho_n E[H_n(X)f(X)]\, E[H_n(Y)g(Y)] \tag{4}$$

where $\rho_n = \int \rho^n\, dQ(\rho)$. $\mathscr{H}$ is unchanged but $\mathscr{M}_\rho$ is the subspace spanned by $U_{\rho_n < \rho}\{(H_n(X), H_n(Y))\}$. Many probability measures Q satisfy $\sum_{n=0}^{\infty} \rho_n^2 = \infty$, that is, P is not φ^2-bounded; but, as is shown above, a useful decomposition of P is still available. For a discussion of the φ^2-bounded case see Sarmanov and Bratoeva [8].

Other examples of φ^2-unbounded bivariate distributions having a discrete canonical decomposition can be found in Eagleson [2] and Griffiths [3] although the decompositions are not pointed out explicitly. Unfortunately such distributions have frequently been excluded from discussion, because they are not φ^2-bounded, when an expression such as (4) was all that was required.

(ii) *A bivariate distribution with continuous canonical decomposition.* Hannan [4] gives a number of examples of bivariate distributions, for pairs of stochastic processes, which have continuous canonical decompositions. A simple example of a bivariate distribution on $\mathscr{B}[0, 1]^2$, the Borel sets in $[0, 1]^2$, having a continuous canonical decomposition is given here.

Let P be defined by

$$P(C) = \int_{\{x \mid (x,x) \in C\}}^{\rho} d\rho + \int_0^1 \frac{1}{\rho} \left(\int_{C \cap [0,\rho]^2} d(u, v) \right) d\rho$$

for $C \in \mathscr{B}[0, 1]^2$. P is a bivariate distribution for the random variables X and Y defined in example (i). The margin spaces are both equivalent to Lebesgue measure restricted to the Borel sets in $[0, 1]$.

For $(f(X), g(Y)) \in \mathscr{H}$

$$E[f(X)g(Y)] = \int_0^1 \rho f(\rho) g(\rho)\, d\rho + \int_0^1 \rho \left(\frac{1}{\rho}\int_0^\rho f(u)\, du\right)\left(\frac{1}{\rho}\int_0^\rho g(u)\, du\right) d\rho$$

$\mathscr{M}_\rho$ is simply the space

$$\mathscr{L}_\rho = \left\{(f(X), f(Y)) \in \mathscr{H} \;\middle|\; \int_0^1 f(x)\, dx = 0 \text{ and } f(x) = 0 \text{ for } x > \rho\right\},$$

if $\rho < 1$, and $\mathscr{M}_1 = \mathscr{L}_1 = \{(f(X), f(Y)) \in \mathscr{H}\}$. The proof of this illustrates the application of the characterisation 2(vii) of $\mathscr{M}_1 \ominus \mathscr{M}_\rho$.

Now $\mathscr{H} = (\mathscr{L}_1 \ominus \mathscr{L}_\rho) \oplus \mathscr{L}_\rho \oplus \mathscr{L}_1^\perp$ so that $(\mathscr{L}_1 \ominus \mathscr{L}_\rho)^\perp = \mathscr{L}_\rho \oplus \mathscr{L}_1^\perp$. $\mathscr{L}_1 \ominus \mathscr{L}_\rho$ is the space $\{(f(X), f(Y)) \in \mathscr{H} \mid f \text{ is constant on } [0, \rho]\}$. Let

$$(f_1(X), g_1(Y)) \in \mathscr{L}_1 \ominus \mathscr{L}_\rho\,,\ (f_2(X), g_2(Y)) \in \mathscr{L}_\rho\,,\ (f_3(X), g_3(Y)) \in \mathscr{L}_1^\perp.$$

It is easy to see that $E[f_1(X)g_2(Y)] = E[f_2(X)g_1(Y)] = 0$ and, noting that $g_3 = -f_3$ and $g_1 = f_1$, $E[f_1(X)g_3(Y) + f_3(X)g_1(Y)] = 0$. Hence

$$E\left[\left(\sum_1^3 f_i(X)\right)\left(\sum_1^3 g_i(Y)\right)\right] = \sum_1^3 E[f_i(X)\,g_i(Y)]. \tag{5}$$

$$\begin{aligned} E[f_1(X)\,g_1(Y)] &= \int_0^1 uf_1^2(u)\,du + \int_0^1 u\left(\frac{1}{u}\int_0^u f_1(v)\,dv\right)^2 du, \\ &> \int_0^\rho uc^2\,du + \rho\int_\rho^1 f_1^2(u)\,du + \int_0^\rho uc^2\,du, \end{aligned}$$

provided $f_1 \not\equiv 0$ and where $f_1^{-1}\{c\} = [0, \rho]$. Thus

$$E[f_1(X)\,g_1(Y)] > \rho\int_0^1 f_1^2(u)\,du = \rho\,|(f_1(X), g_1(Y))|^2$$

$$\begin{aligned} E[f_2(X)\,g_2(Y)] &= \int_0^1 uf_2^2(u)\,du + \int_0^1 u\left(\frac{1}{u}\int_0^u f_2(v)\,dv\right)^2 du \\ &\leqslant \int_0^1 ud\left\{\int_0^u f_2^2(v)\,dv\right\} + \int_0^1 \left\{\int_0^u f_2^2(v)\,dv\right\} du \\ &= \rho\int_0^1 f_2^2(u)\,du = \rho\,|(f_2(X), g_2(Y))|^2. \end{aligned}$$

Considering (5) and the fact that $E[f_3(X)g_3(Y)] \leqslant 0$ it follows that

$$E[f(X)g(Y)] \leqslant \rho\,|(f(X), g(Y))|^2$$

when $(f(X), g(Y)) \in (\mathscr{L}_1 \ominus \mathscr{L}_\rho)^\perp$. This completes the verification that $\mathscr{L}_1 \ominus \mathscr{L}_\rho$ has the properties 2(vii) of $\mathscr{M}_1 \ominus \mathscr{M}_\rho$ for all ρ in [0, 1]. Thus we see that $\mathscr{M}_\rho = \mathscr{L}_\rho$.

Acknowledgment

I would like to thank Dr. G. M. Tallis for introducing the theory of canonical correlation to me and for encouraging me to continue with this work.

REFERENCES

[1] DUNFORD, N. AND SCHWARTZ, J. T. (1963). *Linear operators Part II: Spectral Theory.* Interscience, New York.

[2] EAGLESON, G. K. (1964). Polynomial expansions of bivariate distributions. *Ann. Math. Statist.* **35** 1208–1215.

[3] GRIFFITHS, R. C. (1970). Positive definite sequences and canonical correlation coefficients. *Austral. J. Statist.* **12** 162–165.

[4] HANNAN, E. J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J. Austral. Math. Soc.* **2** 229–242.

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

[6] LANCASTER, H. O. (1969). *The Chi-squared Distribution.* Wiley, New York.

[7] RENYI, A. (1970). *Foundations of Probability.* Holden–Day, San Francisco.

[8] SARMANOV, I. O. AND BRATOEVA, Z. N. (1967). Probabilistic properties of bilinear expansions of Hermite polynomials. *Theor. Probab. Appl.* **12** 470–481.

[9] TYAN, S. AND THOMAS, J. B. (1975). Characterisation of a class of bivariate distribution functions. *J. Multivariate Anal.* **5** 227–235.

[10] HELSON, H. (1948). Remark on measures in almost independent fields. *Studia Math.* **10** 182–183.