The Annals of Statistics
1983, Vol. 11, No. 1. 341–345

STATISTICAL AND ALGEBRAIC INDEPENDENCE

By James D. Malley

National Institutes of Health

Using a simple application of Fubini's theorem, we examine the connection between statistical independence, linear independence of random vectors, and algebraic independence of univariate r.v.'s, where we call a finite set of r.v.'s *algebraically independent* if they satisfy a non-trivial polynomial relationship only with zero probability. As a consequence, we simplify the derivation of a result of Eaton and Perlman (1973) on the linear independence of random vectors, and settle a matrix equation question of Okamoto (1973) concerning the rank of sample covariance-type matrices $S = XAX'$, where X is $p \times n$, and A is $n \times n$, for the case $n \geq p \geq r = \text{rank}(A)$. We also derive a measure-theoretic version of the classical fact that the elementary symmetric polynomials in m indeterminates are algebraically independent. This has applications to sample moments, k-statistics, and U-statistics with polynomial kernels.

1. Introduction. In a multivariate normal setting one makes routine use of the positive definiteness of the sample covariance matrix. A direct proof of this involves study of determinantal equations of the form $|XX'| = 0$ for a matrix X of random column vectors. More generally, Eaton and Perlman (1973) and Okamoto (1973) have studied rank and eigenvalue questions for matrices XAX', where X is $p \times n$ and A is $n \times n$, with the column vectors of X not necessarily normal.

We unify this discussion by deriving some of the results from elementary use of Fubini's theorem, and properties of n-linear functions, and obtain new results, for example, on the algebraic independence of symmetric functions of statistically independent random variables (r.v.'s).

2. n-linear function of random vectors. Let R_m be real m-space. A function $f(X) = f(X_1, \cdots, X_n)$ from $R_p \times \cdots \times R_p$ (n times) to R_1 is said to be n-linear (over R_p) if it is linear in each component: for any $X_1, \cdots, X_n, X_j^* \in R_p$, any j, and $\alpha, \beta \in R_1 : f(X_1, \cdots, \alpha X_j + \beta X_j^*, \cdots, X_n) = \alpha f(X_1, \cdots, X_j, \cdots, X_n) + \beta f(X_1, \cdots, X_j^*, \cdots, X_n)$. (See, for example, Greub, 1980). The following result is standard.

LEMMA 1. *Let $f(X)$ be a non-trivial n-linear function over R_p. Then off a fixed plane in R_p, any specification of $f(X)$ at $X_n = x_n$, $f(X_1, \cdots, X_{n-1}, x_n)$, is a non-trivial $(n-1)$-linear function, and the plane is determined solely by the value of f on the set of products, $e_{i_1} \times \cdots \times e_{i_n}$, of all n-tuples of unit basis vectors in R_p.*

In R_p a *flat* is the set of $x \in R_p$ satisfying $a'x = b$ for some fixed $a \in R_p$, $a \neq 0$, $b \in R_1$. Call a p-variate r.v. *flat-free* if it assigns zero probability to every flat in R_p, and non-atomic (Taylor, 1973, page 237) if it assigns zero probability to any point in R_p. Note that $X \in R_p$ and flat-free implies that X is non-atomic, as is each component of X. Towards a converse of this, a result given in Farrell (1976, page 124) states that if the components are independent and non-atomic, then X assigns zero probability to the boundary of every closed convex set in R_p, hence such X is also flat-free. We can now state:

Received May 1981; revised July 1982.

AMS 1980 *subject classifications.* Primary 62D05; secondary 15A52.

Key words and phrases. Random matrices, statistical independence, algebraic independence, non-atomic measures, flats.

THEOREM 1. *Let X_j, $1 \le j \le n$, be independent, flat-free p-variate* r.v.'s, *and suppose $f(x)$ is a non-trivial n-linear function over R_p. Then $P\{f(X) = 0\} = 0$.*

PROOF. Use induction on n, Fubini's theorem and Lemma 1. For $n = 1$, the result is simply $P\{a'X_1 = 0\} = 0$ which holds for X_1 flat-free. □

NOTE. For the theorem we need not assume that the X_j are identically distributed.

3. Polynomials in random variables. We derive a univariate version of Theorem 1 using the following well-known polynomial fact:

LEMMA 2. *Let $g(t) = g(t_1, \cdots, t_m)$ be a non-trivial polynomial in m indeterminates. Then except for a finite set in R_1, it follows that $g(t_1, \cdots, t_{m-1}, c)$ is a non-trivial polynomial in $(m - 1)$ indeterminates, for $c \in R_1$.*

PROOF. Consider $g(t)$ as an element of the polynomial ring in $(m - 1)$ indeterminates $t_1, \cdots, t_{m-1}$, with coefficients located in the polynomial ring in one indeterminate t_m. Since $g(t) \neq 0$, there exists a term of g of the form $M = M(t_1, \cdots, t_m) = t_1^{a_1} \cdots t_{m-1}^{a_{m-1}} p(t_m)$ with $p(t_m)$ a non-trivial polynomial in t_m, and such that no other term in g has identical exponent type $(a_1, \cdots, a_{m-1})$, a_i positive integers. For $t_m = c$ not one of the finite number of roots of $p(t_m)$, we have $M(t_1, \cdots, t_{m-1}, c) \neq 0$, so that $g(t_1, \cdots, t_{m-1}, c) \neq 0$. □

Let us call a finite set of r.v.'s *algebraically independent* if they can satisfy a non-trivial polynomial relationship only on a set of measure zero. Thus a single, univariate r.v. if non-atomic is trivially algebraically independent. Hence using induction, Lemma 2, and Fubini's theorem again, we get:

THEOREM 2. *Statistically independent, non-atomic univariate* r.v.'s *are algebraically independent.*

Thus for a random sample of a non-atomic, univariate r.v., the sample moments, k-statistics, and U-statistics with polynomial kernels, are all non-atomic. The Theorem also allows us to derive a measure-theoretic version of the classical fact that the m elementary symmetric polynomials, and the m power-sum polynomials, in m indeterminates are algebraically independent.

Take $s_{i,m} = s_{i,m}(u) = s_{i,m}(u_1, \cdots, u_m)$ to be the ith elementary symmetric function in m indeterminates $u = (u_1, \cdots, u_m)$ and let $a_{i,m} = a_{i,m}(u) = a_{i,m}(u_1, \cdots, u_m) = u_1^i + \cdots + u_m^i$ be the ith power-sum polynomial, $1 \le i \le m$.

THEOREM 3. *The elementary symmetric polynomials, as well as the power-sum polynomials, in statistically independent, non-atomic, univariate* r.v.'s *are algebraically independent.*

PROOF. We verify the result for the elementary symmetric polynomials, the power-sum version being completely parallel.

Consider indeterminates $t = (t_1, \cdots, t_n)$ and a non-trivial polynomial $f(t) = f(t_1, \cdots, t_m)$. Re-write $f(s_{1,m}(u), \cdots, s_{m,m}(u))$ as a polynomial $g(u) = g(u_1, \cdots, u_m)$ in the indeterminates $u_1, \cdots, u_m$. If $g(u) \equiv 0$ then $f(t) \equiv 0$, since the $s_{i,m}$ are known to be algebraically independent for (non-random) indeterminates $u_1, \cdots, u_m$.

To conclude then, let $x_1, \cdots, x_m$ be statistically independent, non-atomic, univariate r.v.'s and apply Theorem 2 to the non-trivial $g(u)$ and the random polynomial $g(X_1, \cdots, X_m)$, to get

$$P\{f(s_{1,m}(X_1, \cdots, X_m), \cdots, s_{m,m}(X_1, \cdots, X_m)) = 0\} = P\{g(X_1, \cdots, X_m) = 0\} = 0. \quad \square$$

COROLLARY. *Given a random sample of size m of a non-atomic, univariate* r.v., *the first m raw sample moments, or all the* jth *central sample moments,* $1 < j \le m$, *as well as the first m k-statistics, are algebraically independent.*

PROOF. The first m raw sample moments, based on a sample of size m are, up to a constant $1/m$, exactly the power-sums $a_{i,m}$, $1 \le i \le m$, so the Theorem applies directly.

Next, let $v_1, \cdots, v_m$ be in turn the first m k-statistics, or central sample moments, where we redefine the first central sample moment to be $a_{1,m}$. One can then show that, for polynomial $f(t) = f(t_1, \cdots, t_m)$, $f(v_1, \cdots, v_m)$ can be re-written as $f(v_1, \cdots, v_m) = h(a_{i,m}, \cdots, a_{m,m})$ for polynomial $h(u) = h(u_1, \cdots, u_m)$, such that $f(t)$ non-trivial implies $h(u)$ non-trivial.

To see this, briefly, do an induction on j, the largest subscript appearing non-trivially in $f(t)$, and use the equations connecting the k-statistics, central sample moments and power-sum functions (see Kendall and Stuart, 1977, Section 12.6–Section 12.11). The case for $j = 1$ follows from $k_1 = (x_1 + \cdots + x_m)/m = (a_{1,m})/m$.

Applying Theorem 3 to $h(u)$ now completes the proof. □

For a sample of size $n < m$, the polynomials $u_1^i + \cdots + u_n^i$, $1 \le i \le m$, are no longer algebraically independent and the Corollary is no longer valid: putting $n = 2$, $m = 3$, $\xi_1 = u_1 + u_2$, $\xi_2 = u_1^2 + u_2^2$, $\xi_3 = u_1^3 + u_2^3$ leads to $2\xi_3 - 3\xi_1\xi_2 + \xi_1^3$ identically zero in u_1, u_2.

4. Further application of the theorems. For matrix A we will write $\mathrm{rank}(A) = r(A)$, and p.d. = positive definite, p.s.d. = positive semi-definite.

An immediate application of Theorem 1 is to $f(X) = |X|$ the determinant function of $X = (X_1 | \cdots | X_p)$, $X_j \in R_p$, since f is (skew-symmetric) p-linear. Consequently p independent, flat-free r.v.'s in R_p are linearly independent. Similarly for $n > p$, $X =$ the matrix having the X_j as columns, is of full rank.

A proof of this begins with recalling that $r(X) \le p \Leftrightarrow XX'$ not p.d., or, $|XX'| = 0$. Yet XX' is always p.s.d. Now partition X as $X = (X_1 | X_2)$, where X_1 is $p \times p$ and X_2 is $p \times (n - p)$. Then $a'XX'a = 0 \Leftrightarrow a'X_1X_1'a + a'X_2X_2'a = 0 \Leftrightarrow a'X_1X_1'a = a'X_2X_2'a = 0$ since X_1X_1' and X_2X_2' are p.s.d. By Theorem 1 however X_1X_1' is p.d. (a.e.). Alternatively, simply consider, say, the first p columns of X. By Theorem 1 this submatrix has rank $= p$ (a.e.), so $r(X) = p$(a.e.).

A generalization of the full rank property appears in Eaton and Perlman (1973, page 712). We restate this as:

RESULT EP *(Eaton and Perlman, 1973).* Let $X^* = \begin{pmatrix} X \\ \Gamma \end{pmatrix}$ where (1) $X = (X_1 | \cdots | X_n)$ with $X_j \in R_p$ being statistically independent and flat-free, (2) Γ is constant, $r \times n$, $r(\Gamma) = r$, and (3) $n \ge p + r$. Then X^* has full rank $= p + r$ (a.e.).

We provide a proof which begins as in Eaton and Perlman, uses the notation of Eaton and Perlman, but which concludes using our Theorem 1. Partition

$$X^* \quad \text{as} \quad X^* = \begin{pmatrix} \dot{X} & \ddot{X} \\ \dot{\Gamma} & \ddot{\Gamma} \end{pmatrix},$$

where $\dot{X}$ is $p \times (n - r)$, $\ddot{X}$ is $p \times r$, $\dot{\Gamma}$ is $r \times (n - r)$, and $\ddot{\Gamma}$ is $r \times r$, such that we can assume, with possibly some permutation of the columns of X^*, that $\ddot{\Gamma}$ non-singular. Then X^* has full row rank if and only if $W = \dot{X} - \ddot{X}\ddot{\Gamma}^{-1}\dot{\Gamma}$ does.

For given Γ constant, write $W = W(\dot{X}, \ddot{X})$, and let $I_{\dot{X},\ddot{X}}$ be the indicator function for the set $A_{\dot{X},\ddot{X}} = \{W = W(\dot{X}, \ddot{X})$ has reduced low rank$\}$. Then $P(X^*$ has reduced row rank$) = P(A_{\dot{X},\ddot{X}}) = E(I_{\dot{X},\ddot{X}}) = E\{E(I_{\dot{X},X} | \ddot{X} = \ddot{x})\} = E\{E(I_{\dot{X},\ddot{x}})\}$, since $\dot{X}$ and $\ddot{X}$ are statistically independent, which equals $E[P(A_{\dot{X},\ddot{x}})]$. As the columns of X are flat-free, $W(\dot{X}, \ddot{x})$ has its columns flat-free, since a flat-free random vector remains flat-free upon addition of a constant vector. Applying Theorem 1 as at the beginning of this section, $W(\dot{X}, \ddot{x})$ has full row rank (a.e.) so $P(A_{\dot{X},\ddot{x}}) = 0$. □

Eaton and Perlman use Result EP to show: For $n \geq p + r$, $X_j \in R_p$ statistically independent and flat-free, and a fixed $n \times n$ matrix A, positive semi-definite, $r(A) = n - r$, it follows that $S = XAX'$ is positive definite (a.e.); see also Dykstra (1970). For proof, we follow Eaton and Perlman (1973). Let A have rank $n - r$, $n \geq p + r$, and suppose Γ, $r \times n$, has rows which are a basis set of orthonormal eigenvectors for the zero eigenspace of A. Then XAX' has less than full rank $\Leftrightarrow XA^{1/2}$ less than full rank $\Leftrightarrow \exists\, a \neq 0$, $a'XA^{1/2} = 0 \Leftrightarrow$ for some vector b, $a'X = b'\Gamma$. Thus the matrix $\begin{pmatrix} X \\ \Gamma \end{pmatrix}$ does not have full rank, and this contradicts Result EP, so $|XAX'| > 0$ (a.e.).

We note that in Result EP some condition, such as A being p.s.d., is required, for $\exists\, B$, of full rank and $\{X_j\}$ absolutely continuous such that $S = XBX'$ fails to be even p.s.d. on a set of positive probability. In particular, let $B = I_n - hee'$, h a constant exceeding $1/n$, $e' = (1, \cdots, 1)$ $(1 \times n)$ and let X_j be iid multivariate normal. Then it can be shown that $P(dXBX'\,d' = 0 \text{ some } d \in R_p,\, d \neq 0) > 0$.

Continuing with the $X_j \in R_p$, flat-free and independent, Okamoto (1973) raises the following question: For $n \geq p + r$, $S = XAX'$, X being $p \times n$, $r(A) = r$, and A being $n \times n$, symmetric but not necessarily positive semi-definite, is it true that $r(S) = \min(p, r)$ (a.e.)? In Okamoto (1973) it is shown that the result holds for r.v.'s X_j having a joint distribution absolutely continuous with respect to pn-dimensional Lebesgue measure. A key lemma in Okamoto (1973) is: the solution set in R_m of a non-trivial polynomial equation in m indeterminates has m-dimensional Lebesgue measure zero. From this it follows that a random $p \times n$ matrix having a pn-dimensional density must assign zero probability to the zeros of a non-trivial polynomial in the components of the matrix.

We first partly settle the rank statement here by showing: For $n \geq p \geq r$, $\text{rank}(S) = r$ (a.e.), and then, further, we show the usefulness of Theorem 2 by deriving the key matrix result just mentioned, using a wonderful insight provided by Roger H. Farrell.

Thus, recall that any real symmetric matrix A can be factored as $A = LL'$ where L is $n \times r$, $r(A) = r(L) = r$, $L = (\ell_1 | \cdots | \ell_r)$, $\ell_i = \lambda_i U_i$, $\lambda_i \neq 0$, $1 \leq i \leq r$, each λ_i either pure real or pure imaginary, $U = (U_1 | \cdots | U_r)$ $n \times r$ and real, $U'U = I_r$ (see Rao, 1973, page 40).

Let $X = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, where Y_1 is $r \times n$ and Y_2 is $(p - r) \times n$, so

$$S = XAX' = XLL'X = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} LL'(Y_1' \,|\, Y_2') = \begin{pmatrix} Y_1LL'Y_1' & Y_1'LL'Y_2' \\ Y_2LL'Y_1' & Y_2'LL'Y_2' \end{pmatrix}.$$

Hence $r(S) = r(XLL'X')$ and if we can show $r(Y_1LL'Y_1') = r$ (a.e.) then since $r(X) = p$ (a.e.) for $n \geq p$ by Theorem 1, it would follow that $r \leq r(S) \leq \min\{r(X), r(A)\} = \min(p, r) = r$ (a.e.) as required.

Now $Y_1LL'Y_1'$ is $r \times r$ as is Y_1L. While it is not in general true that $r(AA') = r(A)$ for complex-valued A, it is the case that, for any commutative ring with identity, $|AB| = |A| \cdot |B|$ for square A and B. Hence $Y_1LL'Y_1'$ is non-singular if and only if Y_1L is non-singular. Suppose $\exists\, a$, a complex r-vector, such that $a'Y_1L = 0$. Then $a'Y_1\ell_i = 0$, $1 \leq i \leq r$, $\Rightarrow a'Y_1U_i' = 0$, $1 \leq i \leq r$, since $\ell_i = \lambda_i U_i$, so $\{\mathcal{Re}(a)\}'Y_1U_i' = 0$, and $\{\mathcal{Im}(a)\}'Y_1U_i' = 0$, all i, $1 \leq i \leq r$. Hence we can assume a is real, and show that $a'Y_1U = 0 \Rightarrow a = 0$ (a.e.), $U = (U_1 | \cdots | U_r)$.

Let Γ have rows which are a basis for the orthocomplement of the column space of U; Γ is $(n - r) \times n$. Then $a'Y_1U = 0 \Rightarrow \exists\, b \in R_{n-r}$, such that $a'Y_1 = b'\Gamma$, so $(a', -b')\begin{pmatrix} Y_1 \\ \Gamma \end{pmatrix} = 0$. Since $n - (n - r) = r$ and Y_1 is $r \times n$ we apply Result EP to get (a.e.) $a = 0$, $b = 0$, and Y_1L non-singular. □

Next, the key matrix result of Okamoto (1973) is derived from our Theorem 2. Let the univariate r.v.'s $x_1, \cdots, x_n$ have a joint density $f(t) = f(t_1, \cdots, t_n)$ with respect to n-dimensional Lebesgue measure. Let $y_1, \cdots, y_n$ be *any* independent univariate r.v.'s having a joint density $g(t) = \prod g_i(t_i)$ where g_i, the density of y_i, is strictly positive for all real t_i, and g_i need not equal g_j, and i, j. For any non-trivial polynomial $h(t) = h(t_1, \cdots, t_n)$ let $A \in R_n$ be its hypersurface, $A = \{a \in R_n \mid h(a_1, \cdots, a_n) = 0\}$.

Let P_y be the measure induced in R_n by $y = (y_1, \cdots, y_n)$. Then

$$P\{h(x_1, \cdots, x_n) = 0\} = \int_A f(t)\, dt = \int_A \frac{f(t)}{g(t)} g(t)\, dt = \int_A \frac{f(t)}{g(t)}\, dP_y.$$

But $\int_A dP_y = P\{h(y_1, \cdots, y_n) = 0\} = 0$ by Theorem 2, so using the absolute continuity of the integral $\int dP_y$ we get $\int_A (f(t)/g(t))\, dP_y = 0$ as required.

Staying with this circle of ideas, note that, for every realization of X, $X = x = (x_1 \mid \cdots \mid x_n)$, if $n > p$ then the columns of $X = x$ must be linearly dependent, so $xa = 0$ for some $a \in R_p$. Nonetheless:

THEOREM 4. *Let $X = (X_1 \mid \cdots \mid X_n)$ be a random matrix whose columns X_j, $1 \le j \le n$, are statistically independent and flat-free. Then the rows of X are also flat-free, and for any $a \in R_n$, $a \ne 0$, Xa is flat-free. Hence $P(Xa = 0) = 0$.*

PROOF. We first show that the rows are flat-free. Let $X_j' = (X_{1j}, \cdots, X_{pj})$, $1 \le j \le n$. Since each X_j is flat-free we know the components X_{ij}, $1 \le i \le p$, are non-atomic, and since the X_j are statistically independent, the components X_{ij}, $1 \le j \le n$, i fixed, are also statistically independent. By Theorem 2 then a linear relation $\sum_{j=1}^n a_j X_{ij} = 0$ can obtain only on a null set, so the rows are flat-free.

That Xa is flat-free follows from: X_1, X_2 independent, flat-free $\Rightarrow X_1 + X_2$ is flat-free, and this obtains from Fubini's Theorem.

Finally, $P(Xa = 0) = 0$, since Xa flat-free $\Rightarrow Xa$ non-atomic. □

Acknowledgment. The author wishes to express his gratitude to Roger H. Farrell who carefully read an earlier version of this paper, thereby providing numerous insights and correcting several errors, the joint effect substantially improving this work.

REFERENCES

DYKSTRA, R. L. (1970). Establishing the positive definiteness of the sample covariance matrix. *Ann. Math. Statist.* **41** 2153–2154.

GRUEB, W. (1980). *Multilinear Algebra*, 2nd ed., Springer-Verlag, New York.

EATON, M. L. and PERLMAN, M. D. (1973). The non-singularity of generalized sample covariance matrices. *Ann. Statist.* **1** 710–717.

FARRELL, R. H. (1976). *Techniques of Multivariate Calculation.* Springer-Verlag, New York.

KENDALL, M. and STUART, A.(1969, 1977). *The Advanced Theory of Statistics* (3rd edition, 4th edition) Vol. 1. Hafner, New York

LANG, S. (1965). *Algebra.* Addison-Wesley, Reading, Massachusetts.

OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** 763–765.

TAYLOR, S. J. (1973). *Introduction to Measure and Integration.* Cambridge University Press, Great Britain.

DIVISION OF COMPUTER RESEARCH AND TECHNOLOGY
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MARYLAND 20205