THE CHARACTERISATION OF THE NORMAL DISTRIBUTION

H. O. LANCASTER

(received 10 July 1958, revised 20 August 1959)

1. Introduction

There are a number of well known theorems on the mutual independence of forms, either linear or quadratic, in normal variables. Some of these theorems can only hold when the system of variables is normal or degenerate and so the possibility of certain forms being independent characterises the normal distribution. The theorems on characterisation have usually been proved by consideration of the necessary properties of the characteristic function. Here we shall be considering the characteristic function of the variables but we shall make more use of cumulant theory than previous authors. To do so we have first to prove that the existence of cumulants of all orders is implied by the independence conditions. The basic theorems we use are those from general cumulant theory and the special theorems of Cramér and of Marcinkiewicz. An advantage of the methods of this paper is that it is possible to show that some of the characterisation theorems require neither of these special theorems. For example, spherical symmetry is a very strong condition and so neither theorem is required whereas both theorems are needed for the most general theorems on the independence of two linear forms. Throughout we take the class of normal distributions to include the degenerate normal, that is, the distribution of a sure variable.

2. The Independence of Linear Forms

Lemma 1. Let $F(x)$ be a distribution function and suppose that there exists a positive number, R, such that

$$F(-x) + 1 - F(x) = O(\exp - rx) \text{ as } x \to +\infty \tag{1A}$$

for all positive $r < R$. Then $F(x)$ has finite moments of all orders.

Proof. A proof has been given by Cramér on page 71 of [5] of a related theorem from which this lemma is readily deduced. The essential point is

https://doi.org/10.1017/S1446788700026033 Published online by Cambridge University Press

that for positive integers k, $x^k \exp(-r|x|)$ tends to zero as $|x| \to \infty$. The lemma can be stated in the alternative form that for $N > 0$,

(1B) $$P\{|x| > t\} = O(\exp - rt^N)$$

as $t \to \infty$ implies the existence of all moments of the random variable x, since (1B) is equivalent to

(1C) $$P\{|y| > t\} = O(\exp - rt),$$

where $y = x^N$. Since y has finite moments of all orders by (1A) so has x.

LEMMA 2. Let n be a positive integer greater than unity

(2) $$\varepsilon < n^{-3}$$

implies.

(3) $$n > (1 - \varepsilon)^{-n+1}$$

and (3) holds good for $0 \leqq \varepsilon_j < \varepsilon$.

PROOF.

$$\begin{aligned} \log(1 - \varepsilon)^{-n+1} &= (n-1)(\varepsilon + \tfrac{1}{2}\varepsilon^2 + \tfrac{1}{3}\varepsilon^3 + \cdots) \\ &< (n-1)\varepsilon(1 + \varepsilon + \varepsilon^2 + \cdots) \\ &< 2(n-1)\varepsilon < 2n^{-2} \leqq \tfrac{1}{2}. \end{aligned}$$

So that $(1 - \varepsilon)^{-n+1}$ is always less than $e^{\frac{1}{2}}$ and so less than n. It is obvious that the expression on the right is not increased by replacing ε by $\varepsilon_j < \varepsilon$.

LEMMA 3. Let positive constants ε_j be defined by

(4) $$\varepsilon_s = n^3 \varepsilon_{s-1}^2, \quad \varepsilon_0 = \varepsilon < n^{-3},$$

then

(5) $$\varepsilon_k = n^{-3}(n^3 \varepsilon)^{2^k}$$

PROOF. Suppose it is true for $1, 2, \cdots (s-1)$, then

$$\varepsilon_s = n^3[n^{-3}(n^3\varepsilon)^{2^{s-1}}]^2 = n^{-3}(n^3\varepsilon)^{2^s}$$

and the result follows by mathematical induction.

LEMMA 4. If for $c > 1$ and $\alpha > 0$ and a random variable x,

(6) $$P\{|x| > c^s \alpha\} \leqq \varepsilon_s,$$

as defined in (4), then moments of x of all orders are finite.

PROOF. We define $r > 0$ and $N > 0$ by the equations,

$$n^3 \varepsilon = \exp(-r) \text{ and } N = \log 2/\log c \text{ and } t = c^s$$

https://doi.org/10.1017/S1446788700026033 Published online by Cambridge University Press

and substitute them in (5) and (6) to obtain

$$(7)\qquad \begin{aligned} P\{|x| > t\alpha\} &= P\{|x| > c^s\alpha\} \leqq \varepsilon_s \\ &= n^{-3}(n^3\varepsilon)^{2^s} \\ &= n^{-3}\exp(-2^s r) \\ &= n^{-3}\exp(-t^N r) \\ &= O(\exp(-rt^N)); \end{aligned}$$

hence

$$(8)\qquad P\{|x\alpha^{-1}| > t\} = O(\exp - rt^N)$$

and so finite moments of $x\alpha^{-1}$, and hence of x, of all orders exist by Lemma 1.

We now prove our most important lemma.

LEMMA 5. Let $x_1, x_2, \cdots, x_n$ be a set of n independently distributed random variables, which we write as elements of a vector, $\boldsymbol{x}$. Then either of the following conditions is sufficient for the existence of finite moments of all orders.

(i) there exists a set of m, $(2 \leqq m \leqq n)$, independently distributed random variables, y_i, such that

$$(9\text{A})\qquad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},$$

where $\boldsymbol{A}$ is an $m \times n$ matrix of rank m in which each column contains at least two non-zero elements;

(ii) the x_i are identically distributed and there exists a set of m independently distributed variables, y_i, such that

$$(9\text{B})\qquad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},$$

and $\boldsymbol{A}$ has m rows $(m > 1)$ and at least one column contains two non-zero elements.

PROOF. In either case there is no loss of generality by assuming that $|a_{ij}|$ is not greater than unity for every pair (i, j). Let the least of the non-zero absolute values a_{ij} be a. Choose an $\varepsilon < n^{-3}$ and so obeying (2) and take a positive α to satisfy

$$(10)\qquad P\{|x_j| > \alpha\} < \varepsilon \quad \text{for} \quad j = 1, 2, \cdots, n,$$

which entails

$$(11)\qquad P\{|a_{ij}x| > \alpha\} < \varepsilon \quad \text{for} \quad j = 1, 2, \cdots, n,$$

because of the condition above imposed on the $|a_{ij}|$. Under either hypothesis (i) or (ii) let us choose a pair of variables $y_{i'}$, $y_{i''}$ corresponding to a pair of non-zero coefficients $a_{i'j}$ and $a_{i''j}$. $|y_{i'}| > n\alpha$ is only possible if at least one of the $|x_j|$ is greater than α and since the variables x_j are mutually independent, the events, $|x_j| > \alpha$, are not mutually exclusive. We have, therefore,

$$P\{|y_{i'}| > n\alpha\} < n\varepsilon, \tag{12}$$

and a similar inequality holds for $y_{i''}$.

On the other hand, if $|ax_j| > (2n-1)\alpha$ and $|x_{j'}| \leqq \alpha$ for $j \neq j'$ both $|y_{i'}|$ and $|y_{i''}|$ will each be greater than $n\alpha$. Using the multiplication rule for independent probabilities, we have

$$\begin{aligned} P\{|y_{i'}| > n\alpha,\ |y_{i''}| > n\alpha\} &\geqq P\{|ax_j| > (2n-1)\alpha\} \prod_{j \neq j'} \{P|x_{j'}| < \alpha\} \\ &\geqq (1-\varepsilon)^{n-1} P\{|ax_j| > (2n-1)\alpha\}. \end{aligned} \tag{13}$$

But the hypothesis of independence of $y_{i'}$ and $y_{i''}$ gives

$$P\{|y_{i'}| > n\alpha, \quad |y_{i''}| > n\alpha\} = P\{|y_{i'}| > n\alpha\} P\{|y_{i''}| > n\alpha\}. \tag{14}$$

Combining (13), (14) and (12),

$$(1-\varepsilon)^{n-1} P\{|ax_j| > (2n-1)\alpha\} \leqq (n\varepsilon)^2 \tag{15}$$

and so making use of (3),

$$P\{|x_j| > c\alpha\} < n^3\varepsilon^2 = \varepsilon_1, \tag{16}$$

where $c = (2n-1)/a$ and so $c > 1$.

Hypothesis (i) enabled (16) to be applied to every x_j and so we obtain by repeated application with $c\alpha$, $c^2\alpha$, $c^3\alpha, \cdots$, in place of α the formula (6). For example by writing $c\alpha$ in (16) in place of α, we obtain

$$P\{|ax_j| > c^2\alpha\} < n^3\varepsilon_1^2 = \varepsilon_2. \tag{17}$$

Hypothesis (ii) gives us that every x has the same distribution function and (6) is again true. But (6) assures us of the existence of finite moments of all orders according to Lemma 4.

It is worthy of note that we can obtain a similar result by an alternative method if we are prepared to confine our attention to the class of distributions such that the expectation of $|x_j|^k$ is finite for each j and for one given positive k; k might for example be very small, 10^{-6} say. The independence conditions of two linear forms enable us then to prove that the expectation of $|x_j|^{2k}$ is finite. Repeated application assures the existence of moments of every order. Zinger [18] has used a similar method in his study of the independence of "quasi-polynomial statistics".

3. The Independence of Two Linear Forms

THEOREM 1. Darmois [6], Skitovitch [15], Lukacs and King [13]. *Let $x_1, x_2, \cdots, x_n$ be a set of mutually independent random variables and*

$a_j b_j \neq 0$ for each j. Then if the two linear forms,

$$(18) \qquad l_1 = a_1 x_1 + a_2 x_2 \cdots a_n x_n, \quad \text{and} \quad l_2 = b_1 x_1 + b_2 x_2 \cdots b_n x_n$$

are independent, the x_j are all normally distributed.

PROOF. Condition (i) of Lemma 1 is satisfied and so all cumulants exist. Now it may happen that some ratios b_j/a_j are identical; if so, we form new variables with u_1 the sum of the $a_j x_j$ taken over all values of j, such that $b_j/a_j = r_1$; u_2 is a similar sum corresponding to the ratios, $b_j/a_j = r_2$, and so on.

$$(19) \qquad \left.\begin{aligned} l_1 &= u_1 + u_2 \cdots u_{n'}, \\ l_2 &= r_1 u_1 + r_2 u_2 \cdots r_{n'} u_{n'}, \end{aligned}\right\}$$

where $\{u_j\}$ is a set of mutually independent variables, which have finite cumulants of all orders, and the r_i are pairwise different. We define $K_q^{(j)}$ as the cumulant of order q of the random variable u_j and K_q as an operator signifying the q^{th} cumulant of the specified random variable. Now, the arbitrary real linear form, $\lambda l_1 + \mu l_2$, in the independent variables can be written as the sum of independent random variables in two different ways, namely:

$$(20) \qquad \lambda l_1 + \mu l_2 = \sum_{j=1}^{n'} (\lambda + \mu r_j) u_j.$$

Taking cumulants, and noting the mutual independence of l_1 and l_2,

$$(21) \qquad \begin{aligned} K_q(\lambda l_1 + \mu l_2) &= \lambda^q K_q l_1 + \mu^q K_q l_2 \\ &= \lambda^q \sum_{j=1}^{n'} K_q^{(j)} + \mu^q \sum_{j=1}^{n'} r_j{}^q K_q^{(j)}. \end{aligned}$$

But using (20), and noting the mutual independence of the $\{u_j\}$,

$$(22) \qquad \begin{aligned} K_q(\lambda l_1 + \mu l_2) &= K_q \sum_{j=1}^{n'} (\lambda + \mu r_j) u_j \\ &= \sum_{j=1}^{n'} (\lambda + \mu r_j)^q K_q^{(j)}. \end{aligned}$$

The λ and μ are arbitrary real indeterminates, so that we may take $q > n'$ and equate coefficients in $\lambda^{q-1}\mu,\ \lambda^{q-2}\mu^2, \cdots \lambda^{q-n'}\mu^{n'}$ to obtain

$$(23) \qquad \begin{bmatrix} r_1 r_2 & \cdots & r_{n'} \\ r_1^2 r_2^2 & \cdots & r_{n'}^2 \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot \\ r_1^{n'} r_2^{n'} & \cdots & r_{n'}^{n'} \end{bmatrix} \begin{bmatrix} K_q^{(1)} \\ K_q^{(2)} \\ \cdot\ \cdot \\ K_q^{(n')} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ 0 \end{bmatrix}.$$

The determinant of this matrix is not zero, so that for each variable, u_j, the cumulants of order greater than n' are all zero. A theorem of Marcinkie-

https://doi.org/10.1017/S1446788700026033 Published online by Cambridge University Press

wicz [14], forces all the variables to be normal. But the u_j are linear combinations of the x_j and the theorem of Cramér [4] on the decomposition of a normal variable shows that each x_j is normal.

In order to discuss the degenerate cases rigorously, it is desirable now to note a characterisation theorem for the degenerate normal distribution.

THEOREM 2. *If the variable $l_1 = ax$ is independent of the variable $l_2 = bx$, where a and b are real non-zero constants, then x is a sure variable, (in other words, x is a degenerate normal variable).*

THEOREM 3. *If the sum of two independently distributed random variables is a sure variable, then each of these variables is also a sure variable.*

PROOF. This can be taken as a degenerate case of the Cramér theorem or it can be proved by elementary means. We can now consider the degenerate cases and note some corollaries and a converse of Theorem 1.

COROLLARY 1. If $n = 1$, x_1 is a sure variable. (This is Theorem 2.).

COROLLARY 2. If all the ratios b_j/a_j are identical, then the variables x_j are all sure variables.

PROOF. This follows from Corollary 1.

COROLLARY 3. If the variance of x_j is σ_j^2, then

$$\sum_{j=1}^{m} a_j b_j \sigma_j^2 = 0. \tag{24}$$

PROOF. l_1 is uncorrelated with l_2.

COROLLARY 4. If all the ratios have the same sign then the distributions are all degenerate (each x is a sure variable).

PROOF. (24) cannot hold if $\sigma_j^2 > 0$ for any σ_j.

COROLLARY 5. (Bernstein [1]). If $x_1 + x_2$ is independent of $x_1 - x_2$, then x_1 and x_2 are normal.

PROOF. This follows by specialising the coefficients a_j and b_j and putting $n = 2$ in Theorem 1. Bernstein [1] assumed the existence of the second moments.

Converse to Theorem 1.

If $\{x_j\}$ form a set of mutually independent normal variables with variances, σ_j^2, and if $\sum_{i=1}^{n} a_i b_i \sigma_i^2 = 0$, then $l_1 = \sum a_i x_i$ is independent of $l_2 = \sum b_i x_i$.

PROOF. Skitovitch [15] uses characteristic functions but it may be proved directly or as a corollary to the proposition that in joint normal distributions zero correlation implies independence.

4. Independence with Respect to Two Sets of Axes

The next theorem asserts that if a set of mutually independent variables can be transformed linearly to another set of mutually independent variables then each variable is normally distributed. Some restrictions are placed on the matrix of the transformation to avoid trivial transformations such as a change of scale or a renumbering of the variables. The method of proof is to use Lemma 5 to assure the existence of moments of all orders, to change the scale of the variables so that the variances are either zero or unity. By Theorem 3 above, it is possible then to show that the matrix is reducible since sure variables must transform into sure variables. In fact, we might consider only non-degenerate distributions but this would make Theorem 4 exceptional among the theorems of this paper.

THEOREM 4. (Loève [12]). *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors related by*

$$y = Ax, \quad x = A^{-1}y \tag{25}$$

where $\boldsymbol{A}$ is real, non-singular and of size n, and has at least two non-zero elements in each column. Then if the elements of $\boldsymbol{x}$ form a set of mutually independent variables and if the elements of $\boldsymbol{y}$ form also a mutually independent set, every variable is normally distributed.

PROOF. The conditions satisfy hypothesis (i) of Lemma 5. All moments of the variables of the sets $\{x_j\}$ and $\{y_j\}$ exist. We may therefore centre the variables x_j to have zero means, the y_j also will have zero means.

We suppose that p of the x's and q of the y's are sure variables. We can assume without loss of generality that they are the last p and last q elements of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Now consider the first equality of (25). The last q y's are linear forms in the x's and these must be the last p x's by Theorem 3 above; since $\boldsymbol{A}$ is non-singular $p \geqq q$. Similarly from the second equality of (25) $q \geqq p$ and so $p = q$. Now consider a partition of $\boldsymbol{A}$,

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \text{ where } A_4 \text{ is } p \times p. \tag{26}$$

$\boldsymbol{A}_3$ is a matrix of zeroes. $\boldsymbol{A}_1$ therefore contains at least two non-zero elements in every column and is non-singular. Since $\boldsymbol{A}_2$ is the matrix of coefficients of the sure variables which have been centred, their values are immaterial. It is clear now that we can drop all consideration of the sure variables, which we do. We therefore can assume that p is zero. We have proved that all the remaining variables x_j and y_j have finite second moments which by a change of scale is unity. This is equivalent to pre- and post-multiplication of $\boldsymbol{A}$ by diagonal matrices. We can assume this has been done. But now since the y's have unit variance and are uncorrelated (since they are independent) $\boldsymbol{A}$ is orthogonal and every element is less in absolute value than unity. Let us now write $X_s^{(j)}$ and $Y_s^{(j)}$ for the s^{th} cumulant of the x_j and y_j. We define

$$A^{(s)} = (a_{ij}^{(s)}) = (a_{ij}^s) \tag{27}$$

and take $\boldsymbol{A}^{T(s)}$ for its transpose. Writing now $\boldsymbol{X}_S$ and $\boldsymbol{Y}_S$, vectors whose elements are $X_s^{(j)}$ and $Y_s^{(j)}$, we have, using the cumulant theory and taking $s > 2$,

$$\boldsymbol{Y_S} = \boldsymbol{A}^{(s)} \boldsymbol{X_S} = \boldsymbol{A}^{(s)} \boldsymbol{A}^{T(s)} \boldsymbol{Y_S} = \boldsymbol{B} \boldsymbol{Y_S}, \tag{28}$$

say. This is trivially satisfied if $\boldsymbol{Y}_s = \boldsymbol{0}$. Now,

$$\begin{aligned}\sum_j |b_{ij}| &\leqq \sum_{jk}\sum |a^s_{ik}|\ |a^s_{jk}| \\ &< \sum_{jk}\sum a^2_{ik} a^2_{jk} = \sum_k a^2_{ik} = 1\end{aligned}$$

since every $|a_{uv}| < 1$. If every $Y_s^{(j)}$ is not zero, suppose for definiteness that $Y_s^{(1)}$ is the greatest in absolute value of the s^{th} cumulants. Then

$$|Y_s^{(1)}| \leqq \sum_j |b_{ij}|\ |Y_s^{(j)}| \leqq \sum_j |b_{ij}|\ |Y_s^{(1)}| < Y_s^{(1)}. \tag{29}$$

Thus a contradiction results unless $|Y_s^{(1)}|$ is zero and hence all cumulants vanish for $s > 2$. The variables are thus all normal. Lancaster [11] gave this proof but assumed the existence of all cumulants. The proof requires neither the theorem of Cramér nor that of Marcinkiewicz.

Corollary 1. If $x_1 + x_2$ is independent of $x_1 - x_2$, where x_1 is independent of x_2, then both variables are normal. This was proved by Bernstein [1], under the restrictive condition that the variances were finite and can be deduced from Theorem 1. (see Corollary 5 to Theorems 1 to 3).

Corollary 2. Spherical symmetry of the joint distribution of independent random variables implies that the distributions are identical and normal.

Corollary 3. If x_1 and x_2 are independent and $x_1 \cos\alpha + x_2 \sin\alpha$ is independent of $-x_1 \sin\alpha + x_2 \cos\alpha$ for one value of α not a multiple of $\pi/2$, the distribution is normal. This corollary is sometimes proved using the stronger restriction "for every α."

5. Spherical Symmetry and Independence

We restate corollary 2 of Theorem 4 as

Theorem 5. *If the joint distribution of n independently distributed random variables is spherically symmetrical, then the distributions are all identical in form and normal.*

This can be deduced as a corollary of Theorem 1 or of Theorem 4. We required in either case only to find a rotation which satisfies the condition of the theorem. The deduction from Theorem 4 is preferable since it calls on fewer of the fundamental theorems. The history of this theorem began with an anonymous review by Herschel in the *Edinburgh Review* for 1850.

6. Identically Distributed Variables

The proofs of Theorems 4 and 5 above require neither the theorem of Marcinkiewicz nor that of Cramér but both are required in Theorem 1. If all the variables have the same distribution the conditions on the coefficients can be considerably weakened and the Marcinkiewicz theorem will not be required.

THEOREM 6. *Let $l_1 = \sum_{j=1}^{n} a_j x_j$ and $l_2 = \sum_{j=1}^{n} b_j x_j$ be two independently distributed linear forms in independently and identically distributed random variables, and let the product $a_j b_j$ be non-zero for at least one j, then the distribution is normal.*

PROOF. Lemma 5 enables us to assume the existence of all moments and cumulants. We proceed as before and express the s-th cumulant of an arbitrary linear combination, $\lambda l_1 + \mu l_2$, in two different ways as in (20), then taking the s-th cumulants as before

$$\lambda^s \sum_{j=1}^{n} a_j^s K_s + \mu^s \sum_{j=1}^{n} b_j^s K_s = \sum_{j=1}^{n} (\lambda a_j + \mu b_j)^s K_s \tag{30}$$

where K_s is the s^{th} cumulant and has the same value for each variable. Putting $s = 2$, if $\sum_j a_j b_j \neq 0$, then K_2 is zero and the distribution is the degenerate normal. If K_2 is not zero, $\sum_{j=1}^{n} a_j b_j = 0$. Next taking $s = 2t$, where t is an integer greater than unity, and identifying coefficients in $\lambda^{2t-2}\mu^2$, we find that K_s has a non-zero coefficient and the product is zero. So that every cumulant of even order vanishes and Cramér's theorem shows that the distribution is normal. Since if we took a random variable composed of the difference of two such variables, independently distributed, then it would be symmetrical and its moments of odd order would vanish. If it is known that $\sum_{j=1}^{n} a_j^{s-2u} b_j^{2u}$ does not vanish for at least one value of u for every value of s, then Cramér's theorem would be unnecessary as we could prove that each cumulant of odd order vanishes too. For example, if $x_1 + x_2 + x_3$ is independent of $x_1 + 2x_2 - 3x_3$, then Cramér's theorem is not needed.

7. The Independence of Quadratic Forms

In this section, we shall prove some characterisations of the normal distribution by the independence of two quadratic forms. Since $\boldsymbol{x}^T \boldsymbol{a}\boldsymbol{a}^T \boldsymbol{x}$ is a quadratic form, the theorems and the lemma will apply to the independence of a linear from a quadratic form. We specialize the problem to the consideration of identically distributed variables.

LEMMA 6. Let $x_1, x_2, \cdots, x_n$ be a set of identically distributed and mutually independent random variables. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two non-negative definite real symmetric matrices. Then, if $\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ is distributed independently of $x^T Bx$ and $a_{jj}b_{jj}$ is not zero for at least one j, the distribution of the x's have finite moments of all orders. (Note that without the requirement on the $a_{jj}b_{jj}$, the quadratic forms would be mutually independent regardless of the form of the distribution.)

Proof. Without loss of generality we may assume that the absolute values of the a_{ij} and b_{ij} are less than unity and that a_{11} and b_{11} are both positive, since this involves at most a division by a constant and a renumbering of the variables. As in the proof of Lemma 5, let $\varepsilon < n^{-3}$ be chosen such that $P\{|x| > \alpha\} < \varepsilon$. Under the hypotheses made, it follows from the general theory that any quadratic form can be expressed in the form

$$x^T Ax = a_{11}^{-1}(a_{11}x_1 + \sum_{j=2}^{n} a_{1j}x_j)^2 + w_2 W_2 + w_3 W_3 \cdots w_n W_n \tag{31}$$

where W_j is the square of a linear form in the last $n+1-j$ variables and each w_j is a non-negative constant. We therefore have the inequality

$$x^T Ax \geqq a_{11}^{-1}(a_{11}x + \sum_{j=2}^{n} a_{1j}x_j)^2. \tag{32}$$

Taking a to be $\min(|a_{11}|, |b_{11}|)$, if $|ax_1| > (2n-1)\alpha$ and $|x_2| < \alpha$, $|x_3| < \alpha$, $|x_n| < \alpha$,

$$x^T Ax > n^2\alpha^2 \quad \text{and} \quad x^T Bx > n^2\alpha^2 \tag{33}$$

so that

$$P\{x^T Ax > n^2\alpha^2,\ x^T Bx > n^2\alpha^2\} \geqq P\{|x_1| > (2n-1)\alpha/a\} \prod_{j=2}^{n} P\{|x_j| < \alpha\}. \tag{34}$$

But $P\{x^T Ax > n^2\alpha^2\}$ is less than $n\varepsilon$ since for the inequalities (33) to hold, at least one of the $|x_i|$ must be greater than α and these events are not mutually exclusive because of the independence of the variables x_j.

The rest of the proof follows the lines of the proof of Lemma 5 with hypothesis (ii) holding.

8. The Independence of a Linear from a Quadratic Form

Theorem 7. *Geary's Theorem. If $x_1, x_2 \cdots x_n$ form a set of identically distributed and mutually independent random variables and if the mean is distributed independently of the sample variance, then the distribution is normal.*

Proof. We use Lemma 6 with $x^T aa^T x$ in place of the quadratic form $x^T Ax$ and the elements of a, each $1/\sqrt{n}$. We have $\Sigma(x-\bar{x})^2 = x^T(1-aa^T)x$, so that the sample variance can be represented by $x^T Bx$ where $B = 1 - aa^T$. The conditions of Lemma 6 are fulfilled. All moments exist. We may now complete the proof using Geary's method of equating the bivariate cumulants of $\bar{x}$ and s^2 to the cumulants of the parent distribution. Alternatively we may consider the expectations of $(\boldsymbol{a}^T\boldsymbol{x})^k(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})$, using the independence condition,

$$E(\boldsymbol{a}^T\boldsymbol{x})^k\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} = E(\boldsymbol{a}^T\boldsymbol{x})^k E\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}. \tag{35}$$

Now this is an equation in the moment of order $(k+2)$ in the lower moments. The coefficients are uniquely determined and by inspection we find that the coefficient of the x^{k+2} is not zero. If $E(x) = 0$, $Ex^2 = \sigma^2$, then the equation for $E(x^3)$ or K_3 of the parent distribution is the same for all distributions giving independently distributed mean and variance and similarly with the higher moments. But the normal distribution is one such solution and any other distribution will have the same moments, that is, the normal is the only solution with non-zero variance. We develop this proof in greater detail in Section 9. This proof also clarifies the remark of Geary [7] that in his proof, "it has only been necessary to utilize the condition $\kappa_{i1} = 0$, without taking account of the series $\kappa_{i2} = 0$, $\kappa_{i3} = 0$, etc., in order to establish normality".

In the proof above we have put $E(x) = 0$ for convenience, but it is an inessential step. In the following modification of the theorem we suppose that $E(x) = 0$.

THEOREM 8. *Under the conditions of Theorem* 7 *and* $E(x) = 0$, *let* $\boldsymbol{a}^T\boldsymbol{a} = 1$, *with each element* $a_i \geqq 0$, *then if* $\boldsymbol{a}^T\boldsymbol{x}$ *is independent of* $\boldsymbol{x}^T(\boldsymbol{1}-\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{x}$, the system is normal. (More than one of the elements of $\boldsymbol{a}$ is to be positive.) The proof goes through as in the main theorem with the aid of (35). Some condition such as the restriction of the elements of $\boldsymbol{a}$ to be non-negative is essential. It is indeed sufficient to ensure that $\sum_1^n a_i^k(1-a_i^2)$ is not zero for any integral k as in the following variation.

THEOREM 9. *Under the condition of Theorem* 7, *if* $\boldsymbol{a}$ *is such that* $\boldsymbol{a}^T\boldsymbol{a} = 1$ *and* $\sum_{i=i}^n a_i^k(1-a_i^2)$ *is not zero for any natural number k, and if* $\boldsymbol{a}^T\boldsymbol{a}$ *is distributed independently of* $\boldsymbol{x}^T(\boldsymbol{1}-\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{x}$, *then the distribution is normal* (Zinger, 16).

PROOF. The proof goes through as in Theorem 7.

9. The Independence of two Quadratic Forms

Consideration, similar to those, used in the characterisations of the previous section apply again. But now we find that the equations of the form

$$E(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})^{t'} = E(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t E(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})^{t'} \tag{36}$$

which is analogous to (35), is a homogeneous equation in the moments of order $2(t+t')$ and lower. Indeed there will be in all $(n-1)$ homogeneous equations of the $2n$-th order. It seems likely that these would imply enough conditions on the moments and the coefficients of the matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, to ensure that $\boldsymbol{AB}$ is the null matrix and the parent distribution is normal if the symmetric two point distribution is excluded. But the algebraic complexities appear to be too great to carry this through. For example, Kawada [10], knowing that the distribution was normal, proved that $\boldsymbol{AB}$ was the null matrix if $\boldsymbol{x}^T\boldsymbol{Ax}$ and $\boldsymbol{x}^T\boldsymbol{Bx}$ were mutually independent. But the algebra is very heavy. We shall have to be content with some characterisations which avoid even heavier algebra. Further, it seems difficult to prove that the equations (36) do yield unique solutions to the moments of uneven order. To avoid this difficulty, we now consider only symmetrical parent distributions.

We define the standardised cumulants, C_s of a distribution by

$$C_s = K_s, \quad \text{for} \quad s > 2, \quad C_2 = K_2 - 1 = 0, \quad C_1 = K_1 = 0 \tag{37}$$

where we have first of all standardised the distribution by taking the origin at the mean and changing the scale so that the variance is unity. The further step of taking $C_2 = K_2 - 1$ is to ensure the truth of

LEMMA 7. The standardised cumulants of a normal distribution are all zero.

PROOF. This is obvious from the method of defining the C_j.

LEMMA 8. Let x be a random variable possessing finite moments of all orders, zero mean and unit variance. Then the moment of any order s can be expressed uniquely in terms of the standardised cumulants of order s and lower together with a constant, perhaps zero. In this expansion the coefficient of C_s is unity.

PROOF. In the statistical texts, we find for the moments of a given order expansions in the cumulants of the form

$$\mu_s = K_s + \text{terms in } K_{s-1},\ K_{s-2} \text{ etc.} \tag{38}$$

Now to change this into the required form we note that $C_1 = K_1 = 0$, $C_j = K_j$ for $j > 2$, and we write $(C_2 + 1)$ in place of K_2. (38) is homogeneous in the sense that the subscripts of each term add up to the order s. After the change of notation, the expression is no longer homogeneous but the sum of the subscripts of the C_j in any term is at most equal to s. A constant term, g_s, may appear which has a value *independent of the distribution*. By inspection of the formula relating the μ_s and the cumulants, it is found that the only term which can contribute to the constant term is the term in $K_2^{\frac{1}{2}s}$, when s is even. g_s is the coefficient, in fact, of $K_2^{\frac{1}{2}s}$. It can also be evaluated by specialising the distribution to be normal, in which all the C_j vanish.

$$\begin{aligned} g_s &= 1 \cdot 3 \cdot 5 \cdots (s-1), &&\text{for even } s \\ &= 0 &&\text{for odd } s. \end{aligned} \tag{39}$$

LEMMA 9. The expectation of the powers of a quadratic form, $\boldsymbol{x}^T\boldsymbol{Ax}$ in a set $\{x_i\}$ of mutually independent, identically distributed variables, with zero mean and finite moments of all orders, can be evaluated in terms of the C_j, the standardised cumulants of the distribution.

PROOF. Let us consider $E(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t$, written out in full. From the independence of the variables we can write each term out in terms of the moments of the common distribution and so obtain a homogeneous expression in the moments with the sum of the subscripts of each term equal to $2t$. These can then be converted to expressions in the C_j and perhaps a constant term, $G(t; \boldsymbol{A})$. The value of $G(t; \boldsymbol{A})$ will be independent of the form of the common distribution since the relations used are identities not depending on special relationships between the moments.

LEMMA 10. The expectation of $(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})^{t'}$ can be obtained similarly with conditions as in Lemma 9. A constant term $G(t, t'; \boldsymbol{A}, \boldsymbol{B})$ may appear in the expansions.

PROOF. The proof goes through as in Lemma 9. We assert once more that $G(t, t'; \boldsymbol{A}, \boldsymbol{B})$ is independent of the actual form of the common distribution of the x_i.

LEMMA 11. In an arbitrary distribution subject to the conditions of Lemma 9,

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{O} \tag{40}$$

implies

$$G(t; \boldsymbol{A})\, G(t'; \boldsymbol{B}) = G(t, t'; \boldsymbol{A}, \boldsymbol{B}). \tag{41}$$

PROOF. We have already noted that the functions in (41) are independent of the form of the common distribution. We can obtain numerical results by suitably specialising the distribution. We do so by considering the case where each x is normally distributed. In this case, (37) shows that $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}$ are mutually independent by the theorem of Craig [3]. We have therefore

$$E\{(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})^{t'}\} = E(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^t\, E(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x})^{t'}. \tag{42}$$

Expanding both sides in terms of the C_j and the appropriate constants and remembering that in the normal distribution all the C_j are zero, we obtain (41) which was to be proved. We assert that (41) is true independently of the form of the distribution of the x_i providing it obeys the conditions of Lemma 11, namely $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{O}$.

We are now in a position to prove the main theorem of this section which in effect states that the Craig condition for independence characterises the normal distribution.

THEOREM 10. Let $x_1, x_2, \cdots, x_n$ *be a set of mutually independent, identically*

https://doi.org/10.1017/S1446788700026033 Published online by Cambridge University Press

and symmetrically distributed random variables. Let $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be non-negative matrices such that* $a_{jj}b_{jj} \neq 0$ *for at least one value of* j *and* $\boldsymbol{AB} = \boldsymbol{O}$. *Then, if* $\mathbf{x}^T\boldsymbol{A}\mathbf{x}$ *is stochasticallyi ndependent of* $\mathbf{x}^T\boldsymbol{B}\mathbf{x}$, *the distribution is the normal.*

Proof. The mean of each distribution is zero. Without loss of generality we take the variance to be unity. The proof is by induction. We assume that the C_j vanish for $j = 2, 4, \cdots, 2t$, and prove that C_{2t+2} is also zero. We have that C_2 is zero by definition and C_j for odd j is zero by the hypothesis that the distribution is even. The independence hypothesis gives us, in particular,

$$E(\mathbf{x}^T\boldsymbol{A}\mathbf{x})^t\,\mathbf{x}^T\boldsymbol{B}\mathbf{x} = E(\mathbf{x}^T\boldsymbol{A}\mathbf{x})^t E(\mathbf{x}^T\boldsymbol{B}\mathbf{x}) \tag{43}$$

now expanding both sides in terms of the C_j and constant terms, C_{2t+2} is the C_j with greatest subscript and it occurs only on the left side, where it has a non-zero coefficient, $d = \sum_1^n a_{jj}^t b_{jj}$, and this is positive by the non-negative condition on the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and the hypothesis of the theorem that at least one $a_{jj}b_{jj}$ is not zero. Now the inductive hypothesis gives us that every other C_j occurring in the expansions of the two sides of (43) is zero. There follows,

$$dC_{2t+2} + G(t, 1; \boldsymbol{A}, \boldsymbol{B}) = G(t; \boldsymbol{A})\, G(1; \boldsymbol{B}). \tag{44}$$

But by writing $t' = 1$ in (41) we obtain from (44) by the use of the identity (41),

$$C_{2t+2} = 0. \tag{45}$$

Thus the induction is proved. C_j is zero for every j. The distribution has therefore the same sequence of moments as the normal. This sequence is such that the moment problem is determined and so it follows that the distribution is normal.

In all the following we assume that the x's are a set of mutually independent, identically and symmetrically distributed random variables.

Corollary 1. If $\boldsymbol{A}$ is symmetric real and idempotent and $\mathbf{x}^T\boldsymbol{A}\mathbf{x}$ is distributed independently of $\mathbf{x}^T(\mathbf{1} - \boldsymbol{A})\mathbf{x}$ then the parent distribution is normal.

Proof. If we write $\boldsymbol{B} = \mathbf{1} - \boldsymbol{A}$, then all the conditions of Theorem 10 are satisfied.

Corollary 2. The condition of Cochran [2],

$$|\mathbf{1} - 2\rho\boldsymbol{A}||\mathbf{1} - 2\sigma\boldsymbol{B}| = |\mathbf{1} - 2\rho\boldsymbol{A} - 2\sigma\boldsymbol{B}|, \tag{46}$$

for independence can only hold in a normal system.

Proof. This is equivalent to the condition of Craig [3] given in Theorem 10, as can be proved by matrix methods as in [11].

https://doi.org/10.1017/S1446788700026033 Published online by Cambridge University Press

Corollary 3. If any normalised linear form $\boldsymbol{a}^T\boldsymbol{x}$ is independent of the residual sum of squares, $\boldsymbol{x}^T(\mathbf{1} - \boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{x}$, then the distribution is normal (Zinger [16]).

Proof. We replace $\boldsymbol{A}$ and $\boldsymbol{B}$ in Theorem 9 by $\boldsymbol{a}\boldsymbol{a}^T$ and $\mathbf{1} - \boldsymbol{a}\boldsymbol{a}^T$ respectively.

Corollary 4. If $\boldsymbol{a}^T\boldsymbol{B}$ is a null vector and if $\boldsymbol{a}^T\boldsymbol{x}$ is independent of $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}$, then the distribution is normal. ($\boldsymbol{B}$ is assumed to be non-negative definite and $a_j b_{jj} \neq 0$ for at least one j).

Proof. We write $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}^T$ and note that $\boldsymbol{x}^T\boldsymbol{a}\boldsymbol{a}^T\boldsymbol{x} = (\boldsymbol{a}^T\boldsymbol{x})^2$ is independent of $\boldsymbol{B}$.

Corollary 5. Geisser's characterisation [8]. If $\bar{x}$ is independent of $2^{-1}(n-k)^{-1}\sum_{j=1}^{n-k}(x_{j+k} - x_j)^2$, the distribution is normal.

Proof. It needs only to be verified that the product of the two quadratic forms is zero. Our proof is without restriction in the class of symmetric distributions. Geisser's [8] proof is without restriction if we use our Lemma 6 to prove the existence of all moments.

10. Summary

Some characterisations of the normal distribution by the independence of linear and quadratic forms of special interest in statistical theory have been proved. The method used is alternative to those of previous authors. Here the independence conditions are shown to imply the existence of all moments. Recurrence relations among the moments are then found to determine the normal distribution since the moments uniquely determine the distribution. The theorem on the independence of two linear forms is proved. The proof of the well-known characterisation of Geary [7] is completed, there being no need to specify the existence of the second moment. Some similar theorems on the independence of a general linear form from a quadratic form are proved. The characterisation theorems on quadratic forms lead to what appear to be extremely difficult algebraic computations and so we have been content to prove some theorems with symmetrically distributed variables. Within this class of parent distribution, the well-known Cochran and Craig conditions lead to the characterisation of the normal distribution.

References

[1] Bernstein, S. N., *Trud. Leningrad. Politechn. Institut.* (*Kalinin*), No. 3, (1941) 21–22.
[2] Cochran, W. G., *Proc. Camb. Phil. Soc.* 30 (1934), 178–181.
[3] Craig, A. T., *Ann. Math. Statistics*, 14 (1943), 195–197.
[4] Cramér, H., *Math. Zeit.*, 41 (1936) 405–414.
[5] Cramér, H., *Mathematical Methods of Statistics*, Princeton (1946).
[6] Darmois, G., *Proc. Internat. Statist. Conference*, 1947, Vol. 3A (1951), 231, Washington.

[7] Geary, R. C., *J. Roy. Statist. Soc. Suppl.* 3 (1936), 178—184.
[8] Geisser, S., *Ann. Math. Statistics* 27 (1956), 858—859.
[9] Herschel, J. (Anonymously), *Edinburgh Rev.*, 92 (1850), 1—57.
[10] Kawada, Y., *Ann. Math. Statistics*, 21 (1950), 614—615.
[11] Lancaster, H. O., *J. Roy. Statist. Soc. Ser. B*, 16 (1954), 247—254.
[12] Loève, M. A. note on pages 337—338 of P. Lévy, *Procéssus Stochastiques*, Gauthier-Villars, Paris (1948).
[13] Lukacs, E. and King, E. P., *Ann. Math. Statistics*, 25 (1954) 389—394.
[14] Marcinkiewicz, J., *Math. Zeit.*, 44 (1938), 612—618.
[15] Skitovitch, V. P., *Izvest. Akad. Sc. S.S.S.R., mat. Ser.* 18 (1954), 185—200.
[16] Zinger, A. A., *Uspehi mat. Nauk*, 6 (1951), 171—175.
[17] Zinger, A. A., *Doklady Akad. Nauk, S.S.S.R.*, 110 (1956), 319—322.
[18] Zinger, A. A., *Teoriya Veroyatnostei*, 3 (1958), 265—284.

Department of Mathematical Statistics
University of Sydney