CHARACTERIZATIONS OF NORMALITY BY CONSTANT REGRESSION OF LINEAR STATISTICS ON ANOTHER LINEAR STATISTIC[1]

By Theophilos Cacoullos

New York University

1. Introduction. Let $X_1, \cdots, X_n$ be n independent random variables. Several properties of certain statistics have been used in order to characterize the normality of the X's, for example, the independence of two linear statistics (Bernstein, Kac, Frechet, Darmois, Gnedenko, Skitovich). If the random variables $X_1, \cdots, X_n$ are, in addition, identically distributed (a sample from some population) with distribution function $F(x)$ then various properties imply the normality of F, e.g., if two linear statistics are identically distributed (Markinkiewicz, Linnik), or, if the sample mean $\bar{X}$ is independent of certain polynomial statistics, such as a quadratic statistic which is a multiple of the sample variance s^2 (Geary's theorem). For a detailed discussion of characterizations, we refer to Lukacs-Laha [2], where related references and results needed here are given.

In some cases, a relaxation or a modification of the conditions involved in the characterization theorems was possible. For example, Laha ([2], p. 105) replaced the condition of independence of $\bar{X}$ and s^2 by the property of constant regression of s^2 on $\bar{X}$. This property was also used (see [2], Chapter 6) to characterize other distributions besides the normal.

In [1] normality is characterized in terms of constant regression of the square of a linear statistic V on another linear statistic U. There, though on the one hand, the condition of independence of U and V is relaxed by simply assuming constant regression of V^2 on U, on the other hand, stronger conditions are imposed on the constant coefficients involved in the forms U and V.

In the present note, normality is characterized by the property of constant regression on a linear statistic (i) of a linear statistic and (ii) of a set of linearly independent linear statistics. These characterizations are motivated by the following considerations. It is well-known that two uncorrelated linear forms in jointly normal variables are (statistically) independent; hence, *a fortiori*, either linear form has constant regression on the other. The question then arises as to whether normality can be characterized by the property of constant regression of a linear form on another. This, however, is not true in general in view of the fact (see, e.g., [2], p. 104) that a linear form $U = a_1X_1 + \cdots + a_nX_n$ in independent and identically distributed (iid) random variables $X_1, \cdots, X_n$ has always linear regression on the sum $S = X_1 + \cdots + X_n$ (or on the sample mean $\bar{X}$), namely, $E(U \mid S) = \bar{a}S$, where $\bar{a}$ denotes the average of the a_i; hence, taking $a_1 + \cdots + a_n = 0$ gives constant (zero) regression of U on

Received 3 August 1966; revised 7 July 1967.

[1] Work partially supported by NSF Grant GP-7399.

S (or $\bar{X}$). This points out that we should restrict attention to the case of constant regression of U on some other linear statistic $V = b_1X_1 + \cdots + b_nX_n$ which is not proportional to $\bar{X}$. Indeed, under certain conditions on the constant coefficients a_j and b_j and the assumption that all the moments of the X_i exist, the property of constant regression of U on V implies that the X_i are normal (Theorem 1). The conditions on the b_j can be considerably relaxed if each of $n - 1$ linearly independent statistics has constant regression on V (Theorem 2). Multivariate analogues of these theorems are given in Theorem 3.

2. The results. For our purposes we need the following lemma and definition.

Lemma ([2], p. 105) *A random variable Y, with finite expectation, has constant regression on a random variable X,* i.e., *the relation $E(Y \mid X) = E(Y)$ holds almost everywhere, if, and only if, the relation*

$$E(Ye^{itX}) = E(Y)E(e^{itX}),$$

where i is the imaginary unit, holds for every real t.

Condition A. Let $a = (a_i, \cdots, a_n)'$ and $b = (b_1, \cdots, b_n)'$ be two (column) vectors. The pair of vectors (a, b) is said to satisfy Condition A if whenever the orthogonality relation $a'b = 0$ holds then $a'b^{(s)} \neq 0$ for every integer $s > 1$, where we set $b^{(s)} = (b_1^s, \cdots, b_n^s)'$.

Remark 1. For $n = 2$ it is observed that Condition A is satisfied unless the components of b are all proportional to ± 1 or 0. For $n > 2$ however this is no longer true in general. It would be of interest, if not to characterize, to find at least conditions under which $a'b = 0$ implies $a'b^{(s)} \neq 0$ for all integers $s > 1$. Note also that a necessary condition for the pair (a, b) to satisfy Condition A is that the components of b are not all proportional to 0 or ± 1, i.e., $b \neq \lambda\Delta$ where λ is a non-zero scalar and Δ denotes a non-zero vector with components $0, 1, -1$.

Theorem 1.[2] *Let $X_1, \cdots, X_n$ be a random sample from a univariate population with distribution function $F(x)$, and assume that $F(x)$ has moments of every order. Consider the linear forms*

$$U = a'X, \qquad V = b'X,$$

where $X = (X_1, \cdots, X_n)'$ and suppose (a, b) satisfies Condition A. Then U has constant regression on V, i.e.,

$$E(U \mid V) = E(U) \tag{1}$$

if, and only if, the following two conditions hold:

(i) *the population distribution F is normal,*

(ii) $a'b = 0$.

Proof. The sufficiency of (i) and (ii) is well known. We show their necessity.

[2] Rao gives essentially the same result as Theorem 5 in [4] where several other related characterizations are obtained.

By the lemma condition (1) is equivalent to

$$E(Ue^{itV}) = E(U)E(e^{itV}). \tag{2}$$

Let us assume, without any loss of generality, that

$$a_jb_j \neq 0, \qquad j = 1, \cdots, n, \tag{3}$$

and let

$$Y_j = b_jX_j, \qquad c_j = a_jb_j^{-1}, \qquad j = 1, \cdots, n.$$

Then (2) reduces to the differential equation

$$\textstyle\sum_{j=1}^{n} c_jf_j'(t) \prod_{k\neq j} f_k(t) = iE(U) \prod_{j=1}^{n} f_j(t), \tag{4}$$

where $f_j(t) = f(b_jt)$ denotes the characteristic function (c.f.) of Y_j and $f(t)$ the c.f. of F; $\sum$ stands for $\sum_{j=1}^{n}$. Letting $\varphi(t) = \log f(t)$ denote the cumulant generating functions (cgf) of F and $\varphi_j(t) = \log f_j(t)$ the cgf of Y_j in a neighborhood $|t| < \epsilon$ of the origin where $f(t) \neq 0$, we can write (4) in the form

$$\textstyle\sum c_j\varphi_j'(t) = iE(U). \tag{5}$$

Note that, since F has moments of every order, the cumulants κ_r of F exist for every r and we can differentiate (5) any number of times in the interval $|t| < \epsilon$. Thus differentiating once and then setting $t = 0$, we get

$$\kappa_2 \textstyle\sum c_jb_j^2 = \sigma^2 \sum a_jb_j = 0, \tag{6}$$

by taking also into account the relation

$$\varphi_j^{(r)}(0) = b_j^r\varphi^{(r)}(0) = i^rb_j^r\kappa_r.$$

Since we assume that the distribution F is not degenerate, the variance $\sigma^2 = \kappa_2$ is different than zero; hence (6) yields $\sum a_jb_j = 0$, that is, condition (ii).

Differentiating (5) s times and setting $t = 0$ gives

$$\kappa_{s+1} \textstyle\sum c_jb_j^{s+1} = \kappa_{s+1}a'b^{(s)} = 0, \tag{7}$$

which gives $\kappa_{s+1} = 0$ since, by Condition A, $\sum c_jb_j^{s+1} = \sum a_jb_j^s \neq 0$ for $s > 1$. Thus $\kappa_s = 0$ for $s > 2$, which characterizes normality.

REMARK 2. By Remark 1, the statistic V in Theorem 1 is not a multiple of $\bar{X}$, as already noted in Section 1, nor is it of the more general form $\lambda\Delta'X$. It is interesting at this point to recall Rao's [3] characterization of normality under the hypothesis that $V \neq \lambda\Delta'X$ is the uniformly minimum variance least squares estimate of an estimable parametric function in the usual regression model, when the rank of the so-called design matrix is one, i.e., in the case of an essentially unique estimable parametric function. Rao's characterization involves the solution of the same type of differential equation as in (4), and his proof proceeds like here. This points out an intrinsic relation between the property of constant regression of a linear statistic on another and the property of minimum variance of a linear (least squares) statistic regarded as an estimate of a parametric function. Both properties, in the present context, imply that U (in [3] every linear statistic in the error space) is uncorrelated with V. This in our case follows from the fact that if Y has constant regression on X, then X and Y are uncorrelated.

In order to avoid Condition A on (a, b) and also motivated by Rao's result as just described, we consider, instead of U, $n - 1$ linear statistics which have constant regression on V. More precisely, we have the following.

THEOREM 2[3]. *Let* $X_1, \cdots, X_n$ *be* iid *random variables with* $E(X_1) = 0$ *and* $0 < E(X_1^2) < \infty$. *Suppose that there exist* $n - 1$ *linearly independent statistics*

$$U_j = \alpha_j'X, \qquad j = 1, \cdots, n-1, \quad \text{and} \quad V = b'X$$

such that each U_j *has constant regression on* V, i.e.,

$$E(U_j/V) = 0, \qquad j = 1, \cdots, n-1. \tag{8}$$

Then each X_i *has the normal distribution.*

PROOF. Without loss of generality, let

$$\begin{aligned} \alpha_1' &= (1, 0, \cdots, 0, \alpha_{1n}) \\ \alpha_2' &= (0, 1, \cdots, 0, \alpha_{2n}) \\ &\cdots\cdots\cdots \\ \alpha_{n-1}' &= (0, 0, \cdots, 1, \alpha_{n-1,n}). \end{aligned}$$

By the lemma, condition (8) gives

$$\psi(b_j t) + \alpha_{jn}\psi(b_n t) = 0, \qquad j = 1, \cdots, n-1, \tag{9}$$

where $\psi(t) = f'(t)/f(t)$ and f denotes the c.f. of X_1. Since X_1 is not degenerate $b_n \neq 0$, and we can take $b_n = 1$ without loss of generality. Hence (9) can be written as

$$\psi(t) = d_j\psi(b_j t), \qquad j = 1, \cdots, n-1,$$

and since $b \neq \lambda\Delta$ at least one b_j satisfies $0 < |b_j| \neq 1$. Thus assume without any loss of generality that $d_1 \neq 0$, $0 < |b_1| < 1$ so that

$$\psi(t) = d_1\psi(b_1 t). \tag{10}$$

Furthermore the assumption $0 < E(X_1^2) < \infty$ implies that $\psi'(t)$ exists and is continuous at $t = 0$. Hence from (10) we get

$$\psi'(0)(1 - b_1 d_1) = 0$$

which implies $b_1 d_1 = 1$ since $\psi'(0) \neq 0$. Thus (10) holds with $|b_1 d_1| = 1$ and $|b_1| < 1$. Now by Lemma 1 of [4] it follows that $\psi(t) = ct$ which completes the proof.

REMARK 3. Suppose $b = \lambda\Delta$ and moreover some of the components of b are proportional to $+1$ and some to -1. If in Theorem 2 we assume that the

[3] Originally shown under the assumption that all the moments of X_1 exist. The author thanks the referee for pointing out the present proof based on Rao's recent paper [4].

$(s + 1)$st moment exists, then, by analogy to (7) we obtain

$$\kappa_{s+1}\alpha_j' b^{(s)} = 0, \qquad j = 1, \cdots, n - 1; \tag{11}$$

since $\kappa_2 = E(X_1^2)$, relation (11) with $s = 1$ gives

$$\alpha_j' b = 0, \qquad j = 1, \cdots, n - 1,$$

whereas if s is even (11) implies that the $(s + 1)$st odd cumulant $\kappa_{s+1} = 0$. Thus if X_1 has moments of every order, then all the odd cumulants of X_1 are zero and therefore its distribution is symmetric. Another case in which the distribution of X_1 can be shown to be symmetric is obtained from (7) as a corollary of Theorem 1.

COROLLARY. *Suppose that in addition to the assumptions of Theorem* 1 *the following condition holds*

$$\alpha_j \alpha_k > 0 \quad \textit{for all} \quad j, k = 1, 2, \cdots, n.$$

Then the relation (1) *implies that* F *is symmetric.*

Now we give some multivariate analogues of the preceding results.

THEOREM 3. *Theorems* 1 *and* 2 *hold if the random variables* X_i *are p-vectors, "moments" are replaced by "product moments" and "normal" by "p-variate normal".*

PROOF. Let $t = (t_1, \cdots, t_p)'$ be any constant vector and define $Z_j = t'X_j$ $(j = 1, \cdots, n)$. Then the role of the scalar variables X_j in Theorems 1 and 2 is now played by Z_j, and therefore the Z_j are normal; since this is true for every vector t, the assertion follows by direct application of the well-known characterization of a multivariate normal distribution (see, e.g., [2], p. 30).

REFERENCES

[1] CACOULLOS, T. (1967). Some characterizations of normality. To appear in *Sankhyā Ser. A* **30**.

[2] LUKACS, E. and LAHA, R. G. (1964). *Applications of characteristic functions*. Griffin, London.

[3] RAO, C. R. (1959). Sur une charactérisation de la distribution normale établie d'aprés une propriété optimum des estimations linéaires. *Le Calcul des probabilités et ses applications*, 165–172, Centre National de la Recherche Scientifique, Paris.

[4] RAO, C. R. (1967). On some characterizations of the normal law. *Sankhyā Ser. A* **29** 1–14.