The Annals of Probability
1986, Vol. 14, No. 2, 696–701

SPHERICITY AND THE NORMAL LAW[1]

BY ROBERT H. BERK

Rutgers University

Let $\mathbf{x} = (x_1, \dots, x_n)'$ be a random vector in R^n. Two characterizations of normality are given. One involves the existence of two linear combinations of the $\{x_j\}$ that are independent in every coordinate system. The other, which is actually a consequence of the first, assumes that $\mathbf{x}$ obeys a linear model with spherical errors and involves sufficiency of the least-squares estimator.

1. Two independent linear combinations. The following characterization of the normal law is well known: If $\mathbf{x}$ is spherically distributed ($G'\mathbf{x} \sim \mathbf{x}$ for every $n \times n$ orthogonal matrix G) and has mutually independent coordinates, then the $\{x_j\}$ are iid $N(0, \sigma^2)$. The result (for $n = 3$) is apparently due to Maxwell (1860). It has been subsequently rediscovered, for example by Bartlett (1934). (Bartlett's proof involves an unnecessary smoothness assumption on the common pdf [assumed to exist] of the $\{x_j\}$.) Kac (1939) presented a stronger result: Suppose the coordinates of $\mathbf{x}$ are independent in every coordinate system (i.e., $\mathbf{y} = G'\mathbf{x}$ has independent coordinates for all choices of $n \times n$ orthogonal G). Then the $\{x_j\}$ are mutually independent normal variables with a common variance. (The result Kac states is slightly in error, as he concludes that the $\{x_j\}$ must all have mean zero.) Kac deals explicitly only with the case $n = 2$. Hartman and Wintner (1940) give a correct statement and proof for general n. Their method of proof is quite different from Kac's. The characterization given below is in the spirit of Kac's result. Unlike his, it deals with only two linear combinations of the $\{x_j\}$ which need not be assumed orthogonal. Theorem 1.1 below considers the case $n = 2$, which is then used to get the general result (Theorem 1.2).

THEOREM 1.1. *Let x and y be random variables. If there are coefficients a and b, $a^2 + b^2 > 0$, so that x and $ax + by$ are independent in every coordinate system, then x and y are independent normal variables with the same variance.*

REMARK 1. The hypothesis means that x^* and $ax^* + by^*$ are independent whenever $(x^*, y^*) = (x, y)G$, where G is any 2×2 orthogonal matrix.

REMARK 2. If $a \neq 0$, x and y are degenerate (a.s. constant).

PROOF. The proof is in two parts. We first show that $a \neq 0$ implies x and y are degenerate.

Received June 1984; revised November 1984.

[1]Work supported by NSF Grant DMS 83-01082-01.

AMS 1980 *subject classifications.* Primary 60E99; secondary 62B99.

Key words and phrases. Characterization of normality, sphericity, sufficiency, least-squares estimator.

Clearly $b = 0$ entails that x is degenerate [and then so is y, on applying the orthogonal transformation $(x, y) \to (y, x)$], so we may assume $b = 1$. Suppose now $a \neq 0$. Let $u = x$ and $v = ax + y$, so that $y = v - au$. Let G be a 2×2 orthogonal matrix with elements $\{g_{ij}\}$. Then under G, (x, y) transforms into $(x^*, y^*) = (x, y)G$, so that $x^* = g_{11}x + g_{21}y$ and $y^* = g_{12}x + g_{22}y$. Let $u^* = x^*$ and $v^* = ax^* + y^*$, which are independent by hypothesis. We have also that

$$(1.1)\qquad \begin{aligned} u^* &= (g_{11} - ag_{21})u + g_{21}v, \\ v^* &= (ag_{11} + g_{12} - a^2g_{21} - ag_{22})u + (ag_{21} + g_{22})v. \end{aligned}$$

Let now $A(t)$ and $B(t)$ denote, respectively, the cf's (characteristic functions) of u and v. The independence of u^* and v^* implies that

$$E\exp\{i(su^* + tv^*)\} = E\exp\{isu^*\}E\exp\{itv^*\}.$$

However, $su^* + tv^*$ is a linear combination of u and v, which are also independent. Using this latter fact and (1.1), we obtain the relation

$$\begin{aligned} &A(s[g_{11} - ag_{21}] + t[ag_{11} + g_{12} - a^2g_{21} - ag_{22}])B(sg_{21} + t[ag_{21} + g_{22}]) \\ (1.2)\qquad &= A(s[g_{11} - ag_{21}])B(sg_{21}) + A(t[ag_{11} + g_{12} - a^2g_{21} - ga_{22}]) \\ &\quad \times B(t[ag_{21} + g_{22}]), \end{aligned}$$

which holds for all real s and t and all orthogonal G. Taking in (1.2) $g_{11} = g_{22} = 0$, $g_{21} = 1$, and $g_{12} = \pm 1$ yields

$$(1.3)\quad A(-as + (1 - a^2)t)B(s + at) = A(-as)B(s)A([1 - a^2]t)B(at)$$

and

$$(1.4)\quad A(-as - (1 + a^2)t)B(s + at) = A(-as)B(s)A(-(1 + a^2)t)P(at).$$

We argue next that A and B do not vanish. For suppose A has zeros; let σ be the smallest (in magnitude) such ($\sigma \neq 0$ since $A(0) = 1$ and A is continuous). Then setting $s = -at$ and $-(1 + a^2)t = \sigma$ in (2.4) gives $A(-\sigma/[1 + a^2]) = 0$, which contradicts σ being the smallest zero. Hence A does not vanish. Now suppose B has a zero, say τ. Setting $at = \tau$ in (1.4) then entails (since $A \neq 0$) that $B(s + \tau) = 0$ for all s, which clearly is also a contradiction. Thus B does not vanish either. We may then divide (1.3) by (1.4) and let $p = -as$ to obtain

$$(1.5)\qquad \frac{A(p + (1 - a^2)t)}{A(p - (1 + a^2)t)} = \frac{A([1 - a^2]t)}{A(-(1 + a^2)t)} \quad \text{for all } p, t.$$

If $a^2 = 1$, then (1.5) becomes $A(p - 2t) = A(p)A(-2t)$, which is to say, the Cauchy equation

$$(1.6)\qquad A(p + q) = A(p)A(q).$$

The only continuous (indeed, measurable) solutions of (1.6) are of the form $A(t) = \exp\{\gamma t\}$. Since A is a cf, we must have $A(t) = \exp\{i\lambda t\}$ where λ is real. That is, $u = x$ is degenerate at λ. (1.3) then becomes $B(s + at) = B(s)B(at)$, so also v and hence y is degenerate.

If $a^2 \neq 1$, then, letting $r = (1 + a^2)/(1 - a^2) \neq 1$ and $(1 - a^2)t = \tau$, (1.5) becomes

$$A(\tau)/A(-r\tau) = A(p + \tau)/A(p - r\tau) \quad \text{for all } p, \tau. \tag{1.7}$$

Setting $p = r\tau$ in (1.7) gives $A(\tau)/A(-r\tau) = A(r\tau + \tau)$, so that (1.7) becomes

$$A(p - r\tau)A(r\tau + \tau) = A(p + \tau). \tag{1.8}$$

Since $(p - r\tau) + (r\tau + \tau) = p + \tau$, (1.8) is also Cauchy's equation, so that again x and y are degenerate. Thus $a \neq 0$ implies that x and y are degenerate.

For the second part of the proof, we assume $a = 0$ and show that x and y must be normal with a common variance. This can be concluded from Kac's (1939) result, or from Hartman and Wintner (1940). It also follows from the Darmois–Skitovich theorem (cf. Kagan et al. (1973), page 89). Thus we simply sketch a proof.

When $a = 0$, $u = x$, and $v = y$, taking $g_{11} = g_{12} = g_{21} = -g_{22} = 1$ in (1.2) gives

$$A(s + t)B(s - t) = A(s)A(t)B(s)B(-t), \tag{1.9}$$

which, on interchanging s and t, becomes

$$A(s + t)B(t - s) = A(s)A(t) + B(-s)B(t). \tag{1.10}$$

Again, A and B do not vanish. For suppose σ and τ are the smallest zeros of A and B, respectively ($\sigma = \infty$ if A has no zeros), and suppose $|\sigma| \leq |\tau|$ (the contrary case being similar). Then taking $s = t = \sigma/2$ in (1.9) gives $A^2(\sigma/2)B(\sigma/2)B(-\sigma/2) = 0$, which is a contradiction if σ is finite. Thus A has no zeros. Similarly, B has none [take $t = \tau$ in (1.9)]. Dividing (1.9) by (1.10) and replacing t by $-t$ yields

$$\frac{B(s + t)}{B(-s - t)} = \frac{B(s)}{B(-s)}\,\frac{B(t)}{B(-t)},$$

so that $C(t) = B(t)/B(-t)$ satisfies Cauchy's equation. Thus $C(t) = \exp\{2\delta t\}$ for some δ. Let $b(t) = B(t)\exp\{-\delta t\}$. Then $b(t)/b(-t) = 1$, so that b is even. Similarly, $A(t)/A(-t) = \exp\{2\gamma t\}$ and $a(t) = A(t)\exp\{-\gamma t\}$ is even. (1.9) then entails

$$a(s + t)b(s - t) = a(s)a(t)b(s)b(t) = a(s - t)b(s + t). \tag{1.11}$$

Setting $s = t$ shows that $a = b$ and (1.11) becomes

$$a(s + t)a(s - t) = a^2(s)a^2(t). \tag{1.12}$$

Kac (1939) argues that the only continuous solutions of this equation are of the form $a(t) = \exp\{\alpha t^2\}$, so that $A(t) = \exp\{\gamma t + \alpha t^2\}$. In order that A be a cf, we must have $\gamma = i\lambda$ and $\alpha = -\frac{1}{2}\sigma^2$, where λ and σ are real, so that $x \sim N(\lambda, \sigma^2)$. Since $b = a$, $B(t) = \exp\{\delta t + \alpha t^2\} = \exp\{i\mu t - \frac{1}{2}\sigma^2 t\}$, and $y \sim N(\mu, \sigma^2)$. Since $a = 0$, x and y are independent. □

REMARK 3. The full force of the hypothesis is not used in proving the theorem, as only three particular orthogonal transformations are used.

Theorem 1.1 extends to n dimensions as follows.

THEOREM 1.2. *If* $\mathbf{x}$ *is an n-dimensional random vector for which there are (constant) nonzero vectors* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *so that* $\boldsymbol{\alpha}'\mathbf{x}$ *and* $\boldsymbol{\beta}'\mathbf{x}$ *are independent in every coordinate system, then either the* $\{x_j\}$ *are degenerate or* $\boldsymbol{\alpha} \perp \boldsymbol{\beta}$ *and the coordinates of* $\mathbf{x}$ *are mutually independent normal variables with a common variance.*

PROOF. The hypothesis entails that $\boldsymbol{\alpha}'(G'\mathbf{x})$ and $\boldsymbol{\beta}'(G'\mathbf{x})$ are independent for all $n \times n$ orthogonal G. We can suppose $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = 1$. Let $\mathbf{a} = (1,0,\dots,0)'$ and $\mathbf{b} = (b_1, b_2, 0,\dots,0)'$ in R^n be such that $|\mathbf{b}| = 1$ and $\mathbf{a}'\mathbf{b} = \boldsymbol{\alpha}'\boldsymbol{\beta}$, i.e., $b_1 = \boldsymbol{\alpha}'\boldsymbol{\beta}$. Then there is an $n \times n$ orthogonal matrix G for which $G\boldsymbol{\alpha} = \mathbf{a}$ and $G\boldsymbol{\beta} = \mathbf{b}$. The hypothesis then entails that $\mathbf{a}'\mathbf{x}$ and $\mathbf{b}'\mathbf{x}$, which is to say, x_1 and $b_1x_1 + b_2x_2$ are independent. Theorem 1.1 entails that x_1 and x_2 are degenerate if $\boldsymbol{\alpha}'\boldsymbol{\beta} = b_1 \neq 0$ and that otherwise, x_1 and x_2 are independent normal variables with a common variance. By choice of G, we may replace (x_1, x_2) with any pair (x_i, x_j) and conclude that $x_1,\dots,x_n$ are either all degenerate (if $\boldsymbol{\alpha}'\boldsymbol{\beta} \neq 0$) or are marginally normal with a common variance (and are pairwise independent).

Further, let $\gamma_1x_1 + \gamma_2x_2$ be any linear combination of x_1 and x_2 with $\gamma_1^2 + \gamma_2^2 = 1$. Then there is an orthogonal G so that the first two coordinates of $G'\mathbf{x}$ are $\gamma_1x_1 + \gamma_2x_2$ and x_3. (The first row of G' is $(\gamma_1, \gamma_2, 0,\dots,0)$ and the second is $(0,0,1,0,\dots,0)$.) Thus $\gamma_1x_1 + \gamma_2x_2$ and x_3 are independent. Since every linear combination of x_1 and x_2 is proportional to some $\gamma_1x_1 + \gamma_2x_2$, x_3 is independent of every linear combination of x_1 and x_2. It is clear then that the joint cf of x_1, x_2, and x_3 factors or that x_3 is independent of (x_1, x_2). Continuing this argument, we see that x_{j+1} is independent of $(x_1,\dots,x_j)$ for every $1 \le j \le n-1$, thus that $(x_1,\dots,x_n)$ are mutually independent. □

We have thus the following refinement of Maxwell's result.

COROLLARY 1.3. *If* $\mathbf{x}$ *is spherically distributed in* R^n *and there exist two independent linear combinations* $\boldsymbol{\alpha}'\mathbf{x}$ *and* $\boldsymbol{\beta}'\mathbf{x}$ $(|\boldsymbol{\alpha}| \neq 0 \neq |\boldsymbol{\beta}|)$, *then either* $\mathbf{x} = \mathbf{0}$ *w.p.1 or* $\boldsymbol{\alpha} \perp \boldsymbol{\beta}$ *and the coordinates of* $\mathbf{x}$ *are iid* $N(0, \sigma^2)$.

2. A linear model. We suppose now that $\mathbf{x}$ satisfies the following linear model:

$$\mathbf{x} = V\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where V is a known $n \times p$ matrix (of independent variables, $1 \le p < n$), $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of regression coefficients, and $\boldsymbol{\varepsilon}$ is spherically distributed in R^n. We suppose the distribution of $\boldsymbol{\varepsilon}$ is known, except possibly for a scale factor σ (which can be taken to be the standard deviation of ε_1 when $E\varepsilon_1^2 < \infty$). Such a model is discussed in Berk and Hwang (1984). We show that for σ known, the least-squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is a sufficient statistic iff the $\{x_j\}$ are mutually independent normal variables with variance σ^2.

In discussing sufficiency for this model, we need the following considerations. Let $\mathbf{x}_V$ be the projection of $\mathbf{x}$ into $\langle V \rangle$, the column space of V and let $\mathbf{x}_\perp = \mathbf{x} - \mathbf{x}_V$ be the projection into $\langle V \rangle^\perp$. (To avoid inessential circumlocutions, we assume $\dim\langle V \rangle = \operatorname{rank} V = p$.) Let $\mathbf{u}_V = \boldsymbol{\varepsilon}_V/|\boldsymbol{\varepsilon}_V|$ and $\mathbf{u}_\perp = \boldsymbol{\varepsilon}_\perp/|\boldsymbol{\varepsilon}_\perp|$.

LEMMA 2.1. *For spherical* $\boldsymbol{\varepsilon}$, *the vectors* $\mathbf{u}_V$, $\mathbf{u}_\perp$, *and* $(|\boldsymbol{\varepsilon}_V|, |\boldsymbol{\varepsilon}_\perp|)$ *are mutually independent. Moreover,* $\mathbf{u}_V$ *and* $\mathbf{u}_\perp$ *are uniform on (the surfaces of) the unit spheres* S_V *and* $S_\perp$ *in* $\langle V \rangle$ *and* $\langle V \rangle^\perp$, *respectively.*

PROOF. Let G be an $n \times n$ orthogonal matrix for which $\langle V \rangle$ is an invariant space and which fixes $\langle V \rangle^\perp$ pointwise: $G\langle V \rangle = \langle V \rangle$ and $G\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \langle V \rangle^\perp$. Then $G\boldsymbol{\varepsilon} = G(\boldsymbol{\varepsilon}_V + \boldsymbol{\varepsilon}_\perp) = G\boldsymbol{\varepsilon}_V + \boldsymbol{\varepsilon}_\perp$, showing that $(G\boldsymbol{\varepsilon})_V = G\boldsymbol{\varepsilon}_V$. Since $G\boldsymbol{\varepsilon} \sim \boldsymbol{\varepsilon}$, we have, for A a measurable subset of S_V,

$$P(\mathbf{u}_V \in A \mid |\boldsymbol{\varepsilon}_V| = x, \boldsymbol{\varepsilon}_\perp = \mathbf{y}) = P(G\mathbf{u}_V \in A \mid |\boldsymbol{\varepsilon}_V| = x, \boldsymbol{\varepsilon}_\perp = \mathbf{y}).$$

This shows that the indicated conditional distribution of $\mathbf{u}_V$ is spherically invariant. It is, therefore, the unique spherically invariant (uniform) distribution on S_V. Thus $\mathbf{u}_V$ and $(|\boldsymbol{\varepsilon}_V|, \boldsymbol{\varepsilon}_\perp)$ are independent. Equivalently, $\mathbf{u}_V$ and $(|\boldsymbol{\varepsilon}_V|, |\boldsymbol{\varepsilon}_\perp|, \mathbf{u}_\perp)$ are independent. By reversing the roles of $\boldsymbol{\varepsilon}_V$ and $\boldsymbol{\varepsilon}_\perp$, we have similarly that $\mathbf{u}_\perp$ is uniform on $S_\perp$ and is independent of $(|\boldsymbol{\varepsilon}_V|, |\boldsymbol{\varepsilon}_\perp|, \mathbf{u}_V)$. □

It follows from Lemma 2.1 that with scale known or unknown, $(\mathbf{x}_V, |\mathbf{x}_\perp|)$ is a sufficient statistic for the linear model (2.1): By Lemma 2.1, the omitted information, $\mathbf{x}_\perp/|\mathbf{x}_\perp| = \mathbf{u}_\perp$, is independent of $(\mathbf{x}_V, |\mathbf{x}_\perp|)$ and is ancillary. Note that $\mathbf{x}_V$ is equivalent to $\hat{\boldsymbol{\beta}} = (V'V)^{-1}V'\mathbf{x}$; more precisely, $\mathbf{x}_V = V\hat{\boldsymbol{\beta}}$. (Note too that $|\mathbf{x}_\perp|^2 = |\mathbf{x}|^2 - |V\hat{\boldsymbol{\beta}}|^2$ is the usual residual sum of squares from the model.) When the $\{\varepsilon_j\}$ are iid $N(0, \sigma^2)$ with σ known, $\hat{\boldsymbol{\beta}}$ alone is a sufficient statistic (as then $\mathbf{x}_\perp$ is independent of $\mathbf{x}_V$ and is ancillary). This turns out to be a characteristic property of the normal distribution.

THEOREM 2.2. *For the model* (2.1) *with* $\boldsymbol{\varepsilon}$ *spherical with known scale,* $\hat{\boldsymbol{\beta}}$ *is a sufficient statistic iff the* $\{\varepsilon_j\}$ *are iid normal variables with mean zero.*

To prove Theorem 2.2, we need to establish:

THEOREM 2.3. *Let* $\mathbf{y} \in R^p$ *and* $\mathbf{z}$ *be random vectors and consider the translation family of distributions of* $(\mathbf{y} + \boldsymbol{\beta}, \mathbf{z})$, $\boldsymbol{\beta} \in R^p$. *Then* $\mathbf{y} + \boldsymbol{\beta}$ *is sufficient for* $(\mathbf{y} + \boldsymbol{\beta}, \mathbf{z})$ *iff* $\mathbf{y}$ *and* $\mathbf{z}$ *are independent.*

PROOF. That independence suffices is clear. On the other hand, if $\mathbf{y} + \boldsymbol{\beta}$ is sufficient, then $P(\mathbf{z} \in A \mid \mathbf{y} + \boldsymbol{\beta})$ can be chosen not to depend on $\boldsymbol{\beta}$. (Here A is any measurable subset of range $\mathbf{z}$.) That is, there is a function h on R^p so that we may take $P(\mathbf{z} \in A \mid \mathbf{y} + \boldsymbol{\beta} = \mathbf{t}) = h(\mathbf{t})$, for all $\boldsymbol{\beta}$ and $\mathbf{t}$ in R^p. But then $h(\mathbf{t}) = P(\mathbf{z} \in A \mid \mathbf{y} = \mathbf{t} - \boldsymbol{\beta}) = h(\mathbf{t} - \boldsymbol{\beta})$, showing that h is constant. Thus $P(\mathbf{z} \in A \mid \mathbf{y} = \mathbf{t})$ is constant in $\mathbf{t}$, which means that $\mathbf{y}$ and $\mathbf{z}$ are independent. □

REMARK. The above argument clearly extends to any "transformation parameter" family—i.e., a family generated by a group acting on $\mathbf{y}$ only.

Theorem 2.3 is an analog of Basu's theorem, since it asserts that an ancillary is independent of a sufficient statistic. One does not need bounded completeness—but the choice of ancillary seems somewhat restricted.

PROOF OF THEOREM 2.2. Suppose $\hat{\boldsymbol{\beta}}$ is sufficient. We take $\mathbf{y} + \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\mathbf{z} = \mathbf{x}_{\perp}$ in Theorem 2.3 and conclude that $\hat{\boldsymbol{\beta}}$ and $\mathbf{x}_{\perp}$ are independent, or that $\mathbf{x}_V$ and $\mathbf{x}_{\perp}$ are independent. This implies the existence of nondegenerate independent linear combinations of the $\{\varepsilon_j\}$, so that by Corollary 1.3, the $\{\varepsilon_j\}$ are iid $N(0,\sigma^2)$. □

It seems curious that, except in the normal case, one cannot automatically discard the ancillary $|\mathbf{x}_{\perp}|$.

REFERENCES

BARTLETT, M. S. (1934). The vector representation of a sample. *Proc. Camb. Phil. Soc.* **30** 327–340.

BERK, R. H. and HWANG, J. T. (1984). Optimality of the least squares estimator. Unpublished.

HARTMAN, P. and WINTNER, (1940). On the spherical approach to the normal law. *Amer. J. Math.* **62** 759–779.

KAC, M. (1939). On a characterization of the normal distribution. *Amer. J. Math.* **61** 726–728.

KAGAN, A. M., LINNIK, YU. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

MAXWELL, J. C. (1860). Illustration of the dynamical theory of gases—Part I. On the motions and collisions of perfectly elastic bodies. *Taylor's Phil. Mag.* **19** 19–32.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903