PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 19, Fasc. 2 (1999), pp. 441–452

THE GENERALIZATION OF THE KAC–BERNSTEIN THEOREM

BY

M. P. QUINE AND E. SENETA (SYDNEY)

Abstract. The Skitovich–Darmois Theorem of the early 1950's establishes the normality of independent $X_1, X_2, \ldots, X_n$ from the independence of two linear forms in these random variables. Existing proofs generally rely on the theorems of Marcinkiewicz and Cramér, which are based on analytic function theory. We present a self-contained real-variable proof of the essence of this theorem viewed as a generalization of the case $n = 2$, which is generally called *Bernstein's Theorem*, and also adapt an early little known argument of Kac to provide a direct simple proof when $n = 2$. A large bibliography is provided.

Key words: independence; characterization; normality; Bernstein's theorem; Cramér's theorem; Marcinkiewicz's theorem; characteristic function; Laplace transform; real-variable; real function; moments; cumulants.

1. INTRODUCTION

The Skitovich–Darmois Theorem asserts that if $n \geqslant 2$ is fixed, $X_1, \ldots, X_n$ are independent, and $Y_1 = \sum_{j=1}^n a_j X_j$ is independent of $Y_2 = \sum_{j=1}^n b_j X_j$ for some constants $\{a_j\}, \{b_j\}$ with $a_j b_j \neq 0$, $j = 1, \ldots, n$, then each X_j is normally distributed. This theorem implies Cramér's Theorem (Cramér [6]) through a simple application of the case $n = 4$ (Linnik [25]). On the other hand, proofs of the Skitovich–Darmois Theorem are not self-contained in that they require

(1) the use of Cramér's Theorem (at the very least to cover the case where for some $j \neq k$, $a_j/b_j = a_k/b_k$, since both sums then contain a multiple of $b_j X_j + b_k X_k$), and

(2) the proposition that if a characteristic function $\phi(t) = E(e^{itX})$, t real, has the form $\exp(P(t))$, where $P(t)$ is a polynomial, then the degree of the polynomial is not greater than 2.

This last is a form of Marcinkiewicz's Theorem, which is in terms of a complex variable z instead of t. The complex-variable version is easier to prove directly (e.g., Linnik [26], p. 65); the real variable version is quite long and difficult (Lukacs [28], pp. 213–221; Bryc [5], p. 35), although the jump from real t to complex variable z is sometimes made rather cursorily. As regards (1), the proof of Cramér's Theorem depends on a deep result from the theory of entire functions, Hadamard's factorization theorem, which is stated but not proved in probability monographs (e.g., Linnik [26]). Thus proofs of the Skitovich–Darmois Theorem to a large extent depend on external theorems, whereas an essentially self-contained proof, not heavily dependent on results from entire function theory, for the most part in real variable terms, and avoiding use of the proposition about polynomial exponents, is desirable from a didactic viewpoint.

The essence of the Skitovich–Darmois Theorem is to view it (Darmois [10], p. 6) as an extension of Bernstein's Theorem (the case $n = 2$) by putting aside the possibility that $a_j/b_j = a_k/b_k$ for some $j \neq k$. This enables us to produce, in Section 2, a self-contained proof of the kind desired. Naturally, this proof borrows and interrelates a number of clever arguments to be found in the works of authors such as Skitovich, Lancaster, Lukacs and King, and Dugué, when they address the Skitovich–Darmois setting. There are also novel elements, such as the proof of Lemma 4, and the switch from characteristic functions to Laplace transforms following Lemma 5, in Section 2.2.

In Section 3, which deals with the case $n = 2$, we adapt the largely overlooked real-variable argument of Kac [17] to prove Gnedenko's [16] generalization of Bernstein's Theorem [3]. Our overall treatment in both Section 2 and Section 3 rests heavily on Lemma 2, which is due to Lancaster [22].

The paper includes a large bibliography which, whilst not complete, seeks to illuminate the early published history on this topic, disrupted as it was by World War 2 and its aftermath.

2. THE SKITOVICH–DARMOIS THEOREM

We state our result before proceeding (A restricted version was the purpose of Marcinkiewicz [31].)

THEOREM 1. *Let $n \geqslant 2$ be fixed, $X_1, \ldots, X_n$ be non-degenerate and independently distributed random variables, and suppose that*

$$Y_1 = \sum_{j=1}^{n} X_j \quad \text{and} \quad Y_2 = \sum_{j=1}^{n} b_j X_j$$

are independently distributed, where the constants $\{b_j\}$ satisfy $b_j \neq 0$, $b_j \neq b_k$, $j \neq k$. Then each X_j is normally distributed.

2.1. Real variable arguments. As a first step to a proof of Theorem 1 we follow Skitovich [37] by symmetrizing. Let $(X'_1, \ldots, X'_n)$ be an independent replica of $(X_1, \ldots, X_n)$ and define

$$Y'_1 = \sum_{j=1}^{n} X'_j \quad \text{and} \quad Y'_2 = \sum_{j=1}^{n} b_j X'_j.$$

Then $\tilde{X}_j = X_j - X'_j$, $j = 1, \ldots, n$, are independent and

$$\tilde{Y}_1 = Y_1 - Y'_1 = \sum_{j=1}^{n} (X_j - X'_j), \qquad \tilde{Y}_2 = Y_2 - Y'_2 = \sum_{j=1}^{n} b_j (X_j - X'_j)$$

are also independent. The characteristic functions of the symmetrized variables are of course real-valued, but the independence of the linear forms gives more:

LEMMA 1. *Define*

$$\tilde{\phi}_j(t) = \boldsymbol{E} \exp(it\tilde{X}_j), \qquad -\infty < t < \infty, \ j = 1, \ldots, n.$$

Then $0 < \tilde{\phi}_j(t) \leqslant 1$ *for all real* t.

Proof (based on Skitovich [37]). The independence properties can be expressed as $L(u, v) = R(u, v)$ for $-\infty < u < \infty$, $-\infty < v < \infty$, where

$$L(u, v) = \prod_{j=1}^{n} \tilde{\phi}_j(u + b_j v), \tag{2.1}$$

$$R(u, v) = \prod_{j=1}^{n} \tilde{\phi}_j(u) \cdot \prod_{j=1}^{n} \tilde{\phi}_j(b_j v). \tag{2.2}$$

If the lemma is false, then, by continuity and since $L(0, 0) = R(0, 0) = 1$, there exists a number w such that

$$R(u, v) > 0 \text{ for } |u| < |w| \text{ and } |v| < |w|, \quad \text{and} \quad R(w, w) = 0. \tag{2.3}$$

This entails either $\tilde{\phi}_k(w) = 0$ for some k or $\tilde{\phi}_k(b_k w) = 0$ for some k. In the first case, let $u_1 = (1 - b_k/c) w$ and $v_1 = w/c$, where c is chosen so that $|c| > \max(1, |b_k|)$ and $b_k/c > 0$. Then we have $u_1 + b_k v_1 = w$ and

$$L(u_1, v_1) = \prod_{j \neq k} \tilde{\phi}_j(u_1 + b_j v_1) \cdot \tilde{\phi}_k(u_1 + b_k v_1) = 0,$$

so $R(u_1, v_1) = 0$. This contradicts (2.3), since $|u_1| < |w|$ and $|v_1| < |w|$. On the other hand, if $\tilde{\phi}_k(b_k w) = 0$ for some k, then taking $u_1 = b_k^2 w/c$ and $v_1 = (1 - b_k/c) w$, with c chosen such that $|c| > \max(1, b_k^2)$ and $b_k/c > 0$, we arrive at the same contradiction. ∎

Lemma 1 implies that for $j = 1, \ldots, n$ the second characteristic function $\tilde{\psi}_j(t) = \log \tilde{\phi}_j(t)$ is uniquely defined as a real-valued function for $-\infty < t < \infty$. The following lemma guarantees that we can differentiate $\tilde{\psi}_j(t)$ any number of times (see, e.g., Feller [13], XV.4, Lemma 2).

LEMMA 2. *For* $j = 1, \ldots, n$, $\boldsymbol{E}|\tilde{X}_j|^r < \infty$ *for any* $r \geqslant 1$.

Proof (after Lancaster [22]). Let $\alpha = \min_i |b_i|$ and $\beta = \max_i |b_i|$. Take $0 < \varepsilon < 1$ and choose A so that

$$\boldsymbol{P}(|X_i| > A) < \varepsilon \qquad \text{for } i = 1, 2, \ldots, n.$$

Put $\gamma = (2n-1)\beta/\alpha \geqslant 2n-1$ ($\geqslant 3$ since $n \geqslant 2$). Then

$$\begin{aligned}(1-\varepsilon)^{n-1} \boldsymbol{P}(|X_j| > \gamma A) &\leqslant \boldsymbol{P}(|X_j| > \gamma A, |X_i| \leqslant A \text{ for all } i \neq j) \\ &\leqslant \boldsymbol{P}(|Y_1| \geqslant nA, |Y_2| \geqslant nA\beta),\end{aligned} \tag{2.4}$$

since $|X_j| \leqslant |Y_1| + \sum_{i \neq j} |X_i|$ gives

$$|Y_1| \geqslant |X_j| - \sum_{i \neq j} |X_i| > \gamma A - (n-1)A \geqslant nA,$$

and $|b_j X_j| \leqslant |Y_2| + \sum_{i \neq j} |b_i X_i|$ gives

$$|Y_2| \geqslant |b_j X_j| - \sum_{i \neq j} |b_i X_i| > \alpha\gamma A - \beta(n-1)A \geqslant \beta n A.$$

Now

$$\begin{aligned} P(|Y_1| > nA) &= P\Big(\Big|\sum_{j=1}^{n} X_j\Big| > nA\Big) \\ &\leqslant P\Big(\sum_{j=1}^{n} |X_j| > nA\Big) \\ &\leqslant P\Big(\bigcup_{j=1}^{n} \{|X_j| > A\}\Big) < n\varepsilon \end{aligned} \tag{2.5}$$

by Boole's inequality, and

$$\begin{aligned} P(|Y_2| > nA\beta) &= P\Big(\Big|\sum_{j=1}^{n} b_j X_j\Big| > nA\beta\Big) \\ &\leqslant P\Big(\sum_{j=1}^{n} |b_j X_j| > nA\beta\Big) \\ &\leqslant P\Big(\bigcup_{j=1}^{n} \{|b_j X_j| > A\beta\}\Big) \leqslant P\Big(\bigcup_{j=1}^{n} \{|X_j| > A\}\Big) < n\varepsilon. \end{aligned} \tag{2.6}$$

It follows from (2.4)–(2.6) and the independence of Y_1 and Y_2 that, for $j = 1, 2, \ldots, n$,

$$P(|X_j| > \gamma A) < \frac{n^2 \varepsilon^2}{(1-\varepsilon)^{n-1}}. \tag{2.7}$$

Writing the right-hand side of (2.7) as ε' we have shown that, for γ as defined above,

$$P(|X_j| > A) < \varepsilon \text{ implies } P(|X_j| > \gamma A) < \varepsilon'.$$

If we take $\varepsilon < n^{-3}$, then it follows from Bernoulli's inequality ($(1+x)^a \geqslant 1 + ax$ for $x \geqslant -1$ and $a = 1, 2, \ldots$) that $n(1-\varepsilon)^{n-1} > 1$ for $n \geqslant 2$. Thus $\varepsilon' < n^3 \varepsilon^2 < n^{-3}$. Then, if we put $\varepsilon_0 = \varepsilon$, $\varepsilon_s = n^3 \varepsilon_{s-1}^2$, $s \geqslant 1$, we have proved that, with $\varepsilon_0 < n^{-3}$ and $k \geqslant 0$,

$$P(|X_j| > \gamma^k A) < \varepsilon_k = n^{-3}(n^3 \varepsilon_0)^{2^k} = c g^{2^k}$$

on putting $c = n^{-3}$ and $g = n^3 \varepsilon_0$, so $0 < g < 1$. Finally,

$$E\left(\frac{|X_j|^r}{A^r}\right) = \int_0^\infty P(|X_j|^r > A^r x)\, dx \leqslant \sum_{l=0}^{\infty} P(|X_j| > A l^{1/r})$$

$$\leqslant 1+\sum_{k=0}^{\infty}\sum_{\gamma^k\leqslant t^{1/r}<\gamma^{k+1}} P(|X_j|>\gamma^k A)$$

$$\leqslant 1+\sum_{k=0}^{\infty}\gamma^{r(k+1)}P(|X_j|>\gamma^k A)\leqslant 1+\sum_{k=0}^{\infty}\gamma^{r(k+1)}cg^{2^k}.$$

Since the ratio of the $(k+1)$-st to k-th terms in this sum is $\gamma^r g^{2^{k-1}}\to 0$, D'Alembert's test shows that the sum is finite. ∎

LEMMA 3. *For $j=1,\ldots,n$ there exists a polynomial $\tilde{P}_j(t)$ with real coefficients and of degree at most n, such that*

$$\tilde{\psi}_j(t)=\tilde{P}_j(t),\qquad -\infty<t<\infty.$$

Proof (ideas similar to Lukacs and King [29], pp. 391–392; see also Bryc [5], pp. 76–78). The equality of (2.1) and (2.2) gives

$$\sum_{j=1}^{n}\tilde{\psi}_j(u+b_jv)=\sum_{j=1}^{n}\tilde{\psi}_j(u)+\sum_{j=1}^{n}\tilde{\psi}_j(b_jv). \tag{2.8}$$

It follows from Lemma 2 that each $\tilde{\psi}_j$ has at least n derivatives. Differentiating (2.8) r times, $1\leqslant r\leqslant n$, with respect to v and setting $v=0$ gives

$$\sum_{j=1}^{n}b_j^r\tilde{\psi}_j^{(r)}(u)=\sum_{j=1}^{n}b_j^r\tilde{\psi}_j^{(r)}(0)=\sum_{j=1}^{n}b_j^r i^r\tilde{\kappa}_{jr}, \tag{2.9}$$

where $\tilde{\kappa}_{jr}=(-i)^r\tilde{\psi}_j^{(r)}(0)$ is the r-th cumulant of $\tilde{X}_j$ (Laha and Rohatgi [21], p. 223). If we integrate (2.9) with respect to u, we get

$$\sum_{j=1}^{n}b_j^r\tilde{\psi}_j^{(r-1)}(u)=\sum_{j=1}^{n}b_j^r i^r\tilde{\kappa}_{jr}u+\sum_{j=1}^{n}b_j^r\tilde{\psi}_j^{(r-1)}(0)$$
$$=\sum_{j=1}^{n}b_j^r i^r\tilde{\kappa}_{jr}u+\sum_{j=1}^{n}b_j^r i^{r-1}\tilde{\kappa}_{j,r-1}.$$

Integrating a further $r-1$ times with respect to u, at each stage using the identity $\tilde{\psi}_j^{(r)}(0)=i^r\tilde{\kappa}_{jr}$, we obtain

$$\sum_{j=1}^{n}b_j^r\tilde{\psi}_j(u)=\sum_{j=1}^{n}b_j^r\sum_{s=1}^{r}\tilde{\kappa}_{js}\frac{(iu)^s}{s!}. \tag{2.10}$$

If we denote the right-hand side of (2.10) by $d_r(u)$, it follows that $d_r(u)$ is a polynomial of degree r in u, with real coefficients on account of the present symmetric case with $\tilde{\psi}_j(u)=\tilde{\psi}_j(-u)$ in which $\tilde{\kappa}_{js}=0$ for odd integers s. Thus in the matrix form (2.10) becomes

$$\boldsymbol{B}\tilde{\boldsymbol{\psi}}(u)=\boldsymbol{d}(u), \tag{2.11}$$

where $\tilde{\boldsymbol{\psi}}(u)=(\tilde{\psi}_1(u),\ldots,\tilde{\psi}_n(u))'$, $\boldsymbol{d}(u)=(d_1(u),\ldots,d_n(u))'$ and

$$\boldsymbol{B}=\begin{pmatrix} b_1 & b_2 & \ldots & b_n\\ b_1^2 & b_2^2 & \ldots & b_n^2\\ \cdot & \cdot & \cdot & \cdot\\ b_1^n & b_2^n & \ldots & b_n^n\end{pmatrix}.$$

Since the b_j's are all unequal, $\boldsymbol{B}$ must be non-singular, so it follows from (2.11) that

$$\tilde{\boldsymbol{\psi}}(u) = \boldsymbol{B}^{-1}\boldsymbol{d}(u). \quad \blacksquare \tag{2.12}$$

LEMMA 4. *For* $j = 1, \ldots, n$, $\tilde{X}_j \sim \mathcal{N}(0, 2\sigma_j^2)$.

Proof. Taking $r = 2$ in (2.10) we obtain

$$\sum_{j=1}^{n} b_j^2 \tilde{\psi}_j(t) = -ct^2, \tag{2.13}$$

where $c = \sum b_j^2 \sigma_j^2$, σ_j^2 being the variance of X_j. It follows from (2.13) and Lemma 1 that for each j

$$0 \leqslant -\tilde{\psi}_j(t) \leqslant ct^2/b_j^2, \qquad -\infty < t < \infty. \tag{2.14}$$

In order that (2.14) be consistent with Lemma 3, it is necessary that the degree of the polynomials $\tilde{P}_j(t)$ be at most 2 and normality of the $\tilde{X}_j$'s follows. ■

The normality of the X_j's themselves could now be deduced from Cramér's Theorem, as is done at this point by Kac [17] and Skitovich [37]. Of course, if it were known that the X_j's had symmetric distributions, then the arguments of Section 2.1 could by applied directly to the X_i's themselves. We now show how to establish the normality of the X_j's themselves from Lemma 4, without direct use of Cramér's Theorem.

2.2. Laplace transforms. Lemma 2 is clearly true in terms of the original X_j's, and since $X_j - X_j' \sim \mathcal{N}(0, 2\sigma_j^2)$ from Lemma 4, where X_j and X_j' are independently and identically distributed with characteristic function ϕ_j satisfying $\phi_j(t)\phi_j(-t) = \exp(-\sigma_j^2 t^2)$, it follows that $\phi_j(t) \neq 0$ for any real t, and $\phi_j(t)$ has at least n derivatives. We put $\psi_j(t) = \log\phi_j(t)$, where log refers to the principal branch (since $\phi_j(t)$ may be complex valued even though t is real), so $\psi_j(0) = 0$. The following lemma implies that the X_j's have at most n non-zero cumulants:

LEMMA 5. *For* $j = 1, 2, \ldots, n$ *there exists a polynomial* $P_j(t)$ *of degree at most* n, *such that*

$$\psi_j(t) = P_j(t), \qquad -\infty < t < \infty,$$

where $P_j^{(r)}(0) = \kappa_{jr}$, *the* r-*th cumulant of* X_j.

Proof. We need only mimic the proof of Lemma 3, replacing $\tilde{\psi}_j(t)$ by $\psi_j(t)$, with minor adjustments for non-symmetry. ■

The remainder of our derivation is in terms of the Laplace transform

$$\lambda_j(v) = \boldsymbol{E}\big(\exp(-vX_j)\big), \qquad -\infty < v < \infty,$$

which the next lemma shows is finite.

LEMMA 6. *For* $j = 1, \ldots, n$,

$$0 < \lambda_j(v) < \infty, \qquad -\infty < v < \infty.$$

Proof. According to Lemma 4, $X_j - X'_j \sim \mathcal{N}(0, 2\sigma_j^2)$. Clearly, we can assume without loss of generality that X_j has zero median, that is, $\boldsymbol{P}(X_j < 0) \leqslant \frac{1}{2} \leqslant \boldsymbol{P}(X_j \leqslant 0)$. Then the distribution function F_j of X_j satisfies

$$\begin{aligned} F_j(x) &= \boldsymbol{P}(X_j \leqslant x) = \boldsymbol{P}(X_j \leqslant x,\, X'_j < 0) + \boldsymbol{P}(X_j \leqslant x,\, X'_j \geqslant 0) \\ &\leqslant \boldsymbol{P}(X_j \leqslant x)\boldsymbol{P}(X'_j < 0) + \boldsymbol{P}(X_j - X'_j \leqslant x,\, X'_j \geqslant 0) \\ &\leqslant \tfrac{1}{2}\boldsymbol{P}(X_j \leqslant x) + \boldsymbol{P}(X_j - X'_j \leqslant x). \end{aligned}$$

Writing

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\, du,$$

we obtain

$$F_j(x) \leqslant 2\boldsymbol{P}(X_j - X'_j \leqslant x) = 2\Phi\left(\frac{x}{\sigma_j\sqrt{2}}\right) = O\big(\exp(-x^2/(4\sigma_j^2))\big) \quad \text{as } x \to -\infty.$$

As $x \to \infty$, $1 - F_j(x)$ is similarly bounded. This means we can integrate by parts in

$$1 + v \int_{-\infty}^{0} e^{-vx} F_j(x)\, dx - v \int_{0}^{\infty} e^{-vx} \big(1 - F_j(x)\big)\, dx$$

to get

$$-\infty < \lambda_j(v) = \int_{-\infty}^{0} e^{-vx}\, dF_j(x) + \int_{0}^{\infty} e^{-vx}\, d\big(F_j(x) - 1\big) < \infty. \ \blacksquare$$

It is readily seen that $\lambda_j(v)$ has continuous derivatives of all orders $r \geqslant 1$, with

$$d^r \lambda_j(v)/dv^r = (-1)^r \int_{-\infty}^{\infty} x^r e^{-vx}\, dF_j(x).$$

By Lemma 5, the cumulant generating function $\mathscr{L}_j(v) = \log \lambda_j(v)$ exists for all v since $\lambda_j(v) \neq 0$, and thus has continuous derivatives of all orders. It is clear that $\mathscr{L}_j^{(r)}(0) = (-1)^r \kappa_{jr}$, where κ_{jr} is the r-th cumulant of X_j.

We are now in a position to prove Theorem 1. Recall for the sequel that $\kappa_{j1} = \boldsymbol{E}X_j$ and $\kappa_{j2} = \operatorname{Var} X_j = \sigma_j^2$. It follows from Lemma 5 and the mean value theorem of order $n+1$ that

$$\mathscr{L}_j(v) = \sum_{r=1}^{n} \frac{\kappa_{jr}}{r!}(-v)^r. \tag{2.15}$$

The fact that

$$\mathscr{L}_j''(v) \geqslant 0 \tag{2.16}$$

follows for instance by noting that $\mathscr{L}_j''(v)$ is the variance of the conjugate distribution

$$dG_j(x) = \frac{e^{-vx}\, dF_j(x)}{\int_{-\infty}^{\infty} e^{-vx}\, dF_j(x)}.$$

(This is suggested by an argument of Dugué [11], p. 56; see also Linnik [26], p. 62.)

Lemma 4 implies

$$\mathscr{L}_j(v)+\mathscr{L}_j(-v) = -\sigma_j^2 v^2$$

from which it follows by taking derivatives at $v = 0$ that all even cumulants higher than the second are zero, so that (2.15) reduces to

$$\mathscr{L}_j(v) = \frac{\kappa_{j2}}{2!}v^2 - \sum_{m=0}^{[(n-1)/2]} \frac{\kappa_{j,2m+1}}{(2m+1)!}v^{2m+1}.$$

From this and (2.16) we obtain for $n \geqslant 3$

$$-\kappa_{j2} \leqslant \mathscr{L}_j''(v)-\kappa_{j2} = -\sum_{m=1}^{[(n-1)/2]} \frac{\kappa_{j,2m+1}}{(2m-1)!}v^{2m-1}. \tag{2.17}$$

The right-hand side of (2.17) is an odd function of v, and hence will be large and negative, for either large positive v or large negative v, if $\kappa_{j,2m+1}$ is non-zero for any $m = 1, \ldots, [(n-1)/2]$. But this would contradict the lower bound in (2.17). It follows from Lemma 5 that

$$\psi_j(t) = i\kappa_{j1}t-\frac{\kappa_{j2}}{2}t^2, \quad \text{that is,} \quad X_j \sim \mathscr{N}(\kappa_{j1}, \kappa_{j2}). \ \blacksquare$$

3. ON FORMS OF BERNSTEIN'S THEOREM

In conclusion we indicate a simple direct proof of

THEOREM 2. *Let X_1 and X_2 be non-degenerate and independently distributed random variables and suppose that*

$$Y_1 = pX_1+qX_2 \quad \textit{and} \quad Y_2 = aX_1-bX_2$$

are independently distributed, where p, q, a and b are all real and non-zero. Then X_1 and X_2 are each normally distributed.

The reader will recognize this as the case $n = 2$ of Theorem 1. The case $p = q = a = b = 1$ is known as the celebrated Bernstein's Theorem (after Bernstein [3], who assumed also that X_1 and X_2 had finite, equal variances and positive densities). Bernstein's Theorem was generalized by Gnedenko [16], who proved Theorem 2 in full generality, taking (without loss of generality) $p = q = a = 1$, $b \neq 0, -1$. For a modern proof, see Quine [34], Theorem 1. Our proof, in which passage to logarithms is unnecessary, borrows a little from this, but shows that the Bernstein case is rather special and requires extended treatment. However, such treatment is shown to have already been available, in elegant and simple real variable terms, in Kac [17].

Kac's paper precedes even Bernstein's. From its received date, shortly after his arrival just before World War 2 at Johns Hopkins University on a Polish-Jewish (Parnas Foundation) Fellowship to the U.S., the paper was written by Kac largely in Lwów (then in Poland, now L'viv, in Ukraine; Russian name: L'vov); see Kac [18]. It is possibly due to ongoing disruptions in scientific communications caused by the war, and partly due to its own apparent restrictiveness, that Kac's paper has not received its due within the very large literature emanating from Bernstein's Theorem. There is no mention of it in the Russian papers, or the French sources (Fréchet [14]; Darmois [7]–[10]; Dugué [11], [12]) which deal with the topic in terms of characteristic functions.

Outline of proof for Theorem 2. Using our Lemma 2 (which does not require symmetry of the X_j's) we obtain $E|X_j|^r < \infty$, $r \geqslant 1$. Then, assuming without loss of generality $EX_j = 0$, we obtain, as in Lemma 2 of Quine [34], the equality

$$a\phi_1'(sp)\phi_2(sq) - b\phi_1(sp)\phi_2'(sq) = 0. \tag{3.1}$$

Further, since Y_1 and Y_2 are independent, $0 = \operatorname{Cov}(Y_1, Y_2) = pa\sigma_1^2 - bq\sigma_2^2$, where $\sigma_j^2 = \operatorname{Var} X_j > 0$, and since

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} p & q \\ a & -b \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \tag{3.2}$$

$\tau = pb + aq \neq 0$ (otherwise Y_1 would be a multiple of Y_2). Inverting the matrix in (3.2) gives

$$\tau X_1 = bY_1 + qY_2 \quad \text{and} \quad \tau X_2 = aY_1 - pY_2.$$

Taking characteristic functions, we obtain

$$\begin{aligned} \phi_1(\tau s) &= \phi_1(pbs)\phi_2(qbs)\phi_1(qas)\phi_2(-qbs), \\ \phi_2(\tau s) &= \phi_1(pas)\phi_2(qas)\phi_1(-aps)\phi_2(pbs). \end{aligned} \tag{3.3}$$

Without loss of generality, let us put $p = q = a = 1$, so $\tau = b + 1$, where $b = \sigma_1^2/\sigma_2^2 > 0$. Hence from (3.3) we get

$$\begin{aligned} \phi_1((1+b)s) &= \phi_1(bs)\phi_1(s)\phi_2(bs)\phi_2(-bs), \\ \phi_2((1+b)s) &= \phi_2(bs)\phi_2(s)\phi_1(s)\phi_1(-s). \end{aligned} \tag{3.4}$$

The continuity of $\phi_j(s)$ together with $\phi_j(0) = 1$ implies the existence of $\varepsilon > 0$ such that $|\phi_j(s)| > 0$ for $-\varepsilon < s < \varepsilon$, $j = 1, 2$. Hence from (3.4) we obtain $\phi_j(s) \neq 0$ for any s, $-\infty < s < \infty$, $j = 1, 2$.

Returning to the general formulation, from (3.1) we infer that

$$\frac{d}{ds}\left(\frac{\phi_1^{a/p}(sp)}{\phi_2^{b/q}(sq)}\right) = 0,$$

which leads to

$$\phi_1(t) = \phi_2^{bp/(aq)}(tq/p), \qquad -\infty < t < \infty. \tag{3.5}$$

If we now write

$$Y_2 = (-b)X_2 + aX_1 \quad \text{and} \quad Y_1 = qX_2 - (-p)X_1$$

and apply (3.5) *mutatis mutandis*, we obtain

$$\phi_2(t) = \phi_1^{bp/(aq)}(-ta/b)$$

from which and (3.5), putting $\gamma = bp/(aq)$, we get

$$\phi_2(t) = \phi_2^{\gamma^2}(-t/\gamma). \tag{3.6}$$

Thus, if $\gamma^2 > 1$, we have

$$\phi_2(t) = \phi_2^{\gamma^{2n}}(t/(-\gamma)^n) = \left(1 + \frac{i^2 t^2 \sigma_2^2}{2\gamma^{2n}} + \dots\right)^{\gamma^{2n}} \to \exp\left(\frac{-\sigma_2^2 t^2}{2}\right),$$

which is the characteristic function of $\mathcal{N}(0, \sigma_2^2)$. If $0 < \gamma^2 < 1$, put $\delta = 1/\gamma$ in (3.6) to obtain

$$\phi_2^{\delta^2}(-t/\delta) = \phi_2(t)$$

and proceed as for $\gamma^2 > 1$.

When $\gamma^2 = 1$, the case $\gamma = -1$ has already been dismissed since it corresponds to $\tau = pb + aq = 0$. The case $\gamma = 1$ corresponds to Bernstein's formulation, and (3.6) (and the analogous equation for $\phi_1(t)$) gives

$$\phi_1(t) = \phi_1(-t) \quad \text{and} \quad \phi_2(t) = \phi_2(-t),$$

that is, the distributions of X_1 and X_2 are symmetric about 0, with *real* characteristic functions $\phi_1(t)$ and $\phi_2(t)$. Now, Kac [17] initially assumes that X_1 and X_2 are independent and symmetrically distributed about 0, and that

$$Y_1 = (\cos\beta)X_1 + (\sin\beta)X_2 \quad \text{and} \quad Y_2 = (\sin\beta)X_1 - (\cos\beta)X_2$$

are independent for *every* β, and deduces that X_1 and X_2 are identically normally distributed, as was to be, later, Bernstein's conclusion. In fact, his proof uses the independence assumption only at $\beta = \pi/4$ and $\beta = 3\pi/4$ to show that X_1 and X_2 have the same (real) characteristic function ϕ which satisfies

$$\phi(2\xi) = \phi^4(\xi), \qquad -\infty < \xi < \infty.$$

Using the continuity of ϕ, $|\phi(\xi)| \leqslant 1$ and $\phi(0) = 1$, and the Cauchy method used to deal with the familiar functional equation $\psi(2x) = \psi^2(x)$, $-\infty < x < \infty$, Kac deduces $0 < \phi(\xi) \leqslant 1$, and then $\phi(\xi) = \exp(k\xi^2)$ for some $k < 0$.

We remark that the cases $\beta = \pi/4$ and $\beta = 3\pi/4$ are not in fact different, since both assert the independence of $X_1 + X_2$ and $X_1 - X_2$, and hence to treat our special case $\gamma = 1$, one may 'tap in' directly to Kac's brief argument, condensing it a little more. ∎

Another early paper (Lukacs [27]) also relates to Bernstein's Theorem, although it is concerned with the characterization of the normal distribution function from the independence of the sample mean $\bar{X} = \sum_{i=1}^n X_i/n$, and sample variance $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, where X_i, $i = 1, 2, \dots, n$, are independently and identically distributed with finite variance. This characterization was established under more stringent moment conditions by Geary [15]. Quine [34] showed that the present Lemma 2 can be combined with Lukacs' approach to prove the characterization with no moment assumptions whatsoever. In the case $n = 2$, if we write as with Bernstein, $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$, we see however that $\bar{X} = Y_1/2$ and $S^2 = Y_2^2/2$, so that in this case the characterization amounts to Bernstein's Theorem under the restrictive initial condition that X_1 and X_2 are identically distributed.

REFERENCES

[1] D. Basu, *On the independence of linear functions of independent chance variables*, Bull. Inst. Internat. Statist. 23, Tome 2 (1951), pp. 83–96.

[2] — and R. G. Laha, *On some characterizations of the normal distribution*, Sankhya 13 (1954), pp. 359–362. (See also *Addenda*, ibidem 14 (1954), p. 180.)

[3] S. N. Bernstein, *On a property characterizing the law of Gauss* (in Russian), Trudy Leningradsk. Polytekhn. Inst. 3 (1941), pp. 21–22. (See also Bernstein [4], pp. 394–395.)

[4] — *Sobranie Sochinenii (Collected Works). IV Teoriia Veroiatnostei i Matematicheskaia Statistika [1911–1946]*, Nauka, Moscow 1964.

[5] W. Bryc, *The Normal Distribution: Characterizations with Applications*, Lecture Notes in Statist. 100, Springer, New York 1995.

[6] H. Cramér, *Über eine Eigenschaft der normalen Verteilungsfunktion*, Math. Z. 41 (1936), pp. 405–414.

[7] G. Darmois, *Analyse des liaisons de probabilité*, in: *Proceedings Int. Statist. Conference 1947*, Vol. IIIA, Washington, D.C., 1951, p. 231.

[8] — *Sur une propriété caractéristique de la loi de probabilité de Laplace*, C. R. Acad. Sci. Paris 232 (1951), pp. 1999–2000.

[9] — *Sur diverses propriétés caractéristiques de la loi de probabilité de Laplace–Gauss*, Bull. Inst. Internat. Statist. 33, Tome 2 (1953), pp. 79–82.

[10] — *Analyse générale des liaisons stochastiques*, Rev. Inst. Internat. Statist. 21 (1953), pp. 2–8.

[11] D. Dugué, *Analyticité et convexité des fonctions caractéristiques*, Ann. Inst. H. Poincaré 12 (1951), pp. 45–56.

[12] — *Arithmétique des lois des probabilités*, Mém. Sci. Math., Gauthier-Villars, Paris 1957.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition, Wiley, New York 1971.

[14] M. Fréchet, *Généralisation de la loi de probabilité de Laplace*, Ann. Inst. H. Poincaré 13 (1951), pp. 1–29.

[15] R. C. Geary, *Distribution of Student's ratio for non-normal samples*, J. Roy. Statist. Soc. Supp. 3 (1936), pp. 178–184.

[16] B. V. Gnedenko, *On a theorem of S. N. Bernstein* (in Russian), Izv. Akad. Nauk SSSR Ser. Mat. 12 (1948), pp. 97–100.

[17] M. Kac, *On a characterization of the normal distribution*, Amer. J. Math. 61 (1939), pp. 726–728.

[18] — *Enigmas of Chance. An Autobiography*, Harper and Row, New York 1985.

[19] A. M. Kagan, *The Lukacs-King method applied to problems involving linear forms of independent random variables*, Metron 46 (1988), pp. 5–19.

[20] — Yu V. Linnik and C. R. Rao, *Characterization Problems in Mathematical Statistics*, Wiley, New York 1973.

[21] R. G. Laha and V. K. Rohatgi, *Probability Theory*, Wiley, New York 1979.
[22] H. O. Lancaster, *The characterization of the normal distribution*, J. Austral. Math. Soc. 1 (1960), pp. 368–383.
[23] P. Lévy, *Théorie de l'addition des variables aléatoires* (2nd edition 1954), Gauthier-Villars, Paris 1937.
[24] Yu. V. Linnik, *Remarks concerning the classical derivation of Maxwell's law* (in Russian), Dokl. Akad. Nauk SSSR 85 (1952), pp. 1251–1254.
[25] — *A remark on Cramér's Theorem on the decomposition of the normal law*, Theory Probab. Appl. 1 (1956), pp. 435–436.
[26] — *Razlozhenia veroiatnostnikh zakonov* (*Decompositions of Probability Laws*), Izd. Leningrad. Univ., Leningrad 1960.
[27] E. Lukacs, *A characterization of the normal distribution*, Ann. Math. Statist. 13 (1942), pp. 91–93.
[28] — *Characteristic Functions*, 2nd edition, Griffin, London 1970.
[29] — and E. P. King, *A property of the normal distribution*, Ann. Math. Statist. 25 (1954), pp. 389–394.
[30] E. Lukacs and R. G. Laha, *Applications of Characteristic Functions*, Hafner, New York 1964. (Reprinted as Griffin's Statistical Monograph No. 14, Griffin, London.)
[31] J. Marcinkiewicz, *Sur une propriété de la loi de Gauss*, Math. Z. 44 (1938), pp. 612–618.
[32] S. Mazurkiewicz, *Un théorème sur les fonctions caractéristiques*, Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Sér. A, Numéro Sommaire (1940–1946), pp. 1–3.
[33] P. A. P. Moran, *An Introduction to Probability Theory*, Clarendon, Oxford 1968.
[34] M. P. Quine, *On three characterizations of the normal distribution*, Probab. Math. Statist. 14 (1993), pp. 257–263.
[35] D. Raikov, *On the decomposition of Gauss and Poisson laws* (in Russian), Izv. Akad. Nauk SSSR Ser. Mat. 2 (1938), pp. 91–124.
[36] V. P. Skitovich, *On a property of the normal distribution* (in Russian), Dokl. Akad. Nauk SSSR 89 (1953), pp. 217–219.
[37] — *Linear combinations of independent random variables and the normal distribution law* (in Russian), Izv. Akad. Nauk SSSR Ser. Mat. 18 (1954), pp. 185–200. (Translated in: Selected Translations in Math. Statist. and Probab. 2 (1962), pp. 211–228.)
[38] G. B. Tranquilli, *Sul Teorema di Basu–Darmois*, Giornale dell'Istituto Italiano degli Attuari 29 (1966), pp. 135–152.
[39] A. A. Zinger, *On independence of polynomial and quasi-polynomial statistics* (in Russian), Dokl. Akad. Nauk SSSR 110 (1956), pp. 319–322.
[40] — *Independence of quasi-polynomial statistics and analytical properties of distributions* (in Russian), Teor. Veroyatnost. i Primenen. 3 (1958), pp. 265–284.

School of Mathematics and Statistics
University of Sydney
NSW 2006, Australia

Received on 12.1.1999