IDENTICALLY DISTRIBUTED LINEAR FORMS AND
THE NORMAL DISTRIBUTION

BY

S. G. GHURYE and I. OLKIN

TECHNICAL REPORT NO. 58
JULY 12, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27210

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
IDENTICALLY DISTRIBUTED LINEAR FORMS AND THE NORMAL DISTRIBUTION

by

S. G. Ghurye and I. Olkin

University of Alberta and Stanford University

TECHNICAL REPORT NO. 58

July 12, 1971

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT GP-27210

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
IDENTICALLY DISTRIBUTED LINEAR FORMS AND THE NORMAL DISTRIBUTION

by

S. G. Ghurye and I. Olkin

University of Alberta and Stanford University

Dedicated to George Pólya

0. Introduction.

This problem was first discussed by Pólya in 1923 when he showed that if $X$ and $Y$ are independently and identically distributed (i.i.d.) random variables with a finite second moment and a probability density function which is bounded and Riemann-integrable over every finite interval, and if there exist positive numbers $a, b, c$ such that the distribution of $aX + bY$ is the same as that of $cX$, then $X$ has a normal distribution. This is a remarkable paper, much ahead of its time, which was virtually completely ignored. Recently, however, the problem has received much attention, resulting in its extensions in various directions: the linear form in a pair of variables is replaced by one in a finite number of independent and identically distributed random variables, and next by an infinite series; the scalar random variable is replaced by a random vector; the scalar linear form in random vectors is replaced by a general linear form with matrix coefficients. The most significant extension of Pólya's result is the theorem proved by Linnik (1953) that if $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, and there exist constants $a_1, \ldots, a_n, b_1, \ldots, b_n$, satisfying certain conditions, such that $\Sigma a_i X_i$ and $\Sigma b_i X_i$ have the same distribution, then the common distribution of the $X_i$ is normal.
Linnik's proof depends heavily on the theory of functions of a complex variable, which is one reason why his result has not yet been extended to linear forms in random vectors with matrix coefficients.

In the present paper, we give a survey (with some improvements) of work done on these problems (Section 1), and an extension in the case of the multivariate problem (Section 2).
1. Survey of previous work.

To begin with, we give a chronological list of papers with an indication of their contents; this is followed by more detailed statements of results and outlines of their proofs. We have presented these in as neat a form as possible, correcting some errors that we found in the originals and contributing minor improvements. We use the notation $\mathbf{P}(x)$ to denote the probability law of the random variable (or vector, as the case may be) $\mathbf{X}$.

The next paper after Pólya's which discusses the problem of identically distributed linear forms is that of Marcinkiewicz (1939). This paper deals with a pair of identically distributed linear forms in infinitely many i.i.d. random variables. Whereas Marcinkiewicz assumes the existence of all moments of positive order, the paper by Linnik (1953) gives a complete treatment of the problem of a pair of identically distributed linear forms in a finite number of variables; it is shown here that certain assumptions concerning moments are vital. Laha and Lukacs (1965) study a problem that is intermediate between the problems of Pólya and Marcinkiewicz: it is assumed that there is a linear form, $\sum_{j=1}^{\infty} a_j X_j$, in i.i.d. random variables such that its distribution is the same as that of $X_1$; no assumptions concerning moments is made. Finally, Eaton (1966) considers the multivariate extension: it is assumed that a linear form (with matrix coefficients) in $n$ independent random vectors all having the same distribution, $F$, has the distribution $F$. 
Theorem 1.1. Pólya (1923). Let $X$ be a random variable with a finite second moment and a density function which is bounded and Riemann-integrable over every finite interval. If there exist positive numbers $a, b, c$ such that

$$(1.1) \quad \mathcal{L}(aX) \ast \mathcal{L}(bX) = \mathcal{L}(cX),$$

then $\mathcal{L}(X)$ is normal.

Proof. By equating the variances, we have $a^2 + b^2 = c^2$; setting $\alpha = a/c$ and $\beta = b/c$, we get the equation satisfied by the characteristic function

$$(1.2) \quad \Phi(t) = \Phi(\alpha t) \Phi(\beta t).$$

Iterating (1.2) yields

$$(1.3) \quad \Phi(t) = \prod_{j=1}^{m} \Phi(\gamma_{n,j} t),$$

where $m = 2^n$ and the $\gamma$'s are the $2^n$ individual terms of the type $\alpha^j \beta^{n-j}$ (with unit coefficient) obtained by expanding $(\alpha + \beta)^n$ as a product of $n$ factors. Note that $\sum_{j=1}^{m} \gamma_{n,j}^2 = (\alpha^2 + \beta^2)^n = 1$ and $\max_{1 \leq j \leq m} \gamma_{n,j} \leq (\max(\alpha, \beta))^n \to 0$ as $n \to \infty$. In a neighbourhood of the origin, $\Phi$ is bounded away from 0, and hence $\log \Phi(t) = u(t) + iv(t)$ exists. Since we may assume, without loss of generality, that $E[X] = 0,$
we have that

\[ u(0) = v(0) = 0, \quad u'(0) = v'(0) = 0, \]
\[ v''(0) = 0, \quad u''(0) = -\sigma^2 = \text{Var}(X). \]

From (1.3), we now obtain

\[ (1.4) \quad \log \Phi(t) = \sum_{j=1}^{m} \{ \text{u}(\gamma_{n,j}t) + iv(\gamma_{n,j}t) \} \]
\[ = \frac{1}{2}t^2 \sum_{n,j} \{ \text{u}''(\gamma_{n}\gamma_{j}t) + iv''(\gamma_{n}\gamma_{j}t) \}, \]

where \( 0 < \gamma_{n}, \gamma_{j} < 1 \). Consequently, as \( n \to \infty \), the \( u'' \)-terms converge to \(-\sigma^2\) and the \( v'' \)-terms converge to 0. Hence,

\[ (1.5) \quad \log \Phi(t) = -\frac{1}{2}\sigma^2 t^2. \]

The proof is then completed by showing that the normal density function is the only density function having this characteristic function.||

**Remark (i).** The above proof presents the essential steps which today do not need much explanation; a large part of Polya's proof is taken up with such explanations and justifications, because the techniques involved were, at that time, novel and unfamiliar. In fact, Polya's paper contains an appendix which discusses Lévy's Continuity Theorem for characteristic functions which had just been published (Lévy (1922)). Thus, it is clear
that Pólya's assumption of a bounded and Riemann-integrable density function was due to the fact that the uniqueness of the relation between characteristic functions and the corresponding probability laws was not well-known at the time. On the other hand, the assumption of the finiteness of the second moment could not have been eliminated until several years later, after the theory of infinitely divisible distributions was well-understood.

**Remain (ii).** Vincze (1962) discusses the functional equation (1.2) without restricting \( \Phi \) to be a characteristic function, assuming merely that it is a complex-valued function which is twice differentiable at the origin. He apparently was unaware of Pólya's paper and his proof is similar to that of Pólya. Laha, Lukacs and Renyi (1964) extend the equation (1.2) to \( \Phi(x) = \prod_{1}^{\infty} \Phi(a_jx) \).

**Theorem 1.2.** Laha and Lukacs (1965). Let \( \{X_j\} \) be a sequence of i.i.d., non-degenerate random variables, and let \( \{a_j\} \) be a finite or infinite sequence of real numbers at least two of which are different from 0; if an infinite number of the \( a_j \) are non-zero, it is assumed that the infinite series \( \sum_{j} a_j X_j \) converges stochastically. If there is a real number \( \alpha \neq 0 \) such that (1) \( \sum_{j} a_j^2 \geq \alpha^2 \) and (ii) \( \mathcal{L}(\sum_{j} a_j X_j) = \mathcal{L}(\alpha X) \), then \( \mathcal{L}(X_1) \) is normal.

**Proof.** There are three principal steps in the proof: (1) \(|a_j| < \alpha \) for all \( j \), (2) \( \mathcal{L}(X_1) \) is infinitely divisible, and (3) \( \mathcal{L}(X_1) \) is normal.
We now consider this program. Let $\Phi$ be the characteristic
function of $\alpha X_1$ and set $b_j = a_j/\alpha$. Then $\prod_1^\infty \Phi(b_j t)$ converges, uni-
formly in every finite $t$-interval, to a characteristic function; hence,
$\psi_n(t) = \prod_{n+1}^\infty \Phi(b_j t)$ converges to 1; i.e., to every $\varepsilon$, $\tau > 0$, there
 corresponds a $\nu$ such that

$$|\psi_n(t) - 1| < \varepsilon \text{ for all } |t| \leq \tau \text{ and } n \geq \nu.$$  

Now, $\mathcal{L}(\Sigma a_j X_j) = \mathcal{L}(\alpha X)$ is equivalent to

$$\Phi(t) = \prod_1^\infty \Phi(b_j t) = [\prod_1^n \Phi(b_j t)] \psi_n(t),$$

from which we may conclude that

$$|b_j| < 1 \text{ for all } j.$$  

To see this, suppose the contrary, namely, for some $j$, say $j=1$,
$|b_1| \geq 1$. If $|b_1| > 1$, then we obtain using (1.7), that

$$|\Phi(t)| \leq |\Phi(b_1 t)| \leq \ldots \leq |\Phi(b_1^n t)|.$$  

But this is equivalent to

$$|\Phi(t)| \geq |\Phi(b_1^{-1} t)| \to 1 \text{ as } n \to \infty.$$
Hence $|\Phi(t)| \equiv 1$, which is impossible since $X_\perp$ is non-degenerate. On the other hand, if $|b_\perp| = 1$, again using (1.7), we have

\[(1.10) \quad |\varphi(t)| = |\varphi(t)| |\psi_\perp(t)| .\]

Now $|\varphi(t)| > 0$ in a neighbourhood of 0, and in this neighbourhood, $|\psi_\perp(t)| = 1$. Since at least one of $b_j$, $j = 2, 3, \ldots$, is non-zero, this gives us that $|\Phi(t)| \equiv 1$, throughout a neighbourhood of 0, which is ruled out. This completes the proof of (1).

Next we show that (1.7), together with the conclusion that $|b_j| < 1$ for all $j$, implies that $\Phi$ is an infinitely divisible characteristic function. In other words, we wish to show that, given any $\varepsilon, \tau > 0$, we can represent $\Phi(t)$ as a product, $\prod_j \phi_j(t)$, of characteristic functions such that

\[(1.11) \quad |\phi_j(t)-1| < \varepsilon \text{ for } |t| \leq \tau \text{ and every } j .\]

In order to do this, choose $\nu$ to satisfy (1.6), and iterate (1.7) $k$ times with $n=\nu$. This yields

\[(1.12) \quad \Phi(t) = \left( \prod_{i=1}^{M} \Phi(\gamma_{k,i}, t) \right) \left( \prod_{i=1}^{N} \psi(\delta_{k,i}, t) \right), \quad M = \nu^k, \quad N = \frac{k-1}{\nu-1} \]

where each $\gamma_{k,i}$ is a product of $k$ of the $b_j$, $j=1, 2, \ldots, \nu$, and each $\delta_{k,i}$ is a product of $k$ or fewer $b$'s. Hence, $|\delta_{k,i}| < 1$ and $|\gamma_{k,j}| \leq \beta^k$, where $\beta = \max_{1 \leq j \leq \nu} |b_j| < 1$. Now, given $\varepsilon, \tau > 0$, we can
choose $k$ large enough so that $|\Phi(u)-1| < \varepsilon$ for $|u| < \beta^k \tau$. Consequently, the representation (1.12) satisfies the condition (1.11), which completes Part (2).

To show that $\mathcal{L}(X_1)$ is normal, it is enough to show that $|\Phi(t)| = \exp(-\frac{1}{2} \beta t^2)$, for some $\beta > 0$, since by Cramer's Theorem if the product $\Phi(t) \Phi(-t) = |\Phi(t)|^2$ is the characteristic function of a normal distribution, then the same is true of each of the two factors. Thus we may, without loss of generality, assume that, in (1.7), $\Phi$ is real-valued (and even) and the $b_j$ are all positive (and $< 1$). We now make use of the canonical representation of an infinitely divisible characteristic function:

\begin{equation}
\log \Phi(t) = -\frac{1}{2} \beta t^2 - \int_{x>0} (1-\cos tx) \frac{1+x^2}{x^2} \, dH(x), \tag{1.13}
\end{equation}

where $H(x)$ is non-decreasing and bounded, and $\beta \geq 0$. Let

\begin{equation}
r_n(t) = \log \prod_{j=1}^{n} \Phi(b_j t) = -\frac{1}{2} \beta (\Sigma b_j^2) t^2 - \sum_{j=1}^{n} \int_{x>0} (1-\cos tx) \frac{b_j^2 + x^2}{x^2} \, dH(x), \tag{1.14}
\end{equation}

where

\begin{equation*}
H_n(x) = \begin{cases} 
\sum_{j=1}^{n} \int_{0 \leq u \leq x} \frac{b_j^2 + u^2}{1 + u^2} \, dH(u), & x \geq 0, \\
0, & x < 0.
\end{cases}
\end{equation*}
In the event that the linear form contains only a finite number of non-vanishing terms, there is an \( n \) such that \( f_n = \log \Phi \); otherwise, \( f_n \to \log \Phi \). In the first case, \( H_n = H \) for some \( n \) (because of the uniqueness of the Lévy-Khinchin representation); in the second case, \( H_n \to H \) as \( n \to \infty \), at each continuity point of \( H \) (e.g., see Gnedenko and Kolmogorov, Section 19). In both cases, we have \( H_n \leq H \). However, \( 0 < b_j < 1 \) implies \( (b_j^2 + x^2)/(1 + x^2) > b_j^2 \) and \( H(x/b_j) > H(x) \), so that

\[
(1.15) \quad H(x) \geq \frac{\sum b_j^2}{\sum} H\left(\frac{x}{b_j}\right) \geq \left(\sum b_j^2\right) H(x) .
\]

Consequently, \( \sum b_j^2 \leq 1 \); this, together with the assumption that \( \sum b_j^2 \geq 1 \) implies \( \sum b_j^2 = 1 \). But now (1.15) becomes \( \sum b_j^2 H\left(\frac{x}{b_j}\right) = (\sum b_j^2) H(x) \), or equivalently, \( \sum b_j^2 H\left(\frac{x}{b_j}\right) - H(x) = 0 \). By the monotonicity of \( H \) it follows that \( H\left(\frac{x}{b_j}\right) = H(x) \) and hence \( H(x) \) is constant. Thus from (1.13), \( \log \Phi(t) = -\frac{1}{2} \beta t^2 \).

**Theorem 1.3. Marcinkiewicz (1939).** Let \( \{X_n\} \) be a sequence of i.i.d. random variables having moments of all positive orders, and \( \{a_n\}, \{b_n\} \) be sequences of constants such that one sequence is not a permutation of the other sequence, and let the infinite series \( \sum a_j X_j, \sum b_j X_j \) converge. Then \( L(\sum a_j X_j) = L(\sum b_j X_j) \) if and only if \( L(X_j) \) is normal.

**Remark.** This paper is actually better known for the other results contained in it concerning complex-valued functions of a real variable which cannot be characteristic functions.
We now consider the major paper by Linnik (1953). This paper is exceedingly intricate, long, and troublesome to read. There are also some inconsistencies -- which may be removed by reinterpretation. We have attempted to clarify some of these points in the following.

Let $X_1, \ldots, X_n$ be i.i.d. random variables, whose common distribution is $\mathcal{L}(X)$, and let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be constants. Consider the condition

\begin{equation}
\mathcal{L}(\Sigma a_i X_i) = \mathcal{L}(\Sigma b_i X_i)
\end{equation}

An important step in the discussion is to transform the problem to a symmetric version. Note that (1.16) implies $\mathcal{L}(\Sigma a_i (-X_i)) = \mathcal{L}(\Sigma b_i (-X_i))$.

Furthermore, if $X_i$ and $X'_i$ are i.i.d. with common distribution $G$, then

$$\mathcal{L}(X_i - X'_i) = G(x) \ast [1 - G(-x)] = F(x),$$

so that the c.f. corresponding to $F$ is $f(t) = |g(t)|^2$, which is an even function. As a consequence we may assume that all the coefficients $a_i$ and $b_i$ are non-negative, and use as our starting point

\begin{equation}
f(a_1 t) \ldots f(a_n t) = f(b_1 t) \ldots f(b_n t)
\end{equation}

We assert that $f(u) \neq 0$ on $R_+$ except in the trivial case where $\{a_1, \ldots, a_n\}$ is a permutation of $\{b_1, \ldots, b_n\}$, (in which case every c.f. is a solution). To see this, let $a_1 = \max(a_1, \ldots, a_n, b_1, \ldots, b_n)$ and
suppose that for some $u_0$, $f(u_0) = 0$. In (1.17) let $t = u_0/a_i$, so that

$$0 = \prod_{l=1}^{n} f(\frac{b_i}{a_i} u_0).$$

Consequently, at least one factor on the right must vanish. By iteration we see that $f$ has a sequence of zeroes which converges to 0. But this is a contradiction since $f(0) = 1$.

Thus, if any of the $a_i$ is equal to a $b_j$, the common terms may be removed from (1.17). Consequently, in the canonical form we may assume that none of the $a_i$ is equal to any of the $b_j$. Letting $\alpha_i = \log a_i$ if $a_i > 0$ and $\alpha_i = 0$ if $a_i = 0$, $\beta_i = \log b_i$ if $b_i > 0$ and $\beta_i = 0$ if $b_i = 0$, $u = \exp t$, $\varphi(t) = \log f(u) = \log f(\exp t)$, (1.17) becomes

$$(1.18) \quad \varphi(t+\alpha_1) + \ldots + \varphi(t+\alpha_n) = \varphi(t+\beta_1) + \ldots + \varphi(t+\beta_n).$$

The essence of Linnik's paper is to obtain conditions which guarantee that the only solution of (1.18) is the characteristic function of a normal distribution.

By a direct verification it can be shown that a particular solution of (1.18) is given by

$$e^{\lambda t} P_{\lambda}(t) + e^{\overline{\lambda} t} \overline{P_{\lambda}(t)},$$

where $P_{\lambda}(t) = c_0 + c_1 t + \ldots + c_{\nu-1} t^{\nu-1}$, and where $\lambda$ is a root of multiplicity $\nu$ of the function.
\[ (1.19) \quad \sigma(z) = e^{\alpha_1 z} + \ldots + e^{\alpha_n z} - e^{\beta_1 z} - \ldots - e^{\beta_n z}. \]

It is this entire function of a complex variable which plays a critical role in the solution of the problem. However, the statement of the theorem becomes simpler in terms of

\[ (1.20) \quad \tau(x) = (a_1^2)^x + \ldots + (a_n^2)^x - (b_1^2)^x - \ldots - (b_n^2)^x. \]

**Theorem 1.4 Linnik (1953).** In order for \( \mathcal{I}(\Sigma a_i X_i) = \mathcal{I}(\Sigma b_i X_i) \) to imply \( \mathcal{I}(\mathbf{X}) = \mathcal{N}(\mu, \sigma^2) \) with some \( \mu \) and \( \sigma^2 \), it is necessary and sufficient that the positive zeroes of \( \tau(x) \) satisfy the following conditions:

(i) All but the largest zero are integers and are simple zeroes.

(ii) For the largest zero, \( \xi \), either (1) \( \xi \) is a simple zero and \( \lfloor \xi \rfloor \) (the integral part of \( \xi \)) is odd, or (2) \( \xi \) is an even integer and is either a simple or a double zero.

If (i) and (ii) are satisfied, then \( \mu = 0 \) unless \( \Sigma a_1 = \Sigma b_1 \), and \( \sigma^2 = 0 \) unless \( \Sigma a_1^2 = \Sigma b_1^2 \).

Although the conditions for normality have been provided in terms of conditions on the zeroes of \( \tau(x) \), it quickly becomes apparent that given a set of numbers \( a_1, \ldots, a_n, b_1, \ldots, b_n \), it is, in general, quite troublesome to check whether (i) and (ii) are satisfied. Consequently, it would be advantageous to have some sufficient conditions which guarantee normality. One such set of conditions is provided by the following:

**Lemma 1.5.** If \( a_1^2 \geq \ldots \geq a_n^2, b_1^2 \geq \ldots \geq b_n^2 \), and
(1.21) \[ \sum_{i=1}^{k} a_i^2 \geq \sum_{i=1}^{k} b_i^2, \quad k = 1, \ldots, n-1, \quad \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} b_i^2, \]

with strict inequality for at least one value of \( k \), then the only positive zero of \( \tau(x) \) given by (1.20) is at 1, and it is a simple zero.

Proof. It is well-known that (1.21) holds iff \( \Sigma h(a_i^2) \geq \Sigma h(b_i^2) \) for all convex functions \( h \). Furthermore, because \( a_1^2, \ldots, a_n^2 \) is not a permutation of \( b_1^2, \ldots, b_n^2 \), and \( h(u) = u^m \) is a strictly convex function for \( m > 1 \), and a strictly concave function for \( 0 < m < 1 \), we see that \( \tau(x) > 0 \) for \( x > 1 \) and \( \tau(x) < 0 \) for \( 0 < x < 1 \). Clearly \( \tau(1) = 0 \). The simplicity of the root follows from the fact that \( u \log u \) is strictly convex, so that \( \Sigma a_i^2 \log a_i^2 - \Sigma b_i^2 \log b_i^2 > 0 \).

Applications. Recall that \( \tau(x) = \Sigma(a_i^2)^x - \Sigma(b_i^2)^x \).

1. When \( n=2 \), we must have \( a_1^2 > b_1^2 > b_2^2 > a_2^2 \) if \( a_1^2 + a_2^2 = b_1^2 + b_2^2 \), and hence \( \mathcal{Z}(a_1 X_1 + a_2 X_2) = \mathcal{Z}(b_1 X_1 + b_2 X_2) \) implies \( \mathcal{Z}(X) = N(\mu, \sigma^2) \).

2. Let \( n=3 \), and \( a_1^2 + a_2^2 + a_3^2 = b_1^2 + b_2^2 + b_3^2 \). If \( a_1^2 \) and \( a_2^2 \) are the largest and the smallest of \( a_1^2, a_2^2, a_3^2, b_1^2, b_2^2, b_3^2 \), then (1.21) holds. Hence \( \mathcal{Z}(a_1 X_1 + a_2 X_2 + a_3 X_3) = \mathcal{Z}(b_1 X_1 + b_2 X_2 + b_3 X_3) \) implies \( \mathcal{Z}(X) = N(\mu, \sigma^2) \).

3. For any \( n \), suppose \( a_1^2 + \ldots + a_n^2 = n, b_1 = b_2 = \ldots = b_n = 1 \), and the \( a \)'s are not all equal. Then the conditions (1.21) are satisfied, so that \( \mathcal{Z}(\Sigma a_i X_i) = \mathcal{Z}(\Sigma X_i) \) implies \( \mathcal{Z}(X) = N(\mu, \sigma^2) \).

4. Suppose \( a_1^2 = a_2 = \ldots = a_n = 0 \), then for any \( b_1 \geq \ldots \geq b_n > 0 \) with \( \Sigma b_i^2 = 1 \), (1.21) is satisfied. Thus \( \mathcal{Z}(\Sigma b_i X_i) = \mathcal{Z}(X) \) implies \( \mathcal{Z}(X) = N(\mu, \sigma^2) \).
Remark. On the other hand, it is not difficult to construct large classes of pairs of linear forms which are identically distributed without the common distribution of the individual terms being normal.

Example 1. If \( a_1 > 0, b_1 > 0, i = 1, \ldots, n \) and \( \Sigma a_i = \Sigma b_i, \Sigma a_i^2 = \Sigma b_i^2 \), then \( \mathcal{L}(\Sigma a_i X_i) = \mathcal{L}(\Sigma b_i X_i) \) if \( \mathcal{L}(X) \) is a convolution of a Cauchy and a normal law.

Example 2. If \( n = 3, (a_1^2, a_2^2, a_3^2) = (2c+4, 2, 2), \) and \( (b_1^2, b_2^2, b_3^2) = (c, c, 0) \), where \( 4 < c < 3 + \sqrt{5} \), then \( \tau(2) < \tau(1) = 0 < \tau(3) \), so that \( \tau(x) \) has a zero between 2 and 3; condition (ii) of Theorem 1.4 is not satisfied.
2. A multivariate extension.

An extension to the multivariate case of the original result of Pólya is provided by Eaton (1966):

**Theorem 2.1 Eaton (1966).** Let \( X_1, \ldots, X_n \) be i.i.d. random \( p \)-dimensional row vectors, and let there exist constant, non-singular, symmetric matrices \( A_1, \ldots, A_n \) and a constant row vector \( b \), such that \( \mathcal{Z}(X) = \mathcal{Z}(\Sigma X, A_j + b) \).

If \(-I < A_j < I, j = 1, \ldots, n\), then \( \mathcal{Z}(X) \) is infinitely divisible, whereby \( U > V \) we mean that \( U - V \) is positive definite. In addition

(i) if \( \Sigma A_j^2 > I \), then \( X \) is a constant w.p.l;

(ii) if \( \Sigma A_j^2 = I \) and \( \det(\Sigma A_j^2 - I) = 0 \), then \( \mathcal{Z}(X) \) is normal and its covariance matrix, \( \Gamma \), satisfies the equation \( \Gamma = \Sigma A_j \Gamma A_j^t \).

In the present section we provide a more general result by eliminating the condition that the matrices be symmetric, and also deal with an infinite linear form. The proof is, in its essentials, based on that of Eaton with appropriate modifications.

Let \( \{X_n, n=1,2,\ldots\} \) be a sequence of i.i.d. random \( p \)-dimensional row-vectors, and \( \{A_n, n=1,2,\ldots\} \) a finite or infinite sequence of \( p \times p \) matrices of which at least two, say \( A_1 \) and \( A_2 \), are non-singular. Let \( \mathcal{Z}(X) = \mathcal{Z}(\Sigma X, A_j^t) \), where, if the number of non-vanishing \( A_j \) is infinite, it is assumed that the series converges (in probability, and hence with probability 1). The equation satisfied by the characteristic function is

\[
(2.1) \quad \Phi(t) = \prod_{j} \Phi(tA_j) = \left( \prod_{j=1}^{n} \Phi(tA_j) \right) \psi_n(t),
\]

where \( \psi_n(t) = \prod_{j=n+1}^{\infty} \Phi(tA_j) \).
Lemma 2.2. If one of the $A_j$ has a characteristic root $\lambda$ with $|\lambda| > 1$, then $\mathcal{L}(X)$ is linearly singular; i.e., there exists a non-zero vector $c$ such that $cX'$ is, with probability 1, a constant.

Proof. Denote by $A$ the matrix with the root $\lambda$. From (2.1), we have

$$(2.2) \quad \varphi(t) = \varphi(tA)\psi(t),$$

where $\psi(t)$ is the product of all the $\varphi(tA_j)$ with the exception of the one factor $\varphi(tA)$. By iteration of (2.2) we get, for every positive integer $k$,

$$(2.3) \quad \varphi(t) = \varphi(tA_k)\psi_k(t),$$

where $\psi_k(t) = \prod_{j=0}^{k-1} \varphi(tA_j)$. For the characteristic root $\lambda = \rho e^{i2\pi \theta}$, where $\rho$ is real, there exists a characteristic vector $a = \alpha + i\beta$ with $\alpha \neq 0$ such that $aA^k = \lambda^k a$, $k = 1, 2, \ldots$. Consequently, we have that $\alpha A^k = \rho^k (\alpha \cos 2\pi k \theta - \beta \sin 2\pi k \theta)$, $k = 1, 2, \ldots$. Hence, for all real $s$,

$$(2.4) \quad \varphi(s\rho^{-k}\alpha) = \varphi(s(\alpha \cos 2\pi k \theta - \beta \sin 2\pi k \theta)) \psi_k(s\rho^{-k}\alpha).$$

Now, for any $\theta > 0$, there exists an unbounded sequence of pairs of integers $\{(k_n, \ell_n), n=1, 2, \ldots\}$ such that $|\theta k_n - \ell_n| \to 0$ as $n \to \infty$. Substituting $k_n$ for $k$ in (2.4) gives us
(2.5) \[ \psi_{-k}^{n\alpha} = \varphi[s(\alpha \cos 2\pi k_n - \beta \sin 2\pi k_n)] \psi_{-k}^{n\alpha}, n = 1, 2, \ldots \]

As \( n \to \infty \), the l.h.s. of (2.5) has a limit which is 1 if \( \rho > 1 \), and \( \varphi(s\alpha) \) if \( \rho = 1 \); the first factor on the r.h.s. converges to \( \varphi(s\alpha) \). Furthermore, the second factor on the r.h.s. is a product of characteristic functions, and hence its modulus does not exceed 1.

Consequently, if \( \rho > 1 \), we have \( |\varphi(s\alpha)| \geq 1 \), which means that \( \varphi(s\alpha) = 1 \), so that \( \alpha \) may serve as the vector \( c \) of the lemma. On the other hand, if \( \rho = 1 \), we get

(2.6) \[ \varphi(s\alpha) = \varphi(s\alpha) \lim_{n \to \infty} \psi_{-k}^{n\alpha} \]

Since there is a neighbourhood of the origin in which \( \varphi(s\alpha) \) is bounded away from zero, the second factor on the r.h.s. of (2.6) equals 1 throughout this neighbourhood. But we have here a product of characteristic functions which can equal 1 only by each factor being 1; hence, going back to the definition of \( \psi \), we find that \( \varphi(s\alpha \alpha_j) = 1 \). Since at least one of these \( \alpha_j \), say \( \alpha_1 \), is non-singular, \( \alpha_1 \) may serve as the vector \( c \) of the lemma. ||

Remark. If some of the \( \alpha_j \) have roots of modulus greater than or equal to 1, there exist linearly independent vectors \( a_1, \ldots, a_m \) such that \( a_1 X' \) is, with probability 1, constant. We can make a linear transformation from \( X \) to \( Y = (Y_1, Y_2) \), where \( Y_1 \) is \( m \)-dimensional. Then \( Y_1 \) is, with probability 1, constant; the distribution of \( Y_2 \) has the property
that the distribution of a certain linear form and that of $Y_2$ are identical. Thus the problem reduces to one in fewer than $p$ dimensions. Consequently, we may assume that all roots of the matrices $A_j$ are of modulus less than 1.

**Lemma 2.3.** If every characteristic root of $A_j^tA_j$ is less than 1, for $j = 1, 2, \ldots$, then $\mathcal{L}(X)$ is infinitely divisible.

**Proof.** Recall (2.1) and note that either $\psi_n = 1$ for sufficiently large $n$ (if the linear form is of finite extent) or $\psi_n \to 1$ uniformly over every finite interval as $n \to \infty$. In either case, given $\epsilon, \tau > 0$, there exists a $\nu$ such that $|\psi_n(t) - 1| > \epsilon$ for $\|t\| < \tau$ and $n \geq \nu$.

In order to show that $\Phi$ is infinitely divisible, we have to express it as a product of factors $f_j$ which are characteristic functions such that $|f_j(t) - 1| > \epsilon$ for all $\|t\| < \tau$ and all $j$. For this purpose, choose $n = \nu$ in (2.1) and iterate the relation $k$ times to obtain

\begin{equation}
\Phi(t) = \prod_{j=1}^{M} \Phi(tB_{k,j}) \prod_{j=1}^{N} \psi(tC_{k,j})^\nu, \quad M = \nu^k, \quad N = \frac{\nu - 1}{\nu - 1},
\end{equation}

where each $B_{k,j}$ is a product of $k$ of the $A_j$, and each $C_{k,j}$ is a product of $k$ or fewer $A$'s. Now, let $\lambda^2$ be the largest of the characteristic roots of $A_j^tA_j$, $j = 1, 2, \ldots, \nu$. Then $\|tC_{k,j}\| \leq \|t\|$ and $\|tB_{k,j}\| \leq \lambda^k \|t\|$; we can choose $k$ large enough so that $\lambda^k \tau$ is sufficiently small for $|\Phi(t) - 1| > \epsilon$ whenever $\|t\| \leq \lambda^k \tau$. With such a choice of $k$, (2.7) provides the desired representation for $\Phi$.\|
Lemma 2.4. Given a matrix $A$ and a totally finite measure $G$ on the $p$-dimensional Borel sets, there exists a unique (totally finite) measure $H$ such that

\[
\int_{R_p} (1 - \cos tA') \frac{1 + uu'}{uu'} \, dG(u) = \int_{R_p} (1 - \cos tu') \frac{1 + uu'}{uu'} \, dH(u).
\]

Further, if $h(\theta) = \int_{R_p} (1 + \theta uu')^{-1} \, dH(u)$, then

\[
h(\theta) = \int_{R_p} (1 + \theta uA'Au')^{-1} \frac{1 + uu'}{uu'} \, dG(u).
\]

Proof. There exist non-singular matrices $B$ and $C$ such that $A = C \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} B'$, where $I$ is the identity matrix of dimension equal to the rank of $A = m$, say. Let $s = tC$, $v = uB$, then $tA' = s \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} v' = s_1 v_1'$, where $s = (s_1, s_2)$, $v = (v_1, v_2)$ with $s_1$ and $v_1$ having $m$ components. Consequently,

\[
\int_{R_p} (1 - \cos tA') \frac{1 + uu'}{uu'} \, dG(u) = \int_{R_p} (1 - \cos s_1 v_1') \frac{1 + v(B'B)^{-1} v'}{v(B'B)^{-1} v'} \, dG(vB^{-1}).
\]

Now define the measure $K_1$ on $\mathbb{R}_m$ by

\[
K_1(s) = \int_{S \times R_{p-m}} \frac{v_1 v_1'}{1 + v_1 v_1'} \frac{1 + v(B'B)^{-1} v'}{v(B'B)^{-1} v'} \, dG(vB^{-1}),
\]

and measure $K$ on $\mathbb{R}_p$ by
\[ (2.12) \quad K(S_1 \times S_2) = \begin{cases} K_1(S_1) & \text{if } 0 \in R_m \times S_2, \\ 0 & \text{if } 0 \notin R_m \times S_2. \end{cases} \]

Now the r.h.s. of (2.10) can be re-written as

\[ (2.13) \quad \int_{R_m} (1-\cos s \cdot v_1') \frac{l+v_1'}{v_1' \cdot v_1'} \, dK_1(v_1) = \int_{R_p} (1-\cos s \cdot v') \frac{l+v'}{v' \cdot v'} \, dK(v) \]

\[ = \int_{R_p} (1-\cos t \cdot c \cdot v') \frac{l+v'}{v' \cdot v'} \, dK(v) = \int_{R_p} (1-\cos t \cdot w') \frac{l+w \cdot (c')^{-1} \cdot w'}{w \cdot (c')^{-1} \cdot w'} \, dK(w \cdot (c')^{-1}) \]

\[ = \int_{R_p} (1-\cos t \cdot w') \frac{l+w \cdot (c')^{-1} \cdot w'}{w \cdot (c')^{-1} \cdot w'} \, dH(w), \]

where

\[ H(S) = \int_{S} \frac{uu'}{l+uu'} \frac{l+u \cdot (c')^{-1} \cdot u'}{u \cdot (c')^{-1} \cdot u'} \, dK(u \cdot (c')^{-1}). \]

Hence,

\[ (2.14) \quad h(\theta) = \int_{R_p} (1+\theta uu')^{-1} \, dH(u) = \int_{R_p} (1+\theta uu') \frac{uu'}{l+uu'} \frac{l+u \cdot (c')^{-1} \cdot u'}{u \cdot (c')^{-1} \cdot u'} \, dK(u \cdot (c')^{-1}) \]

\[ = \int_{R_p} (1+\theta v \cdot c' \cdot v')^{-1} \frac{v \cdot c' \cdot v'}{l+\theta v \cdot c' \cdot v'} \, dK(v) \]

Since, by (2.12), there is zero K-measure outside of \( v_2 = 0 \), we can rewrite the r.h.s. of (2.14) as an integral over the m-dimensional space of \( v_1 \).
For this purpose, let \( q(v_1) = v \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} C \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} v' \); then the r.h.s. of (2.14) becomes

\[
(2.15) \quad \int_{R_m} (1+\theta (v_1))^{-1} \frac{q(v_1)}{1+q(v_1)} \frac{l^+ v_1 v'_1}{v_1 v'_1} dk_1(v_1)
\]

\[
= \int_{R_m} (1+\theta (v_1))^{-1} \frac{q(v_1)}{1+q(v_1)} \frac{l^+ v_1 v'_1}{v_1 v'_1} \frac{l^+ v_1 v'_1}{v_1 v'_1} dG(vB^{-1})
\]

\[
= \int_{R_p} (1+\theta uA' Au')^{-1} \frac{u A' Au'}{1+u A' Au'} \frac{l^+ uu'}{uu'} dG(u),
\]

since, on setting \( v = uB \), we get \( q(v_1) = u A' Au' \).

**Remark.** Note that, in particular, we have

\[
H(R_p) = h(0) = \int_{R_p} \frac{l^+ uu'}{uu'} \frac{u A' Au'}{l^+ u A' Au'} dG(u).
\]

**Theorem 2.5.** If all the roots of \( A_j^* A_j \), \( j = 1, 2, \ldots \), are less than 1 and \( E A_j^* A_j - I \) is positive semi-definite, then \( \mathcal{L}(X) \) is normal.

**Proof.** In order to prove that \( \mathcal{L}(X) \) is normal, it is enough to show that \( |\phi(t)| = \exp(-\frac{1}{2} t B t') \), where \( B \) is a real positive semi-definite symmetric matrix. Now, on account of Lemma 2.3, we have the Lévy-Khinchin representation
\begin{align}
(2.16) \quad \log |\hat{\Phi}(t)| &= -\frac{1}{2} t B_{t'} - \int_{\mathbb{R}^p} (1 - \cos t u') \frac{1 + uu'}{uu'} \, dG(u) ,
\end{align}

where \( G \) is a totally finite measure on the Borel sets of \( \mathbb{R}^p \) such that \( G(0) = 0 \), and \( B \) is a positive semi-definite (symmetric) matrix. Hence, we have to show that \( G(\mathbb{R}^p) = 0 \). For this, note that

\begin{align}
(2.17) \quad \log |\hat{\Phi}(tA_j)| &= -\frac{1}{2} t A_j B_{A_j} A_j' - \int_{\mathbb{R}^p} (1 - \cos t A_j u') \frac{1 + uu'}{uu'} \, dG(u) \\
&= -\frac{1}{2} t A_j B_{A_j} A_j' - \int_{\mathbb{R}^p} (1 - \cos t u') \frac{1 + uu'}{uu'} \, dH_j(u) ,
\end{align}

where \( H_j \) is as given in Lemma 2.4 with \( A = A_j \). Now, (2.1) can be written as

\begin{align}
(2.18) \quad \log |\hat{\Phi}(t)| &= \lim_{n \to \infty} \left\{ \sum_{j=1}^{n} \frac{1}{A_j B_{A_j} A_j'} t' - \int_{\mathbb{R}^p} (1 - \cos t u') \frac{1 + uu'}{uu'} \, dG_n(u) \right\} ,
\end{align}

where \( G_n = \sum_{j=1}^{n} H_j \). A necessary condition for (2.18) to hold is that the sequence of measures \( G_n \) converge weakly to \( G \).

Now, let \( \lambda_j (\leq 1) \) be the largest root of \( A_j A_j' \) and let \( M_n = \sum_{j=1}^{n} A_j A_j' \). By assumption, \( M_n \geq I \), so that there is a \( \nu \) such that \( M_n \) is non-singular; let \( \lambda = \max(\lambda_j, j=1,2,\ldots,\nu) \). Then, from Remark 2.2, we have

\begin{align}
(2.19) \quad H_j(\mathbb{R}^p) &= \int_{\mathbb{R}^p} \frac{1 + uu'}{1 + uu' A_j A_j'} \frac{uu'}{uu'} \, dG(u) .
\end{align}
\[ \lambda \int_{R_p} \frac{uA_j' A_j u'}{1+\lambda_j uu'} \frac{uA_j' A_j u'}{uu'} dG(u) \]

\[ = \int_{R_p} \frac{uA_j' A_j u'}{uu'} dG(u) + (1-\lambda_j) \int_{R_p} \frac{uA_j' A_j u'}{1+\lambda_j uu'} dG(u) \]

Hence, for \( n \geq \nu \), we see that

\[ (2.20) \quad G_n(R_p) \geq \int_{R_p} \frac{uM_j u'}{uu'} dG(u) + (1-\lambda) \int_{R_p} \frac{uM_j u'}{1+uu'} dG(u) , \]

where we have ignored the contribution to the second term on the r.h.s. of the terms with \( j > \nu \), and have used the fact that \( \lambda_j \leq \lambda < 1 \), \( j = 1, \ldots, \nu \). Taking the limit as \( n \to \infty \) in (2.20) yields

\[ (2.21) \quad G(R_p) \geq G(R_p) + (1-\lambda) \int_{R_p} \frac{uM_j u'}{1+uu'} dG(u) , \]

since \( uM_j u' \geq uu' \). But \( M_\nu > 0 \) implies that \( uM_j u' > 0 \) for \( u \neq 0 \), so that (2.21) can hold only if \( G(R_p) = 0 \), which completes the proof. \( \Box \)
REFERENCES


