The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 486–494

A NOTE ON A CHARACTERIZATION OF THE MULTIVARIATE NORMAL DISTRIBUTION

By P. R. Fisk

University of Edinburgh

1. Introduction. A well-known characterization of the nonsingular p-variate normal distribution is that it is the only jointly dependent multivariate distribution for which the conditional expectation of each variate is a linear function of the remaining $p-1$ variates and the corresponding conditional distribution depends on the remaining $p-1$ variates only through the conditional mean. An illustration of this for the bivariate case is given by Kendall and Stuart, [5] page 352, on the assumption that all moments exist. Féron and Fourgeaud [3] give a concise proof of this characterization. The purpose of this note is to demonstrate that this characterization of the multivariate normal distribution is valid without making any assumptions on moments of higher order than the first. The method of proof reveals a little more detail than was the case with that of Féron and Fourgeaud. This may enable consequences of the main theorem to be more easily determined.

2. The two-vector case. We are interested in two random vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ of n_1 and n_2 dimensions respectively, where $n_1 \geqq 1$ and $n_2 \geqq 1$. We seek to prove the following theorem.

THEOREM 1. *Necessary and sufficient conditions for the joint distribution of* $\mathbf{z}' = (\mathbf{x}_1' : \mathbf{x}_2')$ *to be an* $(n_1+n_2)-$*variate normal distribution are*

(i) $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are non-degenerate random vectors,*
(ii) *all the first-order absolute moments of* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *exist,*
(iii) *the conditional distribution of* $\mathbf{x}_j$ *given* $\mathbf{x}_k$ $(j = 1, 2; k = 1, 2$ *and* $j \neq k)$ *depends on* $\mathbf{x}_k$ *only through the conditional mean which is* $E(\mathbf{x}_j \mid \mathbf{x}_k) = \mathbf{a}_j + \mathbf{B}_j'\mathbf{x}_k$ *where each row and column of* $\mathbf{B}_j$ *contains at least one non-zero element,*
(iv) *in* (iii) $\mathbf{B}_1\mathbf{B}_2 \neq \mathbf{I}$ *and* $\mathbf{B}_2\mathbf{B}_1 \neq \mathbf{I}$ *where* $\mathbf{I}$ *is an identity matrix of appropriate order.*

PROOF. We lose nothing in generality if we assume $\mathbf{a}_j = \mathbf{0}$ and $\boldsymbol{\mu}_j = \mathbf{0}$, $j = 1, 2$, where $\boldsymbol{\mu}_j = E(\mathbf{x}_j)$.

The necessity is obvious for the class of distribution that we are considering; see, for example, Rao, [6] page 441. We need only demonstrate the sufficiency of the conditions. For this purpose we note that the characteristic function for $\mathbf{z}$ may be written in two equivalent ways corresponding to the equalities

$$\phi(\mathbf{t}_1, \mathbf{t}_2) = E[\exp(i\mathbf{t}_1'\mathbf{x}_1 + i\mathbf{t}_2'\mathbf{x}_2)]$$

$$= \int_{R_2} \exp(i\mathbf{t}_2'\mathbf{x}_2)\, dF(\mathbf{x}_2) \int_{R_1} \exp(i\mathbf{t}_1'\mathbf{x}_1)\, dF(\mathbf{x}_1 \mid \mathbf{x}_2) \tag{1}$$

$$= \int_{R_1} \exp(i\mathbf{t}_1'\mathbf{x}_1)\, dF(\mathbf{x}_1) \int_{R_2} \exp(i\mathbf{t}_2'\mathbf{x}_2)\, dF(\mathbf{x}_2 \mid \mathbf{x}_1). \tag{2}$$

Received July 17, 1969.

The notation used here is standard. The explicit forms of (1) and (2), in the general class of distributions that are being considered here, are

$$\phi(\mathbf{t}_1, \mathbf{t}_2) = \exp[g(\mathbf{t}_1) + H(\mathbf{t}_2 + \mathbf{B}_1 \mathbf{t}_1)] \tag{1'}$$

$$= \exp[h(\mathbf{t}_2) + G(\mathbf{t}_1 + \mathbf{B}_2 \mathbf{t}_2)]. \tag{2'}$$

The expression given in (1′) is derived from the two expectations $E[\exp(i\mathbf{t}_1'\mathbf{x}_1) \mid \mathbf{x}_2] = \exp[g(\mathbf{t}_1) + i\mathbf{t}_1'\mathbf{B}_1'\mathbf{x}_1]$ and $E[\exp(i\mathbf{s}'\mathbf{x}_2)] = \exp[H(\mathbf{s})]$. Similarly the expression given in (2′) is derived from the two expectations $E[\exp(i\mathbf{t}_2'\mathbf{x}_2) \mid \mathbf{x}_1] = \exp[h(\mathbf{t}_2) + i\mathbf{t}_2'\mathbf{B}_2'\mathbf{x}_1]$ and $E[\exp(i\mathbf{s}'\mathbf{x}_1)] = \exp[G(\mathbf{s})]$. From the derivation of (1′) and (2′) and the assumptions of the theorem it can be seen that the functions $g(\cdot)$, $h(\cdot)$, $G(\cdot)$ and $H(\cdot)$, together with all first derivatives of these functions, are zero at the origin. We may also note that the first derivatives of these functions exist everywhere and by definition are continuous everywhere.

Equating terms in (1′) and (2′) we have

$$g(\mathbf{t}_1) + H(\mathbf{t}_2 + \mathbf{B}_1 \mathbf{t}_1) = h(\mathbf{t}_2) + G(\mathbf{t}_1 + \mathbf{B}_2 \mathbf{t}_2). \tag{3}$$

From (3) we can obtain, by putting $\mathbf{t}_2 = \mathbf{0}$ and $\mathbf{t}_1 = \mathbf{0}$ successively,

$$\begin{aligned} g(\mathbf{t}_1) &= G(\mathbf{t}_1) - H(\mathbf{B}_1 \mathbf{t}_1) \\ h(\mathbf{t}_2) &= H(\mathbf{t}_2) - G(\mathbf{B}_2 \mathbf{t}_2) \end{aligned} \tag{4}$$

whence we may write (3) in terms of the functions $G(\cdot)$ and $H(\cdot)$ alone:

$$G(\mathbf{t}_1) + H(\mathbf{t}_2 + \mathbf{B}_1 \mathbf{t}_1) - H(\mathbf{B}_1 \mathbf{t}_1) = H(\mathbf{t}_2) + G(\mathbf{t}_1 + \mathbf{B}_2 \mathbf{t}_2) - G(\mathbf{B}_2 \mathbf{t}_2). \tag{3'}$$

Let us first differentiate both sides of (3′) with respect to $\mathbf{t}_1$ to obtain

$$G^{(1)}(\mathbf{t}_1) + \mathbf{B}_1'[H^{(1)}(\mathbf{t}_2 + \mathbf{B}_1 \mathbf{t}_1) - H^{(1)}(\mathbf{B}_1 \mathbf{t}_1)] = G^{(1)}(\mathbf{t}_1 + \mathbf{B}_2 \mathbf{t}_2). \tag{5}$$

Here $G^{(1)}(\mathbf{u})$ is the vector of first order partial differential coefficients defined by $\partial G(\mathbf{s})/\partial \mathbf{s}$ evaluated at $\mathbf{s} = \mathbf{u}$, with a similar definition for $H^{(1)}(\mathbf{u})$. Putting $\mathbf{t}_1 = \mathbf{0}$ in (5) we get

$$\mathbf{B}_1' H^{(1)}(\mathbf{t}_2) = G^{(1)}(\mathbf{B}_2 \mathbf{t}_2), \qquad \forall \mathbf{t}_2. \tag{6}$$

We may note that if (4) is differentiated with respect to $\mathbf{t}_1$, and if the resultant vector of differential coefficients is evaluated at $\mathbf{t}_1 = \mathbf{B}_2 \mathbf{t}_2$, we will have $g^{(1)}(\mathbf{B}_2 \mathbf{t}_2) = G^{(1)}(\mathbf{B}_2 \mathbf{t}_2) - \mathbf{B}_1' H^{(1)}(\mathbf{B}_1 \mathbf{B}_2 \mathbf{t}_2)$. If $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{I}$ then (6) would indicate that $g^{(1)}(\mathbf{B}_2 \mathbf{t}_2) = \mathbf{0}$, $\forall \mathbf{t}_2$. If we suppose that $n_2 \geqq n_1$ then this implies that $g(\mathbf{t}_1) = \text{constant}$, $\forall \mathbf{t}_1$, and uniform continuity gives the result $g(\mathbf{t}_1) = 0$, $\forall \mathbf{t}_1$. If this is true we would have from (4) that $G(\mathbf{t}_1) = H(\mathbf{B}_1 \mathbf{t}_1)$ whence $h(\mathbf{t}_2) = H(\mathbf{t}_2) - H(\mathbf{B}_1 \mathbf{B}_2 \mathbf{t}_2) = 0$. We see that if $\mathbf{B}_1 \mathbf{B}_2 = \mathbf{I}$ then the conditional distribution of $\mathbf{x}_1$ given $\mathbf{x}_2$ is degenerate at $\mathbf{B}_2'\mathbf{x}_2$ which means that $\mathbf{x}_1$ is an exact linear function of $\mathbf{x}_2$. In these circumstances nothing can be deduced on the distribution of $\mathbf{z}$ from the assumptions of the theorem. It is for this reason that condition (iv) was included in the theorem.

Operating in a similar manner with $\mathbf{t}_2$ we can obtain from (3′)

$$(5') \quad H^{(1)}(\mathbf{t}_2)+\mathbf{B}_2'[G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)] = H^{(1)}(\mathbf{t}_2+\mathbf{B}_1\mathbf{t}_1) \quad \text{and}$$

$$(6') \quad \mathbf{B}_2' G^{(1)}(\mathbf{t}_1) = H^{(1)}(\mathbf{B}_1\mathbf{t}_1), \quad \forall \mathbf{t}_1$$

which are comparable with (5) and (6). From (5), (5′), (6) and (6′) we can obtain the equalities

$$(7) \quad \mathbf{B}_1'[H^{(1)}(\mathbf{t}_2+\mathbf{B}_1\mathbf{t}_1)-H^{(1)}(\mathbf{B}_1\mathbf{t}_1)] = G^{(1)}(\mathbf{B}_2\mathbf{t}_2+\mathbf{B}_2\mathbf{B}_1\mathbf{t}_1)-G^{(1)}(\mathbf{B}_2\mathbf{B}_1\mathbf{t}_1)$$
$$= G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1)$$

and

$$(7') \quad \mathbf{B}_2'[G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)] = H^{(1)}(\mathbf{B}_1\mathbf{t}_1+\mathbf{B}_1\mathbf{B}_2\mathbf{t}_2)-H^{(1)}(\mathbf{B}_1\mathbf{B}_2\mathbf{t}_2)$$
$$= H^{(1)}(\mathbf{t}_2+\mathbf{B}_1\mathbf{t}_1)-H^{(1)}(\mathbf{t}_2).$$

It follows from (6), (6′), (7) and (7′) that

$$\mathbf{B}_1'[H^{(1)}(\mathbf{B}_1\mathbf{t}_1+\mathbf{B}_1\mathbf{B}_2\mathbf{t}_2)-H^{(1)}(\mathbf{B}_1\mathbf{B}_2\mathbf{t}_2)-H^{(1)}(\mathbf{B}_1\mathbf{t}_1)]$$
$$= \mathbf{B}_1'\mathbf{B}_2'[G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1)]$$
$$= G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1)$$

whence

$$(8) \quad (\mathbf{I}-\mathbf{B}_2\mathbf{B}_1)'[G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1)] = \mathbf{0}, \quad \forall \mathbf{t}_1 \text{ and } \mathbf{t}_2.$$

Suppose first that $\mathbf{I}-\mathbf{B}_2\mathbf{B}_1$ is nonsingular, then it is plain from (8) that for all $\mathbf{t}_1$ and $\mathbf{t}_2$ we must have

$$(9) \quad G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1) = \mathbf{0}.$$

Let $n_2 \geqq n_1$ and put $\mathbf{B}_2\mathbf{t}_2 = \mathbf{s}$, then this last equality is equivalent to $G^{(1)}(\mathbf{t}_1+\mathbf{s}) = G^{(1)}(\mathbf{s})+G^{(1)}(\mathbf{t}_1)$ where $\mathbf{s}$ can take all positions in n_1-dimensional space. We note that $G^{(1)}(\cdot)$ is continuous everywhere so the general solution of this last equation is

$$(10) \quad G^{(1)}(\mathbf{t}_1) = \mathbf{C}\mathbf{t}_1$$

(see Aczél [1] page 348) whence from (6) or (6′) we deduce that $H^{(1)}(\mathbf{t}_2)$ has the form

$$(11) \quad H^{(1)}(\mathbf{t}_2) = \mathbf{D}\mathbf{t}_2.$$

Consider now the case when $\mathbf{I}-\mathbf{B}_2\mathbf{B}_1$ is singular. Let $\boldsymbol{\alpha} \neq \mathbf{0}$ be a latent vector of $\mathbf{I}-\mathbf{B}_2\mathbf{B}_1$ corresponding to a zero latent root and let S be a point set in (n_1+n_2)-dimensional space $R_{n_1+n_2}$ for which

$$(12) \quad G^{(1)}(\mathbf{t}_1+\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{B}_2\mathbf{t}_2)-G^{(1)}(\mathbf{t}_1) = \boldsymbol{\alpha}.$$

For the sake of simplicity we suppose that there is only one such $\boldsymbol{\alpha}$. Then for all points in $R_{n_1+n_2}-S$ equation (8) implies that (9) must be true. But $G^{(1)}(\cdot)$ is continuous everywhere so either $S = R_{n_1+n_2}$ or S is empty. In the former case (12) is true for all $\mathbf{t}_1$ and $\mathbf{t}_2$. Define $f(\mathbf{y}) = G^{(1)}(\mathbf{y}) - \boldsymbol{\alpha}$ then (12) is equivalent to $f(\mathbf{t}_1 + \mathbf{s}) = f(\mathbf{t}_1) + f(\mathbf{s})$ where $\mathbf{s}$ is defined above. We deduce from the same argument that led to (10) that $G^{(1)}(\mathbf{y})$ has the form $G^{(1)}(\mathbf{y}) = \boldsymbol{\alpha} + \mathbf{C}\mathbf{y}$. But this conflicts with the property of $G^{(1)}(\mathbf{y})$ that it is null at the origin. We conclude that S is empty and that (10) and (11) are the solutions of (8) for both singular and nonsingular $\mathbf{I} - \mathbf{B}_2\mathbf{B}_1$.

We see from the results given by (10) and (11) that the functions $G(\mathbf{t}_1)$ and $H(\mathbf{t}_2)$ must have the form $G(\mathbf{t}_1) = \mathbf{t}_1'\mathbf{\Sigma}_{11}\mathbf{t}_1$; $H(\mathbf{t}_2) = \mathbf{t}_2'\mathbf{\Sigma}_{22}\mathbf{t}_2$ where $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$ are defined as symmetric matrices. The multivariate normality of $\mathbf{z}$ follows immediately from (4) and (1′). Note that it is not assumed that $\mathbf{\Sigma}_{11}$ or $\mathbf{\Sigma}_{22}$ is nonsingular.

The functions $g(\mathbf{t}_1)$ and $h(\mathbf{t}_2)$ have the form

$$\begin{aligned} g(\mathbf{t}_1) &= \mathbf{t}_1'(\mathbf{\Sigma}_{11} - \mathbf{B}_1'\mathbf{\Sigma}_{22}\mathbf{B}_1)\mathbf{t}_1 \\ h(\mathbf{t}_2) &= \mathbf{t}_2'(\mathbf{\Sigma}_{22} - \mathbf{B}_2'\mathbf{\Sigma}_{11}\mathbf{B}_2)\mathbf{t}_2, \end{aligned} \tag{13}$$

and equality (3) is equivalent to

$$(\mathbf{t}_1' \vdots \mathbf{t}_2') \begin{bmatrix} \mathbf{\Sigma}_{11} \vdots \mathbf{B}_1'\mathbf{\Sigma}_{22} \\ \mathbf{\Sigma}_{22}\mathbf{B}_1 \vdots \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix} = (\mathbf{t}_1' \vdots \mathbf{t}_2') \begin{bmatrix} \mathbf{\Sigma}_{11} \vdots \mathbf{\Sigma}_{11}\mathbf{B}_2 \\ \mathbf{B}_2'\mathbf{\Sigma}_{11} \vdots \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix}.$$

From this last equality we find that $\mathbf{B}_1'\mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{11}\mathbf{B}_2$ and so the conditional variance matrices are from (13) $\mathbf{\Sigma}_{11}(\mathbf{I} - \mathbf{B}_2\mathbf{B}_1)$ and $\mathbf{\Sigma}_{22}(\mathbf{I} - \mathbf{B}_1\mathbf{B}_2)$. We may note that if $\mathbf{I} - \mathbf{B}_2\mathbf{B}_1$ is singular then so also must $\mathbf{I} - \mathbf{B}_1\mathbf{B}_2$ be singular. Thus, if $\boldsymbol{\alpha}$ satisfies $(\mathbf{I} - \mathbf{B}_2\mathbf{B}_1)\boldsymbol{\alpha} = \mathbf{0}$ then $\mathbf{B}_1(\mathbf{I} - \mathbf{B}_2\mathbf{B}_1)\boldsymbol{\alpha} = (\mathbf{I} - \mathbf{B}_1\mathbf{B}_2)\mathbf{B}_1\boldsymbol{\alpha} = \mathbf{0}$. We see that the theorem is true when $\mathbf{x}_1$ or $\mathbf{x}_2$ are separately singularly distributed and also when a linear function of $\mathbf{x}_1$ is exactly equal to a linear function of $\mathbf{x}_2$ with the sole exception of the extreme case excluded by assumption (iv) of the theorem.

Various corollaries of Theorem 1 are possible. Two are given here.

COROLLARY 1. *If* $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$ *are independent random vectors, if* $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ *are independent random vectors and if* $\mathbf{x} = \sum_{j=1}^{m} a_j\mathbf{u}_j$ *and* $\mathbf{y} = \sum_{j=1}^{n} b_j\mathbf{v}_j$ *satisfy the conditions of Theorem* 1, *with all* a_j *and* b_j *non-zero, then* $\mathbf{z}' = (\mathbf{u}_1' : \mathbf{u}_2' : \cdots : \mathbf{u}_m' : \mathbf{v}_1' : \cdots : \mathbf{v}_n')$ *is multivariate normal.*

PROOF. Theorem 1 shows that $(\mathbf{x}' : \mathbf{y}')$ is multivariate normal and Cramér's Theorem 19a ([2] page 113) proves the corollary. A similar proof applies to the next corollary.

COROLLARY 2. *If* $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$ *are independent random vectors and if* $\mathbf{x} = \sum_{j=1}^{m} a_j\mathbf{u}_j$ *and* $\mathbf{y} = \sum_{j=1}^{m} b_j\mathbf{u}_j$ *satisfy the conditions of Theorem* 1, *with all* a_j *and* b_j *non-zero, then* $\mathbf{z}' = (\mathbf{u}_1' : \mathbf{u}_2' : \cdots : \mathbf{u}_m')$ *is multivariate normal.*

It is interesting to compare this second corollary with a theorem of V. P. Skitovich [7] which states that if $\mathbf{y}$ and $\mathbf{x}$ described in Corollary 2 are independent then each $\mathbf{u}_i$ has a multivariate normal distribution. The comparison is more than a curiosity because, as we will now show, a restrictive form of the generalization of Skitovich's theorem given by Ghurye and Olkin [4] may be used to prove, and may be proved by, a restrictive version of Theorem 1. This reciprocal property was first brought to my notice by some comments of David McLaren at one of the Edinburgh statistical seminars.

The restrictive form of Ghurye and Olkin's theorem that will be used here is as follows.

RESTRICTED G–O THEOREM (a). *Let* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *be independent non-degenerate p-dimensional random vectors whose first order moments all exist, and let* $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{C}_1$ *and* $\mathbf{C}_2$ *be nonsingular square matrices of order p. If* $\mathbf{y} = \mathbf{A}_1\mathbf{u}_1 + \mathbf{A}_2\mathbf{u}_2$ *and* $\mathbf{z} = \mathbf{C}_1\mathbf{u}_1 + \mathbf{C}_2\mathbf{u}_2$ *are independent random vectors then* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are each p-variate normal random vectors.*

The condition on first order moments is unnecessary in the direct proof of this theorem as shown by Ghurye and Olkin. The restrictions that we add to Theorem 1 are that $n_1 = n_2 = p$ and $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{I}-\mathbf{B}_2\mathbf{B}_1$ are nonsingular. We will now show that this restricted version of Theorem 1 may be proved using the Ghurye–Olkin theorem. If the conditions of Theorem 1 (restricted) are satisfied then $\mathbf{w}_j = \mathbf{x}_j - \mathbf{a}_j - \mathbf{B}_j'\mathbf{x}_{3-j}$ and $\mathbf{x}_{3-j}$ are independently distributed for $j = 1$ and 2 separately. It follows, in particular, that

$$(14)\qquad (\mathbf{I}-\mathbf{B}_1'\mathbf{B}_2')\mathbf{x}_1 - \mathbf{B}_1'\mathbf{w}_2 = \mathbf{w}_1 + \mathbf{a}_1 + \mathbf{B}_1'\mathbf{a}_2 \quad \text{and} \quad \mathbf{B}_2'\mathbf{x}_1 + \mathbf{w}_2 = \mathbf{x}_2 - \mathbf{a}_2$$

are independently distributed. The nonsingularity of the coefficient matrices for the independent $\mathbf{x}_1$ and $\mathbf{w}_2$ is sufficient, from the Ghurye–Olkin theorem, to demonstrate that $\mathbf{x}_1$ and $\mathbf{w}_2$ are each p-variate normal. Since $\mathbf{w}_2$ is an affine transform of $\mathbf{x}_1$ and $\mathbf{x}_2$ it follows that the latter vectors are jointly distributed in multivariate normal form. Condition (ii) of Theorem 1 is not necessary for this special case.

Conversely, let us suppose that the conditions of the restricted Ghurye–Olkin theorem are satisfied. It follows that the matrix

$$\begin{bmatrix} \mathbf{A}_1 : \mathbf{A}_2 \\ \mathbf{C}_1 : \mathbf{C}_2 \end{bmatrix}$$

is of full-rank and, in particular, $\mathbf{P} = \mathbf{C}_2 - \mathbf{C}_1\mathbf{A}_1^{-1}\mathbf{A}_2$ is nonsingular. Since $\mathbf{y}$ and $\mathbf{z}$ are assumed independent then

$$\boldsymbol{\eta} = \mathbf{A}_1^{-1}\mathbf{y} = \mathbf{u}_1 + \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{u}_2 \qquad \text{and}$$
$$\boldsymbol{\xi} = \mathbf{P}^{-1}\mathbf{z} = \mathbf{P}^{-1}\mathbf{C}_1\mathbf{u}_1 + \mathbf{P}^{-1}\mathbf{C}_2\mathbf{u}_2$$

are also independent. Write $\boldsymbol{\Gamma}_1 = \mathbf{A}_1^{-1}\mathbf{A}_2$, $\boldsymbol{\Gamma}_2 = \mathbf{P}^{-1}\mathbf{C}_1$ and $\boldsymbol{\Gamma}_3 = \mathbf{P}^{-1}\mathbf{C}_2$ then

$$\boldsymbol{\Gamma}_3 - \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1 = \mathbf{I}.$$

We deduce from this that

$$\mathbf{u}_1 = \boldsymbol{\eta} - \boldsymbol{\Gamma}_1\mathbf{u}_2 \quad \text{is independent of} \quad \mathbf{u}_2 \qquad \text{and that}$$
$$\begin{aligned} \boldsymbol{\xi} &= \boldsymbol{\Gamma}_2\mathbf{u}_1 + \boldsymbol{\Gamma}_3\mathbf{u}_2 \\ &= \boldsymbol{\Gamma}_2\boldsymbol{\eta} + \mathbf{u}_2 \quad \text{is independent of} \quad \boldsymbol{\eta}. \end{aligned}$$

This satisfies condition (iii) of Theorem 1. We see that $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and $\mathbf{I}+\mathbf{\Gamma}_2\mathbf{\Gamma}_1 = \mathbf{\Gamma}_3$ are nonsingular thus satisfying assumption (iv) of Theorem 1. This demonstrates that $(\boldsymbol{\eta}':\mathbf{u}_2')$ is jointly distributed in $2p$-variate normal form. Since $\mathbf{u}_1$ is a linear transform of $\boldsymbol{\eta}$ and $\mathbf{u}_2$ it must be distributed in p-variate normal form.

The necessity of the condition on first-order moments for this reciprocal relation to hold may seem to add nothing to the original theorem of Ghurye and Olkin. The arguments given above do show, however, that the Ghurye–Olkin theorem can be extended.

Extension (a) We may remove the restriction that $\mathbf{u}_1$ and $\mathbf{u}_2$ be of the same dimension provided we place suitable restrictions on the matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{C}_1$ and $\mathbf{C}_2$. Suppose that $\mathbf{u}_1$ and $\mathbf{u}_2$ are of dimensions p and q respectively and that $\mathbf{A}_1$ and $\mathbf{C}_2$ are square nonsingular matrices of order p and q respectively. Then the independence of $\mathbf{y}$ and $\mathbf{z}$ defined above is sufficient for the multivariate normality of $\mathbf{u}_1$ and $\mathbf{u}_2$ separately provided none of the rows of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ defined above are null vectors. The nonsingularity of $\mathbf{A}_1$ and $\mathbf{P}$ ensure that each column of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ contains at least one non-zero element.

Extension (b) We may further remove the restriction that $\mathbf{A}_1$ and $\mathbf{C}_2$ be square matrices. Suppose that the orders of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are $(r \times p)$, $(r \times q)$, $(s \times p)$ and $(s \times q)$ respectively with $p+q \geqq r \geqq p$ and $p+q \geqq s \geqq q$. The independence of $\mathbf{y}$ and $\mathbf{z}$ is sufficient for the multivariate normality of $\mathbf{u}_1$ and $\mathbf{u}_2$ separately if the following conditions on $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are satisfied.

(i) the matrices $\mathbf{A}_1'\mathbf{A}_1$, $\mathbf{C}_2'\mathbf{C}_2$ and $\mathbf{P} = \mathbf{C}_2'\mathbf{C}_2 - \mathbf{C}_2'\mathbf{C}_1(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1'\mathbf{A}_2$ are each nonsingular,

(ii) $\mathbf{\Gamma}_1 = (\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1'\mathbf{A}_2$ and $\mathbf{\Gamma}_2 = \mathbf{P}^{-1}\mathbf{C}_2'\mathbf{C}_1$ each have at least one non-zero element in each row.

This second extension can be made equivalent to the first extension by premultiplying the equations in $\mathbf{y}$ and $\mathbf{z}$ by $\mathbf{A}_1'$ and $\mathbf{C}_2'$ respectively whence the conditions required on the matrices become obvious.

3. The multi-vector case. The extension of Theorem 1 to k random vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$, of dimensions $n_1, n_2, \cdots, n_k$ respectively, is not simple in the full generality of that theorem. The extension is relatively easy if we confine our attention to a restricted class of singular and nonsingular distributions in $n(=\sum_{j=1}^{k} n_j)$-dimensional space. For simplicity we will discuss the special case $k = 3$ only. We prove the following special theorem.

THEOREM 2. *Necessary and sufficient conditions for the joint distribution of* $\mathbf{z}' = (\mathbf{x}_1':\mathbf{x}_2':\mathbf{x}_3')$ *to be an n-variate normal distribution are*

(i) $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ *are non-degenerate random vectors,*

(ii) *all the first order absolute moments of* $\mathbf{x}_1$, $\mathbf{x}_2$ *and* $\mathbf{x}_3$ *exist,*

(iii) *for each* $j (j = 1, 2, 3)$ *the conditional distribution of* $\mathbf{x}_j$ *given* $\mathbf{x}_k, \mathbf{x}_l (j \neq k \neq l)$ *depends on the latter two vectors only through the conditional mean which is* $E(\mathbf{x}_j \mid \mathbf{x}_k, \mathbf{x}_l) = \mathbf{a}_j + \mathbf{B}_{kj}'\mathbf{x}_k + \mathbf{B}_{lj}'\mathbf{x}_l$,

(iv) *the vectors can be ordered in such a manner that*

(a) *each row and column of* $\mathbf{B}_{12}$, $\mathbf{B}_{21}$ *and* $(\mathbf{B}'_{13} : \mathbf{B}'_{23})$ *contains at least one non-zero element,*

(b) $\mathbf{I}-\mathbf{B}_{12}\mathbf{B}_{21}$ *and* $\mathbf{I}-\mathbf{B}_{21}\mathbf{B}_{12}$ *are each nonsingular,*

(c) *each row and column of*

$$\begin{bmatrix} \mathbf{I} & : -\mathbf{B}'_{21} \\ -\mathbf{B}'_{12} : & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}'_{31} \\ \mathbf{B}'_{32} \end{bmatrix}$$

contains at least one non-zero element.

Proof. We require the proof of sufficiency only as the necessary conditions are obvious. Consider the joint conditional distribution of $\mathbf{x}_1$ and $\mathbf{x}_2$ given $\mathbf{x}_3$. Within this conditional distribution $\mathbf{x}_1$ and $\mathbf{x}_2$ satisfy the conditions of Theorem 1 hence $(\mathbf{x}_1' : \mathbf{x}_2')$ is conditionally distributed (holding $\mathbf{x}_3$ constant) as an (n_1+n_2)-dimensional normal vector with conditional mean vector

$$E\left[\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \middle| \mathbf{x}_3 \right] = \begin{bmatrix} \mathbf{I} & : -\mathbf{B}'_{21} \\ -\mathbf{B}'_{12} : & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}'_{31} \\ \mathbf{B}'_{32} \end{bmatrix} \mathbf{x}_3 \tag{15}$$

and conditional variance matrix

$$\operatorname{Var}\left[\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \middle| \mathbf{x}_3 \right] = \begin{bmatrix} \boldsymbol{\Sigma}_{11} : \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} : \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Here we have assumed without loss of generality that $\mathbf{a}_j = \mathbf{0}$, $j = 1, 2, 3$, and that the marginal mean vectors are all null. We have also written $\boldsymbol{\Sigma}_{12} = \mathbf{B}'_{21}\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{11}\mathbf{B}_{12}$. The matrices $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the variance matrices in the marginal distributions of $\mathbf{x}_1$ and $\mathbf{x}_2$. If these matrices are nonsingular then $(\mathbf{x}_1' : \mathbf{x}_2')$ is a nonsingular normal random vector from assumption (ivb) of Theorem 2. This latter assumption is necessary for the inverse matrix on the right-hand side of (15) to be invertible. We may note that (15) can be written as

$$E\left[\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \middle| \mathbf{x}_3 \right] = \begin{bmatrix} (\mathbf{I}-\mathbf{B}'_{21}\mathbf{B}'_{12})^{-1}(\mathbf{B}'_{31}+\mathbf{B}'_{21}\mathbf{B}'_{32})\mathbf{x}_3 \\ (\mathbf{I}-\mathbf{B}'_{12}\mathbf{B}'_{21})^{-1}(\mathbf{B}'_{32}+\mathbf{B}'_{12}\mathbf{B}'_{31})\mathbf{x}_3 \end{bmatrix}.$$

It is obvious that either $\mathbf{B}_{31}$ or $\mathbf{B}_{32}$ (but not both) could be null provided condition (ivc) is satisfied.

The vectors $(\mathbf{x}_1' : \mathbf{x}_2')$ and $\mathbf{x}_3$ clearly satisfy the conditions of Theorem 1 hence $\mathbf{z} = (\mathbf{x}_1' : \mathbf{x}_2' : \mathbf{x}_3')$ is multivariate normally distributed. This distribution could be singular but this singularity cannot arise from the singularity of $\mathbf{I}-\mathbf{B}_{12}\mathbf{B}_{21}$ or $\mathbf{I}-\mathbf{B}_{21}\mathbf{B}_{12}$. To this extent the class of distributions to which Theorem 2 applies is more restrictive than the class of distributions to which Theorem 1 applies.

We can see from the above arguments that the appropriate generalization to Theorem 1 is

THEOREM 3. *Necessary and sufficient conditions for the joint distribution of* $\mathbf{z}' = (\mathbf{x}_1' : \mathbf{x}_2' : \cdots : \mathbf{x}_k')$ *to be an n-variate normal distribution are*

(i) $\mathbf{x}_j (j = 1, \cdots, k)$ *are non-degenerate random vectors,*
(ii) *all the first order absolute moments of* $\mathbf{x}_j (j = 1, \cdots, k)$ *exist,*
(iii) *for each* $j (j = 1, \cdots, k)$ *the conditional distribution of* $\mathbf{x}_j$ *given the remaining* $(k-1)$ *vectors depends on the conditioning vectors only through the conditional mean which is*

$$E(\mathbf{x}_j \mid \mathbf{x}_1, \cdots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \cdots, \mathbf{x}_k) = \mathbf{a}_j + \textstyle\sum_{i=1\, i \neq j}^{k} \mathbf{B}'_{ij} \mathbf{x}_i,$$

(iv) *the vectors can be ordered in such a manner that*
(a) *each row and column of* $\mathbf{B}_{12}$, $\mathbf{B}_{21}$ *and* $\mathbf{B}_j' = (\mathbf{B}'_{1j} : \mathbf{B}'_{2j} : \cdots : \mathbf{B}'_{(j-1)j})$ $(j = 1, 2, \cdots, k)$ *contains at least one non-zero element,*
(b) *the matrices*

$$\mathbf{M}_j' = \begin{bmatrix} \mathbf{I} & : -\mathbf{B}'_{21} & : \cdots : & -\mathbf{B}'_{j1} \\ -\mathbf{B}'_{12} : & \mathbf{I} & : \cdots : & -\mathbf{B}'_{j2} \\ \cdots & \cdots & & \cdots \\ -\mathbf{B}'_{1j} : & -\mathbf{B}'_{2j} & : \cdots : & \mathbf{I} \end{bmatrix} \qquad (j = 2, \cdots, k-1)$$

are nonsingular,
(c) *each row and column of* $(\mathbf{B}_{j1} : \mathbf{B}_{j2} : \cdots : \mathbf{B}_{j(j-1)})\mathbf{M}_j^{-1}$ *contains at least one non-zero element.*

We may remark here that the minimal block structure of the conditional expectation is

$$\begin{aligned}
&E_c(\mathbf{x}_1) - \mathbf{B}'_{21}\mathbf{x}_2 &&= \mathbf{a}_1 \\
&-\mathbf{B}'_{12}\mathbf{x}_1 + E_c(\mathbf{x}_2) - \mathbf{B}'_{32}\mathbf{x}_3 &&= \mathbf{a}_2 \\
&-\mathbf{B}'_{13}\mathbf{x}_1 - \mathbf{B}'_{23}\mathbf{x}_2 + E_c(\mathbf{x}_3) - \mathbf{B}'_{43}\mathbf{x}_4 &&= \mathbf{a}_3 \\
&\cdots \quad \cdots \quad \cdots \quad \cdots \\
&-\mathbf{B}'_{1k}\mathbf{x}_1 - \mathbf{B}'_{2k}\mathbf{x}_2 - \mathbf{B}'_{3k}\mathbf{x}_3 - \mathbf{B}'_{4k}\mathbf{x}_4 - \cdots - \mathbf{B}'_{(k-1)k}\mathbf{x}_{k-1} + E_c(\mathbf{x}_k) &&= \mathbf{a}_k.
\end{aligned}$$

Here $E_c(\cdot)$ is the expectation operator with respect to the appropriate conditional probability.

Acknowledgment. I am indebted to Professor I. Olkin for introducing me to Aczél's book and for making me dissatisfied with the less general theorem of Féron and Fourgeaud.

REFERENCES

[1] ACZÉL, J., (1966). *Lectures on Functional Equations and their Applications.* Academic Press, New York.
[2] CRAMÉR, H. (1962). *Random Variables and Probability Distributions*, 2nd ed. Cambridge Univ. Press, London.
[3] FÉRON, R. and FOURGEAUD, C. (1952). Quelques propriétés caractéristiques de la loi de Laplace-Gauss. *Publ. Inst. Statist. Univ. Paris* **1** 44–49.

[4] GHURYE, S. G. and OLKIN, I. (1962). A characterization of the multivariate normal distribution. *Ann. Math. Statist.* **33** 533–541.

[5] KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, 2. Griffin, London.

[6] RAO, C. R. (1965). *Linear Statistical Inference and its Application.* Wiley, New York.

[7] SKITOVITCH, V. P. (1954). Linear forms in independent random variables and the normal distribution. *Izv. Akad. Nauk. SSSR Ser. Mat.* **18** 185–200; (Eng. transl. in *Selected translations in mathematical statistics and probability* **2** 211–228.)