Characterization of Matrix Variate Normal Distributions

A. K. GUPTA AND T. VARGA*

Bowling Green State University

Communicated by C. R. Rao

In this paper, it is shown that two random matrices have a joint matrix variate normal distribution if, conditioning each one on the other, the resulting distributions satisfy certain conditions. A general result involving more than two matrices is also proved. © 1992 Academic Press, Inc.

1. INTRODUCTION

The characterization of multivariate normal distribution through conditional distributions has been studied by many authors in recent years. Results on bivariate normal distribution were obtained by Brucker [1] and Fraser and Streit [2]. Khatri [4] gave characterizations of multivariate normality through regression. In the present paper, their results are generalized for the matrix variate normal distribution.

2. THE MAIN RESULT

In order to derive the results of this paper the following lemma will be useful. It shows how the joint density of two random matrices can be obtained from the conditional densities.

LEMMA 2.1. Let $X \in \mathbb{R}^{p \times n}$ and $Y \in \mathbb{R}^{q \times m}$ be random matrices with joint probability density function $f(X, Y)$. Let $g_1(X)$ and $g_2(Y)$ denote the marginal densities and $h_1(X|Y)$ and $h_2(Y|X)$ be the conditional densities. Assume $f(X, Y)$, $g_1(X)$, $g_2(Y)$, $h_1(X|Y)$, and $h_2(Y|X)$ are defined for all

Received December 5, 1990; revised September 2, 1991.

AMS 1985 subject classifications: 62H10, 62H05.

Key words and phrases: random matrices, conditional distributions, linearly independent, normal distribution.

* Now at the Rose-Hulman Institute of Technology.

80
Suppose there exists $Y_0 \in \mathbb{R}^{q \times m}$ such that $h_2(Y_0 | X) \neq 0$ for all $X \in \mathbb{R}^{p \times n}$. Then

$$f(X, Y) = k \frac{h_2(Y | X) h_1(X | Y_0)}{h_2(Y_0 | X)},$$

where $k$ is a constant.

**Proof.** Let $k = g_2(Y_0)$. Then

$$k \frac{h_2(Y | X) h_1(X | Y_0)}{h_2(Y_0 | X)} = g_2(Y_0) \left( \frac{f(X, Y) / g_1(X)}{f(X, Y_0) / g_2(Y_0)} \right) = f(X, Y).$$

Now we can derive the main results.

**Theorem 2.1.** Let $X: p \times n$, $Y: q \times n$ be random matrices and suppose that

$$Y | X \sim N_q,n(C + DX, \Sigma_2 \otimes \Phi), X | Y = Y_0 \sim N_{p,n}(M, \Sigma_1 \otimes \Phi),$$

where $C: q \times n$, $D: q \times p$, $\Sigma_2: q \times q$, $\Phi: n \times n$, $M: p \times n$, $\Sigma_1: p \times p$, $\Sigma_1 > 0$, $\Sigma_2 > 0$, $\Phi > 0$, and $Y_0$ is a fixed $q \times n$ matrix. Define $B = \Sigma_1 D' \Sigma_2^{-1}$, $A = M - \Sigma_1 D' \Sigma_2^{-1} Y_0$. Then

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q,n} \left( \begin{pmatrix} (I_p - BD)^{-1} (A + BC) \\ (I_q - DB)^{-1} (C + DA) \end{pmatrix}, \begin{pmatrix} (I_p - BD)^{-1} \Sigma_1 & (I_p - BD)^{-1} B \Sigma_2 \\ D(I_p - BD)^{-1} \Sigma_1 & (I_q - DB)^{-1} \Sigma_2 \end{pmatrix} \otimes \Phi \right). \quad (2.1)$$

**Proof.** Let $f(X, Y)$ be the joint probability density function of $X$ and $Y$, $g_1(X)$ the marginal density of $X$, $g_2(Y)$ the marginal density of $Y$, and $h_1(X | Y_0)$, $h_2(Y | X)$ the conditional densities. Throughout the proof $k$ will denote constants that need not be equal. Using Lemma 2.1, we obtain

$$f(X, Y) = k \frac{\text{etr}\{-\frac{1}{2}(Y-(C+DX))' \Sigma_2^{-1}(Y-(C+DX)) \Phi\}}{\text{etr}\{-\frac{1}{2}(Y_0-(C+DX))' \Sigma_2^{-1}(Y_0-(C+DX)) \Phi^{-1}\}} \times \frac{\text{etr}\{-\frac{1}{2}(X-M)' \Sigma_1^{-1}(X-M) \Phi^{-1}\}}{\text{etr}\{-\frac{1}{2}(Y_0-(C+DX))' \Sigma_2^{-1}(Y_0-(C+DX)) \Phi^{-1}\}}$$

$$= k \text{etr}\{-\frac{1}{2}(X' \Sigma_1^{-1} X - 2X' \Sigma_1^{-1} Y + Y' \Sigma_2^{-1} Y + 2Y' \Sigma_2^{-1} C - 2X' D' \Sigma_2^{-1} Y) \Phi\}$$

$$= k \text{etr}\{-\frac{1}{2}(Z' \Omega Z + Z' R) \Phi^{-1}\},$$
Since $f(X, Y)$ is a probability density function, the symmetric matrix $\Omega$ must be positive definite and hence is non-singular. Thus

$$f(Z) = k \text{etr} \left\{ -\frac{1}{2} (Z - (-\frac{1}{2} \Omega^{-1} R))' \Omega (Z - (-\frac{1}{2} \Omega^{-1} R)) \Phi \right\}.$$  

Now using the fact that $D(I_p - BD)^{-1} = (I_q - DB)^{-1} D$ and $B(I_q - DB)^{-1} = (I_p - BD)^{-1} B$ it is easy to see that

$$\Omega^{-1} = \begin{pmatrix} (I_p - BD)^{-1} \Sigma_1 & (I_p - BD)^{-1} B \Sigma_2 \\ D(I_p - BD)^{-1} \Sigma_1 & (I_q - DB)^{-1} \Sigma_2 \end{pmatrix}.$$  

Then

$$-\frac{1}{2} \Omega^{-1} R = \begin{pmatrix} (I_p - BD)^{-1} (A + BC) \\ (I_q - DB)^{-1} (C + DA) \end{pmatrix},$$

which completes the proof. \[\square\]

Using Theorem 2.1 we can derive the following result.

**Corollary 2.1.** Let $X: p \times n$, $Y: q \times n$ be random matrices and suppose that

$$X \mid Y \sim N_{p,n}(A + BY, \Sigma_1 \otimes \Phi), \quad (2.2)$$

$$Y \mid X \sim N_{q,n}(C + DX, \Sigma_2 \otimes \Phi), \quad (2.3)$$

where $A: p \times n$, $B: p \times q$, $\Sigma_1: p \times p$, $\Phi: n \times n$, $C: q \times n$, $D: q \times p$, $\Sigma_2: q \times q$, $\Sigma_1 > 0$, $\Sigma_2 > 0$, $\Phi > 0$. Then

$$\Sigma_2 B = D \Sigma_1, \quad (2.4)$$

and

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q,n} \left( \begin{pmatrix} (I_p - BD)^{-1} (A + BC) \\ (I_q - DB)^{-1} (C + DA) \end{pmatrix}, \begin{pmatrix} (I_p - BD)^{-1} \Sigma_1 & (I_p - BD)^{-1} B \Sigma_2 \\ D(I_q - DB)^{-1} \Sigma_1 & (I_q - DB)^{-1} \Sigma_2 \end{pmatrix} \otimes \Phi \right). \quad (2.5)$$

**Proof.** Let $Y_0 = 0$. Then $X \mid Y = Y_0 \sim N_{p,n}(A, \Sigma_1 \otimes \Phi)$. Using Theorem 2.1 we conclude that $Z$ has distribution (2.1) if we replace $B$ by $B^* = \Sigma_1 D' \Sigma_2^{-1}$ in formula (2.1). It follows from (2.1) that

$$E(X \mid Y) = B^* Y + A.$$
Comparing this with (2.2) we get $B^* = B$, which proves (2.5). Hence (2.5) is established.

Another type of characterization is given in the following theorem.

**Theorem 2.2.** Let $X: p \times n$ and $Y: q \times n$ be random matrices. Suppose that $Y|X \sim N_{q,n}(C + DX, \Sigma \otimes \Phi)$ and $X \sim N_{p,n}(F, \Sigma \otimes \Phi)$, where $C: q \times n$, $D: q \times p$, $\Sigma: q \times q$, $\Phi: n \times n$, $F: p \times n$, $\Sigma_1: p \times p$, $\Sigma_1 > 0$, $\Sigma_2 > 0$, $\Phi > 0$. Then

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q,n} \left( \begin{pmatrix} F \\ \Sigma_1 \\ \Sigma_2 + D \Sigma_1 D' \end{pmatrix} \otimes \Phi \right).$$  

(2.6)

**Proof.** Let $\phi$ be the joint characteristic function of $X$ and $Y$. For $S: p \times n$, $T: q \times n$, we have

$$\phi(S, T) = E(\text{etr}(iS'X + iT'Y)) = E(\text{etr}(iS'X) E_{Y|X}(\text{etr}(iT'Y)))$$

$$= E(\text{etr}(i(S' + T'D)X)) \text{etr}(iCT' - \frac{1}{2} \Sigma_2 T\Phi T')$$

$$= \text{etr}\{i(S'F + T'(C + DF)) - \frac{1}{2}((T'((D + \Sigma_2)(D'))T + S'\Sigma_1, S + 2S'\Sigma_1 D'T/\Phi)\},$$

which proves (2.6).  

Theorem 2.3 is an extension of a multivariate result of Khatri [4] to matrices.

**Theorem 2.3.** Let $X_1, \ldots, X_k$ be random matrices of dimension $p \times n$. Assume that the random variables that are the elements of the matrix $X = (X_1', \ldots, X_k')'$ are not linearly dependent. Suppose that the following conditions are satisfied:

(a) $E(X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k) = A_{i1}X_1 + \cdots + A_{i,i-1}X_{i-1} + A_{i,i+1}X_{i+1} + \cdots + A_{ik}X_k + B_i$, $i = 1, \ldots, k$, where $A_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, k$, $j \neq i$, are $p \times p$ and $B_i$, $i = 1, \ldots, k$, are $p \times n$ constant matrices,

(b) $X_1|X_2, \ldots, X_k$ depends on $X_2, \ldots, X_k$ only through $E(X_1|X_2, \ldots, X_k)$,

(c) $\text{Var}(\text{vec}(X_{i}^{'})|X_1, X_3, \ldots, X_k) = \Sigma \otimes \Phi$, where $\Sigma: p \times p$, $\Phi: n \times n$, $\Sigma > 0$, $\Phi > 0$,

(d) $A_{12}$ and $A_{21}$ are nonsingular,

(e) $$\begin{pmatrix} -I_p & A_{12} & A_{13} & \cdots & A_{1i} \\ A_{21} & -I_p & A_{23} & \cdots & A_{2i} \\ \vdots & & & & \vdots \\ A_{i1} & A_{i2} & A_{i3} & \cdots & -I_p \end{pmatrix} = A_{(i)}$$

is nonsingular for $i = 1, 2, \ldots, k$. 


(f) 

\[
\begin{pmatrix}
A_{i1} A_{i2} \cdots A_{i,i-1} A_{i}^{-1} (i-1) \\
A_{2i} \\
\vdots \\
A_{i-1,i}
\end{pmatrix}
\]

is nonsingular for \(i = 2, 3, \ldots, k\),

(g) with the notation \(A_{ii} = -I_p\), \(i = 1, \ldots, k\),

\[
\begin{pmatrix}
A_{11} & A_{13} & \cdots & A_{1k} \\
A_{21} & A_{23} & \cdots & A_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
A_{i-1,1} & A_{i-1,3} & \cdots & A_{i-1,k} \\
A_{i+1,1} & A_{i+1,3} & \cdots & A_{i+1,k} \\
\vdots & \vdots & \ddots & \vdots \\
A_{k,1} & A_{k,3} & \cdots & A_{k,k}
\end{pmatrix} = A^{(i)}
\]

is nonsingular for \(i = 1, 2, \ldots, k\),

(h)

\[
I_p + (A_{21}A_{23}A_{24}\cdots A_{2k}) \cdot A^{(2)^{-1}}
\]

is nonsingular.

Then \(X\) has a matrix variate normal distribution.

Proof. Define

\[
\bar{X}_i = \text{vec}(X_i), \quad i = 1, \ldots, k, \quad \bar{U}_i = \text{vec}(U_i), \quad i = 1, \ldots, k,
\]

\[
\bar{A}_{ij} = \begin{cases} 
I_p \otimes I_n & \text{if } i = j \\
-A_{ij} \otimes I_n & \text{if } i \neq j
\end{cases}
\]

\[
\bar{\Sigma} = \Sigma \otimes \Phi, \quad \bar{Z}_i = \sum_{j=1}^{k} \bar{A}_{ij} \bar{X}_j + \bar{U}_i.
\]

Then it is easy to see that \(\bar{X}_i, \bar{U}_i, \bar{Z}_i, \bar{A}_{ij}, \bar{\Sigma}\) satisfy the conditions of Theorem 2 of Khatri [4]. From that theorem we conclude that \(\bar{Z} = (\bar{Z}_1^T \bar{Z}_2^T \cdots \bar{Z}_k^T)^T\) has a multivariate normal distribution.
Let $\bar{X} = (\bar{X}_1', \bar{X}_2', \ldots, \bar{X}_k')$. Since $\bar{X} = -(A^{-1}_{(k)} \otimes I_n)(\bar{Z} - \bar{U})$

$$\bar{X} \sim N_{kp}(\mu, \Sigma^*), \quad (2.7)$$

where $\mu = (\mu_1', \mu_2', \ldots, \mu_k')$. The only remaining thing to prove is that $\Sigma^* = S \otimes \Phi$, where $S: (kp) \times (kp)$, $\Phi: n \times n$, because from this the theorem follows immediately.

Let us partition $\Sigma^* = (\Sigma_y^*)$, where $\Sigma_y^*: (pn) \times (pn)$, $i, j = 1, \ldots, k$.

Now by expressing $E(\bar{X}_i | \bar{X}_1, \ldots, \bar{X}_{i-1}, \bar{X}_{i+1}, \ldots, \bar{X}_k)$, first using condition (a) and then using (2.7), and comparing the coefficient matrices of $(\bar{X}_1', \ldots, \bar{X}_{i-1}', \bar{X}_{i+1}', \ldots, \bar{X}_k')$ in the resulting expressions we find

$$\Sigma_{ij} = \sum_{l=1}^{k} \left( A_{il} \otimes I_n \right) \Sigma_{y_{j}}, \quad i, j = 1, 2, \ldots, k, j \neq i. \quad (2.8)$$

Hence

$$(-A_{12} \otimes I_n) \Sigma_{2j} = \sum_{l=1}^{k} \left( A_{il} \otimes I_n \right) \Sigma_{y_{j}}, \quad i, j = 1, 2, \ldots, k, j \neq i. \quad (2.9)$$

Thus, we get

$$\begin{pmatrix}
\Sigma_{1j} \\
\Sigma_{3j} \\
\vdots \\
\Sigma_{kj}
\end{pmatrix} = ((-A^{(j)-1}p^{(j)}) \otimes I_n) \Sigma_{2j}, \quad j, i = 1, 2, \ldots, k. \quad (2.10)$$

where

$$p^{(j)} = \begin{pmatrix}
A_{12} \\
A_{22} \\
\vdots \\
A_{j-1, 2} \\
A_{j+1, 2} \\
\vdots \\
A_{k, 2}
\end{pmatrix}.$$

Let $\tilde{\Sigma}_{21} = (\Sigma_{21}, \Sigma_{23}, \Sigma_{24}, \ldots, \Sigma_{2k})$, $\tilde{\Sigma}_{12} = \Sigma_{21}'$, and

$$\tilde{\Sigma}_{11} = \begin{pmatrix}
\Sigma_{11} & \Sigma_{13} & \Sigma_{14} & \cdots & \Sigma_{1k} \\
\Sigma_{31} & \Sigma_{33} & \Sigma_{34} & \cdots & \Sigma_{3k} \\
\Sigma_{41} & \Sigma_{43} & \Sigma_{44} & \cdots & \Sigma_{4k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\Sigma_{k1} & \Sigma_{k3} & \Sigma_{k4} & \cdots & \Sigma_{kk}
\end{pmatrix}.$$
With this notation, (2.10), for \( j = 2 \), can be written as

\[
\Sigma_{12} = (\{-A^{(2)} - p^{(2)}\} \otimes I_n) \Sigma_{22}. \tag{2.11}
\]

Moreover for \( i = 2 \), (2.8) yields

\[
\Sigma_{21}^{-1} = Q_2 \otimes I_n, \tag{2.12}
\]

where \( Q_2 = (A_{21}, A_{23}, A_{24}, \ldots, A_{2k}) \). Now using (2.11) and (2.12) we get

\[
\text{Var}(X_2|X_1, X_3, ..., X_k) = ((I_p + Q_2 A^{(2)} - p^{(2)} I) \otimes I_n) \Sigma_{22}. \tag{2.13}
\]

Using condition (c) and (2.13), we get \( \Sigma_{22} = ((I_p + Q_2 A^{(2)} - p^{(2)} I) \Sigma) \otimes \Phi \). Therefore \( \Sigma_{22} \) can be written as

\[
\Sigma_{22} = S_{22} \otimes \Phi. \tag{2.14}
\]

From (2.10) and (2.14) we conclude \( \Sigma_{12} = (-A^{(2)} - p^{(2)} S_{22}) \otimes \Phi \). Therefore

\[
\Sigma_{i2} = S_{i2} \otimes \Phi, \quad i = 1, 2, ..., k. \tag{2.15}
\]

From (2.10) and (2.15) we obtain

\[
\begin{pmatrix}
\Sigma_{1j} \\
\Sigma_{2j} \\
\vdots \\
\Sigma_{kj}
\end{pmatrix} = (-A^{(j)} - p^{(j)} S_{rj}) \otimes \Phi, \quad j = 1, 3, ..., k.
\]

Therefore \( \Sigma_{ij} = S_{ij} \otimes \Phi, \quad i = 1, 2, ..., k, \quad j = 1, 2, ..., k \). Hence \( \Sigma^* = S \otimes \Phi \), which completes the proof. \[ \square \]

If in Theorem 2.3 we consider only two matrices, the conditions of the theorem can be weakened as the following result shows.

**Theorem 2.4.** Let \( X \) and \( Y \) be random matrices of dimension \( p \times n \). Assume that the random variables that are the elements of the matrix \( Z = (X_i) \) are not linearly dependent. Suppose that the following conditions are satisfied:

(a)

\[
E(X|Y) = A + BY, \tag{2.16}
\]

\[
E(Y|X) = C + DX, \tag{2.17}
\]

where \( B, D \) are \( p \times p \) and \( A, C \) are \( p \times n \) matrices.
(b) $X \mid Y$ depends on $Y$ only through $E(X \mid Y)$,

(c) 
\begin{align*}
\Var(\text{vec}(Y) \mid X) &= \Sigma \otimes \Phi,
\end{align*}

(2.18)

where $\Sigma: p \times p$, $\Phi: n \times n$, $\Sigma > 0$, $\Phi > 0$,

(d) $B$ and $D$ are nonsingular

(e) $(-I_p \quad B)$ is nonsingular.

Let us define $\Sigma_1 = B\Sigma D^{-1}$ and $\Sigma_2 = \Sigma$. Then $\Sigma_1 > 0$ and

\begin{align*}
Z &= \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p,n} \left( \begin{pmatrix} (I_p - BD)^{-1} (A + BC) \\ (I_q - DB)^{-1} (C + DA) \end{pmatrix}, \\
& \left( (I_p - BD)^{-1} \Sigma_1 \quad (I_p - BD)^{-1} B \Sigma_2 \right)^{\otimes} \Phi \right).
\end{align*}

(2.19)

Proof. First we use Theorem 2.3 with $k = 2$, $X_1 = X$, $X_2 = Y$, $A_{11} = B$, $A_{21} = D$, $B_1 = A$, $B_2 = C$. It is easy to see that conditions (a)–(h) of that theorem are satisfied. Therefore

\begin{align*}
Z &= \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p,n} \left( \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{\otimes} \Phi \right),
\end{align*}

(2.20)

where $S_{11}, S_{12}, S_{21}, S_{22}$ are $p \times p$ and $M_1, M_2$ are $p \times n$ matrices. Then from (2.20) we get

\begin{align*}
E(X \mid Y) &= M_1 - S_{12} S_{22}^{-1} M_2 + S_{12} S_{22}^{-1} Y \\
E(Y \mid X) &= M_2 - S_{21} S_{11}^{-1} M_1 + S_{21} S_{11}^{-1} X.
\end{align*}

Then using (2.16) and (2.17) we get $B = S_{12} S_{22}^{-1}$ and $D = S_{21} S_{11}^{-1}$. Now

\begin{align*}
\Var(\text{vec}(X) \mid Y) &= S_{11.2} \otimes \Phi \\
\Var(\text{vec}(Y) \mid X) &= S_{22.1} \otimes \Phi.
\end{align*}

Then using (2.18) we get $S_{22.1} = \Sigma$. But $S_{11.2}(S_{11}^{-1} S_{12}) = (S_{12} S_{22}^{-1}) S_{22.1}$, and therefore $S_{11.2} = B \cdot \Sigma \cdot D^{-1}$. Hence

\begin{align*}
X \mid Y &\sim N_{p,n}(A + BY, \Sigma \otimes \Phi)
\end{align*}
and

\[ Y|X \sim N_{p,n}(C + DX, \Sigma_2 \otimes \Phi). \]

An application of Corollary 2.1 completes the proof.  

REFERENCES