

A CHARACTERIZATION OF MATRIX VARIATE NORMAL DISTRIBUTION

KHOAN T. DINH¹ and TRUC T. NGUYEN²

¹US Environmental Protection Agency/TS 798
Washington, DC 20460

²Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403-0221

(Received December 31, 1992 and in revised form June 26, 1993)

ABSTRACT. The joint normality of two random vectors is obtained based on normal conditional with linear regression and constant covariance matrix of each vector given the value of the other without assuming the existence of the joint density. This result is applied to a characterization of matrix variate normal distribution.

Key Words and Phrases: characteristic function; eigenvalue; joint distribution; conditional distribution; covariance matrix; norm of a matrix; linear transformation; spectral radius.

AMS 1992 Subject Classification: 62H05.

1. INTRODUCTION AND BASIC RESULTS

Let $X = (X_1, \dots, X_m)'$ and $Y = (Y_1, \dots, Y_n)'$ be two given random vectors. If the joint distribution of X and Y is an $(m + n)$ variate normal distribution, then $X|Y = y$ and $Y|X = x$ both have multivariate normal distributions for all $x \in R^m$ and $y \in R^n$ (Anderson [1]). Brucker [3], then Fraser and Streit [5], Castillo and Galambos [4] considered the bivariate case, Bischoff and Fieger [2] gave a characterization in multivariate case. In these papers the existence of a joint density was accepted. All the above results can be seen in a survey paper of Hamedani [8]. In this note without suppose the existence of a continuous joint density for X and Y we show that if $X|Y = y$ and $Y|X = x$ both have multivariate normal distributions with linear regressions and constant covariance matrices, then the joint distribution of X and Y is a multivariate normal distribution. This result then is extended to characterize a matrix variate normal distribution. A result of Gupta and Varga [6] (Corollary 2.1) can be

considered as a Corollary of this result under some special covariance matrix structure.

The basic results of matrices and linear transformations used in the proof of Theorem 2.1 can be found in Halmos [7] or Young and Gregory [8].

The vector norm we use here is the Euclidean norm, $\|x\| = \left(\sum_{j=1}^n x_j^2\right)^{1/2}$, and

the matrix norm

$$\|A\| = \sup\{\|Ax\| : x \in R^n, \|x\| = 1\} = \sqrt{\lambda_1}, \tag{1.1}$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ are the eigen values of $A'A$. In R^n under the topology corresponding to the metric $d(x,y) = \|x - y\|$ the set $S = \{x : x \in R^n, \|x\| = 1\}$ is a compact set. Consequently, for every $n \times n$ matrix A , $\|Ax\|$ is a real valued continuous function on S . Hence, if A is a nonsingular matrix, there exists s_0 and s_1 of S such that

$$\begin{aligned} 0 < c_0 = \|As_0\| &= \inf\{\|As\| : s \in S\} \leq \|As\| \leq \sup\{\|As\| : s \in S\} \\ &= \|A\| = \|As_1\| = \sqrt{\lambda_1}, \quad \forall s \in S. \end{aligned} \tag{1.2}$$

2. A CHARACTERIZATION OF MULTIVARIATE NORMAL DISTRIBUTION

In this section we give a characterization of multivariate normal distributions based on conditional normality of each vector given the other. This result will be extended to a matrix variate case.

THEOREM 2.1. Let $X = (X_1, \dots, X_m)'$ and $Y = (Y_1, \dots, Y_n)'$ be two random vectors. Suppose that $X|Y = y \sim N(Ay + b, \Sigma_1)$ and $Y|X = x \sim N(Cx + d, \Sigma_2)$ for every $x \in R^m, y \in R^n$, where A is a $m \times n$ matrix, C is a $n \times m$ matrix, $b \in R^m, d \in R^n, \Sigma_1$ is a $m \times m$ positive definite covariance matrix, Σ_2 is a $n \times n$ positive definite covariance matrix, A, b, Σ_1 do not depend on y and C, d, Σ_2 do not depend on x . Then

- (a) $\rho(AC) < 1$, where $\rho(AC)$ denotes the spectral radius of AC .
- (b) $\Sigma_{1,1}C' = A\Sigma_{2,2}$, where $\Sigma_{1,1} = (I_m - AC)^{-1}\Sigma_1, \Sigma_{2,2} = (I_n - CA)^{-1}\Sigma_2$.
- (c) $X \sim N(\mu_1, \Sigma_{1,1}), Y \sim N(\mu_2, \Sigma_{2,2})$, where $\mu_1 = (I_m - AC)^{-1}(Ad + b), \mu_2 = (I_n - CA)^{-1}(Cb + d)$.
- (d) The joint distribution of X and Y is a $(m + n)$ variate normal distribution with covariance matrix $\Sigma = (\Sigma_{i,j}), i, j = 1, 2$, where $\Sigma_{1,1}, \Sigma_{2,2}$ are given in (b), and $\text{Cov}(X, Y) = A\Sigma_{2,2} = \Sigma_{1,1}C'$.

PROOF. From $X|Y = y \sim N(Ay + b, \Sigma_1), \forall y \in R^n$,

$$\phi_{X,Y}(s, t) = \phi_Y(t + A's) \exp\left\{ib's - \frac{1}{2}s' \Sigma_1 s\right\}, \tag{2.1}$$

$\forall s \in R^m, t \in R^n$, and from $Y|X = x \sim N(Cx + d, \Sigma_2), \forall x \in R^m$,

$$\phi_{X,Y}(s, t) = \phi_X(s + C't) \exp\left\{id't - \frac{1}{2}t' \Sigma_2 t\right\}, \quad \forall s \in R^m, t \in R^n. \tag{2.2}$$

Hence

$$\phi_X(s) = \phi_Y(A's) \exp\left\{ib's - \frac{1}{2}s' \Sigma_1 s\right\}, \quad \forall s \in R^m, \tag{2.3}$$

and

$$\phi_Y(t) = \phi_X(C't) \exp\left\{id't - \frac{1}{2}t' \Sigma_2 t\right\}, \quad \forall t \in R^n. \tag{2.4}$$

Substitute (2.4) in (2.3),

$$\phi_X(s) = \phi_X((AC)'s) \exp \left\{ i(Ad + b)'s - \frac{1}{2} s'(A\Sigma_2A' + \Sigma_1)s \right\}, \quad \forall s \in R^m. \quad (2.5)$$

For any $k = 0, 1, 2, \dots$, iterating (2.5) k times,

$$\begin{aligned} \phi_X(s) &= \phi_X((AC)^k s) \exp \left\{ i(Ad + b)' \sum_{j=0}^k (AC)^j s - \frac{1}{2} s' \sum_{j=0}^k ((AC)^j)' (A\Sigma_2A' + \Sigma_1)(AC)^j s \right\}, \\ &\quad \forall s \in R^m. \end{aligned} \quad (2.6)$$

Let $k \rightarrow \infty$, the left side of (2.6) is $\phi_X(s)$, the limit of the right side of (2.6) must exist for all $s \in R^n$. We go to show that

$$\lim_{k \rightarrow \infty} s' \sum_{j=0}^k ((AC)^j)' (A\Sigma_2A' + \Sigma_1)(AC)^j s \quad (2.7)$$

must exist for all $s \in R^m$. Suppose this limit does not exist for a $0 \neq s_0 \in R^m$. Since each $((AC)^j)' (A\Sigma_2A' + \Sigma_1)(AC)^j$ is a nonnegative definite matrix for $j = 0, 1, 2, \dots$, then (2.7) must approach $+\infty$. Consider the characteristic function of the random variable $s_0'X$

$$\begin{aligned} \phi_{s_0'X}(w) &= \phi_X(ws_0) = \phi_X((AC)^k ws_0) \exp \left\{ i(Ad + b)' \sum_{j=0}^k (AC)^j s_0 w \right. \\ &\quad \left. - \frac{1}{2} w^2 s_0' \sum_{j=0}^k ((AC)^j)' (A\Sigma_2A' + \Sigma_1)(AC)^j s_0 \right\}, \quad \forall w \in R, k = 0, 1, 2, \dots \end{aligned}$$

Hence,

$$|\phi_{s_0'X}(w)| \leq \exp \left\{ -\frac{1}{2} w^2 s_0' \sum_{j=0}^k ((AC)^j)' (A\Sigma_2A' + \Sigma_1)(CA)^j s_0 \right\}, \quad (2.8)$$

for all $k = 0, 1, 2, \dots, \forall w \in R$. Then for $w \neq 0$, the limit of the right side of (2.8) is zero when $k \rightarrow \infty$. Consequently, $\phi_{s_0'X}(w) = 0$ for $w \neq 0$ and $\phi_{s_0'X}(0) = 1$. This is a contradiction since $\phi_{s_0'X}(w)$ is a characteristic function. Therefore the limit in (2.7) exists for all $s \in R^m$. From $A\Sigma_2A' + \Sigma_1$ is a symmetric positive definite matrix, then

$$\begin{aligned} c_0^2 \sum_{j=0}^{\infty} \|(AC)^j s\|^2 &\leq \sum_{j=0}^{\infty} s'((AC)^j)' (A\Sigma_2A' + \Sigma_1)(AC)^j s \\ &\leq \|(A\Sigma_2A' + \Sigma_1)^{\frac{1}{2}}\|^2 \sum_{j=0}^{\infty} \|(AC)^j s\|^2, \quad \forall s \in R^m, \end{aligned} \quad (2.9)$$

where $0 < c_0 = \inf\{\|(A\Sigma_2A' + \Sigma_1)^{\frac{1}{2}} s\| : s \in S\}$. From (2.9), the limit in (2.7) exists if and only if $\sum_{j=0}^{\infty} \|(AC)^j s\| < \infty, \forall s \in R^m$. This is equivalent to $\lim_{j \rightarrow \infty} (AC)^j = \emptyset$, and then is equivalent to $\rho(AC) < 1$, where $\rho(AC)$ is the spectral radius of AC . In this case,

$$\lim_{k \rightarrow \infty} \phi_X((AC)^k s) = 1, \quad (2.10)$$

$$\mu_1 = \lim_{k \rightarrow \infty} \sum_{j=0}^k (AC)^j (Ad + b) = [(I - (AC)^k)^{-1}]' (Ad + b) = (I - AC)^{-1} (Ad + b), \tag{2.11}$$

$$\Sigma_{1,1} = \lim_{k \rightarrow \infty} \sum_{j=0}^k ((AC)^j)' (A\Sigma_2 A' + \Sigma_1) (AC)^j, \tag{2.12}$$

and

$$\phi_X(s) = \exp \left\{ i\mu_1' s - \frac{1}{2} s' \Sigma_{1,1} s \right\}, \quad \forall s \in R^m. \tag{2.13}$$

Substituting (2.13) in (2.4)

$$\phi_Y(t) = \exp \left\{ i\mu_2' t - \frac{1}{2} t' \Sigma_{2,2} t \right\}, \quad \forall t \in R^n, \tag{2.14}$$

where

$$\mu_2 = C\mu_1 + d, \tag{2.15}$$

$$\Sigma_{2,2} = C\Sigma_{1,1}C' + \Sigma_2. \tag{2.16}$$

Then if we substitute again (2.14) in (2.3), we will get

$$\mu_1 = A\mu_2 + b, \tag{2.17}$$

$$\Sigma_{1,1} = A\Sigma_{2,2}A' + \Sigma_1. \tag{2.18}$$

The joint characteristic function of X and Y is obtained by substituting (2.13) in (2.2)

$$\phi_{X,Y}(s,t) = \exp \left\{ i\mu_1' s + i\mu_2' t - \frac{1}{2} [s' \Sigma_{1,1} s + t' \Sigma_{2,2} t + s' \Sigma_{1,1} C' t + t' C \Sigma_{1,1} s] \right\}, \tag{2.19}$$

$\forall s \in R^m, t \in R^n$, or by substituting (2.14) in (2.1)

$$\phi_{X,Y}(s,t) = \exp \left\{ i\mu_1' s + i\mu_2' t - \frac{1}{2} [s' \Sigma_{1,1} s + t' \Sigma_{2,2} t + s' A \Sigma_{2,2} t + t' \Sigma_{2,2} A' s] \right\}, \tag{2.20}$$

$\forall s \in R^m, t \in R^n$. Then by the uniqueness of the characteristic function,

$$A \Sigma_{2,2} = \Sigma_{1,1} C'. \tag{2.21}$$

From (2.18) and (2.21),

$$A \Sigma_{2,2} A' = \Sigma_{1,1} C' A' = \Sigma_{1,1} - \Sigma_1.$$

Hence

$$\Sigma_{1,1} = (I_m - AC)^{-1} \Sigma_1, \tag{2.22}$$

and from (2.16) and (2.21),

$$\Sigma_{2,2} = (I_n - CA)^{-1} \Sigma_2. \tag{2.23}$$

Theorem 2.1 is proved.

3. A CHARACTERIZATION OF MATRIX VARIATE NORMAL DISTRIBUTION

Let X be a $k \times n$ random matrix, where the row vectors are $X'_i = (X_{i1}, \dots, X_{in})$, $i = 1, \dots, k$. The notation $\text{Vec}(X') = (X'_{11}, \dots, X'_{kn})'$ is a $kn \times 1$ random vector. X is defined to have a $k \times n$ matrix variate normal distribution with mean matrix μ of dimension $k \times n$ having row vectors μ'_1, \dots, μ'_k and with covariance matrix $(\Sigma_{i,j})$, $i, j = 1, \dots, k$, denoted by $N_{k,n}(\mu, \Sigma)$ if $\text{Vec}(X)$ has a kn -variate normal distribution with mean vector $(\mu'_1, \dots, \mu'_k)'$, where $\mu_i = E(X_i)$, $i = 1, \dots, k$, and with covariance matrix $\Sigma = (\Sigma_{i,j})$, $i, j = 1, \dots, k$, where $\Sigma_{i,j} = \text{Cov}(X_i, X_j)$, $i, j = 1, \dots, k$.

In the following part \otimes denotes the Kronecker product of two matrices and I_k the $k \times k$ identity matrix.

THEOREM 3.1. Let X and Y be two random matrices of dimension $k_1 \times n$ and $k_2 \times n$, respectively. Suppose $X|Y = y \sim N_{k_1,n}(Ay + B, \Sigma_1)$ for all y , where A is a $k_1 \times k_2$ matrix, B is a $k_1 \times n$ matrix, $\Sigma_1 = \Sigma_{i,j}^{(1)}$ is a $k_1 n \times k_1 n$ positive definite covariance matrix, $\Sigma_{i,j}^{(1)} = \text{Cov}(X_i, X_j)$ under the conditional

distribution, $i, j = 1, \dots, k_1$, and A, B, Σ_1 do not depend on y , and $Y|X = x \sim N_{k_2,n}(Cx + D, \Sigma_2)$ for all x , where C is a $k_2 \times k_1$ matrix, D is a $k_2 \times n$ matrix, $\Sigma_2 = (\Sigma_{i,j}^{(2)})$ is a $k_2 n \times k_2 n$ positive definite matrix, $\Sigma_{i,j}^{(2)} = \text{Cov}(Y_i, Y_j)$

under the conditional distribution, $i, j = 1, \dots, k_2$, and A, B, Σ_2 do not depend on x . Then

(a) $\rho(AC) < 1$.

(b) $\Sigma_{1,1}(C' \otimes I_n) = (A \otimes I_n) \Sigma_{2,2}$, where

$$\Sigma_{1,1} = ((I_{k_1} - AC)^{-1} \otimes I_n) \Sigma_1,$$

$$\Sigma_{2,2} = ((I_{k_2} - CA)^{-1} \otimes I_n) \Sigma_2.$$

(c) $X \sim N_{k_1,n}(\mu_1, \Sigma_{1,1})$, $Y \sim N_{k_2,n}(\mu_2, \Sigma_{2,2})$, where $\mu_1 = (I_{k_1} - AC)^{-1}(AD + B)$, $\mu_2 = (I_{k_2} - CA)^{-1}(CB + D)$.

(d) The joint distribution of X and Y is a $(k_1 + k_2) \times n$ matrix variate normal distribution with covariance matrix $(\Sigma_{i,j})$, $i, j = 1, 2$, where $\Sigma_{1,2} = ((A(I_{k_2} - CA)^{-1}) \otimes I_n) \Sigma_2$, $\Sigma_{2,1} = ((C(I_{k_1} - AC)^{-1}) \otimes I_n) \Sigma_1$.

PROOF. From the hypothesis of this theorem and from the fact that $\text{Vec}(AB) = (A \otimes I_n) \text{Vec}(B)$ if B has n columns, $\text{Vec}(X|Y) = \text{Vec}(y) \sim N((A \otimes I_n) \text{Vec}(y) + \text{Vec}(B), \Sigma_1)$ and $\text{Vec}(Y|X) = \text{Vec}(x) \sim N((C \otimes I_n) \text{Vec}(x) + \text{Vec}(D), \Sigma_2)$. Then by Theorem 2.1 and the properties of Kronecker product of matrices (Anderson (1984)) the results of this theorem are obtained.

The following result given by Gupta and Varga [6] can be considered as a Corollary of Theorem 3.1 by giving special structures to covariance matrices of conditional distributions of $X|Y$ and $Y|X$ in Theorem 3.1.

COROLLARY 3.1. Let X and Y be $k_1 \times n$ and $k_2 \times n$ random matrix, respectively, and suppose that

$$X|Y \sim N_{k_1,n}(AY + B, \Sigma_1 \otimes \Phi)$$

$$Y|X \sim N_{k_2,n}(CX + D, \Sigma_2 \otimes \Phi),$$

where $A: k_1 \times k_2$, $B: k_1 \times n$, $\Sigma_1: k_1 \times k_1$, $\Phi: n \times n$, $C: k_2 \times k_1$, $D: k_2 \times n$, $\Sigma_2: k_2 \times k_2$, Σ_1, Σ_2, Φ positive definite. Then

(a) $\rho(AC) < 1$, $\Sigma_1 C' = A \Sigma_2$,

(b) $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{k_1+k_2,n} \left(\begin{pmatrix} (I_{k_1} - AC)^{-1}(AD+B) \\ (I_{k_2} - CA)^{-1}(CB+D) \end{pmatrix}, \begin{pmatrix} (I_{k_1} - AC)^{-1} \Sigma_1 & A(I_{k_2} - CA)^{-1} \Sigma_2 \\ C(I_{k_1} - AC)^{-1} \Sigma_1 & (I_{k_2} - CA)^{-1} \Sigma_2 \end{pmatrix} \otimes \Phi \right)$

If the condition that X and Y have a joint continuous density is accepted, the result of Bischoff and Fieger [2] can be applied to characterize a matrix variate normal distribution without assuming linear regressions. The result is given by the following theorem.

THEOREM 3.2. Let X and Y be two random matrices of dimensions $k_1 \times n$ and $k_2 \times n$, respectively, with a joint continuous density function. If the

conditional distributions of X given $Y = y$ and Y given $X = x$ are both matrix variate normal, then the following statements are equivalent

- (a) X and Y have a joint $(k_1 + k_2) \times n$ matrix variate normal distribution.
- (b) $\Sigma_2(x)$ is constant in $R^{k_1 n}$ where $\Sigma_2(x)$ is the covariance matrix of $Y | X = x$.
- (c) For the minimal eigen value $\lambda(x)$ of the positive definite matrix $\Sigma_2(x)$

$$\tau^2 \lambda(\tau b_j) \rightarrow \infty \text{ as } \tau \rightarrow \infty \text{ for } j = 1, \dots, k_1 n,$$

where $b_1, \dots, b_{k_1 n}$ is an arbitrary, but fixed, basis of $R^{k_1 n}$

Note that without supposing at least a constant covariance for a conditional distribution, the joint normal distribution does not hold [8].

ACKNOWLEDGEMENT. The authors thank the referees for the useful comments and suggestions.

REFERENCES

- [1] Anderson, T. W. An introduction to multivariate statistical analysis, 2nd ed., John Wiley and Sons, New York, (1984).
- [2] Bischoff, W. and Fieger, W. Characterization of the multivariate normal distribution by conditional normal distribution, Metrika **38**(1991), 239-248.
- [3] Brucker, J. A note on the bivariate normal distribution, Commun. Statist. Theor. Meth., **A8**(2)(1979), 175-177.
- [4] Castillo, E. and Galambos, J. Conditional distributions and the bivariate normal distribution, Metrika **36**(1989), 209-214.
- [5] Fraser, D. A. S. and Streit, F. A further note on the bivariate normal distribution, Commun. Statist. Theor. Meth. **A9**(10)(1980), 1097-1099.
- [6] Gupta, A. K. and Varga, T. Characterization of matrix variate normal distributions, Journ. Multi. Analys., Vol. 41, No. 1, 80-88, (1992).
- [7] Halmos, P. R. Finite dimensional vector spaces, Springer-Verlag, New York, (1974).
- [8] Hamadani, G. G. Bivariate and multivariate normal characterization. A Brief Survey. Commun. Statist. Theory Methods **21**(9), 2665-2688, (1992).
- [9] Young, D. M. and Gregory, R. T. A survey of numerical mathematics, Vol. 2., Addison Wesley Co., Massachusetts, (1973).