Some Matrix-Variate Distribution Theory: Notational Considerations and a Bayesian Application
Author(s): A. P. Dawid
Source: *Biometrika*, Vol. 68, No. 1 (Apr., 1981), pp. 265-274
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335827

Biometrika (1981), **68**, 1, *pp.* 265–74
Printed in Great Britain

Some matrix-variate distribution theory: Notational considerations and a Bayesian application

By A. P. DAWID

Department of Mathematics, The City University, London

Summary

We introduce and justify a convenient notation for certain matrix-variate distributions which, by its emphasis on the important underlying parameters, and the theory on which it is based, eases greatly the task of manipulating such distributions. Important examples include the matrix-variate normal, t, F and beta, and the Wishart and inverse Wishart distributions. The theory is applied to compound matrix distributions and to Bayesian prediction in the multivariate linear model.

Some key words: Bayesian prediction; Compound distribution; Extendible; Inverse Wishart; Matrix beta; Matrix F; Matrix t; Multivariate linear model; Random matrix; Rotatable; Scale matrix; Spherical; Wishart.

1. Some matrix distributions

We present below some examples of distributions for random matrices that have appeared in the literature. However, our notation is sometimes nonstandard, for reasons which will be taken up in § 2.

Example 1. Let Z be an $n \times p$ matrix with independent standard normal entries. Then we shall write $Z \sim \mathcal{N}(I_n, I_p)$, where I_n denotes the $n \times n$ identity matrix, and say that Z has a standard matrix normal distribution. More generally, we write $Z \sim \mathcal{N}(I_n, \Sigma)$ if the rows of Z are independent, each having the multivariate normal distribution $N(0, \Sigma)$, and $Z \sim \mathcal{N}(\Lambda, I_p)$ if the columns of Z are independently $N(0, \Lambda)$. Clearly, if $Z \sim \mathcal{N}(\Lambda, I_p)$ then $AZ \sim \mathcal{N}(A\Lambda A', I_p)$, while if $Z \sim \mathcal{N}(I_n, \Sigma)$ then $ZB \sim \mathcal{N}(I_n, B'\Sigma B)$, for any fixed matrices A and B of orders $n_1 \times n$ and $p \times p_1$ respectively.

Example 2. We denote by $W(\nu;\, \Sigma)$ the Wishart distribution with ν degrees of freedom and scale matrix Σ, where Σ is nonnegative-definite symmetric of order $p \times p$. We need not specify explicitly the dimensionality p, since this is implicit in Σ. If desired, we may write $W_p(\nu;\, \Sigma)$, with a similar convention in other examples. As is well known, if $\Psi \sim W(\nu;\, \Sigma)$, then $A\Psi A' \sim W(\nu;\, A\Sigma A')$. The distribution $W(\nu;\, I_p)$ is termed standard Wishart.

If $Z \sim \mathcal{N}(I_n, I_p)$, then by definition, $Z'Z \sim W(n;\, I_p)$. This shows that the Wishart distribution exists for ν a positive integer, and then $W(\nu;\, I_p)$ has, with probability 1, rank ν if $\nu < p$, or rank p if $\nu \geqslant p$. The definition of $W(\nu;\, I_p)$ may be extended, for example by generalizing its density function, to allow arbitrary $\nu > p-1$ (Eaton, 1972), when it has rank p almost surely; however, the distribution is not defined for nonintegral $\nu < p-1$. For $p = 1$, $W(\nu;\, 1)$ is identical with χ^2_ν $(\nu > 0)$.

Example 3. Let $\Psi \sim W(\nu;\, I_p)$ with $\nu > p-1$. Then $\Phi = \Psi^{-1}$ exists almost surely. We say that Φ has a standard inverse Wishart distribution with parameter $\delta = \nu - p + 1$, and write $\Phi \sim IW(\delta;\, I_p)$. The reason for this choice of parameter, rather than ν, will become clear later. Then $IW(\delta;\, I_p)$ exists for any $\delta > 0$, and is almost surely nonsingular. We also write $\Phi \sim IW(\delta;\, \Sigma)$, for $\delta > 0$, Σ nonsingular, if $\Psi = \Phi^{-1} \sim W(\nu;\, \Sigma^{-1})$, with $\nu = \delta + p - 1$. This definition will later be extended to allow singular Σ. For $p = 1$, $IW(\delta;\, 1)$ becomes $(\chi^2_\delta)^{-1}$

Example 4. Let $\Phi \sim IW(\delta;\, I_p)$ and, given Φ, $T \sim \mathcal{N}(I_n, \Phi)$. The induced marginal distribution for T is termed the standard matrix-*t* distribution $T(\delta;\, I_n, I_p)$. This notation differs from that of Dickey (1967), who would denote the above distribution by $T(I_n, I_p, 0, \delta + n + p - 1)$: again the major difference is in the choice of the 'degrees of freedom' parameter. For $n = p = 1$, $T(\delta, 1, 1) = \delta^{-\frac{1}{2}} t_\delta$; to eliminate this discrepancy, an additional scale factor $\delta^{\frac{1}{2}}$ could have been carried along in the matrix definition, but it merely complicates formulae. For $p = 1$, $n > 1$, we get a multivariate t distribution (Cornish, 1954), again scaled by $\delta^{-\frac{1}{2}}$.

The matrix-*t* distribution is fundamental to inference for arbitrary left-spherical distributions, as defined in §4 below, and in no way depends on normality: the above definition is merely a convenient description. For a general derivation, see Dawid (1977).

Example 5. Let $U \sim W(\nu;\, \Phi)$ given Φ, where, marginally, $\Phi \sim IW(\delta; I_p)$ with $\delta > 0$ and $\nu > p-1$ or ν integral. The induced marginal distribution for U is considered by Olkin & Rubin (1964), and is termed multivariate beta II by Tan (1969a), who denotes it by $B_{\mathrm{II}}\{p;\, \frac{1}{2}\nu, \frac{1}{2}(p+\delta-1)\}$, while Dempster (1969, p. 341) uses the notation $G(\nu, p+\delta-1;\, I_p)$. For $p = 1$ the distribution becomes $(\nu/\delta)\, F_{\nu,\delta}$. We shall call the distribution of U above standard matrix-variate F, and write $U \sim F(\nu, \delta;\, I_p)$: note that we again discard some scale parameters from the traditional univariate definition.

It follows from the construction of Example 4 that, if $T \sim T(\delta;\, I_n, I_p)$, then $T'T \sim W(n;\, \Phi)$ given Φ, with $\Phi \sim IW(\delta; I_p)$. That is, $T'T \sim F(n, \delta;\, I_p)$.

Example 6. Let $S_1 \sim W_p(\nu_1;\, \Phi)$, $S_2 \sim W_p(\nu_2;\, \Phi)$, independently, with Φ positive-definite, and let $S = S_1 + S_2$. By sufficiency, the conditional distribution of S_1 given $S = \Sigma$ does not depend on Φ. We write it as $B(\nu_1, \nu_2;\, \Sigma)$, and say that the conditional distribution of S_1 is the matrix-variate beta distribution with ν_1 and ν_2 degrees of freedom and scale matrix Σ. This notation follows Dempster (1969).

We need ν_i to be a positive integer or $\nu_i > p-1$ $(i = 1, 2)$. If ν_1 and ν_2 are integers with $\nu_1 + \nu_2 < p$, then Σ is in the support of S, and so the above construction is admissible, if Σ has rank $\nu_1 + \nu_2$. For $\nu_1 + \nu_2 > p-1$, it is sufficient that Σ have full rank p. In particular, for $\nu_1 + \nu_2 > p-1$ we can define the distribution $B(\nu_1, \nu_2;\, I_p)$, a standard matrix-variate beta distribution. This has been studied by Olkin & Rubin (1964), Tan (1969a), Mitra (1970) and Khatri (1970). Our $B(\nu_1, \nu_2;\, I_p)$ is denoted by $B_{\mathrm{I}}(p;\, \frac{1}{2}\nu_1, \frac{1}{2}\nu_2)$ by Tan, and by $B_p(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)$ by Mitra. For $p = 1$, $B(\nu_1, \nu_2;\, 1)$ becomes the beta distribution $\beta(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)$.

The role of the scale matrix of the general matrix-variate beta distribution is clarified by the following lemma.

LEMMA 1. *Let $S_1 \sim B_p(\nu_1, \nu_2;\, \Sigma)$, and let A be a fixed nonsingular matrix of order $p \times p$. Then $AS_1A' \sim B_p(\nu_1, \nu_2;\, A\Sigma A')$.*

Proof. Start with $S_i \sim W_p(\nu_i;\, \Phi)$, independently for $i = 1, 2$, and $S = S_1 + S_2$, and let $U_i = AS_iA'$ $(i = 1, 2)$, $U = U_1 + U_2 = ASA'$. Then $U_i \sim W_p(\nu_i;\, \Psi)$, independently, where $\Psi = A\Phi A'$. Given $S = \Sigma$, $S_1 \sim B_p(\nu_1, \nu_2;\, \Sigma)$, and we have to show that, in this same conditional distribution, $U_1 = AS_1A' \sim B_p(\nu_1, \nu_2;\, A\Sigma A')$. But the condition $S = \Sigma$ is equivalent to $U = A\Sigma A'$, and so the result follows on applying the construction of Example 6 to the U's.

COROLLARY 1. *If $S_1 \sim B(\nu_1, \nu_2;\, I_p)$, and Q is a fixed $p \times p$ orthogonal matrix, then $Q'S_1Q \sim B(\nu_1, \nu_2;\, I_p)$.*

COROLLARY 2. *Let $S_i \sim W_p(\nu_i;\, \Sigma)$ independently $(i = 1, 2;\, \nu_1 + \nu_2 > p - 1)$, $S = S_1 + S_2$, and let D be a random matrix such that $S = DD'$. Given S, let D and S_1 be independent. Then $D^{-1}S_1(D^{-1})' \sim B(\nu_1, \nu_2;\, I_p)$, independently of D, and hence of S.*

Note in particular, that the independence condition will hold if D is a deterministic function of S, for example the unique lower triangular, positive diagonal, square root of S (Mitra, 1970).

Proof. Conditionally on D, $S_1 \sim B(\nu_1, \nu_2;\, S)$, so that $D^{-1}S_1(D^{-1})' \sim B(\nu_1, \nu_2;\, I_p)$. Since this last distribution does not depend on the conditioning variable D, the result follows.

2. MARGINALIZATION

The notation of § 1 has been chosen so that the standard distributions have the property of consistency under marginalization. Thus let Z^* be a $n^* \times p^*$ submatrix of $Z \sim \mathcal{N}(I_n, I_p)$; then, trivially, $Z^* \sim \mathcal{N}(I_{n^*}, I_{p^*})$.

Now let $\Psi \sim W(\nu;\, I_p)$, and partition Ψ as

$$\begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix},$$

where Ψ_{11} is of order $p^* \times p^*$; and, conformably, $\Phi = \Psi^{-1}$ with components $\Phi_{11}, \Phi_{12}, \Phi_{21}, \Phi_{22}$. Then $\Psi_{11} \sim W(\nu;\, I_{p^*})$, with the same value of ν as for Ψ. We further have that

$$\Phi_{11}^{-1} = \Psi_{11\cdot 2} = \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21} \sim W\{\nu - (p - p^*);\, I_{p^*}\};$$

see, for example, Dempster (1969, Theorem 13.4.2). Thus $\Phi_{11} \sim IW(\delta;\, I_{p^*})$ with the same value of δ as for Φ. This demonstrates that our notations for both the Wishart and the inverse Wishart distribution are consistent under marginalization; this would not have held for the inverse Wishart had we chosen ν as the parameter.

If now T^* is a $n^* \times p^*$ submatrix of $T \sim T(\delta;\, I_n, I_p)$, then $T^* \sim \mathcal{N}(I_{n^*}, \Phi_{11})$ given Φ. Since $\Phi_{11} \sim IW(\delta;\, I_{p^*})$, we find $T^* \sim T(\delta;\, I_{n^*}, I_{p^*})$, again consistently with δ unchanged. Likewise, for a $p^* \times p^*$ submatrix U^* of $U \sim F(\nu, \delta;\, I_p)$ we have $U^* \sim W(n;\, \Phi_{11})$ given Φ, so that $U^* \sim F(\nu, \delta;\, I_{p^*})$, again consistently. The distributions $B(\nu_1, \nu_2;\, I_p)$ are similarly consistent under marginalization, with ν_1, ν_2 unchanged: see, for example, Tan (1969a), Mitra (1970) and § 4 below.

3. EXTENDIBILITY

Consider first, for concreteness, the matrix-t distributions $T(\delta;\, I_n, I_p)$, where $\delta > 0$ is fixed but n and p vary. We can construct a random doubly infinite array $T = T_{\infty,\infty} = (t_{ij}\colon i,j \geqslant 1)$ with the property that, for all (n,p), $T_{n,p} \sim T(\delta;\, I_n, I_p)$, where $T_{n,p}$ denotes the leading $n \times p$ submatrix of T. This construction depends only on the consistency property, and the existence of $T(\delta;\, I_n, I_p)$ for arbitrary n and p: see Dawid (1978). We say that T has the standard infinite matrix-t distribution with parameter δ, and write $T \sim T(\delta)$. Note that we could not have introduced such a distribution had we not singled out δ as the important parameter in our parameterization.

In like fashion we can introduce the following distributions for infinite arrays.

(i) We have $Z = Z_{\infty,\infty} \sim \mathscr{N}$ if $Z_{n,p} \sim \mathscr{N}(I_n, I_p)$, for all n, p.

(ii) We have $\Psi = \Psi_{\infty,\infty} \sim W(\nu)$ if $\Psi_{p,p} \sim W(\nu;\, I_p)$, for all p. Note that ν must now be an integer, since otherwise $W(\nu;\, I_p)$ fails to exist as soon as $p > \nu + 1$. Then $\Psi_{p,p}$ is almost surely singular for $p > \nu$.

(iii) We have $\Phi = \Phi_{\infty,\infty} \sim IW(\delta)$ if $\Phi_{p,p} \sim W(\delta;\, I_p)$, for all p. Any $\delta > 0$ is allowable. Once again, our revised notation is indispensable.

(iv) We have $U = U_{\infty,\infty} \sim F(\nu, \delta)$ if $U_{p,p} \sim F(\nu, \delta;\, I_p)$, for all p. We require $\delta > 0$ and ν integral. Again this distribution is singular for $p > \nu$.

For the matrix-variate beta distribution, however, only finite extension is possible, since $B(\nu_1, \nu_2;\, I_p)$ is undefined for $p \geqslant \nu_1 + \nu_2 + 1$, and will in fact fail to exist as soon as $p \geqslant \nu_i + 1$ if ν_i is nonintegral $(i = 1, 2)$.

In general, suppose we have any family of distributions $\{\Delta_{n,p}\}$ for all leading $n \times p$ submatrices $\{Y_{n,p}\}$ of Y, which are consistent in the above sense. Then we get a distribution Δ for the doubly-infinite array $Y_{\infty,\infty}$ such that $Y_{n,p} \sim \Delta_{n,p}$ for all submatrices. For distributions of symmetric matrices, as in Examples 2, 3 and 5, one would of course restrict to $n = p$ in the foregoing. Many commonly occurring matrix distributions extend to infinity in this way, and it proves useful to consider the finite arrays as submatrices of the infinite one, and choose notation accordingly, as we have done.

4. SPHERICAL SYMMETRY

The standard distributions we have considered, in common with many that arise in statistical applications, possess important symmetry properties (Dawid, 1977, 1978). Let Y be a random $n \times n$ matrix. We say that Y is left-spherical if, for any fixed $n \times n$ orthogonal matrix P, PY has the same distribution as Y. Likewise, Y is right-spherical if YQ has the same distribution as Y, for any fixed $p \times p$ orthogonal Q, while Y is spherical if it is simultaneously left- and right-spherical. Similarly, a random $p \times p$ nonnegative-definite symmetric matrix S will be called rotatable if $Q'SQ$ has the same distribution as S, for any fixed $p \times p$ orthogonal Q.

If $Z \sim \mathscr{N}(I_n, I_p)$, then $PZ \sim \mathscr{N}(PI_n P', I_p) = \mathscr{N}(I_n, I_p)$ since P is orthogonal. Similarly $ZQ \sim \mathscr{N}(I_n, I_p)$. Thus Z is spherical. Likewise, $\Psi \sim W(\nu;\, I_p)$ is rotatable, and therefore so is $\Phi \sim IW(\delta;\, I_p)$. Now let $T \sim T(\delta;\, I_n, I_p)$, and represent T as in Example 4. Then, given Φ, PT has the same distributions as T, and so the same holds marginally. Also, $TQ \sim \mathscr{N}(I_n, \Phi^*)$ given Φ^*, where $\Phi^* = Q'\Phi Q$ has the same distribution as Φ. So T is spherical. It is easily seen that $F(\nu, \delta;\, I_p)$ is rotatable, while the same property for $B(\nu_1, \nu_2;\, I_p)$ has already been established in Corollary 1 to Lemma 1.

The matrix-variate beta distribution is intimately related to the property of left-sphericity. Consider the family $\mathscr{P}$ of all left-spherical distributions for the $n \times p$ matrix Y.

Each of these distributions is invariant under the group of transformations taking Y into PY, for P orthogonal. A maximal invariant under this group is $S = Y'Y$, which is therefore sufficient for $\mathscr{P}$ by a general result of Farrell (1962). Now write $Y' = (Y_1', Y_2')$, where Y_i is of order $n_i \times p$, and define $S_i = Y_i' Y_i$ $(i = 1, 2)$. Then $S = S_1 + S_2$ and, by sufficiency, the distribution of S_1 given S is the same for all left-spherical distributions. Taking, in particular, $Y \sim \mathscr{N}(I_n, \Sigma)$, we have $S_i \sim W_p(n_i; \Sigma)$ independently $(i = 1, 2)$, so that this common distribution for S_1 given S must be $B(n_1, n_2; S)$.

Another important representation of $B(n_1, n_2; I_p)$ with $p \leqslant n = n_1 + n_2$ now follows on taking $Y \sim \Upsilon_{n,p}$, the unique spherical uniform distribution over the set $O_{n,p}$ of $n \times p$ matrices having orthonormal columns. For a definition and some properties of this distribution, see Dawid (1977). Since, in this case, $S = I_p$ almost surely, the marginal distribution of $S_1 = Y_1' Y_1$ is $B(n_1, n_2; I_p)$. This property can be used to define the distribution (Dempster, 1969; Khatri, 1970). If we use it, the consistency property of the $B(n_1, n_2; I_p)$ distributions, for varying $p \leqslant n = n_1 + n_2$, follows readily from a corresponding property of the $T_{n,p}$ distributions: if Y^* is a uniformly distributed $n \times n$ orthogonal matrix, and Y contains the first p columns of Y^*, then $Y \sim \Upsilon_{n,p}$ $(p \leqslant n)$.

5. Transposition

An important property of spherical distributions is the following (Dawid, 1977).

Theorem 1. *If the $n \times n$ matrix Y is spherical, then Y' has the same distribution as Y.*

Now suppose we have a distribution Δ for $Y_{\infty,\infty}$ such that any leading submatrix $Y_{n,p}$ is spherical; or equivalently, a consistent family $\{\Delta_{n,p}\}$ such that each $\Delta_{n,p}$ is spherical. In this case we have the following theorem.

Theorem 2. *If $Y \sim \Delta_{n,p}$, then $Y' \sim \Delta_{p,n}$.*

Proof. Supposing, without loss of generality, $n \geqslant p$, we can regard Y as the first p columns of Y^*, where $Y^* \sim \Delta_{n,n}$. Then Y' constitutes the first p rows of $Y^\dagger = Y^{*\prime}$. But $Y^\dagger \sim \Delta_{n,n}$ by Theorem 1, so that $Y' \sim \Delta_{p,n}$ by consistency.

Clearly, Theorem 2 continues to hold for finite consistent families, so long as $\Delta_{m,m}$ for $m = \max\{n, p\}$ exists. As an application of Theorem 2, let $T \sim T(\delta; I_n, I_p)$; then $T' \sim T(\delta; I_p, I_n)$ and so we can represent T' by $T' \sim \mathscr{N}(I_p, \Lambda)$ given Λ, with $\Lambda \sim IW(\delta; I_n)$. It follows that an alternative representation of T is $T \sim \mathscr{N}(\Lambda, I_p)$ with $\Lambda \sim IW(\delta; I_n)$ (Dickey, 1967). Note that, in this representation, T has, given Λ, independent $N(0, \Lambda)$ columns, in contrast with the original definition of Example 4, which gives T conditionally independent rows. The equivalence of these representations is not immediately obvious. Note also that our parameterization means that we use the same value of δ in both representations. Taking $n = 1$, we see that we can obtain a scaled multivariate t row vector $t' = (t_1, \ldots, t_p)$ either by taking $t_i \sim N(0, \phi)$, independently given ϕ, with $\phi^{-1} \sim \chi^2_\delta$; or by taking $t \sim N(0, \Lambda)$ given Λ, with $\Lambda^{-1} \sim W(\delta + p - 1, I_p)$ (Ando & Kaufman, 1965; Dempster, 1969, Corollary 13·6·1).

Dawid (1978) showed, extending the above result, that there is a one-to-one correspondence between infinite spherical distributions $\{\Delta\}$ for $Y_{\infty,\infty}$ and infinite rotatable distributions $\{\Pi\}$, such that we can represent $\Delta_{n,p}$ either by $Y_{n,p} \sim \mathscr{N}(I_n, \Phi)$ given Φ, with $\Phi \sim \Pi_p$, or by $Y_{n,p} \sim \mathscr{N}(\Lambda, I_p)$ given Λ, with $\Lambda \sim \Pi_n$.

6. SCALE-PARAMETER EXTENSIONS

Suppose we have a left-spherical distribution Δ for the $\infty \times p$ matrix Y, with $p \leqslant \infty$, or equivalently, a consistent family $\{\Delta_n\}$ of left-spherical distributions for $\{Y_n\}$, where Y_n comprises the first n rows of Y. Let A be a fixed $m \times n$ matrix, and $X = AY_n$. Then the distribution of X depends on n and A only through $H = AA'$. For let B be an $m \times N$ matrix such that $BB' = AA' = H$, where, without loss of generality, $N \geqslant n$. Then the inner products between the vectors constituting the rows of B, and of the $m \times N$ matrix $A^* = (A\,0)$, are the same, so that each of these two ordered sets of m row vectors has the same configuration in N-dimensional space R^N. It follows that there exists an $N \times N$ orthogonal matrix P such that $B = A^*P$: this may be established, for example, by considering the Gram–Schmidt orthonormalization procedure applied to the two sets of vectors. Thus $BY_N = A^*PY_N$, which, by left-sphericity of Y_N, has the same distribution as $A^*Y_N = AY_n$.

The above theory demonstrates that the following definition is admissible. Let Δ be an infinite left-spherical distribution, and H an arbitrary nonnegative-definite symmetric matrix of arbitrary order $m \times m$. Then the distribution $\Delta(H)$ is defined as that of AY_n, where A, of order $m \times n$, satisfies $AA' = H$, but n and A need not be further specified.

A similar construction holds for scale-modified right-spherical distributions. For an infinite spherical distribution Δ for $Y_{\infty,\infty}$ corresponding to the consistent family $\{\Delta_{n,p}\colon n, p \geqslant 1\}$, we get doubly scaled distributions of the form $\Delta(H, K)$, with H and K being $m \times m$ and $q \times q$ arbitrary nonnegative-definite symmetric: $\Delta(H, K)$ is the distribution of $AY_{n,p}B$, where A, of order $m \times n$, and B, of order $p \times q$, satisfy $AA' = H$, $B'B = K$. For a rotatable distribution Π for $S_{\infty,\infty}$ we likewise get $\Pi(K)$ as the distribution of $B'S_{p,p}B$.

This notation for scale-parameter distributions agrees with that of our examples, but extends it in the sense that we now have no difficulty in defining $\Phi \sim IW(\delta;\, K)$, say, for singular K. Such a distribution is itself singular, with the column-space of Φ equalling, almost surely, that of K, so that the true dimensionality of this distribution is $\operatorname{rank}(K) < q$. This again demonstrates the importance of having a notation which avoids explicit dependence on dimensionality.

Similarly, we can define without difficulty the scaled matrix t distribution $T(\delta;\, H, K)$, in which H or K may be singular. Such singular matrix t distributions have been considered in a University of Wisconsin Technical Report by W. Y. Tan, and reported by Tan (1969b) and Johnson & Kotz (1972, Chapter 37). However, owing to the absence of suitable notation, Tan was driven to a complicated construction. In the full rank case, the notation of Dickey (1967) corresponding to our $T(\delta; H, K)$ would be $T(H^{-1}, K, 0, \delta + n + p - 1)$. This creates obvious difficulties if H is to be allowed to be singular, and less obvious ones, on account of the inappropriate 'degrees of freedom' parameter, if K is singular.

The catalogue of scale extensions of our standard examples of infinite distributions now consists of: $\mathcal{N}(H, K)$, $W(\nu;\, K)$ for ν an integer, $IW(\delta;\, K)$ for $\delta > 0$, $T(\delta;\, H, K)$ for $\delta > 0$, and $F(\nu, \delta;\, K)$ for ν an integer, $\delta > 0$. Similarly, we can define $B(\nu_1, \nu_2;\, K)$ so long as $\operatorname{rank}(K) < \nu_1 + \nu_2 + 1$ and also, for ν_i nonintegral, $\operatorname{rank}(K) < \nu_i + 1$; likewise $W(\nu;\, K)$ and $F(\nu, \delta;\, K)$ for nonintegral ν and K of suitably low rank.

Clearly, if Δ is left-spherical, and $Y \sim \Delta(H)$, then $AY \sim \Delta(AHA')$, with similar formulae in the other cases. Our scaled distributions are consistent in the sense that, for $Y_m \sim \Delta(H)$, with Δ left-spherical, we have $Y_{m_1} \sim \Delta(H_1)$ $(m_1 < m)$, with H_1 the leading $m_1 \times m_1$ submatrix of H; etc.

If $Y \sim \Delta(H, K)$, with Δ spherical, then, if these moments exist, $E(Y) = 0$ and $\mathrm{cov}(Y) = \lambda K \otimes H$ for some scalar λ: that is to say, $\mathrm{cov}(y_{ij}, y_{uv}) = \lambda h_{iu} k_{jv}$. For $S \sim \Delta(K)$, with Δ rotatable, we have $E(S) = \alpha K$,

$$\mathrm{cov}(s_{ij}, s_{uv}) = \beta(k_{iu} k_{jv} + k_{iv} k_{ju}) + \gamma k_{ij} k_{uv}.$$

7. Compound distributions

Let Π, Ξ be rotatable distributions, and construct a new rotatable distribution Ω by the requirement: $\Psi \sim \Omega(I_p)$ if, given Φ, $\Psi \sim \Xi(\Phi)$, where marginally $\Phi \sim \Pi(I_p)$. These distributions are consistent. We say that Ω is produced by compounding Ξ with Π, and write $\Omega = \Xi \circ \Pi$. The following result is related to that cited at the end of § 5.

Theorem 3. *We have* $\Xi \circ \Pi = \Pi \circ \Xi$.

Before showing this, we need two lemmas.

Lemma 2. *Let Φ be a $p \times p$ rotatable matrix, and let $\lambda_1 \geqslant \ldots \geqslant \lambda_p \geqslant 0$ be the eigenvalues of Φ, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. Then Φ has the same distribution as $P\Lambda P'$, where $P \sim \Upsilon_{p,p}$, independently of Λ.*

The proof of Lemma 2 is entirely analogous to that of Representation 2 of Dawid (1977).

Corollary. *The distribution of a rotatable matrix is determined by that of its eigenvalues.*

Lemma 3. *Let Φ be a $p \times p$ rotatable matrix. Then there exists a spherical distribution for a $p \times p$ matrix A such that AA' has the same distribution as Φ.*

We can take, for example, $A = P\Phi^{\frac{1}{2}} Q$, where $\Phi^{\frac{1}{2}}$ is the symmetric square root of Φ, and P, $Q \sim \Upsilon_{p,p}$, all independently.

Proof of Theorem 3. Let A, B be independent spherical $p \times p$ matrices with $AA' \sim \Pi(I_p)$, $BB' \sim \Xi(I_p)$, and let $\Psi_1 = ABB'A'$. Conditioning on A, we get $\Psi_1 \sim \Xi(\Phi)$, where $\Phi = AA' \sim \Pi(I_p)$ marginally. Thus $\Psi_1 \sim \Xi \circ \Pi(I_p)$. Now consider $\Psi_2 = B'A'AB$. This is rotatable, with the same eigenvalues as Ψ_1. Consequently, Ψ_2 has the same distribution as Ψ_1. But, since A is spherical, $A'A \sim \Pi(I_p)$ by Theorem 1, and similarly $B'B \sim \Xi(I_p)$. Conditioning on B, we find $\Psi_2 \sim \Pi \circ \Xi(I_p)$.

Note that the extendibility of the distributions is not needed for this proof.

We now outline two applications. Applying Theorem 3 to Example 5, we see that an alternative representation of $U \sim F(\nu, \delta;\, I_p)$ is $U \sim IW(\delta;\, \Lambda)$ given Λ, where marginally $\Lambda \sim W(\nu;\, I_p)$. It follows that, if $U \sim F(\nu, \delta;\, I_p)$, where $\nu > p-1$, then we can represent U^{-1} by $U^{-1} \sim W(\delta+p-1;\, \Gamma)$ given Γ, where $\Gamma = \Lambda^{-1} \sim IW(\nu-p+1;\, I_p)$. Thus, from the original definition, $U^{-1} \sim F(\delta+p-1, \nu-p+1;\, I_p)$.

Let $Z \sim \mathcal{N}(I_n, \Sigma)$ independently of $S \sim W(\nu;\, \Sigma)$, with Σ nonsingular of order $p \times p$, and $\nu > p-1$. Clearly the distribution of $U = ZS^{-1}Z'$ does not depend on Σ, so that we take $\Sigma = I_p$. We have, given Z, $U \sim IW(\nu-p+1;\, \Lambda)$, where $\Lambda = ZZ' \sim W(p;\, I_n)$. Thus

$U \sim F(p, \nu-p+1;\, I_n)$. For $n = 1$ we get, effectively, Hotelling's T^2, with $U \sim \{p/(\nu-p+1)\}\, F_{p,\nu-p+1}$.

Another application of the above lemmas yields the following result.

THEOREM 4. *Let* $U \sim F(\nu, \delta;\, I_p)$ *and* $V = (I_p+U)^{-1}$. *Then* $V \sim B(\delta+p-1, \nu;\, I_p)$.

Proof. Let $S_1 \sim W(\delta+p-1;\, I_p)$, $S_2 \sim W(\nu;\, I_p)$ independently, with $S_1^{-1} = AA'$, and A spherical. Then, since $A'A \sim IW(\delta;\, I_p)$, we can take $U = A'S_2 A$. The eigenvalues of U are the same as those of $AA'S_2 = S_1^{-1} S_2$, so that the eigenvalues of $(I_p+U)^{-1}$ are the same as those of $(I+S_1^{-1} S_2)^{-1} = S^{-1} S_1$, where $S = S_1+S_2$. These in turn are the same as those of $D^{-1} S_1 (D^{-1})'$ where $S = DD'$. But, by Corollary 2 to Lemma 1, we may choose D so that $D^{-1} S_1 (D^{-1})' \sim B(\delta+p-1, \nu;\, I_p)$. The result now follows from the Corollary to Lemma 2.

COROLLARY. *Let* $V \sim B(\nu_1, \nu_2;\, I_p)$ $(\nu_1 > p-1)$, *and* $U = V^{-1} - I_p$. *Then* $U \sim F(\nu_2, \nu_1-p+1;\, I_p)$.

8. PREDICTIVE DISTRIBUTIONS FOR THE MULTIVARIATE LINEAR MODEL

Suppose p variables are measured on n individuals, yielding data displayed as a $n \times p$ matrix X. The multivariate linear model postulates

$$X = \Gamma M + Z, \quad Z \sim \mathcal{N}(H, \Sigma), \tag{1}$$

where the parameters are M and Σ, of respective orders $m \times p$ and $p \times p$, and the $n \times m$ design matrix Γ is known, as is H. Normally, $H = I_n$.

Assume a prior distribution of the form,

$$M \,|\, \Sigma \sim M^* + \mathcal{N}(H^*, \Sigma), \quad \Sigma \sim \Pi(K), \tag{2}$$

the latter marginally, where M^*, H^* and K are known, and Π is rotatable. The standard conjugate distribution has this form with $\Pi = IW(\delta)$, but has various restrictive features (Press, 1972, §8·6·2); these are, unfortunately, mostly shared by our extension.

Suppose we partition $X = (X_1\ X_2)$, with X_i of order $n \times p_i$. We observe X_1 and wish to find the conditional predictive distribution of X_2. The direct way of proceeding would be as follows, in which, without loss of generality, we take $M^* = 0$.

First, for the model (1), standard normal theory implies that, given X_1,

$$X_2 \sim \Gamma(M_2 - M_1 B) + X_1 B + \mathcal{N}(H, \Sigma_{22\cdot 1}), \tag{3}$$

where we have partitioned $M = (M_1\ M_2)$ with M_i of order $m \times p_i$, and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

with Σ_{ij} of order $p_i \times p_j$, and defined $B = \Sigma_{11}^{-1}\Sigma_{12}$, $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. If Σ_{11} is singular, minor changes are required to the theory. We now have to marginalize (3) with respect to the posterior distribution for its parameters, $\Gamma(M_2 - M_1 B)$, B and $\Sigma_{22\cdot 1}$, based on observation of X_1 with distribution

$$X_1 \sim \Gamma M_1 + \mathcal{N}(H, \Sigma_{11}). \tag{4}$$

A first attempt at simplification would be to note that, given Σ, X is an independent sum of $\Gamma M \sim \mathcal{N}(\Gamma H^* \Gamma', \Sigma)$ and of $Z \sim \mathcal{N}(H, \Sigma)$. Thus, with $Q = H + \Gamma H^* \Gamma'$, given Σ,

$$X \sim \mathcal{N}(Q, \Sigma), \quad \Sigma \sim \Pi(K). \tag{5}$$

the latter marginally.

Comparing (5) with (1) and (2), we have succeeded in eliminating the design term ΓM. Thus we now have to marginalize the distribution $X_2 \sim X_1 B + \mathcal{N}(Q, \Sigma_{22\cdot 1})$ with respect to the posterior distribution for $(B, \Sigma_{22\cdot 1})$ given X_1, when $X_1 \sim \mathcal{N}(Q, \Sigma_{11})$. However, this programme may still be quite complicated. For the conjugate case it is straightforward, but lengthy.

An alternative revealing approach, which is often simpler, is as follows. Since Π is rotatable, (5) defines X as having the scale-modified spherical distribution $\Delta(Q, K)$, where Δ corresponds to Π as at the end of § 5. For example, if $\Pi = IW(\delta)$, the standard conjugate case, then $\Delta = T(\delta)$. Our problem is therefore equivalent to the general one of finding the conditional distribution of X_2 given X_1 when $X = (X_1\, X_2) \sim \Delta(Q, K)$. One way of doing this is by using the alternative representation of $\Delta(Q, K)$: given Λ,

$$X \sim \mathcal{N}(\Lambda, K), \tag{6}$$

where $\Lambda \sim \Pi(Q)$. For this representation, the conditional distributions are, given (X_1, Λ),

$$X_2 \sim X_1 D + \mathcal{N}(\Lambda, K_{22\cdot 1}), \tag{7}$$

where we have partitioned K as we did Σ, and defined $D = K_{11}^{-1} K_{12}$, $K_{22\cdot 1} = K_{22} - K_{21} K_{11}^{-1} K_{12}$. To obtain the distribution of X_2 given X_1 only, we marginalize (7) with respect to the posterior distribution of Λ given X_1. This is easily obtained from the model $X_1 \sim \mathcal{N}(\Lambda, K_{11})$ and prior $\Lambda \sim \Pi(Q)$. Indeed, a sufficient statistic for Λ based on X_1, assuming K_{11} nonsingular, is easily seen to be $S_1 = X_1 K_{11}^{-1} X_1'$, distributed as $W(p_1; \Lambda)$, so that the problem reduces to a general Bayesian analysis for the Wishart distribution; a simple transformation allows us to take $Q = I_n$, making the prior rotatable.

In this approach, the model is again simplified, but, more important, the same parameter Λ governs the marginal distribution of X_1, or S_1, and the conditional distribution of X_2 given X_1. As an example, if we have a conjugate prior, so that $\Sigma \sim IW(\delta;\, K)$ in (2), then all we need do is find the posterior distribution for Λ given S_1, when $S_1 \sim W(p_1; \Lambda)$ and *a priori* $\Lambda \sim IW(\delta; Q)$. This is standard, and gives $\Lambda \sim IW(\delta + p_1;\, Q + S_1)$. Insertion into (7) yields the predictive distribution for X_2 given only X_1, namely

$$X_2 \sim X_1 D + T(\delta + p_1;\, Q + X_1 K_{11}^{-1} X_1', K_{22\cdot 1}). \tag{8}$$

More generally, inspection of (7) immediately shows that, whenever the prior has the form (2), the best linear predictor of X_2 based on X_1 is $\Gamma(M_2^* - M_1^* D) + X_1 D$; interestingly, apart from the constant term, this does not involve the quantities (Γ, H) which occur in the model (1), nor (M^*, H^*) in (2), but only the scale matrix K of the prior distribution for Σ. In all cases, the predictive distribution of the discrepancy $X_2 - X_1 D$ depends on X_1 through $S_1 = X_1 K_{11}^{-1} X_1'$ only.

The above approach has as parameter Λ, an $n \times n$ covariance matrix, in contrast to the preceding approach involving the $p \times p$ matrix Σ. Normally $n \geqslant p$, so that analytical simplification is bought at the cost of increased computation; but if it should happen that $n < p$, then the computational burden is also eased by this method of analysis. The case

$n = 1$ is of some special interest, corresponding to prediction of unobserved from observed variables on a single individual.

This work has benefited from many valuable discussions with J. M. Dickey and J. B. Kadane.

REFERENCES

ANDO, A. & KAUFMAN, G. M. (1965). Bayesian analysis of the independent multinormal process—neither mean nor precision known. *J. Am. Statist. Assoc.* **60**, 347–58.

CORNISH, E. A. (1954). The multivariate t-distribution associated with a set of normal deviates. *Aust. J. Phys.* **7**, 531–42.

DAWID, A. P. (1977). Spherical matrix distributions and a multivariate model. *J. R. Statist. Soc.* B **39**, 254–61.

DAWID, A. P. (1978). Extendibility of spherical matrix distributions. *J. Mult. Anal.* **8**, 559–66.

DEMPSTER, A. P. (1969). *Elements of Continuous Multivariate Analysis*. Reading, Mass: Addison-Wesley.

DICKEY, J. M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.* **38**, 511–8.

EATON, M. L. (1972). *Multivariate Statistical Analysis*. Institute of Mathematical Statistics, University of Copenhagen.

FARRELL, R. H. (1962). Representation of invariant measures. *Ill. J. Math.* **6**, 447–67.

JOHNSON, N. L. &. KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

KHATRI, C. G. (1970). A note on Mitra's paper 'A density-free approach to the matrix variate beta distribution'. *Sankhyā* A **32**, 311–8.

MITRA, S. K. (1970). A density-free approach to the matrix variate beta distribution. *Sankhyā* A **32**, 81–8.

OLKIN, I. &. RUBIN, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist.* **35**, 261–9.

PRESS, S. J. (1972). *Applied Multivariate Analysis*. New York: Holt, Rinehart & Winston.

TAN, W. Y. (1969a). Note on the multivariate and the generalized multivariate beta distributions. *J. Am. Statist. Assoc.* **64**, 230–41.

TAN, W. Y. (1969b). Some results on multivariate regression analysis. *Nanta Math.* **3**, 54–71.

[*Received December* 1979. *Revised April* 1980]