JOURNAL OF MULTIVARIATE ANALYSIS **63**, 73–87 (1997)
ARTICLE NO. MV971689

Wishart and Pseudo-Wishart Distributions and Some Applications to Shape Theory

José A. Díaz-García*

Universidad Autónoma Agraria Antonio Narro

Ramón Gutierrez Jáimez

University of Granada

and

Kanti V. Mardia

University of Leeds

Suppose that $X \sim \mathcal{N}_{N\times m}(\mu, \Sigma, \Theta)$. An expression for the density function is given when $\Sigma \geqslant 0$ and/or $\Theta \geqslant 0$. An extension of Uhlig's result (Uhlig [17]) is expanded for the singular value decomposition of a matrix Z of order $N \times m$ when the rank $(Z) = q \leqslant \min(N, m)$. This paper fills an important gap in unifying, for the first time, all Wishart and pseudo-Wishart distributions, whether central or noncentral, whether singular or nonsingular, and applying them in shape analysis. In particular, the shape density and the size-and-shape cone density are obtained for the singular general case.

1. INTRODUCTION

Let $X \sim \mathcal{N}_{N\times m}(\mu, \Sigma, I_N)$, where Σ is positive definite, $\Sigma > 0$ of order $m \times m$ and $N \geqslant m$. Then, the random matrix $V = X'X$ has a noncentral/nonsingular Wishart distribution $\mathcal{W}_m(N, \Sigma, \Omega)$, with a matrix of noncentrality parameters defined by $\Omega = \Sigma^{-1}\mu'\mu$ (see Muirhead [12, p. 441], and

Received May 21, 1996; revised January 30, 1997.

AMS 1991 subject classification: Primary 62H10; secondary 62H11, 62E15, 62E17.

Key words and phrases: singular matrix distributions, Wishart distribution, Pseudo-Wishart distribution, Normal matrix variate distribution, Noncentral distributions, singular values, Stiefel manifold, shape, size-and-shape.

* This work was supported in part by the CONACYT, UAAAN, BANCO DE MEXICO and by the ANUIES. Dr. Díaz-García is currently a visiting professor in the Dept. of Statistics at the University of Granada, Spain.

0047-259X/97 $25.00

Srivastava and Khatri [14, p. 84]). Recently, Uhlig [7] extended this result to the central case when $N<m$, calling it a Wishart singular distribution. Goodall and Mardia [9], based on the QR decomposition of matrix $X=HT$, found the so-called size-and-shape density distribution of T for the central and noncentral case when $N<m$ and $N>m$. Only under the latter condition is it possible to determine the Wishart density, since when $N<m$, T is singular (see Graybill [10, p. 194]) and the Jacobian of the transformation $V=T'T$ is zero (see Muirhead [12, p. 60], and Srivastava and Khatri [14, p. 38]).

In this paper we study all the possible cases arising on defining the Wishart matrix, considering all the possible combinations among the following conditions: $N \geqslant m$, $N<m$, $\Sigma>0$, $\Sigma \geqslant 0$, $\Theta>0$, $\Theta \geqslant 0$, when $X \sim \mathcal{N}_{N \times m}(\mu, \Sigma, \Theta)$, with rank$(\Sigma)$, $r(\Sigma)=r \leqslant m$, $r(\Theta)=k \leqslant N$. Adopting the classification proposed by Srivastava and Khatri [14, p. 72] we obtain

$$\text{distribution} \begin{cases} \text{Wishart } r(\Sigma) \leqslant r(\Theta) & \begin{cases} \text{nonsingular} & r(\Sigma)=m \\ \text{singular} & r(\Sigma) \leqslant m \end{cases} \\ \text{pseudo-Wishart } r(\Sigma) > r(\Theta) & \begin{cases} \text{nonsingular} & r(\Sigma)=m \\ \text{singular} & r(\Sigma) \leqslant m \end{cases} \end{cases}$$

In this context, the singular distributions are an immediate consequence of the singular distribution of the normal matrix. An extension of the representation of the normal singular density given by Khatri (see Rao [15, p. 527]) is studied in Theorem 2.1 for the matrix case. The Jacobian calculation proposed by Uhlig [17] for the singular value decomposition, SVD, is generalised to the case of a matrix with order $N \times m$ and a rank $q \leqslant \min(N, m)$ (see Theorem 3.1). In Theorem 3.3 and Corollary 3.1 we unify, for the first time, the noncentral and central densities, respectively, for all the cases arising when defining a Wishart or pseudo-Wishart matrix (singular or nonsingular).

The results obtained are fundamental and very useful in different areas of multivariate statistics, such as, in the computation of likelihood estimators of μ and Σ when $X \sim \mathcal{N}_{N \times m}(\mu, \Sigma, \Theta)$, $\Sigma m \times m$, $\Theta N \times N$ known with $r(\Sigma)<m$ and/or $r(\Theta)<N$. That is, when there is a linear dependence between the variables (columns of X) and/or there is a linear dependence between the sample elements (rows of X). When vectors X_i, $i=1, ..., N$ are a sample of a normal distribution $\mathcal{N}_m(\mu, \Sigma)$ with $r(\Sigma)<m$ and independent $\Theta=I_N$, and using the representation of the normal singular vector density, Rao [15, p. 535] has demonstrated that the maximum likelihood estimators are the same for both the singular and nonsingular cases.

Srivastava and Khatri [14] studied the general multivariate linear model when there is a linear dependence between the sample and a correlation among its elements, that is, $r(\Theta) \leqslant N$, proposing in this case BLUE estimators for μ and Σ. In the problem of proving the multivariate linear hypothesis, the densities obtained here can be used to extend the results of the multivariate beta distribution (see Muirhead [12, p. 449], and Uhlig [17]) given in Díaz and Gutiérrez [4]. Similarly, the pseudo-Wishart densities play a fundamental role for updating a Bayesian posterior when tracking a time-varying variance-covariance matrix (see Uhlig [17]), thereby allowing the study of the singular case.

Thus, a great number of classic results in multivariate analysis can now be systematically studied using the theory of Wishart and pseudo-Wishart singular distributions. In particular, the unified Wishart and pseudo-Wishart distributions enable the development of an alternative approach to the proposed by Goodall and Mardia [8] and [9] in shape analysis. In other words, rather than considering QR decomposition, it is now possible to use singular value decomposition. As the density of $S = X'\Theta^{-}X$ exists for all S, whether singular or nonsingular, this may be considered the size-and-shape density from which we may obtain the shape density that defines the transformation $V = S/v$, where $v = \|S\| = (\operatorname{tr} S^2)^{1/2}$. In an analogous fashion, given the volume expression (dS) as a function of L and W_1 (see Theorems 3.1 and 3.2), it is possible to determine the density function of the nonzero eigenvalues of S, known as size-and-shape cone density in shape analysis terms, (see Goodall [7], and Goodall and Mardia [9]). This approach, together with other associated results, is extended by Díaz-García and Gutiérrez [5] to the case in which the matrix representation of the shape has an elliptically contoured distribution.

2. SINGULAR NORMAL DISTRIBUTION

The Stiefel manifold denoted by $V_{m,N}$ defines the matrix group such that if $H \in V_{m,N}$, $H'H = I_m$. Specifically, if $m = N$, $V_{m,N}$ will be noted as $\mathcal{O}(m)$. The next result with respect to the singular density of the normal random matrix is based on the SVD (see Remark 2.1) of matrices Σ and Θ. Using the complete decomposition, if Y is a matrix of oder $N \times m$ and of rank q, let us write $Y = H\Delta P'$, where $H \in \mathcal{O}(N)$

$$\Delta = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

D is diagonal matrix with $D_{11} > D_{22} > \cdots > D_{qq} > 0$ and $P \in \mathcal{O}(m)$. Alternatively, $Y = H_1 D P_1'$, where D is as above, $H_1 \in V_{q,N}$ and $P_1 \in V_{q,m}$ (see Rao [15, p.42]). This representation is called by some authors the non-singular representation of the singular value decomposition (see Uhlig [17]).

Remark 2.1. Note that when Y is symmetric, $H = P$. Thus, the spectral decomposition given by $Y = P\Delta P'$ is obtained. In an analogous fashion to the SVD, we can speak in terms of the nonsingular representation of the spectral decomposition, which occurs when Y is singular (see 1c.3(i) and (v) in Rao [15, pp. 39 and 42–43]).

THEOREM 2.1. *Let* $X \sim \mathcal{N}_{N\times m}(\mu, \Sigma, \Theta)$, *with* $\Sigma m\times m$, $r(\Sigma) = r < m$ *or* $\Theta N\times N$, $r(\Theta) = k < N$. *This distribution will be called a normal singular distribution and will be denoted as*

$$X \sim \mathcal{N}^{k,r}_{N\times m}(\mu, \Sigma, \Theta)$$

omitting the supra-index when $r = m$ *and* $k = N$. *In addition, its density function is given by*

$$\frac{1}{(2\pi)^{rk/2}(\prod_{i=1}^{r}\lambda_i^{k/2})(\prod_{j=1}^{k}\delta_j^{r/2})}\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-}(X-\mu)'\,\Theta^{-}(X-\mu)\right) \tag{1}$$

$$\left.\begin{aligned} H'_2XP'_1 &= H'_2\mu P'_1\\ H'_1XP'_2 &= H'_1\mu P'_2\\ H'_2XP'_2 &= H'_2\mu P'_2\end{aligned}\right\}\quad a.s., \tag{2}$$

where A^{-} *is a symmetric generalised inverse,* λ_i *and* δ_j *are the nonzero eigenvalues of* Σ *and* Θ *respectively. Let* $H = (H_1 \vdots H_2) \in \mathcal{O}(N)$ *and* $P = (P'_1 \vdots P'_2) \in \mathcal{O}(m)$ *be matrices associated with the spectral descomposition of matrices* Σ *and* Θ *respectively with* $H_1 \in V_{k,N}$, $H_2 \in V_{N-k,N}$, $P_1 \in V_{m,r}$ *and* $P_2 \in V_{m,m-r}$.

Proof. Let P_1 and H_1 correspond to the nonzero eigenvalues of Σ and Θ, respectively. Then,

$$P'\Sigma P = \begin{pmatrix} P_1\\ P_2\end{pmatrix}\Sigma(P'_1\ \ P'_2) = \begin{pmatrix} P_1\Sigma P'_1 & P_1\Sigma P'_2\\ P_2\Sigma P'_1 & P_2\Sigma P'_2\end{pmatrix} = \begin{pmatrix} D_\Sigma & 0\\ 0 & 0\end{pmatrix} = \Delta_\Sigma$$

$$H'\Theta H = \begin{pmatrix} H'_1\\ H'_2\end{pmatrix}\Theta(H_1\ \ H_2) = \begin{pmatrix} H'_1\Theta H_1 & H'_1\Theta H_2\\ H'_2\Theta H_1 & H'_2\Theta H_2\end{pmatrix} = \begin{pmatrix} D_\Theta & 0\\ 0 & 0\end{pmatrix} = \Delta_\Theta.$$

Consider now the transformation

$$\begin{pmatrix} H'_1\\ H'_2\end{pmatrix} X(P'_1\ \ P'_2) = \begin{pmatrix} H'_1XP'_1 & H'_1XP'_2\\ H'_2XP'_1 & H'_2XP'_2\end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12}\\ Y_{21} & Y_{22}\end{pmatrix} = Y.$$

Then

$$\begin{aligned} &E(Y_{12}) = H'_1 \mu P'_2 & & \mathrm{Cov}(\mathrm{vec}\ Y'_{12}) = H'_2 \Theta H_2 \otimes P_1 \Sigma P'_1 = 0 \\ &E(Y_{21}) = H'_2 \mu P'_1 & \text{and} \quad & \mathrm{Cov}(\mathrm{vec}\ Y'_{21}) = H'_1 \Theta H_1 \otimes P_2 \Sigma P'_2 = 0 \\ &E(Y_{22}) = H'_2 \mu P'_2 & & \mathrm{Cov}(\mathrm{vec}\ Y'_{22}) = H'_2 \Theta H_2 \otimes P_2 \Sigma P'_2 = 0, \end{aligned}$$

where, if A is a $p \times q$ matrix then by $\mathrm{vec}(A)$ we mean the $pq \times 1$ vector formed by stacking the columns of A under each other. In addition,

$$E(Y_{11}) = H'_1 \mu P'_1 \quad \text{and} \quad \mathrm{Cov}(\mathrm{vec}\ Y'_{11}) = H'_1 \Theta H_1 \otimes P_1 \Sigma P'_1$$

and therefore

$$Y_{11} \sim \mathcal{N}_{k \times r}(H'_1 \mu P'_1, P_1 \Sigma P'_1, H'_1 \Theta H_1).$$

Thus, the joint distribution of Y or, equivalently, of X is given by

$$\frac{\mathrm{etr}(-\frac{1}{2}(P_1 \Sigma P'_1)^{-1}(Y_{11} - H'_1 \mu P'_1)'(H'_1 \Theta H_1)^{-1}(Y_{11} - H'_1 \mu P'_1))}{(2\pi)^{rk/2} |P_1 \Sigma P'_1|^{k/2} |H'_1 \Theta H_1|^{r/2}} \tag{3}$$

$$\left.\begin{aligned} Y_{12} &= H'_1 \mu P'_2 \\ Y_{21} &= H'_2 \mu P'_1 \\ Y_{22} &= H'_2 \mu P'_2 \end{aligned}\right\} \quad \text{a.s.} \tag{4}$$

Now observe that $|P_1 \Sigma P'_1| = |D_\Sigma| = \prod_{i=1}^r \lambda_i$ and $|H'_1 \Theta H_1| = |D_\Theta| = \prod_{j=1}^k \delta_j$, where λ_i and δ_j are the nonzero eigenvalues of Σ and Θ respectively. Further, $\Sigma = P'_1 D_\Sigma P_1$ and $\Theta = H_1 D_\Theta H'_1$, and then $\Sigma^- = P'_1 D_\Sigma^{-1} P_1 = P'_1 (P_1 \Sigma P'_1)^{-1} P_1$ and $\Theta^- = H_1 D_\Theta^{-1} H'_1 = H_1 (H'_1 \Theta H_1)^{-1} H'_1$; hence

$$\begin{aligned} &\mathrm{tr}(-\tfrac{1}{2}(P_1 \Sigma P'_1)^{-1}(Y_{11} - H'_1 \mu P'_1)'(H'_1 \Theta H_1)^{-1}(Y_{11} - H'_1 \mu P'_1)) \\ &\quad = \mathrm{tr}(-\tfrac{1}{2} P'_1 (P_1 \Sigma P'_1)^{-1} P_1 (X - \mu)' H_1 (H'_1 \Theta H_1)^{-1} H'_1 (X - \mu)) \\ &\quad = \mathrm{tr}(-\tfrac{1}{2} \Sigma^- (X - \mu)' \Theta^- (X - \mu)) \end{aligned}$$

from which we can obtain the desired result. ∎

It is note that this density is valid for any combination between r and k. Especially if $r = m$ or $k = N$, then $\Sigma^- = \Sigma^{-1}$ or $\Theta^- = \Theta^{-1}$ or $\prod_{i=1}^r \lambda_i = |\Sigma|$ or $\prod_{j=1}^k \delta_j = |\Theta|$, thereby obtaining the nonsingular multivariate normal density.

Before continuing, it is important to make some comments on the singular density of a matrix.

Remark 2.2. The density specified by (1) and (2) is interpreted as the density on the subspace determined by (2). The density defined by (3) and (4) is analogously interpreted.

Remark 2.3. Given that the random matrices Y_{12}, Y_{21} and Y_{22} are degenerate, that is, they satisfy (4), a measure for matrix Y is given by the cartesian product $C_1\{dY_{12}\} \times C_2\{dY_{21}\} \times C_3\{dY_{22}\} \times \lambda_{k \times r}\{dY_{11}\}$, in which $C_i\{\cdot\}$ is the count measure. In this case, there is only one point specified in (4), for $i = 1, 2, 3$ and $\lambda_{k \times r}\{\cdot\}$ is the Lebesgue measure on $\mathsf{R}^{k \times r}$. This is a special case of the Hausdorff measure, described by $h_{k \times r}(dY)$ (see Billingsley [1, pp. 208–218]). Thus, taking into consideration the notation used in Billingsley, the set $A \in \mathsf{R}^{N \times m}$ is defined by (4).

Remark 2.4. Note that the representations of densities (1)–(2) and (3)–(4) are not unique, as P and H are not, either. Nevertheless, for any fixed P and H, the resulting densities may be employed, for example, to determine the probability of an event or to estimate the parameters μ, Σ or Θ, without these results depending on the selected P and H matrices.

Remark 2.5. Finally, note that the generalised symmetric inverse is not unique. This may be, for example, a reflexive, generalised or Moore–Penrose inverse (see Remark 2.1 and problem 28 in Rao [15, pp. 76–77]).

3. WISHART AND PSEUDO-WISHART DISTRIBUTIONS

Consider now the definition of the Wishart matrix. It is obvious that its singularity is given by two reasons: (i) that $r < m$ and/or (ii) that $r > k$. Although other authors refer to both cases as a singular Wishart distribution, we have opted for Srivastava and Khatri's classification, where case (ii) is called a pseudo-Wishart distribution, with the purpose of differentiating the cause of the singularity of the Wishart matrix. According to the normal random matrix, (ii) implies that the number of observations is less than the number of study variables or that there is a linear dependence between the sample elements up to when $k < r$. In contrast, case (i) actually occurs when there is a linear dependence between the study variables. As will be seen below, the singularity of the pseudo-Wishart matrix can occur due to either motive or to both. This fact is very important, since when defining multivariate beta and F distributions, they exist at present only under the singularity of the type (ii) Wishart distributions (see Díaz-García and Gutiérrez [4]). On defining the Wishart (pseudo-Wishart) matrix, the sample is considered to be uncorrelated; nevertheless, as mentioned above, certain authors do take them as being correlated (see Srivastava and Khatri [4]). The same situation occurs in the analysis of time series.

Let us therefore suppose that $Y \sim \mathcal{N}^{k,r}_{N \times m}(\mu. \Sigma, \Theta)$ and define the Wishart matrix as

$$S = Y'\Theta^{-} Y,$$

where Θ^{-} is a symmetric generalised inverse of Θ (see Remark 2.5).

Then, if S has a noncentral Wishart distribution, we denote

$$S \sim \mathcal{W}^{q}_{m}(k, \Sigma, \Omega)$$

and if S has a noncentral pseudo-Wishart distribution, we denote

$$S \sim \mathcal{P}\mathcal{W}^{q}_{m}(k, \Sigma, \Omega),$$

where q is the rank of S and is written only if the distribution is singular, m is the dimension of matrix S, k is the degree of freedom, and Ω is the matrix of noncentrality parameters defined by $\Omega = \Sigma^{-}\mu'\Theta^{-}\mu$. In the case of the nonsingular pseudo-Wishart distribution, k by definition will also denote its rank.

The possible cases that can present themselves in practice when defining the random Wishart or Pseudo-Wishart matrix are shown in Table I.

Note that if $r(S) = q < m$, the densities do not exist on the space of the symmetric $m \times m$ matrices. However, the density does exist on a manifold of rank-q positive semidefinite $m \times m$ matrices with q distinct positive eigenvalues; the manifold is written as $\mathcal{S}^{+}_{m,q}$ (see Uhlig [17]). It is plain to see that these twelve practical cases can be classified within the four distributions proposed by Srivastava and Khatri [14].

TABLE I

Case	X	$r = r(\Sigma)$	$k = r(\Theta)$	$q = r(S)$
1	$N \geqslant m$	m	N	m
2		m	$m \leqslant k < N$	m
3		m	$k < m$	k
4		m	$k < m$	k
5		$r < m$	$r \leqslant k < N$	r
6		$r < m$	$k < r$	k
7	$N < m$	m	N	N
8		$N < r < m$	N	N
9		$r \leqslant N$	N	r
10		m	$k < N$	k
11		$k < r < m$	$k < N$	k
12		$r \leqslant k$	$k < N$	r

The next result is an extension of Theorem 5 in Uhlig [17]:

THEOREM 3.1. *Let Z be a matrix of order $N \times m$ with $r(Z) = q \leqslant \min(N, m)$ and let $Z = H_1 D W_1'$ be the nonsingular part of the SVD, where $H_1 \in V_{q,N}$, D is a diagonal matrix with $D_{11} > D_{22} > \cdots > D_{qq} > 0$ and $W_1 \in V_{q,m}$. Then,*

$$(dZ) = 2^{-q} |D|^{N+m-2q} \prod_{i<j}^{q} (D_{ii}^2 - D_{jj}^2)(dD)(H_1' \, dH_1)(W_1' \, dW_1),$$

where $(D) = \bigwedge_{i=1}^{q} dD_{ii}$ and $(H_1' \, dH_1)$ defines the unnormalized invariant probability measure on $V_{q,N}$ (see Muirhead [12, pp. 67–72]).

Proof. Let $H_2 N \times (N-q)$ matrix be such that $H = (H_1 \vdots H_2) = (h_1, ..., h_q \vdots h_{q+1}, ..., h_N) \in \mathcal{O}(N)$ where $h_1, ..., h_q$ are the columns of H_1, $h_{q+1}, ..., h_N$ are the columns of H_2 and W_2 $m \times (m-q)$ matrix such that $W = (W_1 \vdots W_2) = (w_1, ..., w_q \vdots w_{q+1}, ..., w_m) \in \mathcal{O}(m)$ where $w_1, ..., w_q$ are the columns of W_1, $w_{q+1}, ..., w_N$ are the columns of W_2. Then, given that

$$dZ = dH_1 D W_1' + H_1 \, dD W_1' + H_1 D dW_1'$$

we obtain

$$H' dZ W = \begin{pmatrix} H_1' dH_1 D + dD + D dW_1' W_1 & D dW_1' W_2 \\ H_2' dH_1 D & 0 \end{pmatrix}.$$

In addition, $(H' dZ W) = |H'|^m \, |W|^N (dZ) = (dZ)$, and therefore the exterior product will be given by the exterior product on the right-hand side of the above expression. Thus, proceeding analogously to Uhlig [7] we see that

$$(H_2' dH_1 D) = |D|^{N-q} \bigwedge_{i=1}^{q} \bigwedge_{j=q+1}^{N} h_j' \, dh_i$$

$$\begin{aligned}(H_1' dH_1 D &+ dD + D dW_1' W_1) \\ &= \bigwedge_{i<j}^{q} (D_{ii}^2 - D_{jj}^2) \bigwedge_{i=1}^{q} dD_{ii} \wedge \bigwedge_{i=1}^{q} \bigwedge_{j=i+1}^{q} h_j' \, dh_i \wedge \bigwedge_{i=1}^{q} \bigwedge_{j=i+1}^{q} w_j' \, dw_i\end{aligned}$$

$$(W_2' \, dW_1 D)' = |D|^{m-q} \bigwedge_{i=1}^{q} \bigwedge_{j=q+1}^{m} w_j' \, dw_i.$$

The product of the right-hand side of the last three expressions gives us

$$|D|^{N+m-2q} \prod_{i<j}^{q} (D_{ii}^2 - D_{jj}^2)(dD)(H_1' \, dH_1)(W_1' \, dW_1)$$

Finally, we must divide by 2^q due to the arbitrary assignment of signs to the columns of H_1, since Z is the image of 2^q decompositions. We thereby obtain the desired result. ▮

From Theorem 2 in Uhlig [17] we can ascertain that, given $S \in \mathcal{S}^+_{m,q}$,

$$S = Z'Z = W_1 D H'_1 H_1 D W'_1 = W_1 L W'_1$$

with $W_1 \in V_{q,m}$ and $L = D^2$ diagonal matrix, $l_1 > \cdots > l_q > 0$, then

$$(dS) = 2^{-q} \prod_{i=1}^{q} l_i^{m-q} \prod_{i<j}^{q} (l_i - l_j)(W'_1\, dW_1) \wedge \bigwedge_{i=1}^{q} dl_i \tag{5}$$

THEOREM 3.2. *With the assumptions of Theorem* 3.1,

$$(dZ) = 2^{-q}\, |L|^{(N-m-1)/2} (dS)(H'_1\, dH_1),$$

where $S = W_1 L W'_1$.

Note that when $q = m$, $|L| = |W'_1 S W_1| = |S|$ since $W_1 \in \mathcal{O}(m)$.

Proof. The proof is a close copy of the proof for Muirhead's Theorem 2.1.14 (see Muirhead [12, p. 66]), taking into account as well that $L = D^2$ and

$$\bigwedge_{i=1}^{q} dl_i = 2^q \bigwedge_{i=1}^{q} D_{ii} \bigwedge_{i=1}^{q} dD_{ii}. \quad ▮$$

We are now in a position to establish in a unified fashion the densities of the noncentral singular or nonsingular Wishart or pseudo-Wishart matrix.

THEOREM 3.3. *Let us suppose that* $\mathcal{N}^{k,r}_{N\times m}(\mu, \Sigma, \Theta)$, *with* $r(\Sigma) = r \leqslant m$, $r(\Theta) = k \leqslant N$ *and let* $q = \min(r, k)$; *then the density of* $S = Y'\Theta^- Y$ *is given by*

$$\frac{\pi^{k(q-r)/2}\, |L|^{(k-m-1)/2}}{2^{kr/2}\, \Gamma_q[\frac{1}{2}k](\prod_{i=1}^{r} \lambda_i^{k/2})} \operatorname{etr}\left(-\frac{1}{2}\Sigma^- S - \frac{1}{2}\Omega\right) {}_0F_1\left(\frac{1}{2}k; \frac{1}{4}\Omega\Sigma^- S\right) \tag{6}$$

$$P_2 S P'_2 = P_2 \mu' \Theta^- \mu P'_2 \quad a.s., \tag{7}$$

where $S = W_1 L W'_1$, Σ^- *is a symmetric generalised inverse of* Σ (*see Remark* 2.5), $\Omega = \Sigma^- \mu' \Theta^- \mu$, Θ^- *is a symmetric generalised inverse of* $\Theta = Q'Q$ *with* $Q k \times N$ *matrix*, $r(Q) = k$ (*see Remark* 2.5), $P'\Sigma P = \Delta_\Sigma$ (*see Theorem* 2.1), *and* ${}_0F_1(\cdot)$ *is a hypergeometric function with a matrix argument* (*see Muirhead* [12, p. 258]).

Proof. First observe that $\Theta = Q'Q$ with Q of order $k \times N$ of $r(Q)$ such that

$$(Q^-)'\,\Theta Q^- = (Q^-)'\,(Q'Q)\,Q^- = (QQ^-)'\,(QQ^-) = (QQ^-)(QQ^-) = QQ^- = I_k$$

since $r(Q) = r(Q) = k$ and QQ^- is of order $k \times k$.

Now define $X = (Q^-)'\,Y$ and then

$$X \sim \mathcal{N}^{k,r}_{k \times m}((Q^-)'\,\mu, \Sigma, (Q^-)'\,\Theta Q^-) = \mathcal{N}^{k,r}_{k \times m}(\mu_x, \Sigma, I_k),$$

where $\mu_x = (Q^-)'\,\mu$. Therefore,

$$S = Y'\Theta^- Y = Y'(Q'Q)^- Y = Y'Q^-(Q^-)'\,Y = ((Q^-)'\,Y)'\,(Q^-)'\,Y) = X'X.$$

Now consider the SVD of matrix $X = U_1 D W'_1$,

$$S = Y'\Theta^- Y = X'X = W_1 D U'_1\, U_1 D W'_1 = W_1 L W'_1$$

where $L = D^2$. By Theorem 3.2,

$$(dX) = 2^{-q}\,|L|^{(k-m-1)/2}(dS)(U'_1\, dU_1),$$

where $q = r(Y) = r(X) = r(S) = \min(r, k)$. From Theorem 2.1, the density of X is given by

$$\frac{1}{(2\pi)^{rk/2}(\prod_{i=1}^{r} \lambda_i^{k/2})}\, \operatorname{etr}\left(-\frac{1}{2}\,\Sigma^-(X-\mu_x)'\,(X-\mu_x)\right)(dX)$$

$$XP'_2 = \mu_x P'_2 \quad \text{a.s.}$$

Considering only the nondegenerate part of X, the joint density of U_1, S, W_1 is

$$\frac{2^{-q}\,|L|^{(k-m-1)/2}}{(2\pi)^{rk/2}(\prod_{i=1}^{r} \lambda_i^{k/2})}\, \operatorname{etr}\left(-\frac{1}{2}\,\Sigma^- S - \frac{1}{2}\,(\Sigma^- \mu'_x \mu_x)\right)$$

$$\times \operatorname{etr}(\Sigma^- W_1 D U'_1 \mu_x)(dS)(U'_1 dU)$$

The following integral is taken from Muirhead [12, p. 262] or James [11]. Note that it is the same if $k \geqslant r$ or $k < r$, Goodall and Mardia [9]:

$$\int_{U_1 \in V_{q,k}} \operatorname{etr}(\mu_x \Sigma^- W_1 D U_1')(U_1' dU_1)$$

$$= \frac{2^q \pi^{qk/2}}{\Gamma_q(\frac{1}{2}k)} \, {}_0F_1\left(\frac{1}{2}k; \frac{1}{4}\mu_x \Sigma^- W_1 L W_1' \Sigma^- \mu_x'\right)$$

$$= \frac{2^q \pi^{qk/2}}{\Gamma_q(\frac{1}{2}k)} \, {}_0F_1\left(\frac{1}{2}k; \frac{1}{4}\mu_x \Sigma^- S \Sigma^- \mu_x'\right)$$

$$= \frac{2^q \pi^{qk/2}}{\Gamma_q(\frac{1}{2}k)} \, {}_0F_1\left(\frac{1}{2}k; \frac{1}{4}(Q^-)' \mu \Sigma^- S \Sigma^- \mu' Q^-\right)$$

$$= \frac{2^q \pi^{qk/2}}{\Gamma_q(\frac{1}{2}k)} \, {}_0F_1\left(\frac{1}{2}k; \frac{1}{4}\Omega \Sigma^- S\right);$$

this latter expression is obtained by substituting $\mu_x = (Q^-)'\mu$, $\Omega = \Sigma^- \mu' \Theta^- \mu$ and observing that the matrix argument of ${}_0F_1$, when this is developed in zonal polynomials, is valid, since $\Omega\Sigma^- = \Sigma^- \mu' \Theta^- \mu \Sigma^-$ and S are symmetric matrices (see Takemura [16, p. 25], and Davis [3]).

If we now analyse the degenerate part, $P((X - \mu_x) P_2' = 0) = 1$, we see that

$$P_2 X' X P_2' = P_2 \mu_x' \mu_x P_2' = P_2 \mu' \Theta^- \mu P_2' = P_2 S P_2',$$

then

$$P_2 S P_2' = P_2 \mu' \Theta^- \mu P_2' \quad \text{a.s.} \quad \blacksquare$$

Remark 3.1. Density (6) is interpreted as the density on subspace (7), with respect to the volume given in (5) (see Remark 2.3).

If the ranks of Σ and Θ are chosen appropriately in Theorem 3.3, we will obtain the corresponding expressions for each of the densities for the singular or nonsingular Wishart or pseudo-Wishart cases. For example, if $q = r = m$, $\Theta = I_N$, we get the nonsingular, noncentral Wishart density. In that case, $\prod_{i=1}^m \lambda_i = |\Sigma|$, $W_1 \in \mathcal{O}(m)$, $|L| = |S|$ and

$${}_0F_1(\tfrac{1}{2}k; \tfrac{1}{4}\Omega\Sigma^- S) = {}_0F_1(\tfrac{1}{2}N; \tfrac{1}{4}\Omega\Sigma^{-1} S).$$

This is the only specific case studied in the literature and presented in Muirhead [12, p. 442] of James [11], among others.

Below we apply this result to the central case. It is plain to see that this contains the particular density case proposed by Uhlig [17] (see Theorem 6), for which only $q = k = n$ and $r = m$ are taken.

COROLLARY 1. *Suppose that* $Y \sim \mathcal{N}^{k,r}_{N\times m}(0, \Sigma, \Theta)$, *with* $r(\Sigma)=r\leqslant m$, $r(\Theta)=k\leqslant N$ *and let* $q=\min(r,k)$, *then the density of* $S=Y'\Theta^{-}Y$ *is given by*

$$\frac{\pi^{k(q-r)/2}\,|L|^{(k-m-1)/2}}{2^{kr/2}\Gamma_q[\frac{1}{2}k](\prod_{i=1}^{r}\lambda_i^{k/2})}\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-}S\right)$$

$$P_2SP_2'=kP_2\Sigma P_2'=0 \quad a.s.,$$

where $S=W_1LW_1'$, Θ^{-} *is a symmetric generalised inverse of* Θ *and* Σ^{-} *is a generalised inverse of* Σ (*see Remark* 2.5 *and Remark* 3.1).

Finally, note that all the properties of the distributions defined here coincide in general with the nonsingular cases and can be verified with the aid of the respective characteristic functions, since in both cases these functions are valid for the densities studied here (see Srivastava and Khatri [14, pp. 42, 84], and Eaton [6, pp. 304–306]).

4. SHAPE THEORY

Two cases occur within shape theory when the Cholesky decomposition of the $X=HT$ matrix is considered (see Goodall and Mardia [9]). If $S=X'X$ is nonsingular, the Wishart distribution is used; however, if S is singular, the $T'T$ distribution is used. When the SVD is applied to matrix X, the density of S for all possible cases is obtained (see Theorem 3.3) and the above mentioned distinction need not be made.

In this section, suppose that X is invariant to transformations of the shape Y with the density $S=X'\Theta^{-}X$ as given in Theorem 3.3 (see Goodall and Mardia [9]) and Σ is nonsingular. We then obtain the following results:

THEOREM 4.1. *Let* $q=\min(m,k)$, *then the shape density of the matrix* $V=S/v$ *is given by*

$$\frac{\pi^{k(q-m)/2}\,|W_1'VW_1|^{(k-m-1)/2}\mathbf{J}(\mathbf{u})}{2^{mk/2}\Gamma_q(\frac{1}{2}k)\,|\Sigma|^{k/2}}$$

$$\times\operatorname{etr}\left(-\frac{1}{2}\Omega\right)\sum_{f=0}^{\infty}\sum_{\lambda}\frac{2^{f+km/2}\Gamma(f+km/2)}{(\operatorname{tr}\Sigma^{-1}V)^{f+km/2}(k/2)_\lambda}\,\frac{C_\lambda(\frac{1}{4}\Omega\Sigma^{-1}V)}{f!},$$

where $S=Y'\Theta^{-}Y$, $v=\|S\|=\operatorname{tr}S^2$, $C_\lambda(\cdot)$ *denotes the zonal polynomial* (*see James* [11]), W_1 *defined by Theorem* 3.3 *and* $(k/2)_\lambda$ *is the generalized hypergeometric coefficient.*

Proof. First, let $q<m$. Then S has $m-q$ rows (or columns) that are linearly dependent, without there being any loss of generality, and suppose the first q rows to be linearly independent. Therefore, accoding to Nel [13, pp. 157–169], and Goodall and Mardia [9], for $q=\min(m,k)$ it follows that $J(S\to v, V)=v^{((m-q)q+q(q+1)/2)-1}\,\mathbf{J}(\mathbf{u})$. Expanding this in zonal polynomials ${}_0F_1$ in the density of S (see James [11]) and performing the change of variable, the joint density of v and V is given by

$$\frac{\pi^{k(q-m)/2}\,|W_1VW_1'|^{(k-m-1)/2}\,v^{(km-2)/2}\,\mathbf{J}(\mathbf{u})}{2^{mk/2}\,\Gamma_q(\tfrac{1}{2}k)\,|\Sigma|^{k/2}}$$

$$\times\operatorname{etr}\left(-\frac{1}{2}(\Omega+v\Sigma^{-1}V)\right)\sum_{f=0}^{\infty}\sum_{\lambda}\frac{v^f C_\lambda(\tfrac{1}{4}\Omega\Sigma^{-1}V)}{(k/2)_\lambda\, f!}.$$

Adding the exponents of v, we then have

$$\int_0^\infty v^{f+(km-2)/2}\operatorname{etr}\left(\frac{v}{2}\Sigma^{-1}V\right)dv=\frac{2^{f+km/2}\,\Gamma(f+km/2)}{(\operatorname{tr}\Sigma^{-1}V)^{f+km/2}},$$

producing the result. ∎

Finally, considering the notation and the invariant polynomials proposed by Davis [3], and Chikuse [2], it is possible to obtain a general expression for the size-and-shape cone density.

THEOREM 4.2. *The size-and-shape cone density or joint density of l_i, $i=1,2,\ldots,q$ is*

$$\frac{\pi^{k(q-m)/2}\prod_{i=1}^{q} l_i^{(k+m-1)/2-q}\prod_{i<j}(l_i-l_j)}{2^{mk/2}\,\Gamma_q(\tfrac{1}{2}k)\,\Gamma_m(\tfrac{1}{2}q)\,|\Sigma|^{k/2}\operatorname{etr}(\tfrac{1}{2}\Omega)}$$

$$\times\sum_{\kappa,\lambda}^{\infty}\sum_{\phi\in\kappa,\lambda}\frac{\theta_\phi^{\kappa,\lambda}}{(k/2)_\lambda\,C_\phi(I_m)}\,\frac{C_\phi^{\kappa,\lambda}(-\tfrac{1}{2}\Sigma^{-1},\tfrac{1}{4}\Omega\Sigma^{-1})\,C_\phi(L)}{l!\,f!}$$

with $q=\min(k,m)$, $S=W_1LW_1'$ and $\theta_\phi^{\kappa,\lambda}$ and $C_\phi^{\kappa,\lambda}$ given in Davis [3].

Proof. Expressing the volume (dS) in terms of L and W_1 (see Theorem 3.1), the joint density of L and W_1 is given by

$$\frac{\pi^{k(q-m)/2}\prod_{i=1}^{q} l_i^{(k+m-1)/2-q}\prod_{i<j}(l_i-l_j)}{2^{mk/2+q}\,\Gamma_q(\tfrac{1}{2}k)\,\Gamma_m(\tfrac{1}{2}q)\,|\Sigma|^{k/2}\operatorname{etr}(\tfrac{1}{2}\Omega)}$$

$$\times\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}W_1LW_1'\right){}_0F_1\left(\frac{1}{2}k;\frac{1}{4}\Omega\Sigma^{-1}W_1LW_1'\right)(W_1'\,dW_1)\,dL$$

with $dL = \bigwedge_{i=1}^{q} dl_i$. Thus, to find the density of $l_1, l_2, ..., l_q$ it is necessary to evaluate the integral

$$I = \int_{W_1 \in V_{q,m}} \operatorname{etr}(-\tfrac{1}{2}\Sigma^{-1} W_1 L W_1')\, {}_0F_1(\tfrac{1}{2}k; \tfrac{1}{4}\Omega\Sigma^{-1} W_1 L W_1')(W_1' dW_1).$$

For Lemma 9.5.3 in Muirhead [12, p. 397]

$$I = \frac{2^q \pi^{qm/2}}{\Gamma_m(\frac{1}{2}q)} \int_{\mathcal{O}(m)} \operatorname{etr}(-\tfrac{1}{2}\Sigma^{-1} W_1 L W_1')\, {}_0F_1(\tfrac{1}{2}k; \tfrac{1}{4}\Omega\Sigma^{-1} W_1 L W_1')(dW),$$

where (dW) defines the normalized invariant probability measure on $\mathcal{O}(m)$ (see Muirhead [12, p. 72]). Now, on expanding ${}_0F_1$ and the exponential in zonal polynomials (see James [11]), and using the notation of Davis [3] for the sums

$$I = \frac{2^q \pi^{qm/2}}{\Gamma_m(\frac{1}{2}q)} \sum_{\kappa,\lambda}^{\infty} \frac{1}{(k/2)_\lambda\, l!\, f!} \times \int_{\mathcal{O}(m)} C_\kappa\left(-\frac{1}{2}\Sigma^{-1} W_1 L W_1'\right) C_\lambda\left(\frac{1}{4}\Omega\Sigma^{-1} W_1 L W_1'\right)(dW).$$

Applying 4.13 in Davis [3]

$$I = \frac{2^q \pi^{qm/2}}{\Gamma_m(\frac{1}{2}q)} \sum_{\kappa,\lambda}^{\infty} \sum_{\phi \in \kappa,\lambda} \frac{1}{(k/2)_\lambda\, l!\, f!} \frac{C_\kappa(-\frac{1}{2}\Sigma^{-1}, \frac{1}{4}\Omega\Sigma^{-1})\, C_\phi^{\kappa,\lambda}(L, L)}{C_\phi(I_m)}$$

The result follows from 5.1 in Davis [3]. ▮

Goodall [7], and Goodall and Mardia [9] present some applications of the results given here, using the Cholesky decomposition.

ACKNOWLEDGMENTS

The authors are grateful for the suggestions made by two anonymous referees. This research has been supported in part by DGICYT, grant PB94–1041.

REFERENCES

[1] Billingsley, P. (1979). *Probability and Measure*. Wiley, New York.

[2] Chikuse, Y. (1980). Invariant polynomials with matrix arguments and their applications. In *Multivariate Statistical Analysis* (R. P. Gupta, Ed.), pp. 53–68. North-Holland, Amsterdam.

[3] Davis, A. W. (1980). Invariant polynomials with two matrix arguments, extending the zonal polynomials. In *Multivariate Analysis V* (P. R. Krishnaiah, Ed.), pp. 287–299. North-Holland, Amsterdam.

[4] Díaz-García, J. A., and Gutiérrez, J. R. (1997). Proof of the conjectures of H. Uhlig on the singular multivariate beta and the jacobian of a certain matrix transformation. *Ann. Statist.* **25** u.5, to appear.

[5] Díaz-García, J. A., and Guttiérrez, J. R. (1995). Singular and nonsingular distributions in shape theory and the elliptical models. Submitted for publication.

[6] Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York.

[7] Goodall, C. R. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *J. Roy. Statist. Soc. Ser. B* **53** 285–339.

[8] Goodall, C. R., and Mardia, K. V. (1992). The noncentral Bartlett decompositions and shape densities. *J. Multivariate Anal.* **40** 94–108.

[9] Goodall, C. R., and Mardia, K. V. (1993). Multivariate aspects of Shape Theory. *Ann. Statist.* **21** 848–866.

[10] Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics.* Wadsworth, Belmont, CA.

[11] James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 473–501.

[12] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[13] Nel, D. G. (1989). On matrix differentiation in statistics. *South African Statist. J.* **14** 137–193.

[14] Srivastava, M. S., and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, Amsterdam.

[15] Rao, C. R. (1973). *Linear Statistical Inference and its Applications (2nd ed.).* Wiley, New York.

[16] Takemura, A. (1984). *Zonal Polynomials.* Institute of Mathematical Statistics Lecture Notes Monograph Series (S. S. Gupta, Séries Ed.).

[17] Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions. *Ann. Statist.* **22** 395–405.