Ann. Inst. Statist. Math.
Vol. 47, No. 4, 767–783 (1995)

APPROXIMATING BY THE WISHART DISTRIBUTION

TÕNU KOLLO[1] AND DIETRICH VON ROSEN[2]

[1] *Department of Mathematical Statistics, University of Tartu, J. Liivi Street 2, EE-2400 Tartu, Estonia*
[2] *Department of Mathematics, Uppsala University, Box 480, S-751 06 Uppsala, Sweden*

(Received January 24, 1994; revised February 6, 1995)

Abstract. Approximations of density functions are considered in the multivariate case. The results are presented with the help of matrix derivatives, powers of Kronecker products and Taylor expansions of functions with matrix argument. In particular, an approximation by the Wishart distribution is discussed. It is shown that in many situations the distributions should be centred. The results are applied to the approximation of the distribution of the sample covariance matrix and to the distribution of the non-central Wishart distribution.

Key words and phrases: Density approximation, Edgeworth expansion, multivariate cumulants, Wishart distribution, non-central Wishart distribution, sample covariance matrix.

1. Introduction

In statistical approximation theory the most common tool for approximating the density or the distribution function of a statistic of interest is the Edgeworth expansion or related expansions like tilted Edgeworth (e.g. see Barndorff-Nielsen and Cox (1989)). Then a distribution is approximated by the standard normal distribution using derivatives of its density function. However, for approximating a skewed random variable it is natural to use some skewed distribution. This idea was elegantly used by Hall (1983) for approximating a sum of independent random variables with the chi-square distribution.

The same ideas are also valid in the multivariate case. For different multivariate statistics Edgeworth expansions have been derived on the basis of the multivariate normal distribution, $N_p(0, \Sigma)$ (e.g. see Traat (1986), Skovgaard (1986), McCullagh (1987), Barndorff-Nielsen and Cox (1989)), but it seems more natural in many cases to use multivariate approximations via the Wishart distribution. Most of the test-statistics in multivariate analysis are based on functions of quadratic forms. Therefore, it is reasonable to believe, at least when the statistics are based on normal samples, that we could expect good approximations for these statistics using the Wishart density.

In this paper we are going to obtain the Wishart-approximation for the density function of a symmetric random matrix. In Section 2 basic notions and formulas for the probabilistic characterization of a random matrix will be given. Section 3 includes a general relation between two different density functions. The obtained relation will be utilized in Section 4 in the case when one of the two densities is the Wishart density. In particular, the first terms of the expansion will be written out. In Section 5 we present two applications of our results and consider the distribution of the sample covariance matrix as well as the non-central Wishart distribution.

2. Moments and cumulants of a random matrix

In the paper we are, systematically, going to use matrix notations. Most of the results will be presented using notions like vec-operator, Kronecker product, commutation matrix and matrix derivative. Readers, not very familiar with these concepts, are referred to the book by Magnus and Neudecker (1988), for example. Now we present those definitions of matrix derivatives which will be used in the subsequent.

For a $p \times q$-matrix X and a $m \times n$-matrix $Y = Y(X)$ the matrix derivative $\frac{dY}{dX}$ is a $mn \times pq$-matrix:

$$\frac{dY}{dX} = \frac{d}{dX} \otimes \operatorname{vec} Y,$$

where

$$\frac{d}{dX} = \left(\frac{\partial}{\partial X_{11}}, \ldots, \frac{\partial}{\partial X_{p1}}, \frac{\partial}{\partial X_{12}}, \ldots, \frac{\partial}{\partial X_{p2}}, \ldots, \frac{\partial}{\partial X_{1q}}, \ldots, \frac{\partial}{\partial X_{pq}} \right),$$

i.e.

$$\frac{dY}{dX} = \frac{d}{d \operatorname{vec}' X} \otimes \operatorname{vec} Y. \tag{2.1}$$

Higher order derivatives are defined recursively:

$$\frac{d^k Y}{dX^k} = \frac{d}{dX} \frac{d^{k-1} Y}{dX^{k-1}}. \tag{2.2}$$

When differentiating with respect to a symmetric matrix X we will instead of (2.1) use

$$\frac{dY}{d\Delta X} = \frac{d}{d \operatorname{vec}' \Delta X} \otimes \operatorname{vec} Y, \tag{2.3}$$

where ΔX denotes the upper triangular part of X and

$$\operatorname{vec} \Delta X = (X_{11}, X_{12}, X_{22}, \ldots, X_{1p}, \ldots, X_{pp})'.$$

There exists a $(p^2 \times \frac{1}{2}p(p+1))$-matrix G which is defined by the relation

$$G' \operatorname{vec} T = \operatorname{vec} \Delta T.$$

Explicitly the block-diagonal matrix G is given by $p \times i$-diagonal blocks G_{ii}:

$$G_{ii} = (e_1, e_2, \ldots, e_i) \qquad i = 1, 2, \ldots, p, \tag{2.4}$$

where e_i is the i-th unit vector, i.e. e_i is the i-th column of I_p. An important special case of (2.3) is when $Y = X$. Replacing $\operatorname{vec} \Delta X$ by $G' \operatorname{vec} X$ we get by definition (2.1) the following equality

$$\frac{dX}{d\Delta X} = (I_{p^2} + K_{p,p} - (K_{p,p})_d)G,$$

where $(K_{p,p})_d$ stands for the commutation matrix $K_{p,p}$ where the off-diagonal elements have been put to 0. To shorten the expressions we shall use the notation

$$H_{p,p} = I_{p^2} + K_{p,p} - (K_{p,p})_d, \tag{2.5}$$

where the indices may be omitted, if dimensions can be understood from the text. Hence our derivative equals the product

$$\frac{dX}{d\Delta X} = HG.$$

Note, that the use of HG is equivalent to the use of the duplication matrix (see Magnus and Neudecker (1988)).

For a random $p \times q$-matrix X the characteristic function is defined as

$$\varphi_X(T) = E[\exp(i \operatorname{tr}(T'X))],$$

where T is a $p \times q$-matrix. The characteristic function can also be presented through the vec-operator;

$$\varphi_X(T) = E[\exp(i \operatorname{vec}' T \operatorname{vec} X)], \tag{2.6}$$

which is a useful relation for differentiating. In the case of a symmetric $p \times p$-matrix X the nondiagonal elements of X appear twice in the exponent and so definition (2.6) gives us the characteristic function for $X_{11}, \ldots, X_{pp}, 2X_{12}, \ldots, 2X_{pp-1}$. However, it is more natural to present the characteristic function of a symmetric matrix for X_{ij}, $1 \leq i \leq j$, solely, which has been done for the Wishart distribution in Muirhead (1982), for example. We shall define the characteristic function of a symmetric $p \times p$-matrix X, using the elements of the upper triangular part of X:

$$\varphi_X(T) \equiv \varphi_{\Delta X}(\Delta T) = E[\exp(i \operatorname{vec}' \Delta T \operatorname{vec} \Delta X)]. \tag{2.7}$$

Moments $m_k[X]$ and cumulants $c_k[X]$ of X can be found from the characteristic function by differentiation, i.e.

$$m_k[X] = \frac{1}{i^k} \frac{d^k \varphi_X(T)}{dT^k} \bigg|_{T=0} \tag{2.8}$$

and

$$c_k[X] = \frac{1}{i^k} \frac{d^k \ln \varphi_X(T)}{dT^k}\bigg|_{T=0}$$

Following Cornish and Fisher (1937) we call the function

$$\psi_X(T) = \ln \varphi_X(T)$$

the cumulative function of X.

Applying the matrix derivative (2.1) and higher order matrix derivatives, (2.2), we get the following formulae for the moments:

$$\begin{aligned} m_1[X] &= E[\operatorname{vec}' X], \\ m_k[X] &= E[(\operatorname{vec} X)^{\otimes k-1} \operatorname{vec}' X], \qquad k \geq 2 \end{aligned}$$

where $a^{\otimes k}$ stands for $\underbrace{a \otimes a \otimes \cdots \otimes a}_{k \text{ times}}$ and $a^{\otimes 0} = 1$. The last statement can easily be proved using mathematical induction. In fact, the proof repeats the deduction of an analogous result for random vectors (e.g. see Kollo (1991)). Moreover, the following equalities are valid for the central moments $\bar{m}_k[X]$:

$$\bar{m}_k[X] = E[(\operatorname{vec}(X - E[X]))^{\otimes k-1} \operatorname{vec}'(X - E[X])], \qquad k = 1, 2, \ldots \tag{2.9}$$

which can be obtained as the derivatives of the characteristic function of $X - E[X]$. Using (2.7), similar results can be stated for a symmetric matrix.

To shorten notations we will use the following conventions:

$$\begin{aligned} c_k[\Delta X] &= c_k[\operatorname{vec} \Delta X], \\ m_k[\Delta X] &= m_k[\operatorname{vec} \Delta X], \\ \bar{m}_k[\Delta X] &= \bar{m}_k[\operatorname{vec} \Delta X], \end{aligned}$$

as well as

$$\begin{aligned} E[\Delta X] &= E[\operatorname{vec} \Delta X], \\ D[\Delta X] &= D[\operatorname{vec} \Delta X]. \end{aligned}$$

Finally we note that as in the univariate case there exist relations between cumulants and moments, and expressions for the first three will be utilized later:

$$\begin{aligned} c_1[X] &= m_1[X], \\ c_2[X] &= \bar{m}_2[X] = D[X], \\ c_3[X] &= \bar{m}_3[X]. \end{aligned}$$

3. Relation between two densities

Results in this section are based on Taylor expansions. If $g(X)$ is a scalar function of a $p \times q$-matrix X, we can present the Taylor expansion of $g(X)$ at the point X_0 in the following form (Kollo (1991)):

$$(3.1)\quad g(X) = g(X_0) + \sum_{k=1}^{m} \frac{1}{k!}(\operatorname{vec}'(X - X_0))^{\otimes k-1} \left.\frac{d^k g(X)}{dX^k}\right|_{X=X_0} \operatorname{vec}(X - X_0) + R_m,$$

where R_m stands for the remainder term. For the characteristic function of X we have from (3.1), using (2.6) and (2.8):

$$\varphi_X(T) = 1 + \sum_{k=1}^{m} \frac{i^k}{k!}(\operatorname{vec}' T)^{\otimes k-1} m_k[X] \operatorname{vec} T + R_m.$$

For the cumulative function

$$(3.2)\qquad \psi_X(T) = \sum_{k=1}^{m} \frac{i^k}{k!}(\operatorname{vec}' T)^{\otimes k-1} c_k[X] \operatorname{vec} T + R_m.$$

If X is symmetric, $\operatorname{vec} T$ will be changed by $\operatorname{vec} \Delta T$;

$$\varphi_{\Delta X}(\Delta T) = 1 + \sum_{k=1}^{m} \frac{i^k}{k!}(\operatorname{vec}' \Delta T)^{\otimes k-1} m_k[\Delta X] \operatorname{vec}(\Delta T) + R_m,$$
$$\psi_{\Delta X}(\Delta T) = \sum_{k=1}^{m} \frac{i^k}{k!}(\operatorname{vec}' \Delta T)^{\otimes k-1} c_k[\Delta X] \operatorname{vec}(\Delta T) + R_m.$$

Let X and Y be two $p \times q$ random matrices with densities $f_X(X)$ and $f_Y(Y)$, corresponding characteristic functions $\varphi_X(T)$, $\varphi_Y(T)$ and cumulative functions $\psi_X(T)$ and $\psi_Y(T)$. Our aim is to present the more complicated density function, say $f_Y(Y)$, through the simpler one, $f_X(X)$. In the univariate case, the problem was examined by Cornish and Fisher (1937) who obtained the principal solution to this problem and used it in the case when $X \sim N(0,1)$. Finney (1963) generalized the idea to the multivariate case and gave a general expression of the relation between two densities. In his paper Finney applied the idea in the univariate case, presenting one density through another. From later presentations we mention McCullagh (1987) and Barndorff-Nielsen and Cox (1989) who with the help of tensor notations briefly consider generalized formal Edgeworth expansions. Tan (1979) utilized Finney's (1963) work when approximating the non-central Wishart distribution with the Wishart distribution. One main difference between the approach in this paper and Finney (1963) and Tan (1979) is that we use matrix representations of moments and cumulants as well as matrix derivatives. This makes all computations much simpler and enables us to derive results in explicit form. When comparing the approach in this paper with the coordinate free tensor approach, this is a matter of taste which one to prefer. The tensor notation approach, as put forward by McCullagh (1987), gives compact expressions. However, these can sometimes be difficult to apply in real calculations and the approximation of the inverted Wishart distribution may serve as such an example.

To establish our results, in particular, for random symmetric matrices, we need some properties of Fourier transforms. The basic relation is given in the next lemma. The proof will, however, be omitted since the lemma is a direct generalization of a result for random vectors (e.g. see Traat (1986), or Kollo (1991)).

LEMMA 3.1. *The density $f_X(X)$ and the characteristic function $\varphi_X(T)$ of a $p \times q$-matrix X are connected by the following relation;*

$$\varphi_X(T)(i \operatorname{vec} T)^{\otimes k-1} i \operatorname{vec}' T = (-1)^k \int_{R^{pq}} e^{i \operatorname{vec}' T \operatorname{vec} X} \frac{d^k f_X(X)}{dX^k} dX,$$

where the integral is calculated elementwise and the derivatives are supposed to exist.

If X is symmetric, then

$$\begin{aligned}\varphi_{\Delta X}(\Delta T)&(i \operatorname{vec} \Delta T)^{\otimes k-1} i \operatorname{vec}' \Delta T \\ &= (-1)^k \int_{R^{p(p+1)/2}} e^{i \operatorname{vec}' \Delta T \operatorname{vec} \Delta X} \frac{d^k f_{\Delta X}(\Delta X)}{d\Delta X^k} d\Delta X.\end{aligned}$$

The lemma supplies us with the following crucial result.

COROLLARY 3.1. *If X is a $p \times q$-matrix and a is an arbitrary nonrandom $(pq)^k$-vector, then the Fourier transform (inverse transform) of the product $a'(i \operatorname{vec} T)^{\otimes k} \varphi_X(T)$ equals*

$$(-1)^k a' \operatorname{vec} \frac{d^k f_X(X)}{dX^k} = (2\pi)^{-pq} \int_{R^{pq}} \varphi_X(T) a'(i \operatorname{vec} T)^{\otimes k} e^{-i \operatorname{vec}' T \operatorname{vec} X} dT.$$

If X is symmetric and a an arbitrary $(\frac{1}{2}p(p+1))^k$-vector, then

$$\begin{aligned}(-1)^k a' \operatorname{vec} &\frac{d^k f_{\Delta X}(\Delta X)}{d\Delta X^k} \\ &= (2\pi)^{-p(p+1)/2} \int_{R^{p(p+1)/2}} \varphi_{\Delta X}(\Delta T) a'(i \operatorname{vec} \Delta T)^{\otimes k} e^{-i \operatorname{vec}' \Delta T \operatorname{vec} \Delta X} d\Delta T.\end{aligned}$$

PROOF. Lemma 3.1 states that $\varphi_X(T)(i \operatorname{vec} T)^{\otimes k-1} i \operatorname{vec}' T$ is the Fourier transform of $(-1)^k \frac{d^k f_X(X)}{dX^k}$. After vectorizing we get that $\varphi_X(T)(i \operatorname{vec} T)^{\otimes k}$ is the Fourier transform of the vector $(-1)^k \operatorname{vec} \frac{d^k f_X(X)}{dX^k}$. Then the inverse Fourier transform is given by

$$(-1)^k \operatorname{vec} \frac{d^k f_X(X)}{dX^k} = (2\pi)^{-pq} \int_{R^{pq}} \varphi_X(T)(i \operatorname{vec} T)^{\otimes k} e^{-i \operatorname{vec}' T \operatorname{vec} X} dT.$$

Premultiplying this equality with a' gives the statement of the corollary. The symmetric case is treated similarly. □

Now we are able to present the main result of the section.

THEOREM 3.1. *Let $f_X^{(k)}(X)$ denote the k-th derivative $\frac{d^k f_X(X)}{dX^k}$. If Y and X are two random $p \times q$-matrices the density $f_Y(X)$ can be presented through the density $f_X(X)$ by the following formal equality*:

$$\begin{aligned} f_Y(X) = f_X(X) &- (E[Y] - E[X])' \operatorname{vec} f_X^{(1)}(X) \\ &+ \frac{1}{2} \operatorname{vec}'\{D[Y] - D[X] + (E[Y] - E[X])(E[Y] - E[X])'\} \operatorname{vec} f_X^{(2)}(X) \\ &- \frac{1}{6}\{\operatorname{vec}'(c_3[Y] - c_3[X]) + 3 \operatorname{vec}'(D[Y] - D[X]) \otimes (E[Y] - E[X])' \\ &\qquad + (E[Y] - E[X])'^{\otimes 3}\} \operatorname{vec} f_X^{(3)}(X) + \cdots . \end{aligned}$$

PROOF. Using the expansion (3.2) of the cumulative function we have

$$\psi_Y(T) - \psi_X(T) = \sum_{k=1}^{\infty} \frac{i^k}{k!} (\operatorname{vec}' T)^{\otimes k-1} (c_k[Y] - c_k[X]) \operatorname{vec} T$$

and thus

$$\varphi_Y(T) = \varphi_X(T) \prod_{k=1}^{\infty} \exp\left\{ \frac{1}{k!} (i \operatorname{vec}' T)^{\otimes k-1} (c_k[Y] - c_k[X]) i \operatorname{vec} T \right\}.$$

By using series expansion of the exponential function we obtain, after ordering the terms according to i^k, the following equality

$$\begin{aligned} \varphi_Y(T) = \varphi_X(T) \Big\{ 1 &+ i(c_1[Y] - c_1[X]) \operatorname{vec} T \\ &+ \frac{i^2}{2} \operatorname{vec}' T \{c_2[Y] - c_2[X] \\ &\qquad + (c_1[Y] - c_1[X])'(c_1[Y] - c_1[X])\} \operatorname{vec} T \\ &+ \frac{i^3}{6} (\operatorname{vec}' T^{\otimes 2} \{c_3[Y] - c_3[X] + (c_1[Y] - c_1[X])' \\ &\qquad \otimes (c_1[Y] - c_1[X])'(c_1[Y] - c_1[X])\} \operatorname{vec} T \\ &\quad + 3(c_1[Y] - c_1[X]) \operatorname{vec} T \operatorname{vec}' T \\ &\quad \times (c_2[Y] - c_2[X]) \operatorname{vec} T) + \cdots \Big\}. \end{aligned}$$

Repeatedly applying the equality

$$\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$$

we obtain

$$\begin{aligned}\varphi_Y(T) = \varphi_X(T)\Big\{&1 + i(c_1[Y] - c_1[X])\operatorname{vec} T \\ &+ \frac{i^2}{2}\operatorname{vec}'\{c_2[Y] - c_2[X] \\ &\qquad + (c_1[Y] - c_1[X])'(c_1[Y] - c_1[X])\}(\operatorname{vec} T)^{\otimes 2} \\ &+ \frac{i^3}{6}(\operatorname{vec}'(c_3[Y] - c_3[X]) \\ &\qquad + 3\operatorname{vec}'(c_2[Y] - c_2[X]) \otimes (c_1[Y] - c_1[X]) \\ &\qquad + (c_1[Y] - c_1[X])^{\otimes 3})(\operatorname{vec} T)^{\otimes 3} + \cdots\Big\}.\end{aligned}$$

This equality can be inverted by applying the inverse Fourier transform given in Corollary 3.1. The characteristic functions turn then into density functions and taking into account that $c_1[\cdot] = E[\cdot]'$ and $c_2[\cdot] = D[\cdot]$ the theorem is established. □

For symmetric matrices we can rephrase the theorem in the following way.

COROLLARY 3.2. *If Y and X are symmetric random $p \times p$-matrices, then the density $f_{\Delta Y}(\Delta X)$ can be presented through the density $f_{\Delta X}(\Delta X)$ by the following formal equality*:

$$\begin{aligned}(3.3)\quad f_{\Delta Y}(\Delta X) = f_{\Delta X}(\Delta X) &- (E[\Delta Y] - E[\Delta X])'\operatorname{vec} f^{(1)}_{\Delta X}(\Delta X) \\ &+ \frac{1}{2}\operatorname{vec}'\{D[\Delta Y] - D[\Delta X] + (E[\Delta Y] - E[\Delta X]) \\ &\qquad \times (E[\Delta Y] - E[\Delta X])'\}\operatorname{vec} f^{(2)}_{\Delta X}(\Delta X) \\ &- \frac{1}{6}\{\operatorname{vec}'(c_3[\Delta Y] - c_3[\Delta X]) \\ &\qquad + 3\operatorname{vec}'(D[\Delta Y] - D[\Delta X]) \otimes (E[\Delta Y] - E[\Delta X])' \\ &\qquad + (E[\Delta Y] - E[\Delta X])'^{\otimes 3}\}\operatorname{vec} f^{(3)}_{\Delta X}(\Delta X) + \cdots.\end{aligned}$$

In the following we shall use the notation

$$f_X(X) \equiv f_{\Delta X}(\Delta X)$$

for the density function of a symmetric random matrix X analogously to the characteristic function in (2.7).

4. On Wishart approximation

If in Theorem 3.1 $f_X(X)$ is a normal density we shall get, as a special case, a matrix Edgeworth expansion of the density function $f_Y(X)$. We shall, however, not deal with the Edgeworth expansion in this paper. For us the starting point is Corollary 3.2 and we are going to assume that X is Wishart distributed. To get an expansion for $f_{\Delta Y}(\Delta X)$ we have to replace the derivatives and cumulants for X with the explicit expressions for moments and cumulants of the Wishart distribution.

Let W be a $p \times p$ Wishart distributed matrix with n degrees of freedom, $W \sim W_p(\Sigma, n)$. If $\Sigma > 0$, the matrix W has the density function

$$f_W(W) = \begin{cases} \dfrac{1}{2^{pn/2}\Gamma_p\left(\frac{n}{2}\right)|\Sigma|^{n/2}}|W|^{(n-p-1)/2}e^{-\operatorname{tr}(\Sigma^{-1}W)/2}, & W > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where the multivariate gamma function $\Gamma_p(\frac{n}{2})$ takes the value

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n+1-i)\right).$$

The characteristic function of the nonrepeated elements of W equals (e.g. see Muirhead (1982))

$$\varphi_W(T) = |I_p - iM(T)\Sigma|^{-n/2}, \tag{4.2}$$

where

$$M(T) = \sum_{i,j\in I} t_{ij}(e_ie_j' + e_je_i'),$$

e_i is the i-th column of I_p and $I = \{i,j; 1 \le i \le j \le p\}$. Furthermore, we need the first derivatives of the Wishart density. Straightforward calculations yield

LEMMA 4.1. *The derivative* $\frac{d^k f_W(W)}{d\Delta W^k}$ *is of the form*

$$\frac{d^k f_W(W)}{d\Delta W^k} = (-1)^k L_k(W,\Sigma) f_W(W), \qquad k = 0,1,2,\ldots \tag{4.3}$$

where $f_W(W)$ *is the density of the Wishart distribution* $W_p(\Sigma, n)$. *For* $k = 0,1,2,3$ *the matrices* $L_k(W,\Sigma)$ *are of the form*

$$L_0(W,\Sigma) = 1,$$

$$L_1(W,\Sigma) = -\frac{1}{2}\operatorname{vec}'(sW^{-1} - \Sigma^{-1})HG, \tag{4.4}$$

$$L_2(W,\Sigma) = -\frac{1}{2}G'H\{s(W^{-1}\otimes W^{-1}) \tag{4.5}$$

$$-\frac{1}{2}\operatorname{vec}(sW^{-1}-\Sigma^{-1})\operatorname{vec}'(sW^{-1}-\Sigma^{-1})\}HG,$$

$$\begin{aligned}L_3(W,\Sigma)=\frac{1}{2}(HG\otimes HG)'\Bigg\{\Bigg(&\frac{1}{2}\{\operatorname{vec}(sW^{-1}-\Sigma^{-1})\otimes I_{p^2}\\&+I_{p^2}\otimes\operatorname{vec}(sW^{-1}-\Sigma^{-1})\}\\&-(I_p\otimes K_{p,p}\otimes I_p)\{(I_{p^2}\otimes\operatorname{vec}W^{-1})\\&+(\operatorname{vec}W^{-1}\otimes I_{p^2})\}\Bigg)s(W^{-1}\otimes W^{-1})\\&-\frac{1}{2}\operatorname{vec}\Bigg\{\frac{1}{2}\operatorname{vec}(sW^{-1}-\Sigma^{-1})\operatorname{vec}'(sW^{-1}-\Sigma^{-1})\\&-s(W^{-1}\otimes W^{-1})\Bigg\}\operatorname{vec}'(sW^{-1}-\Sigma^{-1})\Bigg\}HG,\end{aligned}\tag{4.6}$$

where $s=n-p-1$, $K_{p,p}$ is the commutation matrix and G and H are defined by (2.4) *and* (2.5), *respectively.*

In order to apply Corollary 3.2 to the Wishart distribution we need also expressions for the first three cumulants of the Wishart distribution. These matrices can be found by differentiating the cumulative function $\psi_W(T)$ where $W\sim W_p(\Sigma,n)$. From (4.2) we obtain

$$\psi_W(T)=-\frac{n}{2}\ln|I_p-iM(T)\Sigma|.\tag{4.7}$$

The expectation and covariance of W are well known and equal

$$E[\Delta W]=n\operatorname{vec}'\Delta\Sigma,\tag{4.8}$$

$$D[\Delta W]=nG'(I_{p^2}+K_{p,p})(\Sigma\otimes\Sigma)G.\tag{4.9}$$

To find the third order cumulant of ΔW we have to take the third order derivative from the cumulative function (4.7). It follows from (2.5) and the definition of $M(T)$ in (4.2) that

$$\frac{dM(T)}{d\Delta T}=(I_{p^2}+K_{p,p})G$$

and then after some calculations we obtain

$$\begin{aligned}c_3[\Delta W]=n(G'(I_{p^2}+K_{p,p})\otimes G')&(I_p\otimes K_{p,p}\otimes I_p)\\\times(\Sigma\otimes\Sigma\otimes\operatorname{vec}\Sigma&+\operatorname{vec}\Sigma\otimes\Sigma\otimes\Sigma)(I_{p^2}+K_{p,p})G.\end{aligned}\tag{4.10}$$

A minor complication with the Wishart approximation is that the derivatives of the density of a Wishart distributed matrix increase with n. If the differences between cumulants are small this will not matter. However, in the general case, it seems wise to adjust the expansion so that derivatives decrease with n. Indeed, some authors have not observed this negative property of the Wishart distribution.

One way to overcome the problem is to translate the Wishart matrix so that a centred version is obtained, i.e.

$$V = W - n\Sigma.$$

From (4.1) it follows that the matrix V has the density function

$$(4.11) \qquad f_V(V) = \begin{cases} \dfrac{1}{2^{pn/2}\Gamma_p\left(\frac{n}{2}\right)|\Sigma|^{n/2}} \\ \qquad \times |V + n\Sigma|^{(n-p-1)/2} e^{-\operatorname{tr}(\Sigma^{-1}(V+n\Sigma))/2}, & V + n\Sigma > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The first cumulant of V equals zero and the other cumulants are identical to the corresponding cumulants of the Wishart distribution. In particular, the second and third order cumulants are equal to those given by (4.9) and (4.10).

In Theorem 4.1 given below we are going to present the density function $f_Y(X)$ through the centred Wishart density $f_V(X)$ on the basis of Corollary 3.2. Once again it follows that expressions for the first derivatives are needed. The derivatives of $f_V(X)$ can easily be obtained by simple transformations of $L_i(X, \Sigma)$ if we take into account the expressions of the densities (4.1) and (4.11). Analogously to Lemma 4.1 we have

Lemma 4.2. *Let $V = W - n\Sigma$ where $W \sim W_p(\Sigma, n)$. Then*

$$(4.12) \qquad f_V^{(k)}(V) = \frac{d^k f_V(V)}{d\Delta V^k} = (-1)^k L_k^\star(V, \Sigma) f_V(V),$$

where

$$L_k^\star(V, \Sigma) = L_k(V + n\Sigma, \Sigma), \qquad k = 0, 1, 2, \ldots.$$

The matrices $L_k(V, \Sigma)$, $k = 1, 2, 3$ are given by (4.4)–(4.6) and for $n \gg p$

$$(4.13) \qquad L_1^\star(V, \Sigma) \approx -\frac{1}{2n} \operatorname{vec}'(B_1) HG,$$

$$(4.14) \qquad L_2^\star(V, \Sigma) \approx -\frac{1}{2n} G'HB_2HG - \frac{1}{4n^2} G'H \operatorname{vec}(B_1)\operatorname{vec}'(B_1)HG,$$

where

$$B_1 = \Sigma^{-1}V^{1/2}\left(\frac{1}{n}V^{1/2}\Sigma^{-1}V^{1/2} + I_p\right)^{-1} V^{1/2}\Sigma^{-1},$$

$$B_2 = (V/n + \Sigma)^{-1} \otimes (V/n + \Sigma)^{-1}.$$

For $k = 3, 4, \ldots$ the matrix $L_k^\star(V, \Sigma)$ is of order $n^{-(k-1)}$.

PROOF. The first statement of the lemma follows directly from (4.3) in Lemma 4.1 if we replace W with the expression $V + n\Sigma$ since $W = V + n\Sigma$. For $L_1^\star(W, \Sigma)$ we have

$$\begin{aligned} L_1^\star(V,\Sigma) &= -\frac{1}{2}\operatorname{vec}'\{(n-p-1)(V+n\Sigma)^{-1} - \Sigma^{-1}\}HG \\ &= -\frac{1}{2}\operatorname{vec}'\left\{\frac{n-p-1}{n}(V/n+\Sigma)^{-1} - \Sigma^{-1}\right\}HG. \end{aligned}$$

If $n \gg p$ we have

$$L_1^\star(V,\Sigma) \approx -\frac{1}{2}\operatorname{vec}'\{(V/n+\Sigma)^{-1} - \Sigma^{-1}\}HG$$

and using the matrix equality (e.g. see Srivastava and Khatri (1979), p. 7)

$$(A + BEB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + E^{-1})^{-1}B'A^{-1},$$

where A and E are positive definite matrices of orders p and q, respectively, and B is a $p \times q$-matrix, we get that

$$L_1^\star(V,\Sigma) \approx -\frac{1}{2n}\operatorname{vec}'\left\{\Sigma^{-1}V^{1/2}\left(\frac{1}{n}V^{1/2}\Sigma^{-1}V^{1/2} + I_p\right)^{-1}V^{1/2}\Sigma^{-1}\right\}G.$$

Hence, (4.13) has been proved.

For $k = 2$ we obtain in a similar way ($s = n - p - 1$)

$$\begin{aligned} L_2^\star(V,\Sigma) = &-\frac{1}{2}G'H\Big\{s(V+n\Sigma)^{-1} \odot (V+n\Sigma)^{-1} \\ &\qquad - \frac{1}{2}\operatorname{vec}(s(V+n\Sigma)^{-1} - \Sigma^{-1}) \\ &\qquad \times \operatorname{vec}'(s(V+n\Sigma)^{-1} - \Sigma^{-1})\Big\}HG \\ \approx &-\frac{1}{2n}G'H(V/n+\Sigma)^{-1} \odot (V/n+\Sigma)^{-1}HG \\ &- \frac{1}{4n^2}G'H(\operatorname{vec} B_1 \operatorname{vec}' B_1)HG. \end{aligned}$$

Thus relation (4.14) is established.

To complete the proof we remark that from (4.6) we have, with the help of (4.13) and (4.14), that $L_3^\star(V, \Sigma)$ is of order n^{-2}. From the recursive definition of the matrix derivative the last statement of the lemma is established. □

Now we shall formulate the result about representation of a density $f_Y(X)$ through the Wishart distribution $W_p(\Sigma, n)$.

THEOREM 4.1. *Let W, Y and V be $p \times p$ random symmetric matrices with $W \sim W_p(\Sigma, n)$ and $V = W - n\Sigma$. Then for the density $f_Y(X)$ the following formal expansion holds:*

$$
\begin{aligned}
(4.15)\quad f_Y(X) = f_V(X)\Big\{ & 1 + E[\Delta Y]' \operatorname{vec} L_1^\star(X,\Sigma) \\
& + \frac{1}{2}\operatorname{vec}'(D[\Delta Y] - D[\Delta V] \\
& \qquad + E[\Delta Y]E[\Delta Y]') \operatorname{vec} L_2^\star(X,\Sigma) \\
& + \frac{1}{6}(\operatorname{vec}'(c_3[\Delta Y] - c_3[\Delta V]) \\
& \quad + 3\operatorname{vec}'(D[\Delta Y] - D[\Delta V]) \otimes E[\Delta Y]' \\
& \quad + E[\Delta Y]'^{\otimes 3}) \operatorname{vec} L_3^\star(X,\Sigma) + \cdots \Big\} \qquad X > 0.
\end{aligned}
$$

PROOF. The theorem follows directly from (3.3) if applying Lemma 4.2. □

5. Applications

In the following we are going to utilize Theorem 4.1 when considering an approximation of the density of the matrix nS by the density of a centred Wishart distribution, where S is the sample covariance matrix. The possibility of approximating the distribution of the sample covariance matrix with the Wishart distribution was probably first noted by Tan (1980), but he did not present explicit expressions in general. Only in the two-dimensional case formulas for the approximation were derived.

THEOREM 5.1. *Let $Z = Z_1, \ldots, Z_n$ be a sample of size n from a p-dimensional population with $E[Z_i] = \mu$, $D[Z_i] = \Sigma$ and finite central moments: $\bar{m}_k[Z_i] < \infty$, $k = 3, 4, \ldots$; let S denote the sample covariance matrix. Then the density function $f_{S^\star}(X)$ of $S^\star = n(S - \Sigma)$ has the following representation through the centred Wishart density $f_V(X)$, where $V = W - n\Sigma$ and $W \sim W_p(\Sigma, n)$:*

$$
\begin{aligned}
(5.1)\qquad f_{S^\star}(X) = f_V(X)\Big\{ & 1 - \frac{1}{4}\operatorname{vec}'(G'\{\bar{m}_4[Z_i] - \operatorname{vec}\Sigma\operatorname{vec}'\Sigma \\
& - (I_{p^2} + K_{p,p})(\Sigma \otimes \Sigma)\}G) \\
& \times \operatorname{vec}(G'H((V/n + \Sigma)^{-1} \otimes (V/n + \Sigma)^{-1})HG) \\
& + O\left(\frac{1}{n}\right)\Big\} \qquad X > 0,
\end{aligned}
$$

where G and H are defined by (2.4) *and* (2.5), *respectively.*

PROOF. To obtain (5.1) we have to insert the expressions of $L_k^\star(X, \Sigma)$ and cumulants $c_k[\Delta S^\star]$ and $c_k[\Delta V]$ in (4.15) and examine the result. At first let us remark that $c_1[\Delta S^\star]=0$ and in (4.15) all terms including $E[\Delta Y]$ vanish. By Kollo and Neudecker ((1993), Appendix I)

$$\begin{aligned} D[\sqrt{n}\operatorname{vec} S] &= \bar{m}_4[Z_i] - \operatorname{vec}\Sigma\operatorname{vec}'\Sigma + \frac{1}{n-1}(I_{p^2}+K_{p,p})(\Sigma\otimes\Sigma) \\ &= \bar{m}_4[Z_i] - \operatorname{vec}\Sigma\operatorname{vec}'\Sigma + O\left(\frac{1}{n}\right). \end{aligned}$$

Hence, by the definition of G, we have

$$D[\Delta S^\star] = nG'(\bar{m}_4[Z_i] - \operatorname{vec}\Sigma\operatorname{vec}'\Sigma)G + O(1). \tag{5.2}$$

In Lemma 4.2 we have shown that for $n \gg p$, $L_2^\star(X,\Sigma)$ is of order n^{-1} and in the approximation we can neglect the second term in (4.14). Multiplying vectors in (4.15) and using (4.9), (4.14) and (5.2) give us

$$\begin{aligned} &\frac{1}{2}\operatorname{vec}'(D[\Delta S^\star] - D[\Delta V])\operatorname{vec} L_2^\star(X,\Sigma) \\ &\quad = -\frac{1}{4}\operatorname{vec}'(G'\{\bar{m}_4[Z_i] - \operatorname{vec}\Sigma\operatorname{vec}'\Sigma - (I_{p^2}+K_{p,p})(\Sigma\otimes\Sigma)\}G) \\ &\qquad \times \operatorname{vec}(G'H((V/n+\Sigma)^{-1}\otimes(V/n+\Sigma)^{-1})HG) + O\left(\frac{1}{n}\right). \end{aligned}$$

To complete the proof we have to show that in (4.15) the remaining part of the sum within curly brackets is $O(\frac{1}{n})$. Let us first show that the term including $L_3^\star(X,\Sigma)$ in (4.15) is of order n^{-1}. From Lemma 4.2 we have that $L_3^\star(X,\Sigma)$ is of order n^{-2}. From (4.10) it follows that the cumulant $c_3[\Delta V]$ is of order n. Traat (1984) has found a matrix M_3, which is independent of n, such that

$$c_3(\operatorname{vec} S) = n^{-2}M_3 + O(n^{-3}).$$

Therefore

$$c_3[\Delta S^\star] = nK_3 + O(1),$$

where the matrix K_3 is independent of n. Thus, the difference of the third order cumulants is of order n and multiplying it with $\operatorname{vec} L_3^\star(X,\Sigma)$ gives that the product is $O(n^{-1})$.

All the other terms in (4.15) are scalar products of vectors which dimensionality does not depend on n. Thus, when examining the order of these terms it is sufficient to consider products of $L_k^\star(X,\Sigma)$ and differences of cumulants $c_k[\Delta S^\star]-c_k[\Delta V]$. Remember that it was shown in Lemma 4.2 that $L_k^\star(X,\Sigma)$, $k\geq 4$ is of order n^{-k+1}. Furthermore, from (4.7) and properties of sample cumulants of k-statistics it follows that the differences $c_k[\Delta S^\star]-c_k[\Delta V]$, $k\geq 2$ are of order n. Then from the construction of the formal expansion (3.3) we have that for $k=2p$, $p=2,3,\ldots$, the term including $L_k^\star(X,\Sigma)$, is of order $n^p\times n^{-2p+1}=n^{-p+1}$, where the main term of the cumulant differences is the term where the second order cumulants have been multiplied p times. Hence, the $L_4^\star(X,\Sigma)$-term, i.e. the expression including $L_4^\star(X,\Sigma)$ and the product of $D[\Delta S^\star] - D[\Delta V]$ with itself, is $O(n^{-1})$, the $L_6^\star(X,\Sigma)$-term is $O(n^{-2})$, etc.

For $k = 2p+1$, $p = 2,3,\dots$ the order of the $L_k^\star(X,\Sigma)$-term is determined by the product of $L_k^\star(X,\Sigma)$ and the $(p-1)$ products of the differences of the second order cumulants and a difference of the third order cumulants. So the order of the $L_k^\star(X,\Sigma)$-term ($k = 2p+1$) is $n^{-2p} \times n^{p-1} \times n = n^{-p}$. Thus, the $L_5^\star(X,\Sigma)$-term is $O(n^{-2})$, the $L_7^\star(X,\Sigma)$-term is $O(n^{-3})$ and so on. The presented arguments complete the proof. □

Our second application is about the non-central Wishart distribution. It turns out that Theorem 4.1 gives a very convenient way to describe the non-central Wishart density. Previously the approximation of the non-central Wishart distribution by the Wishart distributions has, among others, been considered by Steyn and Roux (1972) and Tan (1979). Both Steyn and Roux (1972) and Tan (1979) perturbed the covariance matrix in the Wishart distribution so that moments of the Wishart distribution and the non-central Wishart distribution should be close to each other. Moreover, Tan (1979) based his approximation on Finney's (1963) approach but never explicitly calculated the derivatives of the density. It was not considered that the density is dependent of n. Although our approach is a matrix version of Finney's there is a fundamental difference with the approach in Steyn and Roux (1972) and Tan (1979). Instead of perturbing the covariance matrix we use the idea of centring the non-central Wishart distribution. Indeed, as shown below, this will also simplify the calculations because we are now able to describe the difference between the cumulants in a convenient way, instead of treating the cumulants of the Wishart distribution and non-central Wishart distribution separately.

Let $Y \sim W_p(\Sigma, n, \mu)$, i.e. the non-central Wishart distribution with a non-centrality parameter $\Sigma^{-1}\mu\mu'$. If $\Sigma > 0$ the matrix Y has the characteristic function (see Muirhead (1982))

$$\varphi_Y(T) = \varphi_W(T) e^{-\operatorname{tr}(\Sigma^{-1}\mu\mu')/2} e^{\operatorname{tr}(\Sigma^{-1}\mu\mu'(I_p - iM(T)\Sigma)^{-1})/2}, \tag{5.3}$$

where $M(T)$ and $\varphi_W(T)$, the characteristic function of $W \sim W_p(\Sigma, n)$, are given by (4.2).

We shall consider centred versions of Y and W again, where $W \sim W_p(\Sigma, n)$. Let $Z = Y - n\Sigma - \mu\mu'$ and $V = W - n\Sigma$. Since we are interested in the differences $c_k[Z] - c_k[V]$, $k = 1,2,3,\dots$ we can, by the similarity of the characteristic functions $\varphi_Z(T)$ and $\varphi_V(T)$, obtain by (5.3) the difference of the cumulative functions

$$\begin{aligned}\psi_Z(T) - \psi_V(T) = & -\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\mu\mu') - i\frac{1}{2}\operatorname{tr}\{M(T)\mu\mu'\} \\ & + \frac{1}{2}\operatorname{tr}\{\Sigma^{-1}\mu\mu'(I - iM(T)\Sigma)^{-1}\}.\end{aligned}$$

After expanding the matrix $(I - iM(T)\Sigma)^{-1}$ we have

$$\psi_Z(T) - \psi_V(T) = \frac{1}{2}\sum_{j=2}^{\infty} i^j \operatorname{tr}\{\Sigma^{-1}\mu\mu'(M(T)\Sigma)^j\}. \tag{5.4}$$

From (5.4) it follows that $c_1[Z]-c_1[V]=0$, which, of course, must be true because $E[Z]=E[V]=0$. In order to obtain the difference of the second order cumulants we have to differentiate (5.4) and obtain

$$
\begin{aligned}
(5.5)\qquad c_2[Z]-c_2[V] &= \frac{1}{2}\frac{d^2tr(\mu\mu' M(T)\Sigma M(T))}{d\Delta T^2}\\
&= G'(I_p\otimes\mu\mu'\Sigma + I_p\otimes\Sigma\mu\mu')(I_{p^2}+K_{p,p})G.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
(5.6)\quad c_3[Z]-c_3[V] = {} & (G'(I_{p^2}+K_{p,p})\otimes G')(I_p\otimes K_{p,p}\otimes I_p)(I_{p^4}+K_{p^2,p^2})\\
&\times\{\Sigma\otimes\Sigma\otimes\operatorname{vec}\mu\mu' + (\Sigma\otimes\mu\mu'+\mu\mu'\otimes\Sigma)\otimes\operatorname{vec}\Sigma\}\\
&\times(I_{p^2}+K_{p,p})G.
\end{aligned}
$$

Hence the next theorem will easily follow.

THEOREM 5.2. *Let $Z=Y-n\Sigma-\mu\mu'$, where $Y\sim W_p(\Sigma,n,\mu)$ and $V=W-n\Sigma$, where $W\sim W_p(\Sigma,n)$. Then*

$$
\begin{aligned}
(5.7)\quad f_Z(X) = f_V(X)\Big\{1 &+ \frac{1}{2}\operatorname{vec}'\{(I_p\otimes\mu\mu'\Sigma+I_p\otimes\Sigma\mu\mu')(I_{p^2}+K_{p,p})\}\\
&\times(G\otimes G)\operatorname{vec}L_2^\star(X,\Sigma)\\
&+\frac{1}{6}\operatorname{vec}'(c_3[\Delta Z]-c_3[\Delta V])\\
&\times\operatorname{vec}L_3^\star(X,\Sigma)+o(n^{-2})\Big\}\qquad X>0,
\end{aligned}
$$

where $L_k^\star(X,\Sigma)$, $k=2,3$ are given by (4.14), (4.6) *and* (4.12), $(c_3[\Delta Z]-c_3[\Delta V])$ *is determined by* (5.6) *and G is defined by* (2.4).

PROOF. The proof follows from (4.15) if we replace the difference of the second order cumulants by (5.5) and take into account that by Lemma 4.2 $L_k^\star(X,\Sigma)$ is of order n^{-k+1}, $k\geq 3$, and that the differences of cumulants $c_k[\Delta Z]-c_k[\Delta V]$ do not depend on n. □

For an approximation of order n^{-1} we get from Theorem 5.2 the following

COROLLARY 5.1.

$$
\begin{aligned}
f_Z(X) = f_V(X)\Big\{1 &- \frac{1}{4n}\operatorname{vec}'\{(I_p\otimes\mu\mu'\Sigma+I_p\otimes\Sigma\mu\mu')(I_{p^2}+K_{p,p})\}\\
&\times(GG'H\otimes GG'H)\\
&\times\operatorname{vec}((V/n+\Sigma)^{-1}\otimes(V/n+\Sigma)^{-1})+o\left(\frac{1}{n}\right)\Big\}\qquad X>0.
\end{aligned}
$$

PROOF. The statement follows from (5.7) if we omit the $L_3^\star(X,\Sigma)$-term, which is of order n^{-2} and use the n^{-1} term from (4.14) for $L_2^\star(X,\Sigma)$. □

Acknowledgements

The authors are extremely thankful to the referees whose remarks have avoided several misprints and mistakes which occurred in the first version of the paper. Tõnu Kollo is indepted to the Swedish Institute for covering a stay at the Uppsala University and to the Estonian Research Foundation for their support. Dietrich von Rosen gratefully acknowledges the support from the Swedish Natural Research Council.

References

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Cornish, E. A. and Fisher, R. A. (1937). Moments and cumulants in the specification of distributions, *Internat. Statist. Rev.*, **5**, 307–322.

Finney, D. J. (1963). Some properties of a distribution specified by its cumulants, *Technometrics*, **5**, 63–69.

Hall, P. (1983). Chi-squared approximations to the distribution of a sum of independent random variables, *Ann. Probab.*, **11**, 1028–1036.

Kollo, T. (1991). *Matrix Derivative in Multivariate Statistics*, Tartu University Press, Tartu (in Russian).

Kollo, T. and Neudecker, H. (1993). Asymptotics of eigenvalues and unit-length eigenvalues of sample variance and correlation matrices, *J. Multivariate Anal.*, **47**, 283–300.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

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

Skovgaard, I. (1986). On the multivariate Edgeworth expansions, *Internat. Statist. Rev.*, **54**, 169–186.

Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New York.

Steyn, H. S. and Roux, J. J. J. (1972). Approximations for the noncentral Wishart distribution, *South African Journal of Statistics*, **6**, 165–173.

Tan, W. Y. (1979). On the approximation of noncentral Wishart distribution by Wishart distribution, *Metron*, **37**, 50–58.

Tan, W. Y. (1980). On approximating multivariate distributions, *Multivariate Statistical Analysis, Proceedings of the Research Seminar at Dalhousie University*, 237–249, North Holland, Amsterdam.

Traat, I. (1984). Moments of the sample covariance matrix, *Proceedings of the Computer Centre of Tartu State University*, **51**, 108–126 (in Russian).

Traat, I. (1986). Matrix calculus for multivariate distributions, *Acta et Commentationes Universitatis Tartuensis*, **733**, 64–84.