The noncentral Wishart as an exponential family, and its moments

Gérard Letac\textsuperscript{a,*}, Hélène Massam\textsuperscript{b}

\textsuperscript{a}Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062, Toulouse, France
\textsuperscript{b}Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada

Available online 18 April 2008

Abstract

While the noncentral Wishart distribution is generally introduced as the distribution of the random symmetric matrix $Y_1^*Y_1 + \cdots + Y_n^*Y_n$ where $Y_1, \ldots, Y_n$ are independent Gaussian rows in $\mathbb{R}^k$ with the same covariance, the present paper starts from a slightly more general definition, following the extension of the chi-square distribution to the gamma distribution. We denote by $\gamma(p, a; \sigma)$ this general noncentral Wishart distribution: the real number $p$ is called the shape parameter, the positive definite matrix $\sigma$ of order $k$ is called the shape parameter and the semi-positive definite matrix $a$ of order $k$ is such that the matrix $\omega = \sigma a \sigma$ is called the noncentrality parameter. This paper considers three problems: the derivation of an explicit formula for the expectation of $\text{tr}(Xh_1) \cdots \text{tr}(Xh_m)$ when $X \sim \gamma(p, a, \sigma)$ and $h_1, \ldots, h_m$ are arbitrary symmetric matrices of order $k$, the estimation of the parameters $(a, \sigma)$ by a method different from that of Alam and Mitra [K. Alam, A. Mitra, On estimated the scale and noncentrality matrices of a Wishart distribution, Sankhyä, Series B 52 (1990) 133–143] and the determination of the set of acceptable $p$'s as already done by Gindikin and Shanbag for the ordinary Wishart distribution $\gamma(p, 0, \sigma)$.

\copyright 2008 Elsevier Inc. All rights reserved.

\textit{AMS subject classifications:} primary 62H10; secondary 60B11

\textit{Keywords:} Noncentral Wishart; Noncentrality; Natural exponential families

1. Introduction

Let $\mathcal{S}_k$ be the real linear space of real symmetric matrices of order $k$, let $\mathcal{P}_k \subset \mathcal{S}_k$ be the cone of positive definite matrices and let $\overline{\mathcal{P}}_k \subset \mathcal{S}_k$ be the cone of semi-positive definite matrices. For
\( p > 0, a \in \mathcal{P}_k \) and \( \sigma \in \mathcal{P}_k \) we define here the general noncentral Wishart distribution \( \gamma(p, a; \sigma) \) on \( \mathcal{P}_k \) by its Laplace transform, that is, for all \( s \in \mathcal{P}_k \), by

\[
\int_{\mathcal{P}_k} e^{-\text{tr}(st)} \gamma(p, a; \sigma)(dt) = \frac{1}{\text{det}(1 + \sigma s)} e^{-\text{tr}(s(1 + \sigma s)^{-1} \sigma a \sigma)}.
\]

In the following the transpose of any matrix \( a \) will be denoted by \( a^* \). Consider independent normal random variables \( Y_1, \ldots, Y_n \) in \( \mathbb{R}^k \) written as column vectors with the same nonsingular covariance matrix \( \Sigma \) but possibly different means \( M = [m_1, \ldots, m_n] \) (\( n \) columns, \( k \) rows). The distribution of \( Y_1 Y_* + \cdots + Y_n Y_* \) is, following the notation in [14], \( W_k(n, M, \Sigma) \) which is the \( \gamma(p, a; \sigma) \) for \( p = n/2, \sigma = 2 \Sigma \) and \( a \) given by \( \sigma a \sigma = MM^* \). The relationship between \( W_k(n, M, \Sigma) \) and \( \gamma(p, a; \sigma) \) is similar to that of the chi-square and gamma distributions. While many aspects of \( W_k(n, M, \Sigma) \) have been much studied, the following topics about the \( \gamma(p, a; \sigma) \) distribution need attention:

1. The values of \( p \) for which \( \gamma(p, a; \sigma) \) exists. We will show that it does exist if and only if

\[
p \in \Lambda = \left\{ \frac{1}{2}, 2, \ldots, k - \frac{1}{2} \right\} \cup \left( k - \frac{1}{2}, \infty \right).
\]

(Propositions 2.2 and 2.3 in Section 2 for the crucial case \( \gamma(p, a; I_k) \) denoted by \( \gamma(p, a) \) and called the standard noncentral Wishart distribution).

2. The computation of moments of \( \gamma(p, a; \sigma) \). We observe that for fixed \( (p, a) \) in \( \Lambda \times \mathcal{P}_k \) the family \( F(p, a) = \{ \gamma(p, a; \sigma); \sigma \in \mathcal{P}_k \} \) is a natural exponential family (NEF). From this remark we obtain in Section 4 reasonably simple expressions for the moments in terms of \( p, a, \sigma \).

3. We give a compact presentation of the second moments of \( W_k(n, M, \Sigma) \), previously calculated by Alam and Mitra [1] for instance, by giving explicitly the variance function

\[
m \mapsto V_{F(p, a)}(m) = V(m)
\]

of the NEF \( F(p, a) \). From this, we design a new method of estimation of \( (a, \sigma) \) (thus of \( MM^*, \Sigma \) for \( W_k(n, M, \Sigma) \)). Our estimators \( \hat{\sigma} \) and \( \hat{a} \) are defined by

\[
\hat{\sigma} = \hat{m} + (m^2 - p \hat{v})^{1/2}, \quad \hat{\sigma} \hat{a} \hat{\sigma} = \hat{m} - p \hat{\sigma}
\]

where

\[
\hat{m} = \frac{1}{N} (X_1 + \cdots + X_N), \quad \hat{v} = \frac{1}{N - 1} \sum_{j=1}^N (X_j - \bar{X}_N) \times \text{tr} (X_j - \bar{X}_N)
\]

and \( X_1, \ldots, X_N \) are \( N \) iid symmetric matrices with the same unknown distribution \( \gamma(p, a; \sigma) \).

This text was started in 2005 (and posted on the web) as a collection of results on the noncentral Wishart for colleagues and students. We obtained a lot of our information from Johnson and Kotz [10]. Our main sources for the ordinary Wishart distributions are Muirhead [14] and Faraut and Korányi [4]. Another source of inspiration (in Sections 4 and 6) has been the coordinate-free approach of multivariate statistics described in [3]. While preparing this text for the web, we realized that the three questions mentioned above either had no satisfactory treatment in the literature or deserved some new consideration. This is how the present paper came to be. We have ignored the natural generalization to Euclidean Jordan algebras, which is rather automatic and academic (except in the Hermitian case: see for example Kang and Alouini [8]).
The standard noncentral Wishart laws are introduced in Section 2 while the general noncentral Wishart laws are defined in Section 3. In Section 4 we compute the moments of the general noncentral Wishart. In Section 5, we complete the proof of statement 1 above, using tools similar to the ones introduced in Section 4. Section 6 is devoted to the description of estimation methods for the parameters of the general noncentral Wishart. It includes a modern presentation of two interesting and little known methods given by Alam and Mitra [1].

2. The standard noncentral Wishart and Gaussian laws

We gather some classical facts in the next proposition. A referee kindly pointed out that these are due originally to Khatri [9].

Proposition 2.1. Let $Z$ be a random variable in $\mathbb{R}^k$ with distribution $N(0, I_k)$ and let $m$ be in $\mathbb{R}^k$. We write the vectors of $\mathbb{R}^k$ as column vectors. Then for $s$ a symmetric matrix of order $k$ such that $s + I_k$ is in $\mathcal{P}_k$, we have

$$
\mathbb{E}(e^{-\frac{1}{2}(Z+m)^*s(Z+m)}) = \frac{1}{\det(I_k + s)^{1/2}} e^{-\frac{1}{2}m^*(I_k + s)^{-1}sm}
$$

$$
\mathbb{E}(e^{-\frac{1}{2} tr(s(Z+m)(Z+m)^*)}) = \frac{1}{\det(I_k + s)^{1/2}} e^{-\frac{1}{2} tr((I_k + s)^{-1}smm^*)}.
$$

(1)

Proof. We use the fact that the Laplace transform of $N(0, \Sigma)$ is $\exp\frac{\theta^*\Sigma\theta}{2}$ for the particular values $\Sigma = (I_k + s)^{-1}$ and $\theta = -sm$. After observing that $s(I_k + s)^{-1}s - s = -(I_k + s)^{-1}s$ we get the result. □

More generally consider $n$ independent random variables $Z_1, \ldots, Z_n$ in $\mathbb{R}^k$ with the same distribution $N(0, I_k)$ and let $m_1, \ldots, m_n$ be in $\mathbb{R}^k$. A consequence of Proposition 2.1 is that

$$
\mathbb{E}(e^{-\frac{1}{2} tr(s \sum_{j=1}^n (Z_j + m_j)(Z_j + m_j)^*)}) = \frac{1}{\det(I_k + s)^{n/2}} e^{-\frac{1}{2} tr((I_k + s)^{-1}s(m_1m_1^* + \cdots + m_nm_n^*))}.
$$

(2)

This leads to the following question: if $p > 0$ and if $a$ is in $\overline{\mathcal{P}_k}$, does there exist a probability distribution $\gamma(p, a)$ on $\mathcal{P}_k$ with Laplace transform

$$
\frac{1}{\det(I_k + s)^p} e^{-tr((I_k + s)^{-1}sa)}?
$$

A detailed answer will be given in Proposition 2.3. To prove the existence of $\gamma(p, a)$ we now need to introduce the zonal polynomials.

Given a symmetric real matrix $x = (x_{ij})_{1 \leq i, j \leq k}$ of order $k$ for $1 \leq m \leq k$ we denote $\Delta_m(x) = \det(x_{ij})_{1 \leq i, j \leq m}$. Consider a sequence of integers $\kappa = (m_1, \ldots, m_k)$ such that $m_1 \geq m_2 \geq \cdots \geq m_k \geq 0$. We denote $|\kappa| = m_1 + m_2 + \cdots + m_k$ and we denote by $E_m$ the set of $\kappa$ such that $m = |\kappa|$. Let

$$
\Delta_\kappa(x) = (\Delta_1(x))^{m_1-m_2} (\Delta_2(x))^{m_2-m_3} \cdots (\Delta_{k-1}(x))^{m_{k-1}-m_k} (\Delta_k(x))^{m_k}.
$$

We remark that $\Delta_\kappa(x) > 0$ for $x \in \mathcal{P}_k$. Also, for $z = (z_1, \ldots, z_k) \in \mathbb{R}^k$, $z_j > (j-1)/2$, $j = 1, \ldots, k$, we define the function $z \mapsto \Gamma_k(z)$ by

$$
\Gamma_k(z) = (\pi)^{k(k-1)/4} \prod_{j=1}^k \Gamma\left(z_j - \frac{j-1}{2}\right).
$$
If \( p \) is real, with a traditional abuse of notation we write \( \Gamma_k(z + p) \) for \( \Gamma_k(z_1 + p, \ldots, z_k + p) \). In particular for \( p > (k - 1)/2 \) we have

\[
\Gamma_k(p) = (\pi)^{k(k-1)/4} \prod_{j=1}^{k} \Gamma \left( p - \frac{j-1}{2} \right)
\]

which is equal to (see \cite{14}, Th. 2.1.12)

\[
\Gamma_k(p) = \int_{\mathcal{P}_k} e^{-\text{tr}(x)} \left( \det x \right)^{p-\frac{k+1}{2}} \, dx.
\] (3)

Here \( dx \) is the Lebesgue measure on the linear space of symmetric matrices \( x = (x_{ij})_{1 \leq i, j \leq k} \) defined by \( dx = \prod_{1 \leq i \leq j \leq k} dx_{ij} \).

For \( \kappa = (m_1, \ldots, m_k) \) with \( m_1 \geq m_2 \geq \cdots \geq m_k \geq 0 \), this leads to the notation,

\[
(p)_{\kappa} = \frac{\Gamma_k(\kappa + p)}{\Gamma_k(p)}.
\]

The zonal polynomial \( C_\kappa(x) \) of parameter \( \kappa \) is defined by the following integral on the group \( \mathcal{O}(k) \) of orthogonal matrices of order \( k \) with respect to the Haar measure \( du \) (normalized in order to have total mass one):

\[
C_\kappa(x) = C_\kappa \int_{\mathcal{O}(k)} \Delta_\kappa(u^{-1}xu) \, du,
\]

where \( C_\kappa \) is a complicated normalizing constant which can be found in the last line of page 234 in \cite{4}, or in formula (18) on page 237 of \cite{14}. Its exact value is not important here. Different presentations of zonal polynomials are given in \cite{14}, Ch. 7 and in \cite{17}. The zonal polynomial \( C_\kappa(x) \) is a homogeneous polynomial of degree \( |\kappa| \) with respect to the entries \( x_{ij} \) of the symmetric matrix \( x \). We emphasize the fact that by definition, \( C_\kappa(x) \) takes positive values on \( \mathcal{P}_k \). We note that if \( a \in \mathcal{P}_k \) and if \( x \) is symmetric of order \( k \) then \( a^{1/2}x a^{1/2} \) is also symmetric. Another important remark is that for any \( v \) in the orthogonal group \( \mathcal{O}(k) \) we have

\[
C_\kappa(x) = C_\kappa(v^{-1}xv).
\] (4)

This is clear from the definition of Haar probability. A consequence of (4) is that actually, \( C_\kappa(x) \) depends only on the eigenvalues of \( x \).

The zonal polynomials satisfy many remarkable formulas. We mention here some of them:

\[
e^{\text{tr}x} = \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{1}{m!} C_\kappa(x)
\] (5)

\[
det(I_k - x)^{-p} = \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{(p)_{\kappa}}{m!} C_\kappa(x)
\] (6)

\[
(p)_{\kappa} (\det s)^{-p} C_\kappa(s^{-1}) = \int_{\mathcal{P}_k} e^{-\text{tr}(sx)} C_\kappa(x) (\det x)^{p-\frac{k+1}{2}} \frac{dx}{\Gamma_k(p)}
\] (7)

\[
\frac{C_\kappa(t)C_\kappa(a)}{C_\kappa(I_k)} = \int_{\mathcal{O}(k)} C_\kappa(a^{1/2}u^{-1}tu a^{1/2}) \, du.
\] (8)
The unspecified normalizing constant $C_\kappa$ above is chosen such that these formulas hold. Notice that a statement equivalent to (5) is that

$$(\text{tr} \, x)^m = \sum_{\kappa \in E_m} C_\kappa(x)$$

because $C_\kappa$ is a homogeneous polynomial of degree $|\kappa|$. A consequence is that for $\kappa \in E_m$ and for $x \in \overline{P_k}$ we have

$$0 \leq C_\kappa(x) \leq (\text{tr} \, x)^m. \tag{9}$$

The following proposition is crucial for the proof of the existence of the $\gamma(p, a)$ distribution for $p > (k - 1)/2$. From Proposition 2.1 we can already deduce immediately that $\gamma(p, a)$ exists for the discrete values $p \in \{n/2, \ n = 1, 2, \ldots\}$.

**Proposition 2.2.** Let $p > (k - 1)/2$ and $a \in \overline{P_k}$. Then

$$\gamma(p, a)(dt) = e^{-\text{tr} \, (t+a)}(\det t)^{p-\frac{k+1}{2}} \left( \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{C_\kappa(a^{1/2}ta^{1/2})}{m!(p)_\kappa} \right) 1_{P_k}(t) \frac{dt}{\Gamma_k(p)}$$

is a probability on $P_k$ such that for $I_k + s \in P_k$

$$\int_{P_k} e^{-\text{tr} \, (st)} \gamma(p, a)(dt) = \frac{1}{\det(I_k + s)^p} e^{-\text{tr} \, ((I_k+s)^{-1}a)}. \tag{10}$$

In particular if $X \sim \gamma(p, a)$ and $Y \sim \gamma(q, b)$ are independent we have $X + Y \sim \gamma(p+q, a+b)$.

**Proof.** Since $t \in P_k$, $\Delta_k(t)$ is positive. Moreover, from the definition of $C_\kappa(t)$ the number $C_\kappa(a^{1/2}ta^{1/2})$ is non-negative and therefore $\gamma(p, a)(dt)$ is a positive measure. Suppose first that $a \in P_k$. For $\kappa \in E_m$ we make the change of variable $x = a^{1/2}ta^{1/2}$ in the integral

$$I_\kappa(a) = e^{-\text{tr} \, a} \int_{P_k} e^{-\text{tr} \, ((s+I_k)t)}(\det t)^{p-\frac{k+1}{2}} \frac{C_\kappa(a^{1/2}ta^{1/2})}{m!(p)_\kappa} \frac{dt}{\Gamma_k(p)}.$$

The Jacobian of $t \mapsto a^{1/2}ta^{1/2}$ yields $dx = (\det a)^{(k+1)/2}dt$. Using formula (7) we get

$$I_\kappa(a) = e^{-\text{tr} \, a} \det(I_k + s)^{-p} C_\kappa(a^{1/2}(I_k + s)^{-1}a^{1/2}) \frac{m!}{\Gamma_k(p)}.$$

Suppose now that $a$ is singular in $\overline{P_k}$. Remark that for $n \geq 1$ we have $a_n = a + \frac{1}{n}I_k \in P_k$. Inequality (9) implies that

$$0 \leq C_\kappa(a_n^{1/2}x a_n^{1/2}) \leq (\text{tr} \, (a_n x))^m \leq (\text{tr} \, (a_1 x))^m.$$

From the dominated convergence theorem, it follows that $\lim_{n \to \infty} I_\kappa(a_n) = I_\kappa(a)$ and this implies that (11) holds even for $a$ singular. Summing up all equalities (11) over $m$ and $\kappa$ and using (5) we easily obtain (10). We can verify that $\gamma(p, a)$ is a probability measure by substituting $s = 0$ in (10). \Box

Let $p$ be in the so-called Gindikin set $\Lambda$ of order $k$ defined by

$$\Lambda = \left\{ \frac{1}{2}, \ldots, \frac{k-1}{2} \right\} \cup \left( \frac{k-1}{2}, \infty \right)$$

and let $a$ be in $\overline{P_k}$. We define the standard noncentral Wishart distribution $\gamma(p, a)$ on $\overline{P_k}$ as the unique probability such that (10) holds. For $p > (k - 1)/2$ its existence is given by
Proposition 2.2. If \( p \) is the half integer \( n/2 \), the existence of \( \gamma(p, a)(dr) \) as well a Gaussian interpretation comes from (2). Actually, these values of \( p \in \Lambda \) are the only ones for which \( \gamma(p, a) \) does exist. More specifically

**Proposition 2.3.** For \( p > 0 \) and \( a \) in \( \mathcal{P}_k \), there exists a probability \( \gamma(p, a) \) such that (10) holds if and only if \( p \) is in \( \Lambda \).

This result was proved for \( a = 0 \) by Gindikin [7]. Shanbag [16] proved it again for \( a = 0 \) and Peddada and Richards [15] proved the same result once more for \( a = 0 \) but also for \( a \) of rank one. Of course Proposition 2.3 implies that \( \gamma(p, a) \) is not infinitely divisible, except for \( k = 1 \). Since its proof for \( a \neq 0 \) requires some notations which will be introduced in the following sections, we postpone it to Section 5.

### 3. The general noncentral Wishart distributions

In the first subsection we describe the action of certain linear transformations of \( \mathcal{S}_k \) on the standard noncentral Wishart laws. We then consider the NEF generated by \( \gamma(p, a) \) and prove that this NEF coincides with the set of images of \( \gamma(p, a) \) by these linear transformations. In the second subsection we clarify the link between our general noncentral Wishart and the classical noncentral Wishart distributions issued from Gaussian variables. We shall pay special attention to the correspondence between the different possible parameterizations given in the literature.

#### 3.1. The general noncentral Wishart natural exponential family

**Proposition 3.1.** Let \( b \) be a nonsingular real matrix of order \( k \) and denote by \( b^* \) its transposed matrix. Denote by \( b = qu \) its polar decomposition, that is \( q \) is symmetric and positive definite, while \( u \) is orthogonal of order \( k \). Let \( p \in \Lambda \) and let \( a \in \mathcal{P}_k \). Define \( \sigma \) and \( \omega \) to be \( \sigma = q^2 \) and \( \omega = \sigma a \).

1. The image of the \( \gamma(p, a) \) distribution by \( t \mapsto utu^* \) is \( \gamma(p, uau^*) \).
2. The image of the \( \gamma(p, qaq) \) distribution by \( t \mapsto qtq \) is a distribution denoted by \( \gamma(p, a; q^2) = \gamma(p, a; \sigma) \) with the following Laplace transform:

\[
\int_{\mathcal{P}_k} e^{-\operatorname{tr}(st)} \gamma(p, a; \sigma)(dr) = \frac{1}{\det(I_k + qsq)^p} e^{-\operatorname{tr}((I_k + qsq)^{-1}qsqaq)} \quad (12)
\]

\[
= \frac{1}{\det(I_k + \sigma s)^p} e^{-\operatorname{tr}(s(I_k + \sigma s)^{-1}\omega)} \quad (13)
\]

3. The image of the \( \gamma(p, a) \) by \( t \mapsto btb^* \) is the \( \gamma(p, q^{-1}uau^*q^{-1}; \sigma) \) distribution.
4. The NEF generated by \( \gamma(p, a) \) is

\[
 F(\gamma(p, a)) = \{ \gamma(p, a; \sigma); \ \sigma \in \mathcal{P}_k \}.
\]

5. Suppose \( X \sim \gamma(p, a; \sigma) \). If \( c \) is a positive constant, then \( cX \) follows the \( \gamma(p, c^2, c\sigma) \) distribution, with the corresponding \( \omega \) changed into \( c\omega \). If \( X \sim \gamma(p, a, \sigma) \) and \( Y \sim \gamma(q, b, \sigma) \) are independent we have \( X + Y \sim \gamma(p + q, a + b, \sigma) \).

If \( z \) is a column vector of \( \mathbb{R}^k \) denote \( \lambda_z = z^* \sigma z \). Then the distribution of the real random variable \( z^* Xz \) is a noncentral gamma distribution (i.e. one-dimensional noncentral Wishart distribution) with Laplace transform

\[
\mathbb{E}(e^{-s_1 z^* Xz}) = \frac{1}{(1 + s_1 \lambda)^p} e^{-\frac{s_1}{\lambda^2} z^* \omega z},
\]
thus with shape parameter $p$, with scale parameter $\lambda_z$ and with parameter

$$a_z = \lambda_z^{-2} z^* \omega_z = \frac{z^* \omega_z}{(z^* \sigma z)^2}.$$

In particular the mean $m_z$ and the variance $v_z$ of $z^*X_z$ are respectively

$$m_z = p z^* \sigma z + z^* \omega z, \quad v_z = z^* \sigma z (p z^* \sigma z + 2 z^* \omega z). \quad (14)$$

**Proof.**

1. We write

$$\int_{\mathcal{P}_k} e^{-tr(sut^*)} \gamma(p, a)(dt) = \int_{\mathcal{P}_k} e^{-tr(u^*sut)} \gamma(p, a)(dt)$$

$$= \frac{1}{\det(I_k + s)^p} e^{-tr((I_k + u^*su)^{-1} u^*su)}$$

$$= \frac{1}{\det(I_k + s)^p} e^{-tr(u(a^*a + u^*su)^{-1} u^*su)}$$

$$= \frac{1}{\det(I_k + s)^p} e^{-tr((I_k + s)^{-1} u^*su)}.$$

2. The proof of (12) is standard: the Laplace transform of $X \sim \gamma(p, qaq)$ is as given by (10) with $a$ replaced by $qaq$. The Laplace transform of $qXq$ is equal to $E(e^{-\langle x, qXq \rangle}) = E(e^{-\langle qsq, X \rangle})$ and is therefore obtained by replacing $s$ by $qsq$ and $a$ by $qaq$ in (10). Then (12) follows immediately. To pass from (12) to (13) we write

$$\det(I_k + qsq) = \det((q^{-1} + qs)q) = \det(q^{-1} + sq)$$

$$= \det(I_k + q^2 s) = \det(I_k + s \sigma s)$$

and by standard manipulations using the inverse of products and the circular property of traces we have

$$\text{tr} ((I_k + qsq)^{-1} qsqqaq) = \text{tr} ((q(q^{-2} + s)q)^{-1} qsqqaq)$$

$$= \text{tr} (q^{-1}(q^{-2} + s)^{-1} q^{-1} qsqqaq)$$

$$= \text{tr} ((\sigma^{-1} + s)^{-1} \sigma \sigma a)$$

$$= \text{tr} ((\sigma^{-1}(I_k + \sigma s))^{-1} \sigma \sigma a)$$

$$= \text{tr} ((I_k + \sigma s)^{-1} \sigma \sigma a)$$

$$= \text{tr} (\sigma (s(I_k + \sigma s)^{-1} \sigma a))$$

$$= \text{tr} (s(I_k + \sigma s)^{-1} \omega).$$

The only subtlety is in equality $\dagger$ where we use the fact that $x(I_k + x)^{-1} = (I_k + x)^{-1} x$ for any square matrix $x$ and we apply it to $x = s \sigma s$.

3. Since the transformation $\varphi_b : t \mapsto btb^*$ satisfies $\varphi_b = \varphi_q \circ \varphi_u$ we observe from part 1 that the image of $\gamma(p, a)$ is the $\gamma(p, uau^*)$ distribution which we rewrite as $\gamma(p, q^{-1} uaq^* q^{-1} q)$. By definition of $\gamma(p, a; \sigma)$, its image by $\varphi_q$ is indeed the $\gamma(p, q^{-1} uaq^* q^{-1}; \sigma)$ distribution.

4. The probability $\mu$ belongs to $F(\gamma(p, a))$ if and only if there exists $s_0$ with $I_k + s_0 \in \mathcal{P}_k$ such that for $I_k + s + s_0 \in \mathcal{P}_k$ one has

$$\int_{\mathcal{P}_k} e^{-\text{tr}(st)} \mu(dt) = \left(\frac{\det(I_k + s_0 + s)}{\det(I_k + s_0)}\right)^p e^{-\text{tr}((I_k + s_0 + s)^{-1}((s_0 + s)a))} e^{\text{tr}((I_k + s_0)^{-1} s_0 a)}.$$
We rewrite the second member by introducing \( q = (I_k + s_0)^{-1/2} \) or \( s_0 = q^{-2} - I_k \) and we obtain the right-hand side of (12) which proves the result.

5. If \( s_1 \) is a positive number, apply formula (12) to the symmetric matrix \( s = s_1 z z^* \). In order to obtain the Laplace transform \( \mathbb{E}(e^{-s_1 z X z^*}) \), we choose an orthonormal basis \( e = (e_1, \ldots, e_k) \) of \( \mathbb{R}^k \) such that \( q z = \lambda^{1/2} e_1 \). With such a choice the representative matrices \( M \) and \((I_k + M)^{-1}\) of the endomorphisms \( s_1 q z z^* q \) and \((I_k + q s q)^{-1} q s q \) in the above basis is simply, by blocks

\[
M = \begin{bmatrix} s_1 \lambda & 0 \\ 0 & 0 \end{bmatrix}, \quad (I_k + M)^{-1} = \begin{bmatrix} s_1 \lambda & 0 \\ 0 & 0 \end{bmatrix}.
\]

Thus \( \det(I_k + q s q) = 1 + s_1 \lambda \) and \( \text{tr}((I_k + q s q)^{-1} q s q a q) = \frac{s_1 \lambda}{1 + s_1 \lambda} \text{tr} e_1 e_1^* q a q \). Since

\[
\text{tr} e_1 e_1^* q a q = \frac{1}{\lambda} \text{tr} q z^* z q a q = \frac{1}{\lambda} z^* \sigma a \sigma z
\]

we obtain the desired results. \( \square \)

The \( \gamma(p, a; \sigma) \) distribution is called a general noncentral Wishart distribution. For reasons appearing in Proposition 3.2 below, the parameter \( \omega = \sigma a \sigma \) is called the noncentrality parameter of the distribution \( \gamma(p, a; \sigma) \). Clearly the parameterization by the pair \((a, \sigma)\) could be replaced by the parameterization by \((\omega, \sigma)\) which seems more practical. However, the calculation of the density below or of the moments in Section 4 shows that our choice has many advantages.

By definition, for \( p > (k - 1)/2 \) the density of the general \( \gamma(p, a; \sigma) \) is obtained by taking the image by \( t \mapsto x = q t q \). \( q = \sigma 1/2 \), of the standard \( \gamma(p, q a q) \) distribution. Thus for \( p > (k - 1)/2 \), \( \gamma(p, a; \sigma) \) (dx) is equal to

\[
\begin{align*}
\text{(det } \sigma)^{-p} e^{-\text{tr}(q^{-1} x q^{-1} + q a q)} (\text{det } x)^{p - \frac{k + 1}{2}} \\
\times \left( \sum_{m=0}^{\infty} \sum_{k \in E_m} C_k (q^{1/2} a^{1/2} q^{-1/2} x q^{-1/2} a^{1/2} q^{1/2}) \right) \frac{1_{p_k}(x)}{m!(p)_k} \frac{\text{dx}}{\Gamma_k(p)}.
\end{align*}
\]

Note that another presentation of the argument of the exponential is

\[
\text{tr} (q^{-1} x q^{-1} + q a q) = \text{tr}(x \sigma^{-1} + a \sigma).
\]

### 3.2. The Gaussian case and the various parameterizations

**Proposition 3.2.** Let \( Y_1, \ldots, Y_n \) be independent normal random variables valued in \( \mathbb{R}^k \) such that \( Y_j \sim N(m_j, \Sigma) \). We assume that \( Y_j \) and \( m_j \) are written as row vectors. Denote by \( M \) the matrix with \( n \) columns and \( k \) rows defined by \( M = [m_1, \ldots, m_n] \). Then the random \((k, k)\) matrix

\[
W = Y_1 Y_1^* + \cdots + Y_n Y_n^*
\]

follows the general noncentral Wishart distribution \( \gamma(p, a; \sigma) \) whose parameters are given by

\[
p = n/2, \quad \sigma = 2 \Sigma \quad w = \sigma a \sigma = M M^*.
\]

**Proof.** We write \( Y_j = \Sigma^{1/2} (Z_j + \Sigma^{-1/2} m_j) \) where \( Z_j \sim N(0, I_k) \). Thus

\[
W = \Sigma^{1/2} \left( \sum_{j=1}^n \frac{(Z_j + \Sigma^{-1/2} m_j)(Z_j + \Sigma^{-1/2} m_j)^*}{\Sigma^{1/2}} \right) \Sigma^{1/2}.
\]
To apply (2) we observe that
\[
\sum_{j=1}^{n} (\Sigma^{-1/2} m_j)(\Sigma^{-1/2} m_j)^* = \sum_{j=1}^{n} \Sigma^{-1/2} m_j m_j^* \Sigma^{-1/2} = \Sigma^{-1/2} M M^* \Sigma^{-1/2}.
\]

Thus (2) gives
\[
\mathbb{E}(e^{-\frac{1}{2} \text{tr} (\Sigma^{-1/2} s \Sigma^{-1/2} W)}) = \mathbb{E}(e^{-\frac{1}{2} \text{tr} (s \Sigma^{-1/2} W \Sigma^{-1/2})}) = \frac{1}{\det(I_k + s) n/2} e^{-\frac{1}{2} \text{tr} ((I_k + s)^{-1} \Sigma^{-1/2} M M^* \Sigma^{-1/2})}.
\]

Now let \( \theta = \frac{1}{2} \Sigma^{-1/2} s \Sigma^{-1/2} \) that is \( s = 2 \Sigma^{1/2} \theta \Sigma^{1/2} \). The Laplace transform of \( W \) is then
\[
\mathbb{E}(e^{-\text{tr} (\theta W)}) = \frac{1}{\det(I_k + 2 \Sigma^{1/2} \theta \Sigma^{1/2}) n/2} \times \exp\{- \text{tr} ((I_k + 2 \Sigma^{1/2} \theta \Sigma^{1/2})^{-1} \Sigma^{1/2} \theta \Sigma^{1/2} \Sigma^{-1/2} M M^* \Sigma^{-1/2})\}
\]
which is the right-hand side of (12) for \( q = \sqrt{2} \Sigma^{1/2} \) which completes the proof. \( \Box \)

4. The moments and the variance function

In this section we compute the moments of \( \gamma(p, a; \sigma) \). Special attention is paid to the second moments. At the end of the section, we recall and offer a new presentation of a result by Alam and Mitra \[1\] giving the second order moment \( E((X - E(X))^2) \) for \( X \) noncentral Wishart. This result can be derived as a particular case of one of our results in this section. It is also crucial to the understanding of the estimation methods of \( a \) and \( \sigma \) given in that same paper and that we will recall further in Section 6.

4.1. Leibnitz formula

First we observe that the exponential family \( F(\gamma(p, a)) \) can be generated by the unbounded positive measure \( \mu(dt) = e^{\text{tr}(a t)} \gamma(p, a)(dt) \) whose Laplace transform is defined on \( -\mathcal{P}_k \) as
\[
L_\mu(\theta) = \int_{\mathcal{P}_k} e^{\text{tr}(\theta x)} \mu(dx) = \frac{1}{(-\theta)^{p}} e^{\text{tr}(a(-\theta)^{-1})}.
\]

We define the following two functions of \( \theta \in -\mathcal{P}_k \):
\[
\sigma = \sigma(\theta) = (-\theta)^{-1}, \quad k_\mu(\theta) = \text{tr}(a\sigma) + p \log \det \sigma.
\]

Note that \( k_\mu = \log L_\mu \) is the cumulant generating function of \( F(\gamma(p, a)) = F(\mu) \) when \( \mu \) is taken as the generating measure of the NEF. With this notation the element \( P(\theta, \mu)(dt) \) of the exponential family is exactly
\[
P(\theta, \mu) = \gamma(p, a; \sigma)
\]
as can be checked by (12) and
\[
\int_{\mathcal{P}_k} e^{-\text{tr} st} P(\theta, \mu)(dt) = \frac{L_\mu(\theta - s)}{L_\mu(\theta)}.
\]
Let us now recall some general facts about the moments of a multivariate exponential family generated by a measure \( \mu \) on some finite-dimensional real linear space \( E \). If \( \Theta(\mu) \) (contained in the dual \( E^* \) of \( E \)) is the interior of the domain of existence of \( L_\mu \) and if \( P(\theta, \mu)(dx) = \frac{e^{\langle \theta, x \rangle}}{L_\mu(\theta)(dx)} \in F(\mu) \) is the distribution indexed by \( \theta \) then the \( n \)th differential of the Laplace transform \( L_\mu \) in the directions \( h_1, \ldots, h_n \) has the following probabilistic interpretation

\[
L_\mu^{(n)}(\theta)(h_1, \ldots, h_n) = L_\mu(\theta) \int_E \langle h_1, x \rangle \ldots \langle h_n, x \rangle P(\theta, \mu)(dx)
\]  

(18)

where \( \langle h, x \rangle \) is the value taken by the linear form \( h \in E^* \) on the vector \( x \in E \). Thus this formula gives moments of \( P(\theta, \mu) \). In our case, \( E \) is the space of symmetric matrices of order \( k \), and \( E^* \) is identified to \( E \) by writing \( \langle \theta, x \rangle = \text{tr}(\theta x) \).

Let us also mention a general fact about the \( n \)th differential of the product of two real functions \( f \) and \( g \) defined on an open subset of a finite-dimensional real linear space \( F \): there exists in this context a formula similar to the Leibnitz formula. For \( F = \mathbb{R} \) the classical Leibnitz formula is:

\[
(f g)^{(n)}(\theta) = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} f^{(j)}(\theta)g^{(n-j)}(\theta).
\]

In general, let us fix a finite-dimensional real linear space \( F \). If \( (h_1, \ldots, h_n) \in F^n \) and if \( T \subset \{1, \ldots, n\} \) we denote \( h_T = (h_i)_{i \in T} \) and \( T' = \{1, \ldots, n\} \setminus T \). With these notations the general Leibnitz formula is

\[
(f g)^{(n)}(\theta)(h_1, \ldots, h_n) = \sum_{T \subset \{1, \ldots, n\}} f^{(|T|)}(\theta)(h_T)g^{(|T'|)}(\theta)(h_{T'}).
\]

(19)

### 4.2. Two differentials of order \( n \)

The Laplace transform \( L_\mu \) in (15) is of the form \( L_\mu(\theta) = f(\theta)g(\theta) \), with \( \theta \in F = E^*, f(\theta) = e^{\text{tr}(a\sigma(\theta))} = e^{\langle a, \sigma \rangle} \) and \( g(\theta) = e^{p \log \det \sigma(\theta)} \). Therefore in order to compute the moments of the \( \gamma(p, a; \sigma) \) distribution, we shall need to compute the differentials \( f^{(n)}(\theta)(h_1, \ldots, h_n) \) and \( g^{(n)}(\theta)(h_1, \ldots, h_n) \) of \( f \) and \( g \).

The differential \( g^{(n)} \) is known, (see [5,6,12]) and we shall only recall the result. Let \( S_n \) denote the group of permutations \( \pi \) of \( \{1, \ldots, n\} \) and \( C(\pi) \) the set of cycles of the permutation \( \pi \). We write \( m(\pi) \) for the number of cycles in \( \pi \) and define

\[
r_\pi(\sigma)(h_1, \ldots, h_n) = \prod_{c \in C(\pi)} \text{tr} \left( \prod_{j \in c} \sigma h_j \right).\]

Then the differential is

\[
g^{(n)}(\theta)(h_1, \ldots, h_n) = g(\theta) \sum_{\pi \in S_n} p^{m(\pi)} r_\pi(\sigma)(s_1, \ldots, s_k).
\]

(20)

The differential of \( f(\theta) = e^{\text{tr}(a\sigma)} \) is new. To derive it, we first need to recall the following two differentials

\[
\sigma'(\theta)(h) = \sigma h \sigma
\]

\[
(\log \det \sigma(\theta))'(h) = \text{tr} (\sigma h).
\]
Then for each $\pi$ in $S_n$, we introduce a quantity close to $r_\pi(\sigma)(h_1, \ldots, h_n)$ by simply replacing formally $A = \text{tr} \left( \prod_{j \in e} \sigma h_j \right)$ by $B = \text{tr} \left( \sigma a \prod_{j \in e} \sigma h_j \right)$. However this hardly makes sense for the following reason. Suppose that the cycle $c$ is $(2, 5, 4)$ that is, the permutation changes $2$ into $5$, $5$ into $4$ and $4$ into $2$. Then $A = \text{tr} \left( \sigma h_2 \sigma h_5 \sigma h_4 \right)$. The same cycle could also have been written as $(5, 4, 2)$ and the property of commutativity of traces implies that $A$ does not change, i.e $A$ depends on the cycle, not on its particular representation. Things are different for $B$, and the two numbers $\text{tr} \left( \sigma a \sigma h_2 \sigma h_5 \sigma h_4 \right)$ and $\text{tr} \left( \sigma a \sigma h_5 \sigma h_4 \sigma h_2 \right)$ do not coincide.

As a consequence, for a given integer $n$ we introduce the following set $P_n$ of objects. An element $P$ of $P_n$ consists of two things

- a partition $T = (T_1, \ldots, T_q)$ of $\{1, \ldots, n\}$ into nonvoid subsets (the order of the sequence $T_1, \ldots, T_q$ does not matter).
- a permutation $\pi_j$ of $T_j$ for each $j = 1, \ldots, q$.

Thus the information about $P$ is $q$ and the $q$ pairs $(T_j, \pi_j)$. For instance, the set $\{1, 2, 3\}$ has 5 partitions

$$
T^{(1)} = (\{1\}, \{2\}, \{3\})
$$

$$
T^{(2)} = (\{1, 2\}, \{3\})
$$

$$
T^{(3)} = (\{1\}, \{2, 3\})
$$

$$
T^{(4)} = (\{2\}, \{1, 3\})
$$

$$
T^{(5)} = (\{1, 2, 3\}).
$$

Thus $T^{(1)}, \ldots, T^{(5)}$ generate respectively $1, 2, 2, 6$ elements of $P_3$ and $P_3$ has $13$ elements. As another example, the partition of the set $\{1, 2, 3, 4, 5, 6\}$ given by $(\{1, 3\}, \{2, 4, 5\}, \{6\})$ generates $2! \times 3! \times 1 = 12$ elements of $P_6$. One can see that $P_4$ and $P_5$ have $73$ and $501$ elements respectively.

The following functions $s_\pi$ indexed by $P \in P_n$ are the analog of $r_\pi$. If $P$ is given by $q$ and $(T_j, \pi_j)$, $j = 1, \ldots, q$ we define

$$
s_\pi(\sigma)(h_1, \ldots, h_n) = \prod_{j=1}^q \left( \text{tr} \left( \sigma a \prod_{i \in T_j} \sigma h_{\pi_j(i)} \right) \right).
$$

**Proposition 4.1.**

$$
(e^{\text{tr}(a\sigma)})^{(n)}(\theta)(h_1, \ldots, h_n) = e^{\text{tr}(a\sigma)} \sum_{P \in P_n} s_\pi(\sigma)(h_1, \ldots, h_n).
$$

**The proof is by induction on $n$.**

**Example 1.** We compute the first three differentials of $\theta \mapsto f(\theta) = e^{\text{tr}(a\sigma)}$. For simplicity, we write $a' = \sigma a$ and $h'_j = \sigma h_j$. Thus from the previous proposition we get

$$
\frac{1}{f(\theta)} f'(\theta)(h_1) = \text{tr} \left( a' h'_1 \right)
$$

$$
\frac{1}{f(\theta)} f''(\theta)(h_1, h_2) = \text{tr} \left( a' h'_1 \right) \text{tr} \left( a' h'_2 \right) + \text{tr} \left( a' h'_1 h'_2 \right) + \text{tr} \left( a' h'_2 h'_1 \right)
$$

$$
\frac{1}{f(\theta)} f'''(\theta)(h_1, h_2) = \text{tr} \left( a' h'_1 \right) \text{tr} \left( a' h'_2 \right) \text{tr} \left( a' h'_3 \right) + \text{tr} \left( a' h'_1 h'_2 \right) \text{tr} \left( a' h'_3 \right) + \text{tr} \left( a' h'_2 h'_1 \right) \text{tr} \left( a' h'_3 \right)
$$

$$
+ \text{tr} \left( a' h'_2 h'_1 \right) \text{tr} \left( a' h'_3 \right) + \text{tr} \left( a' h'_1 h'_3 \right) \text{tr} \left( a' h'_2 \right) + \text{tr} \left( a' h'_3 h'_1 \right) \text{tr} \left( a' h'_2 \right)
$$
Consider the symmetric matrix for each symmetric and according to we have about the mean and the covariance. The linear endomorphism of Furthermore the covariance operator Proposition 4.2.

We now reformulate the results This enables us to compute the covariance Example 2. We now compute the first three differentials of \( g(\theta) = e^{p \log \det \sigma(\theta)} = \frac{1}{\det(-\theta)^p} \) still using the notation \( h'_j = \sigma h_j \). From (20) and according to (18) we have

\[
\frac{1}{g(\theta)} g'(\theta)(h_1) = p \text{ tr}(h'_1)
\]

\[
\frac{1}{g(\theta)} g''(\theta)(h_1, h_2) = p^2 \text{ tr}(h'_1) \text{ tr}(h'_2) + p \text{ tr}(h'_1 h'_2)
\]

\[
\frac{1}{g(\theta)} g'''(\theta)(h_1, h_2) = p^3 \text{ tr}(h'_1) \text{ tr}(h'_2) \text{ tr}(h'_3) + p \text{ tr}(h'_1 h'_2 h'_3) + p \text{ tr}(h'_2 h'_1 h'_3)
\]

\[
+ p^2 \text{ tr}(h'_1 h'_2) \text{ tr}(h'_3) + p^2 \text{ tr}(h'_1 h'_3) \text{ tr}(h'_2) + p^2 \text{ tr}(h'_2 h'_3) \text{ tr}(h'_1)
\]

4.3. Moments of orders 1, 2, 3

We now combine the results in the two examples above, using the Leibnitz formula in order to obtain the first two moments of the noncentral Wishart random variable \( X \sim P(\theta, \mu) = \gamma(p, a; \sigma) \) as in (17). Recall that \( P(\theta, \mu)(dx) = e^{\mu(\theta,x)} L_{\mu(\theta)}(dx) \), where \( \mu \) is as in (15) with cumulant generating function \( k_\mu \) as in (16), that \( p \in \Lambda_k a \in \mathcal{P}_k \) and \(-\theta \in \mathcal{P}_k\) is positive definite. We write \( \sigma = (-\theta)^{-1} h'_j = \sigma h_j \) and \( a' = \sigma a \). With these notations, the first two moments are

\[
\mathbb{E}(\text{ tr}(X h_1)) = k'_\mu(\theta)(h_1) = \frac{L'_\mu(\theta)(h_1)}{L_\mu(\theta)} = p \text{ tr}(h'_1) + \text{ tr}(a' h'_1)
\]

\[
\mathbb{E}(\text{ tr}(X h_1) \text{ tr}(X h_2)) = \frac{L''_\mu(\theta)(h_1, h_2)}{L'_\mu(\theta)} = \text{ tr}(a' h'_1) \text{ tr}(a' h'_2) + \text{ tr}(a' h'_1 h'_2) + \text{ tr}(a' h'_2 h'_1)
\]

\[
+ p \text{ tr}(h'_1) \text{ tr}(a' h'_2) + p \text{ tr}(h'_2) \text{ tr}(a' h'_1) + p^2 \text{ tr}(h'_1) \text{ tr}(h'_2) + p \text{ tr}(h'_1 h'_2).
\]

This enables us to compute the covariance \( k''_\mu(\theta) \) of \( X \) as

\[
\mathbb{E}[\text{ tr}((X - \mathbb{E}(X)) h_1) \text{ tr}((X - \mathbb{E}(X)) h_2)] = \text{ tr}(a' h'_1 h'_2) + \text{ tr}(a' h'_2 h'_1) + p \text{ tr}(h'_1 h'_2).
\]

We now reformulate the results (21) and (23) about the mean and the covariance. The linear space \( S_k \) of real symmetric matrices of dimension \( k \) is equipped with the Euclidean structure \( \langle h_1, h_2 \rangle = \text{ tr}(h_1 h_2) \).

Proposition 4.2. Let \( X \) follow the noncentral Wishart \( \gamma(p, a; \sigma) \) as above. Then

\[
\mathbb{E}(X) = m = k'_\mu(\theta) = p \sigma + \sigma a \sigma = p \sigma + \omega.
\]

Furthermore the covariance operator \( k''_\mu(\theta) = \mathbb{E}((X - \mathbb{E}(X)) \otimes (X - \mathbb{E}(X))) \), as an endomorphism of \( S_k \) is given by the linear map

\[
h \mapsto \sigma a h \sigma + \sigma h \sigma a + p \sigma h \sigma = \omega h \sigma + \sigma \omega + p \sigma h \sigma = mh \sigma + \sigma hm - p \sigma h \sigma.
\]

Proof. Consider the symmetric matrix \( v = \mathbb{E}(X) - p \sigma - \sigma a \sigma \). From (21) for each symmetric matrix \( h \) we have \( \text{ tr}(v h) = 0 \) (recall that \( \text{ tr}(\mathbb{E}(X) h) = \mathbb{E}(\text{ tr}(X h)) \)). Now taking \( h = v \)
yields $\text{tr} v^2 = 0$. Since $v^2$ is a semi-positive definite matrix, this implies that $v = 0$ and (24) is proved.

Consider the covariance $c = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X)))$ of $X$. From (23) it follows that

\[
\langle c(h_1), h_2 \rangle = \text{tr} (\sigma \sigma h_1 \sigma h_2) + p \text{tr} (\sigma h_1 \sigma h_2) = \langle \sigma \sigma h_1 \sigma + \sigma h_1 \sigma a \sigma + p \sigma h_1 \sigma, h_2 \rangle.
\]

Since this is true for all $h_2$, $c(h_1) = \sigma \sigma h_1 \sigma + \sigma h_1 \sigma a \sigma + p \sigma h_1 \sigma$ which is the first expression of the covariance operator in (25). Replacing $\sigma a \sigma$ by $\omega = m - p \sigma$ gives the other expressions.

We will now display the third moment $\frac{1}{\mathbb{L}_m(\theta)} L^m(\theta)(h_1, h_2, h_3)$ which is the following sum of 27 monomials (the fourth moment is the sum of 267 monomials):

\[
\mathbb{E} \left[ \prod_{i=1}^{3} \text{tr}(Xh_i) \right] = \text{tr} (a'h'_1) \text{tr} (a'h'_2) \text{tr} (a'h'_3) + \text{tr} (a'h'_1) \text{tr} (a'h'_2) \text{tr} (a'h'_3)
\]

for a noncentral Wishart random variable is only obtained as the sum of a series of zonal polynomials:

**Proposition 4.3.** If $X$ has the standard noncentral Wishart distribution $\gamma(p, a)$ then for $p - k + 1 > 0$ we have

\[
\left( p - \frac{k + 1}{2} \right) \mathbb{E}(X^{-1}) = I_k - e^{-\text{tr} a} a^{1/2} \left( \sum_{m=0}^{k} \sum_{\kappa \in E_m} \frac{m_k(p, a)}{m!(p)_\kappa} \right) a^{1/2}
\]
where the symmetric matrix $m_\kappa(p,a)$ is
\[ m_\kappa(p,a) = \int_{\mathcal{P}_k} C_\kappa'(a^{1/2}xa^{1/2})(\det x)^{p-k+1/2} e^{-\text{tr} x} \, dx. \]

**Proof.** Let
\[ M(x) = \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{C_\kappa(x)}{m!(p)_\kappa}. \]
From Proposition 2.2, the density of $X$ is then
\[ f(x) = M(a^{1/2}xa^{1/2})e^{-\text{tr}(a+x)}(\det x)^{p-k+1/2} \frac{1}{\Gamma_k(p)}. \]
Now apply Stokes' formula to $f$ and to the closure of $\mathcal{P}_k$ to claim that for any symmetric matrix $h$
\[ \int_{\mathcal{P}_k} f'(x)(h) \, dx = \int_{\mathcal{P}_k} \frac{f'(x)(h)}{f(x)} f(x) \, dx = 0. \]
Since
\[ \frac{f'(x)(h)}{f(x)} = \left(p - \frac{k+1}{2}\right) \text{tr}(x^{-1}h) - \text{tr} h + \frac{M'(a^{1/2}xa^{1/2})(a^{1/2}xa^{1/2})}{M(a^{1/2}xa^{1/2})}, \]
we obtain
\[ \left(p - \frac{k+1}{2}\right) \mathbb{E}(X^{-1}) = I_k - e^{-\text{tr} a} a^{1/2} \]
\[ \times \left( \int_{\mathcal{P}_k} M'(a^{1/2}xa^{1/2})e^{-\text{tr} x}(\det x)^{p-k+1/2} \frac{1}{\Gamma_k(p)} \right) a^{1/2}, \]
which proves (26). □

### 4.5. The variance function

We now calculate the variance function of the noncentral Wishart, i.e., the NEF generated by $\mu$ indexed by $p$ and $a$ and with Laplace transform (15). We have to express $\sigma$ as a function of $m$ where $m$ and $\sigma$ are related by (24), a strange second degree equation. To do so, we need the following lemma.

**Lemma 4.4.** Let $a$ and $b$ be in the set $\mathcal{P}_k$ of positive definite real symmetric matrices of order $k$. Then there exists one and only one matrix $x \in \mathcal{P}_k$ such that $xax = b$. This matrix is
\[ x = a^{-1/2}(a^{1/2}ba^{1/2})^{1/2}a^{-1/2}. \]

**Proof.** Clearly $x = a^{-1/2}(a^{1/2}ba^{1/2})^{1/2}a^{-1/2}$ is a solution and therefore such a matrix exists. Let us prove it is unique. If $y \in \mathcal{P}_k$ is another solution, then $a^{1/2}ya^{1/2} \in \mathcal{P}_k$ is a square root of $a^{1/2}ba^{1/2}$. Since the square root in $\mathcal{P}_k$ is unique we have $a^{1/2}ya^{1/2} = a^{1/2}xa^{1/2}$ which implies $x = y$. □
Proposition 4.5. Let \(a\) be semi-positive definite. The variance function of the natural exponential family generated by the measure \(\mu\) of (15) is
\[
V(m)(h) = m h \sigma + \sigma h m - p \sigma h \sigma
\]
where \(\sigma\) is as follows. If \(a\) is invertible we have
\[
\sigma = -\frac{p}{2} a^{-1} + a^{-1/2} \left( a^{1/2} ma^{1/2} + \frac{p^2}{4} I_k \right)^{1/2} a^{-1/2}.
\]
If \(a\) is not invertible, orthonormal coordinates in \(\mathbb{R}^k\) are chosen such that \(a, m\) and \(\sigma\) are written by blocks \(k_1 \times k_1, k_1 \times k_2, k_2 \times k_1, k_2 \times k_2\) with \(k_1 + k_2 = k\)
\[
a = \begin{bmatrix} a_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad m = \begin{bmatrix} m_1 & m_{12} \\ m_{21} & m_2 \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{21} & \sigma_2 \end{bmatrix}\]
where \(a_1\) is invertible. Then \(\sigma_1\) is obtained from \(a_1\) and \(m_1\) by the formula (28). Furthermore \(\sigma_{12} = (\sigma_1 a_1 + p I_{k_1})^{-1} m_{12}\) and \(\sigma_2 = \frac{1}{p}(m_2 - \sigma_{21} m_1 \sigma_{12})\), where \(\sigma_{21}\) is the transposed matrix of \(\sigma_{12}\).

Finally, with the notation \(\omega = \sigma a \sigma\) we also write \(m = \omega + p \sigma\) and
\[
V(m)(h) = \omega h \sigma + \sigma h \omega + p \sigma h \sigma.
\]

Proof. Let \(m \in \mathcal{P}_k\). We compute \(\sigma \in \mathcal{P}_k\) such that (24) holds. For this we write \(\sigma = x - \frac{p}{2} a^{-1}\) from which it follows that \(x a x = m + \frac{p^2}{4} a^{-1}\). Apply the previous lemma to \(a\) and \(b = m + \frac{p^2}{4} a^{-1}\) and we get
\[
\sigma = -\frac{p}{2} a^{-1} + x = -\frac{p}{2} a^{-1} + a^{-1/2} \left( a^{1/2} ma^{1/2} + \frac{p^2}{4} I_k \right)^{1/2} a^{-1/2}.
\]
Since the variance function is the endomorphism \(V(m)\) defined by \(m \mapsto m h \sigma + \sigma h m - p \sigma h \sigma\), the desired result is obtained for the invertible case. When \(a\) is singular, it is easier to study the equation \(\sigma a \sigma + p \sigma = m\) when the coordinates are chosen so that \(a = \begin{bmatrix} a_1 & 0 \\ 0 & 0 \end{bmatrix}\) with \(a_1\) invertible. Note that if \(a = 0\) we get back the variance function of the central Wishart distribution \(V(m)(h) = \frac{1}{p} m h m\). Finally (29) is easily obtained from (27).

4.6. The formula of Alam and Mitra and its extension

Alam and Mitra [1] have written a remarkable paper which has not received the attention it deserves. They first prove formula (31), or rather the equivalent (32) below, which is a particular case of the following more general result.

Proposition 4.6. Let \(X\) be a noncentral Wishart random variable \(\gamma(p, a, \sigma)\) as defined by (12), with the notation \(\omega = \sigma a \sigma\) and \(m = p \sigma + \omega\). Then for all symmetric matrices \(h\) of order \(k\)
\[
\mathbb{E}((X - m)h(X - m)) = \frac{1}{2} \left[ wh \sigma + \sigma h \omega + p \sigma h \sigma + w \text{tr}(h \sigma) \right. \\
\left. + \sigma \text{tr}(h \omega) + p \sigma \text{tr}(h \sigma) \right].
\]
Remark. For $h = I_k$, this yields the following formula first proved by Alam and Mitra [1]

$$
\mathbb{E}((X - m)^2) = \frac{1}{2} [w \sigma + \sigma \omega + p \sigma^2 + w \text{tr}(\sigma) + \sigma \text{tr}(\omega) + p \sigma \text{tr}(\sigma)]
$$

(31)

$$
= \frac{1}{2} [(\omega + \sigma)^2 + (\omega + \sigma)\text{tr}(\omega + \sigma) + (p - 1)(\sigma^2 + \sigma \text{tr}\sigma)]
$$

$$
= \frac{1}{2} p [m^2 + m \text{tr} m - \omega^2 - \omega \text{tr} \omega].
$$

(32)

Formula (32) is (2.1) in [1] but with different notations. Let us comment on the importance of (30) and of the innocent looking (32). Up to now, we have been considering a noncentral Wishart random variable $X$ only as an element of the linear space of symmetric matrices of order $k$. We have never used the multiplicative structure (or rather the Jordan algebra structure $x \circ y = xy + yx$) of symmetric matrices. In this proposition, the squares $X^2$ play an important role. For the ordinary Wishart distributions, we find in the literature not only expressions like $\mathbb{E}(\langle h, X \rangle \ldots \langle h, X \rangle)$ (recall that $\langle h, X \rangle = \text{tr}(hX)$) but also expressions like

$$
\mathbb{E}(\langle h_1, X^{\alpha_1} \rangle \ldots \langle h_n, X^{\alpha_n} \rangle)
$$

where $\alpha_1, \ldots, \alpha_n$ are arbitrary integers, that is expressions involving powers of $X$ (see [5,6,12]). For noncentral Wishart, formula (31) given by Alam and Mitra [1] is the first of this type (with $n = 1$ and $\alpha_1 = 2$).

To prove (30) we need a result of linear algebra that we will state without proof. Let $L_s(S_k)$ be the space of symmetric endomorphisms of the Euclidean space $S_k$ namely the set of linear maps $f$ of $S_k$ into itself such that for any $u$ and $v$ in $S_k$ we have $\text{tr}(f(u)v) = \text{tr}(uf(v))$. To each $y \in S_k$ we associate the elements $y \otimes y$ and $\mathbb{P}(y)$ of $L_s(S_k)$ defined respectively by

$$
h \mapsto (y \otimes y)(h) = y \text{tr}(yh), \quad h \mapsto \mathbb{P}(y)(h) = yhy
$$

which are two important examples of elements of $L_s(S_k)$. Observe that $y \otimes y$ is not the Kronecker product $y \otimes y$ of $y$ by itself. Indeed $y \otimes y$ is the representative matrix in a natural basis of the operator $\varphi$ sending the space of $(n,n)$ real matrices into itself by $\varphi(M) = yMy$. If $h$ is symmetric, $\varphi(h) = \mathbb{P}(h)$.

If $x$ and $y$ are in $S_k$ one can even consider $(x \otimes y + y \otimes x) = ((x + y) \otimes (x + y) - x \otimes x - y \otimes y)(h) = x \text{tr}(yh) + y \text{tr}(xh)

$$
\mathbb{P}(x,y)(h) = (\mathbb{P}(x + y) - \mathbb{P}(x) - \mathbb{P}(y))(h) = xhy + yhx.
$$

With these notations, (27) and (29) could be rewritten with $Y = X - m$ as

$$
V(m) = \mathbb{E}(Y \otimes Y) = \mathbb{P}(m, \sigma) - p\mathbb{P}(\sigma) = \mathbb{P}(\omega, \sigma) + p\mathbb{P}(\sigma).
$$

(33)

Finally, equation (30) that we aim to prove is

$$
\mathbb{E}(\mathbb{P}(Y)) = \frac{1}{2} [\mathbb{P}(\omega, \sigma) + (\omega \otimes \sigma + \sigma \otimes \omega) + p(\mathbb{P}(\sigma) + \sigma \otimes \sigma)].
$$

(34)

Now, $L_s(S_k)$ is itself a linear space, and the result that we are going to admit is the following (see [2] Lemma 6.1 and [11] Prop. 3.1 for a proof).
Proposition 4.7. There exists a unique endomorphism $\Psi$ of $L_{s}(S_{k})$ such that for all $y \in S_{k}$ one has

$$\Psi(y \otimes y) = \mathbb{P}(y).$$  \hfill (35)

Furthermore

$$\Psi(\mathbb{P}(y)) = \frac{1}{2}(y \otimes y + \mathbb{P}(y)).$$  \hfill (36)

Proof of Proposition 4.6. The proof is now very easy. Using (35) and (33), we have

$$E(\mathbb{P}(Y)) = E(\Psi(Y \otimes Y)) = \Psi(E(Y \otimes Y))$$
$$= \Psi(\mathbb{P}(\omega, \sigma)) + p \Psi(\mathbb{P}(\sigma))$$
$$= \Psi(\mathbb{P}(\omega + \sigma)) - \Psi(\mathbb{P}(\omega)) - \Psi(\mathbb{P}(\sigma)) + p \Psi(\mathbb{P}(\sigma)).$$

Now applying (36) we get (31) under the form (34). To pass from (31)–(32) use $m = p\sigma + \omega$. \hfill \Box

5. Proof of Proposition 2.3

First we need another variation of the Leibnitz formula. Let $\theta \mapsto f(\theta)$ and $\theta \mapsto g(\theta)$ be sufficiently differentiable real functions defined on the same open subset of $\mathbb{R}^{n}$. For $j = 1, \ldots, n$ we write $D_{j} = \frac{\partial}{\partial \theta_{j}}$. Then for $a = (a_{1}, \ldots, a_{n}) \in \mathbb{N}^{n}$

$$D_{1}^{a_{1}} \ldots D_{n}^{a_{n}}(fg)(\theta) = \sum \left(\begin{array}{c} a_{1} \\ i_{1} \\ \vdots \\ a_{n} \\ i_{n} \end{array}\right) \right) D_{1}^{i_{1}} \ldots D_{n}^{i_{n}}(f)(\theta) D_{1}^{a_{1} - i_{1}} \ldots D_{n}^{a_{n} - i_{n}}(g)(\theta)$$  \hfill (37)

where the sum is taken for all $i = (i_{1}, \ldots, i_{n}) \in \mathbb{N}^{n}$ such that $i_{j} \leq a_{j}, j = 1, \ldots, n$.

Let us now prove the proposition. We imitate the proof of Gindikin’s theorem due to Shanbag [16]. Let $a \in \mathbb{P}_{k}$. Suppose that there exist $p > 0$ and a positive measure $\mu_{p}(dt)$ on $\mathbb{P}_{k}$ such that for all $\theta \in -\mathbb{P}_{k}$ one has

$$\frac{1}{(-\theta)^{p}} e^{\text{tr}(a(-\theta)^{-1})} = \int_{\mathbb{P}_{k}} e^{\text{tr}(\theta t)} \mu_{p}(dt).$$  \hfill (38)

We show that $p \in \Lambda$.

Let $Q$ be any real polynomial on the space of real symmetric matrices of order $k$. Then we have

$$Q \left( \frac{\partial}{\partial \theta} \right) \frac{1}{(-\theta)^{p}} e^{\text{tr}(a(-\theta)^{-1})} = \int_{\mathbb{P}_{k}} Q(t) e^{\text{tr}(\theta t)} \mu_{p}(dt).$$

Suppose that the maximal degree of $Q$ is $n$. Then there exists a real polynomial $P_{Q}$ on $\mathbb{R}$ with respect to $p$ such that

$$Q \left( \frac{\partial}{\partial \theta} \right) \frac{1}{(-\theta)^{p}} e^{\text{tr}(a(-\theta)^{-1})} = \frac{1}{(-\theta)^{n+p}} e^{\text{tr}(a(-\theta)^{-1})} P_{Q}(p).$$  \hfill (39)

Let us insist on the fact that the coefficients of $P$ depend on $\theta$ and $a$. This result can be shown by using Leibnitz formula (37) applied to the usual pair $f(\theta) = e^{\text{tr}(a\sigma(\theta))} = e^{\langle a, \sigma \rangle}$ and
\(g(\theta) = (\det \sigma(\theta))^p\), using induction on \(n\). We now apply (39) to the polynomial \(Q(t) = \det t\) whose degree is \(k\) to obtain

\[
\frac{1}{(-\theta)^{k+p}} e^{\text{tr}(a(-\theta)^{-1})} P_Q(p) = \int_{\mathbb{P}_k} (\det t) e^{\text{tr}(\theta t)} \mu_p(dt). \tag{40}
\]

Note that the right-hand side of (40) is \(\geq 0\). Note also that this right-hand side is 0 for \(p = 0, 1/2, \ldots, (k-1)/2\) since \(\mu_0 = \delta_0\) and since \(\mu_p(dt)\) is concentrated on singular matrices in \(k-1\) other cases. Now the left-hand side of (40) has the same sign as \(P_Q(p)\) which is a polynomial of degree \(\leq k\) with at least zeros at \(p = 0, 1/2, \ldots, (k-1)/2\). Furthermore, Proposition 2.2 shows that \(P_Q(p) > 0\) for \(p > (k-1)/2\). Thus deg \(P_Q = k\), and the zeros of \(P_Q\) are all real and simple. Also \((-1)^i P_Q(p) > 0\) for \(\frac{k-1-i}{2} < p < \frac{k-i}{2}\) and \(i = 1, \ldots, k-1\). Now, assume that a positive measure \(\mu_p\) exists and that \(p \notin A\). Then we would have \(P_Q(p) > 0\) and therefore there would exist an even \(i \in \{1, \ldots, k-1\}\) such that \(\frac{k-1-i}{2} < p < \frac{k-i}{2}\). For \(k = 2\) this is impossible. For \(k \geq 3\) we observe that if \(\mu_p\) exists, then

\[
\mu_{p+\frac{1}{2}} = \mu_p * \mu_p \tag{40.1}
\]

also exists, as can be seen using the Laplace transform. But now \(P_Q(p + \frac{1}{2}) < 0\) which is a contradiction.

To complete the proof, suppose that there exists \(p \notin A\) such that a probability \(\gamma(p, a)\) on \(\mathbb{P}_k\) exists and such that for \(I_k + s \in \mathbb{P}_k\) one has

\[
\int_{\mathbb{P}_k} e^{-\text{tr}(st)} \gamma(p, a)(dt) = \frac{1}{\det(I_k + s)^p} e^{-\text{tr}((I_k + s)^{-1}sa)}. \tag{40.2}
\]

Defining \(\mu_p(dt) = e^{\text{tr}(t+a)} \gamma(p, a)(dt)\) we see that (38) holds. This contradiction ends the proof. \(\square\)

6. The estimation of \(\sigma\) and \(a\)

Let \(X_1, \ldots, X_N\) be \(N\) iid observations from the noncentral Wishart distribution \(\gamma(p, a; \sigma)\) as defined by (12). If \(p\) and \(a\) are known, and \(\sigma\) is unknown we have an natural exponential family. However, if \(a\) is unknown, this is not longer true and our model is not even a general exponential family. We therefore devise below new methods of estimation. All methods will be of the following type: given some functions \(\alpha\) and \(\beta\) of \((a, \sigma)\) we invert them to obtain

\[
a = f(\alpha, \beta) \quad \sigma = g(\alpha, \beta). \tag{6.1}
\]

Each method then chooses \((\alpha, \beta)\) along with estimators \((\hat{\alpha}, \hat{\beta})\) which are, of course, functions of the observations \(X_1, \ldots, X_N\). These estimators are generally unbiased, but not always. We now plug \((\hat{\alpha}, \hat{\beta})\) into \(f(\alpha, \beta)\) and \(g(\alpha, \beta)\) in order to get the estimators

\[
\hat{a} = f(\hat{\alpha}, \hat{\beta}) \quad \hat{\sigma} = g(\hat{\alpha}, \hat{\beta}). \tag{6.2}
\]

6.1. \(\sigma\) unknown and \(a\) known

In this case, the model is a natural exponential family. The estimation of \(\sigma = (-\theta)^{-1}\) is easy since the methods of natural exponential families are now available, and we can find the
maximum likelihood estimator. We take
\[ \hat{m} = \bar{X}_N = \frac{1}{N}(X_1 + \cdots + X_N) \]
as an estimator of \( m \) and if \( a \) is invertible, we plug this value of \( m \) into (28) and we get a reasonable estimator for \( \sigma \). If \( a \) is singular, we use the formula following (28). Let us note, for further use, that the distribution of \( \bar{X}_N \) is \( \gamma(Np, N^2a, \sigma/N) \). Thus \( \omega = \sigma a \sigma \) is unchanged and from (32), we have
\[ E((\bar{X}_N - m)^2) = \frac{1}{2Np}(m^2 + m \text{tr} m - \omega^2 - \omega \text{tr} \omega). \]

6.2. \( \sigma \) known and \( a \) unknown

The classical method is to use \( m = p\sigma + \sigma a\sigma = p\sigma + \omega \) to obtain the estimate
\[ \hat{a} = \sigma^{-1}\bar{X}_N\sigma^{-1} - p\sigma^{-1}. \]
This is specially popular when \( \sigma = I_k \) since \( \hat{a} = \bar{X}_N - pI_k \). See [13] for variants and properties. This estimator is not always semi-positive definite.

6.3. Method 1 of Alam and Mitra for \( a \) and \( \sigma \) unknown

This method uses \( \alpha = m \) and \( \beta = \omega^2 + \omega \text{tr} \omega \) (with the usual notation \( \omega = \sigma a \sigma \)). A remarkable observation of [1] is the fact that \( \omega \mapsto \omega^2 + \omega \text{tr} \omega \) is a bijection of \( \mathbb{P}_k \) onto itself, as a consequence of the following lemma.

Lemma 6.1. If \( \omega \) is a semi-positive definite matrix and if \( \beta = \omega^2 + \omega \text{tr} \omega \) then \( \nu = \text{tr} \omega \) is a function of \( \beta \) alone.

Proof of the Lemma. We have that
\[ \beta + \frac{\nu^2}{4}I_k = \left(\omega + \frac{\nu}{2}I_k\right)^2 \]
which leads to
\[ \left(\beta + \frac{\nu^2}{4}I_k\right)^{1/2} = \omega + \frac{\nu}{2}I_k. \] (41)
Taking the trace of both sides we see that the following function on \( \mathbb{R} \)
\[ h(x) = -x \left(1 + \frac{k}{2}\right) + \text{tr} \left(\beta + \frac{x^2}{4}I_k\right)^{1/2} \]
is zero for \( x = \nu \). Since \( x \mapsto h(x) \) is easily proved to be convex (for \( \beta \) is semi-positive definite) and since \( \lim_{x \to -\infty} h(x) = -\infty \) then \( \nu \) is the only root of \( h(x) = 0 \). This shows that \( \nu = \nu(\beta) \) is a function of \( \beta \) alone. □

We now describe the two functions \( a = f(\alpha, \beta) \) \( \sigma = g(\alpha, \beta) \) corresponding to the first method of Alam and Mitra. We observe that (41) gives \( \omega \) as a function of \( \beta \):
\[ \omega = \omega(\beta) = -\frac{\nu(\beta)}{2}I_k + \left(\beta + \frac{\nu(\beta)^2}{4}I_k\right)^{1/2}. \]
Since $\sigma = \frac{1}{p}(m - \omega)$ and $a = \sigma^{-1}\omega\sigma^{-1}$ we get

$$\sigma = f(\alpha, \beta) = \frac{1}{p}(\alpha - \omega(\beta)), \quad a = g(\alpha, \beta) = f(\alpha, \beta)^{-1}\omega(\beta)f(\alpha, \beta)^{-1}.$$  

The second part of the method is to choose the estimators $\hat{\alpha}$ and $\hat{\beta}$. For $\hat{\alpha}$ we take $\bar{X}_n$. For $\hat{\beta}$ we first recall formula (32) that we write under the form

$$\beta = \omega^2 + \omega \text{ tr } \omega = m^2 + m \text{ tr } m - 2p \mathbb{E}((X - m)^2).$$

Consider now the following estimators of $\mathbb{E}((X - m)^2)$ and of $m^2 + m \text{ tr } m$ respectively defined by

$$\hat{\beta}_1 = \frac{1}{N - 1} \sum_{i=1}^{N} (X_i - \bar{X}_N)^2,$$

$$\hat{\beta}_2 = \frac{1}{N(N - 1)} \sum_{i \neq j} (X_i X_j + X_i \text{ tr } X_j).$$

They are both unbiased. For $\hat{\beta}_1$ write $(X_i - \bar{X}_N)^2 = (X_i - m + m - \bar{X}_N)^2$ and for $\hat{\beta}_2$ observe that $X_i$ and $X_j$ are independent and that $\mathbb{E}(\text{ tr } X_j) = \text{ tr } \mathbb{E}(X_j)$. Thus $\hat{\beta}_3 = \hat{\beta}_2 - 2p \hat{\beta}_1$ is an unbiased estimator of $\beta$. It is not necessarily semi-positive definite. For this reason, let us write $\hat{\beta}_3 = u \text{ diag}(c_1, \ldots, c_k) u^{-1}$ such that $u$ is an orthogonal matrix and such that the eigenvalues of $\hat{\beta}_1$ satisfy $c_1 > \cdots c_j > 0 > c_{j+1} > \cdots > c_k$. The estimator $\hat{\beta}$ that we finally consider and that we plug into $f$ and $g$ is simply the semi-positive definite matrix

$$\hat{\beta} = u \text{ diag}(c_1, \ldots, c_j, 0, \ldots, 0) u^{-1}.$$  

What is the cost of this method? Essentially:

1. The calculation of $\hat{\beta}$ from $\hat{\beta}_3$, that is, the diagonalization of a symmetric matrix.
2. The calculation of $v(\hat{\beta})$ (which is the trace of the estimator of $\omega$), for instance by the Newton approximation.

The cost of the calculation of $(\hat{\beta} + \frac{v(\hat{\beta})^2}{4} I_k)^{1/2}$ after the calculations 1 and 2 above are completed, is negligible.

6.4. Method 2 of Alam and Mitra for $a$ and $\sigma$ unknown

The second method is based on (14). Its aesthetic value is diminished by the fact that it is not free of coordinates as the first method was. For any column vector $z \in \mathbb{R}^k$ define $\alpha_z = m_z$ and $\beta_z = v_z$ with the notations of (14). One can express $z^*\sigma z$ and $z^*\omega z$ with respect to $\alpha_z$ and $\beta_z$ since from (14) we have

$$z^*\sigma z = \frac{1}{p}(m_z + (m_z^2 - p v_z)^{1/2}) = \frac{1}{p}(\alpha_z + (\alpha_z^2 - p \beta_z)^{1/2}),$$

$$z^*\omega z = (m_z^2 - p v_z)^{1/2} = (\alpha_z^2 - p \beta_z)^{1/2}.$$  

In the formula above, Alam and Mitra [1] replace $\alpha_z$ and $\beta_z$ by the unbiased estimators

$$\hat{\alpha}_z = z^* \bar{X}_N z = \text{ tr } \bar{X}_N zz^*,$$

$$\hat{\beta}_z = \frac{1}{N - 1} \sum_{i=1}^{N} (z^* X_i z - z^* \bar{X}_N z)^2 = \frac{1}{N - 1} \sum_{i=1}^{N} \text{ tr } (X_i - \bar{X}_N) z z^*.$$
Let us temporarily write
\[ f(z z^*) = \frac{1}{p}(\tilde{\alpha} z + (\tilde{\alpha} z^2 - p \tilde{\beta} z)^{1/2}), \quad g(z z^*) = (\tilde{\alpha} z^2 - p \tilde{\beta} z)^{1/2}. \] (42)

Alam and Mitra [1, Eqs. (2.9) and (2.10)], define the estimators \( \hat{\sigma} \) and \( \hat{\omega} \) by
\[ z^* \hat{\sigma} z = \text{tr}(\hat{\sigma} z z^*) = f(z z^*), \quad z \hat{\omega} z^* = \text{tr}(\hat{\omega} z z^*) = g(z z^*). \] (43)

There are, in fact, no \( \hat{\sigma} \) or \( \hat{\omega} \) which satisfy these equations for all \( z \in \mathbb{R}^k \). This would imply in particular that the function \( h \mapsto g(h) \) defined on the space \( S_k \) of symmetric matrices satisfying for semi-positive \( h \)
\[ g(h) = \left( \text{tr}(X_N h)^2 - \frac{1}{N - 1} \sum_{i=1}^N [\text{tr}(X_i - X_N) h]^2 \right)^{1/2} \]
is a linear form on \( S_k \). This is clearly not true for most \( (X_1, \ldots, X_N) \). To see this, simply square both sides of the equation above and use Kronecker’s law of inertia for the quadratic forms. However, if we assume (as Alam and Mitra [1] eventually do) that (43) is true only for \( z \)’s of the form \( z = e_i + e_j \) where \( e = (e_1, \ldots, e_k) \) is the canonical basis of \( \mathbb{R}^k \), then (43) defines \( \hat{\sigma} \) and \( \hat{\omega} \) completely. As one can see, this method is linked to the special basis \( e \). Its advantage is the simplicity of formulas (42).

6.5. Our method for \( a \) and \( \sigma \) unknown

To estimate \( a \), we use (25) which says that \( V(m)(h) = mh \sigma + \sigma hm - p \sigma^2 h \sigma \). We let \( h = I_k \) and for simplicity, denote
\[ v = V(m)(I_k) = m \sigma + \sigma m - p \sigma^2. \]

By definition we have
\[ v = \mathbb{E}((X - \mathbb{E}(X)) \otimes (X - \mathbb{E}(X))(I_k) = \mathbb{E}((X - \mathbb{E}(X)) \times \text{tr} (X - \mathbb{E}(X))). \]
But the equality \( v = m \sigma + \sigma m - p \sigma \) can be rewritten as \( (\sigma - \frac{1}{p} m)^2 = \frac{1}{p} m^2 - \frac{1}{p} v \) or
\[ \sigma = \frac{1}{p}(m + (m^2 - pv)^{1/2}). \]

Thus finding an estimate of \( v \) will give us an estimate of \( \sigma \). This, in turn, will lead to an estimate of \( a \) via (24) since
\[ a = \sigma^{-1} m \sigma^{-1} - p \sigma^{-1}. \]

We now suggest the following unbiased estimator for the matrix \( v \)
\[ \hat{v} = \frac{1}{N - 1} \sum_{j=1}^N [(X_j - X_N) \times \text{tr} (X_j - X_N)] \]
which leads to the estimators \( \hat{\sigma} \) and \( \hat{a} \) for \( \sigma \) and \( a \) respectively defined by
\[ \hat{\sigma} = \frac{1}{p}(\hat{m} + (\hat{m}^2 - p \hat{v})^{1/2}), \quad \hat{a} = \hat{\sigma}^{-1} \hat{m} \hat{\sigma}^{-1} - p \hat{\sigma}^{-1}. \]
This method belongs to the general framework mentioned in the introduction of Section 6, for
\((\alpha, \beta) = (m, \nu)\). Here the functions \(f\) and \(g\) are
\[
g(\alpha, \beta) = \frac{1}{p}(\alpha + (\alpha^2 - p\beta)^{1/2}), \quad f(\alpha, \beta) = g(\alpha, \beta)^{-1}\alpha g(\alpha, \beta)^{-1} - pg(\alpha, \beta)^{-1}.
\]

For \(\sigma\) to exist, the symmetric matrix \(\hat{m}^2 - p\hat{v}\) has to be semi-positive definite and this condition
is not necessarily fulfilled for every sample \(X_1, \ldots, X_N\). Estimator \(\hat{e}_3\) for the first method of [1]
was suffering from the same defect and we could force \(\hat{m}^2 - p\hat{v}\) to be semi-positive definite by
the same trick which was replacing \(\hat{e}_3\) by \(\hat{\sigma}\). The properties of these estimators have yet to be
studied. A positive side of the proposed method is that it uses only one square root of a matrix,
namely \(\frac{1}{p}(\hat{m}^2 - \hat{v})^{1/2}\). For a known, the classical method of estimation of parameters for a natural
exponential family was leading to the estimator of \(\sigma\) equal to
\[
-\frac{p}{2}a^{-1} + a^{-1/2}
\left(a^{1/2}\hat{m}a^{1/2} + \frac{p^2}{4}I_k\right)^{1/2} a^{-1/2}
\]
involving two roots in the nonsingular case and was leading to an even more complicated formula
in the singular case.

6.6. A closer look at our estimation method

As said above, a weak point of our method is the fact that our estimators are undefined when
\(U_N = \hat{m}^2 - p\hat{v}\) is not in the cone \(\overline{P}_k\) of semi-positive definite matrices. Clearly from the law
of large numbers we have \(\lim_N \Pr(U_N \in \overline{P}_k) = 1\). For relatively small \(N\), we are now going to
give a method for approximating \(\Pr(U_N \in \overline{P}_k)\). To do so, we first state a classical result.

**Proposition 6.2.** Let \(E\) be an Euclidean space with scalar product \(\langle x, y \rangle_E\) and let \(S(E)\) be the
space of symmetric linear endomorphisms of \(E\) endowed with the scalar product \(\langle a, b \rangle_{S(E)} = \text{tr}(ab)\).
Let \(\mu\) be a probability on \(E\) with mean \(m\) and covariance \(V\) with exponential moments.
Let \(X_1, \ldots, X_N, \ldots\) be a sequence of iid random variables of \(E\) with distribution \(\mu\). Let
\(\bar{X}_N = \frac{1}{N}(X_1 + \cdots + X_N)\) and
\[
S_N^2 = \frac{1}{N-1} \sum_{j=1}^N (X_j - \bar{X}_N) \otimes (X_j - \bar{X}_N)
\]
be the empirical mean and empirical covariance. Then the distribution of the random variable
\(\sqrt{N}(X_N - m, S_N^2 - V)\) of \(E \times S(E)\) converges to the distribution of a centered normal variable
\((Z_1, Z_2)\) of \(E \times S(E)\) such that \(Z_1\) and \(Z_2\) are independent and such that
1. the covariance of \(Z_1\) is \(V\),
2. the covariance of \(Z_2\) is the linear operator \(\mathbb{P}(V)\) on \(S(E)\) defined by \(\mathbb{P}(V)(s) = VsV\). \(\square\)

We now apply this result to the particular case where the Euclidean space \(E\) is the space \(S_k\)
of symmetric matrices and \(\mu\) is the noncentral Wishart distribution \(\gamma(p, a; \sigma)\). In that case,
\[
U_N := \hat{m}^2 - p\hat{v} = \bar{X}_N^2 - pS_N^2(I_k).
\]
Let us introduce the linear operator \(\mathbb{L}(m)\) of \(S_k\) into itself defined by
\[
\mathbb{L}(m)(s) = ms + sm.
\]
Proposition 6.3. The distribution of the random variable $\sqrt{N}(U_N - m^2 + pv) = Z_N$ in $S_k$ converges to the distribution of a centered normal variable

$$Z = Z_1m + mZ_1 - pZ_2(I_k) = \mathbb{L}(m)(Z_1) - pZ_2(I_k)$$

of $S_k$ where $(Z_1, Z_2)$ are as in Proposition 6.2. In particular the covariance of $Z$ is

$$\mathbb{L}(m)V(m)\mathbb{L}(m) + \frac{p^2}{2}[v \otimes v + (\text{tr} v)V(m)].$$

Proof. We first observe that

$$A_N := \sqrt{N}(X_N^2 - m^2) = \sqrt{N}(X_N - m)^2 + \mathbb{L}(m)(\sqrt{N}(X_N - m)).$$

From Proposition 6.2 this implies that the distribution of $A_N$ converges to the distribution of $\mathbb{L}(m)(Z_1)$. Similarly from Proposition 6.2 the distribution of $B_N := \sqrt{N}(S_N^2(I_k) - V(m)(I_k))$ converges to the distribution of $Z_2(I_k)$. Thus the distribution of $Z_N = A_N - pB_N$ converges to the distribution of $\mathbb{L}(m)(Z_1) - pZ_2(I_k)$. The only point left is the calculation of the variance. We use the following lemma without proof. □

Lemma 6.4. Let $F_1$ and $F_2$ be two Euclidean spaces and $f : F_2 \rightarrow F_1$ be a linear application. Let $f^* : F_1 \rightarrow F_2$ be the adjoint application. Let $Y$ be a random variable valued in $F_2$ with covariance operator $\Sigma$. Then the covariance operator of $X = f(Y)$ is $f^* \Sigma f^*$. □

We now perform the calculation of a particular adjoint in Lemma 6.5 below. If $F_1$ is a Euclidean space and if $x_0$ and $x_1$ are in $F_1$, recall that the operator $x_0 \otimes x_1$ on $F_1$ is the linear map from $F_1$ to itself defined by

$$x \mapsto (x_0 \otimes x_1)(x) = \langle x_1, x \rangle F_1 x_0.$$

Note that if $y : F_1 \rightarrow F_1$ is any linear map, then

$$\text{tr} [(x_0 \otimes x_1)y] = \langle x_1, y(x_0) \rangle_{F_1}. \quad (44)$$

Lemma 6.5. Suppose that $F_2 = S(F_1)$ is the linear space of symmetric operators on the Euclidean space $F_1$. We endow $F_2$ with the Euclidean structure $\langle a, b \rangle_{F_2} = \text{tr} (ab)$, we fix $x_0 \in F_1$ and we define $f : F_2 \rightarrow F_1$ by $y \mapsto f(y) = y(x_0)$. Then the adjoint of $f$ is

$$f^*(x) = \frac{1}{2}(x_0 \otimes x + x \otimes x_0).$$

Proof. We note $g(x) = \frac{1}{2}(x_0 \otimes x + x \otimes x_0)$. Note that $g(x)$ is a symmetric operator on $F_1$, thus an element of $F_2$. We now write for $x \in F_1$ and $y \in F_2$

$$\langle g(x), y \rangle_F_2 = \text{tr} \left[ \frac{1}{2}(x_0 \otimes x + x \otimes x_0) y \right]$$

$$= \frac{1}{2} \langle x, y(x_0) \rangle_{F_1} + \frac{1}{2} \langle x_0, y(x) \rangle_{F_1} \quad (45)$$

$$= \langle x, y(x_0) \rangle_{F_1} \quad (46)$$

$$= \langle x, f(y) \rangle_{F_1},$$

where equality (45) uses (44) and where equality (46) uses the fact that $y$ is a symmetric operator. This shows that $g = f^*$. □
End of the proof of Proposition 6.3. We apply Lemma 6.4 to \( F_1 = F_2 = S_k \), to \( f = \mathbb{L}(m) \) and to \( Y = Z_1 \) whose covariance \( \Sigma \) is \( V(m) \). We observe that \( \mathbb{L}(m) \) is a symmetric operator of the Euclidean space \( S_k \), that is \( \mathbb{L}(m) = \mathbb{L}(m)^* \). To see this we write for any \( x \) and \( x' \) in \( S_k \):

\[
\langle x, \mathbb{L}(m)(x') \rangle_{S_k} = \text{tr} [x(\mathbb{L}(m)(x'))] \\
= \text{tr} (xm'x') + \text{tr} (xx'm) \\
= \text{tr} (xm'x') + \text{tr} (mxx') \\
= \langle \mathbb{L}(m)(x), x \rangle_{S_k}.
\]

Lemma 6.4 implies that \( \mathbb{L}(m)(Z_1) = Z_1m + mZ_1 \) has variance \( \mathbb{L}(m)V(m)\mathbb{L}(m) \).

We now apply Lemma 6.5 to \( F_1 = S_k \) (and thus to \( F_2 = S(F_1) \)), to \( f(y) = y(I_k) \) and to \( Y = Z_2 \) whose covariance \( \Sigma \) is \( \mathbb{P}(V(m)) \) from Proposition 6.2. We see that the covariance of \( Z_2(I_k) \) is \( \mathbb{P}(V(m))f^* \). To compute \( f^* \) we apply Lemma 6.5 to \( x_0 = I_k \). Recall now that we denote \( v = V(m)(I_k) \). The operator \( f\mathbb{P}(V(m))f^* \) on \( S_k \) changes the symmetric matrix \( x \) of order \( k \) into the symmetric matrix of order \( k \) equal to

\[
f\mathbb{P}(V(m))f^*(x) = \frac{1}{2} f \left[ \mathbb{P}(V(m))[I_k \otimes x + x \otimes I_k] \right] \\
= \frac{1}{2} f \left[ V(m)[I_k \otimes x + x \otimes I_k] \right] V(m) \\
= \frac{1}{2} V(m)[I_k \otimes x + x \otimes I_k] V(m)(I_k) \\
= \frac{1}{2} V(m)[I_k \otimes x + x \otimes I_k](v) \\
= \frac{1}{2} V(m)(I_k \text{tr}(xv) + x \text{tr}(I_kv)) \\
= \frac{1}{2} v \text{tr}(xv) + (\text{tr} v)V(m)(x) \\
= \frac{1}{2} [v \otimes v + (\text{tr} v)V(m)](x).\]

Thus \( f\mathbb{P}(V(m))f^* = \frac{1}{4} [v \otimes v + (\text{tr} v)V(m)] \). Now since \( Z_1 \) and \( Z_2 \) are independent, the covariance operator of \( Z = Z_1m + mZ_1 - pZ_2(I_k) \) is simply obtained by adding the covariance operators of \( Z_1m + mZ_1 \) and of \( -pZ_2(I_k) \) which leads us to the desired result. Proposition 6.3 is proved. □

Remark. A shorter presentation of the variance of \( Z \) can be obtained by using the symbol \( \mathbb{P}(a,b) \) (where \( a \) and \( b \) are not necessarily symmetric \((k,k)\) matrices) for the endomorphism of \( S_k \) defined by

\[
h \mapsto \mathbb{P}(a,b)(h) = ahb + b^*ha^*.\]

This extends the notation introduced in Section 4.6 to nonsymmetric matrices. With this notation, the covariance operator of \( Z \) in Proposition 6.3 is written

\[
\mathbb{P}(m^3, \sigma) + \mathbb{P}(m^2, m\sigma) + \mathbb{P}(m\sigma m, m - \sigma) + \mathbb{P}(m\sigma m, m - \sigma)
+ \frac{p^2}{2}(\text{tr} v)\mathbb{P}(m, \sigma) - \frac{p^3}{2}(\text{tr} v)\mathbb{P}(\sigma) + \frac{p^2}{2}v \otimes v.
\]
Acknowledgment

The second author was supported by NSERC Discovery Grant A8946.

References