A tutorial on non central Wishart distributions.

G. Letac, H. Massam

October 18, 2004

Contents

This text is made for our friends and students who are already aware of the basic facts of the ordinary Wishart distributions. From our point of view, these basic facts can be gathered from two sources: statistics (and our reference is the book by Muirhead) and Jordan algebras (the book to read is Faraut Koranyi). When we have to choose between two normalizations, we choose Muirhead. Some of the statements of section 6 and 9 are possibly new. Comments at `letac@cict.fr` or `massamh@mathstat.yorku.ca` are welcome.

1 The non central χ^2 and the non central gamma

Proposition 1.1. Let Z be a $N(0,1)$ real variable and m a real number. Then for $s > -1$ we have

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

Proof. We just apply the fact that the Laplace transform of $N(0,\sigma^2)$ is $e^{\frac{\sigma^2\theta^2}{2}}$ by specializing to $\sigma^2 = \frac{1}{1+s}$ and $\theta = -ms$. We get

$$\begin{aligned}
\mathbb{E}(e^{-\frac{s}{2}(Z+m)^2}) &= \frac{e^{-\frac{sm^2}{2}}}{(1+s)^{1/2}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}(1+s)-msz}(1+s)^{1/2}\frac{dz}{(2\pi)^{1/2}} \\
&= \frac{e^{-\frac{sm^2}{2}}}{(1+s)^{1/2}} e^{\frac{s^2m^2}{2(1+s)}} \\
&= \frac{1}{(1+s)^{1/2}} e^{-\frac{sm^2}{2(1+s)}}.
\end{aligned}$$

For $p > 0$, consider the entire function f_p defined by

$$f_p(z) = \sum_{m=0}^{\infty} \frac{z^m}{m!\Gamma(m+p)}. \tag{1}$$

It is related to the classical Bessel function by the formula $(\frac{z}{2})^p f_p(-\frac{1}{4}z^2) = J_p(z)$ which implies (see Watson (1966)) that f_p has an infinity of simple zeros on $(-\infty, 0)$ and no other zeros in the complex plane. Some readers will prefer to use the generalized hypergeometric symbol and write $f_p(z) = \Gamma(p)\,{}_0F_1(-;p;z)$.

Proposition 1.2. Let $a > 0$, let $N, X_1, \ldots, X_n, \ldots$ be independent random variables such that N is Poisson distributed with mean a and such that X_j has density $e^{-x}\mathbf{1}_{(0,\infty)}(x)$ and define

$$Y = X_1 + \cdots + X_N$$

with the convention that $Y = 0$ if $N = 0$. Then $\mathbb{E}(e^{-sY}) = e^{-\frac{sa}{1+s}}$. Furthermore the law of Y is

$$\nu_a(dy) = e^{-a}\delta_0(dy) + ae^{-a-y}f_2(ay)\mathbf{1}_{(0,\infty)}(y)dy.$$

Proof. We just condition by N to compute $\mathbb{E}(e^{-sY})$, we use $\mathbb{E}(e^{-sX_j}) = (1+s)^{-1}$ and we get the result for $\mathbb{E}(e^{-sY})$. For computing the distribution of Y we use the fact that the distribution of $X_1+\cdots+X_n$ (for $n \neq 0$) has density $\frac{x^{n-1}}{(n-1)!}e^{-x}\mathbf{1}_{(0,\infty)}(x)$. The distribution of Y being a mixing of the distributions of $X_1+\cdots+X_n$ with respective weights $e^{-a}\frac{a^n}{n!}$ this leads easily to the result. $\square$

Definition 1.1. Let p and $a \geq 0$. The distribution $\gamma(p,a)$ is the distribution on $(0,\infty)$ defined by

$$\int_0^\infty e^{-st}\gamma(p,a)(dt) = \frac{1}{(1+s)^p}e^{-\frac{sa}{1+s}}.$$

For $\sigma > 0$, denote by $\gamma(p,a;\lambda)$ the image of $\gamma(p,a\lambda)$ by $t \mapsto \lambda t$. Thus

$$\int_0^\infty e^{-st}\gamma(p,a;\lambda)(dt) = \frac{1}{(1+\lambda s)^p}e^{-\frac{\lambda^2 sa}{1+\lambda s}}.$$

We say that $\gamma(p,a;\lambda)(dt)$ is a noncentral gamma distribution with shape, non centrality and scale parameters p,a,λ. Note $\gamma(p,0;\lambda)$ is the ordinary gamma distribution with scale parameter λ and shape parameter p. We say that $\gamma(p,a) = \gamma(p,a;1)$ is standard. For reason appearing in part 3 of the next proposition, $\gamma(n/2, 2a; 2))$ is called the standard non central χ^2 distribution with drift a and n degrees of freedom. We gather some properties of $\gamma(p,a;\lambda)$:

Proposition 1.3. We fix p and $a \geq 0$.

1. We have $\gamma(p,0) * \nu_a = \gamma(p,a)$ and thus $\gamma(p,a)$ exists. It satisfies

   $$\gamma(p,a) * \gamma(p',a') = \gamma(p+p', a+a')$$

   and $\gamma(p,a)$ is infinitely divisible.

2. For $a > 0$ we have

   $$\begin{aligned}\gamma(p,a)(dt) &= \frac{e^{-t-a}}{\Gamma(p)}[t^{p-1} + a\int_0^t (t-y)^{p-1}f_2(ay)dy]\mathbf{1}_{(0,\infty)}(t)dt \qquad (2)\\ &= e^{-t-a}t^{p-1}f_p(at)\mathbf{1}_{(0,\infty)}(t)dt. \qquad (3)\end{aligned}$$

3. If $Z_1,\ldots,Z_n$ are independent $N(0,1)$ random variables and if $m_1,\ldots,m_n$ are real numbers then the distribution of

   $$\frac{1}{2}((Z_1+m_1)^2+\cdots+(Z_n+m_n)^2)$$

   is $\gamma(n/2, \frac{1}{2}(m_1^2+\cdots+m_n^2))$.

4. The natural exponential family generated by $\gamma(p,a)$ is the family

   $$F(\gamma(p,a)) = \{\gamma(p,a,\lambda);\ \lambda > 0\}.$$

Its domain of the means is $(0, \infty)$ and its variance function is

$$V(m) = \frac{p^2 + 2ma}{a^2}(am + \frac{p^2}{4})^{1/2} - \frac{4p}{a}(am + \frac{p^2}{4}). \tag{4}$$

In terms of the parameters (p, a, λ) the mean and the variance of $\gamma(p, a, \lambda)$ are respectively

$$m = a\lambda^2 + p\lambda, \quad V(m) = 2a\lambda^3 + p\lambda^2. \tag{5}$$

Proof. Part 1 is obvious. Since $\gamma(p, a)$ is the convolution between ν_a and an absolutely continuous distribution, thus $\gamma(p, a)$ is also absolutely continuous and its density is

$$\int_0^t \frac{e^{-t+y}}{\Gamma(p)}(t-y)^{p-1}\nu_a(dy)$$

which leads to 2. Then we simply replace f_2 by the defining series and we get easily 3. Part 3 is an obvious consequence of Proposition 1.1. For getting part 4, write the Laplace transform for $\theta < 1$

$$L_{\gamma(p,a)}(\theta) = \int_0^\infty e^{\theta t}\gamma(p, a)(dt) = \frac{1}{(1-\theta)^p} e^{\frac{a\theta}{1-\theta}}$$

and recall that $F(\gamma(p, a))$ is the set of all probabilities μ such that there exists $\theta_0 < 1$ satisfying in a suitable interval $L_\mu(\theta) = \frac{L_{\gamma(p,a)}(\theta+\theta_0)}{L_{\gamma(p,a)}(\theta_0)}$. Thus

$$L_\mu(\theta) = (1 - \frac{\theta}{1-\theta_0})^{-p} \exp\left[\frac{a}{(1-\theta_0)^2}\frac{\theta}{1-\frac{\theta}{1-\theta_0}}\right].$$

Denoting $\lambda = \frac{1}{1-\theta_0}$ we see that $L_\mu = L_{\gamma(p,a,\lambda)}$ which is the desired result. The computation of the variance function is standard: denote $k = \log L_{\gamma(p,a)}$ and for simplification $X = 1/(1-\theta)$. Then

$$\begin{aligned} k'(\theta) &= aX^2 + pX = m \\ k''(\theta) &= 2aX^3 + pX^2 = V(m). \end{aligned}$$

Computing X from the first equality and carrying it in the second one gives the result. These last two formulas also give 5. One remarks that this family belongs to the Babel class of exponential families classified in the Rio lectures notes on exponential families (1991). $\square$

2 Noncentral Wishart and Gaussian laws

We denote by $\mathcal{P}_k$ the cone of positive definite symmetric matrices of order k and by $\overline{\mathcal{P}}_k$ its closure, that is the cone of semi-positive definite symmetric matrices.

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 line vectors. Then we have for a symmetric matrix s of order k such that $s + I_k$ is in $\mathcal{P}_k$

$$\begin{aligned}\mathbb{E}(e^{-\frac{1}{2}(Z+m)s(Z+m)^T}) &= \frac{1}{\det(I_k+s)^{1/2}}e^{-\frac{1}{2}m(I_k+s)^{-1}sm^T}\\ \mathbb{E}(e^{-\frac{1}{2}\operatorname{tr}(s(Z+m)^T(Z+m))}) &= \frac{1}{\det(I_k+s)^{1/2}}e^{-\frac{1}{2}\operatorname{tr}((I_k+s)^{-1}sm^Tm)}.\end{aligned} \tag{6}$$

Proof. We apply the fact that the Laplace transform of $N(0, \Sigma)$ is $\exp\frac{\theta\Sigma\theta^T}{2}$ by specializing to $\Sigma = (I_k + s)^{-1}$ and $\theta = -ms$. After observing that $s(I_k+s)^{-1}s - s = -(I_k + s)^{-1}s$ we get that

$$\begin{aligned}\mathbb{E}(e^{-\frac{1}{2}(Z+m)s(Z+m)^T}) &= \frac{e^{-\frac{msm^T}{2}}}{\det(I_k+s)^{1/2}}\int_{\mathbb{R}^k} e^{-\frac{1}{2}z(I_k+s)z^T - msz^T}(\det I_k+s)^{1/2}\frac{dz}{(2\pi)^{k/2}}\\ &= \frac{e^{-\frac{msm^T}{2}}}{\det(I_k+s)^{1/2}}e^{\frac{1}{2}ms(I_k+s)^{-1}sm^T}\\ &= \frac{1}{\det(I_k+s)^{1/2}}e^{-\frac{1}{2}m(I_k+s)^{-1}sm^T}. \quad \square\end{aligned}$$

More generally considering 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

$$\mathbb{E}(e^{-\frac{1}{2}\operatorname{tr}(s\sum_{j=1}^n(Z_j+m_j)^T(Z_j+m_j))}) = \frac{1}{\det(I_k+s)^{n/2}}e^{-\frac{1}{2}\operatorname{tr}((I_k+s)^{-1}s(m_1^Tm_1+\cdots+m_n^Tm_n))}. \tag{7}$$

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

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

A detailed answer will be given in Proposition 3.2. This natural question cannot be solved as simply as in the case $k = 1$ since no extension to $\mathbb{R}^k$ of the above Proposition 1.2 is available: more specifically one can prove that in the case $k > 1$ there is no positive measure μ_p on $\overline{\mathcal{P}}_k$ such that for all $s \in \mathcal{P}_k$ one can write

$$\int_{\mathcal{P}_k} e^{-\operatorname{tr}(sx)}\mu_p(dx) = (\operatorname{tr} s)^{-p}$$

and this easily shows that $\exp(-\operatorname{tr}((I_k + s)^{-1}sa))$ will not be associated to an infinitely divisible distribution as it was the case for $k = 1$. Equality 8 below shows that the correct generalization of the one dimensional case replaces trace by determinant in the above formula. As a substitute to Proposition 1.2 we introduce the zonal polynomials in the next section.

3 Density of the standard non central Wishart.

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 $\kappa = (m_1, \dots, m_k)$ such that $m_1 \geq m_2 \geq \dots \geq m_k \geq 0$. We denote $|\kappa| = m_1 + m_2 + \dots + m_k$ and E_m denotes the set of κ such that $m = |\kappa|$. We now introduce

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

We remark that $\Delta_\kappa(x) > 0$ for $x \in \mathcal{P}_k$. We also introduce some useful notation. The function $z \mapsto \Gamma_{\mathcal{P}_k}(z)$ is defined for $z = (z_1, \dots, z_k) \in \mathbb{R}^k$ such that furthermore $z_j > (j-1)/2$ for $j = 1, \dots, k$ by the following formula

$$\Gamma_{\mathcal{P}_k}(z) = (\pi)^{\frac{k(k-1)}{4}} \prod_{j=1}^{k} \Gamma(z_j - \frac{j-1}{2}).$$

If p is real a traditional abuse of notation writes $\Gamma_{\mathcal{P}_k}(z+p)$ for $\Gamma_{\mathcal{P}_k}(z_1+p, \dots, z_k+p)$. In particular for $p > (k-1)/2$ we have

$$\Gamma_{\mathcal{P}_k}(p) = (\pi)^{\frac{k(k-1)}{4}} \prod_{j=1}^{k} \Gamma(p - \frac{j-1}{2}).$$

This leads to the notation, for $\kappa = (m_1, \dots, m_k)$ with $m_1 \geq m_2 \geq \dots \geq m_k \geq 0$:

$$(p)_\kappa = \frac{\Gamma_{\mathcal{P}_k}(\kappa + p)}{\Gamma_{\mathcal{P}_k}(p)}$$

The normalization of $\Gamma_{\mathcal{P}_k}$ has been chosen to insure the validity of the formula

$$\Gamma_{\mathcal{P}_k}(p) = \int_{\mathcal{P}_k} e^{-\operatorname{tr}(sx)} (\det x)^{p-\frac{k+1}{2}} dx \tag{8}$$

where dx is the Lebesgue measure on the linear space of symmetric matrices $x = (x_{ij})_{1\leq i,j\leq k}$ defined by $dx = \prod_{1\leq i\leq j\leq k} dx_{ij}$. This is the choice made by the statisticians: see Muirhead page 62. However, Faraut and Koranyi make a different choice of Lebesgue measure: they equip the symmetric matrices with the Euclidean structure $\langle x, y\rangle = \operatorname{tr}(xy)$ which induces a different Lebesgue measure, giving mass 1 to the unit cube. For this reason the factor $(\pi)^{\frac{k(k-1)}{4}}$ in $\Gamma_{\mathcal{P}_k}$ is replaced in Faraut Koranyi page 123 by $(2\pi)^{\frac{k(k-1)}{2}}$ for still getting 8 with this other Lebesgue measure dx.

The zonal polynomial $C_\kappa(x)$ of parameter κ is defined by the following integral on the group $\mathbb{SO}(k)$ of orthogonal matrices of order k with determinant 1 with respect to the Haar measure du (normalized in order to have total mass one):

$$C_\kappa(x) = C_\kappa \int_{\mathbb{SO}(k)} \Delta_\kappa(u^{-1}xu)du,$$

where C_κ is a complicated normalizing constant which can be found on the last line of page 234 in Faraut Koranyi, or on formula (18) on page 237 of Muirhead. This is a homogeneous polynomial of degree $|\kappa|$ with respect to the entries x_{ij} of the symmetric matrix x. We insist on the fact that by definition, it takes positive values on $\mathcal{P}_k$. If $a \in \overline{\mathcal{P}}_k$ and if x is symmetric of order k then $a^{1/2}xa^{1/2}$ is also symmetric.

An other important remark is that for any v in the orthogonal group $\mathbb{O}(k)$ we have

$$C_\kappa(x) = C_\kappa(v^{-1}xv). \tag{9}$$

For $\det v = 1$ this is clear from the definition of Haar probability. For $\det v = -1$ enough is to see that $v^{-1}xv = x$ when $v = \mathrm{diag}(-1, 1, 1, \ldots, 1)$. A consequence of 9 is that actually, $C_\kappa(x)$ depends only on the eigenvalues of x. For $k = 2$ this enables us to compute in Section 4 the value of $C_\kappa(x)$ up to the cursed constant C_κ.

They satisfy many remarkable formulas. A selection is the following:

$$\begin{aligned}
e^{\operatorname{tr} x} &= \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{1}{m!} C_\kappa(x) && (10)\\
\det(I_k - x)^{-p} &= \sum_{m=0}^{\infty} \sum_{\kappa \in E_m} \frac{(p)_\kappa}{m!} C_\kappa(x) && (11)\\
(p)_\kappa (\det s)^{-p} C_\kappa(s^{-1}) &= \int_{\mathcal{P}_k} e^{-\operatorname{tr}(sx)} C_\kappa(x) (\det x)^{p-\frac{k+1}{2}} \frac{dx}{\Gamma_{\mathcal{P}_k}(p)} && (12)\\
.\frac{C_\kappa(t) C_\kappa(a)}{C_\kappa(I_k)} &= \int_{\mathbb{SO}(k)} C_\kappa(a^{1/2}u^{-1}tua^{1/2}) du. && (13)
\end{aligned}$$

We accept these formulas without proof: the unspecified normalizing constant C_κ above is chosen such that these formulas hold. Notice that a consequence of 10 is that

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

This comes from the fact C_κ is homogeneous of degree $|\kappa|$. A consequence is that for $\kappa \in E_m$ and for $x \in \overline{\mathcal{P}}_k$ we have

$$0 \leq C_\kappa(x) \leq (\operatorname{tr} x)^m. \tag{14}$$

Proposition 3.1. Let $p > (k-1)/2$ and $a \in \overline{\mathcal{P}}_k$. Then

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

is a probability on $\mathcal{P}_k$ such that for $I_k + s \in \mathcal{P}_k$ one has

$$\int_{\mathcal{P}_k} e^{-\operatorname{tr}(st)} \gamma(p, a)(dt) = \frac{1}{\det(I_k + s)^p} e^{-\operatorname{tr}((I_k+s)^{-1}sa)}. \tag{15}$$

Proof. The fact that $\gamma(p,a)(dt)$ is a positive measure comes from the fact that for $t\in\mathcal{P}_k$ we have $\Delta_\kappa(t)>0$ and therefore, from the definition of $C_\kappa(t)$ we have $C_\kappa(a^{1/2}ta^{1/2})>0$. Suppose first $a\in\mathcal{P}_k$. For $\kappa\in E_m$ we do the change of variable $x=a^{1/2}ta^{1/2}$ in the integral

$$I_\kappa(a)=e^{-\operatorname{tr}a}\int_{\mathcal{P}_k}e^{-\operatorname{tr}((s+I_k)t)}(\det t)^{p-\frac{k+1}{2}}\frac{C_\kappa(a^{1/2}ta^{1/2})}{m!(p)_\kappa}\frac{dt}{\Gamma_{\mathcal{P}_k}(p)}.$$

A classical result about the jacobian implies $dx=(\det a)^{(k+1)/2}dt$. Using formula 12 we get

$$I_\kappa(a)=e^{-\operatorname{tr}a}\det(I_k+s)^{-p}\frac{C_\kappa(a^{1/2}(I_k+s)^{-1}a^{1/2})}{m!}. \qquad (16)$$

Suppose now that a is singular in $\overline{\mathcal{P}}_k$. Remark that for $n\geq 1$ we have $a_n=a+\frac{1}{n}I_k\in\mathcal{P}_k$. The inequality 14 implies that

$$0\leq C_\kappa(a_n^{1/2}xa_n^{1/2})\leq(\operatorname{tr}(a_nx))^m\leq(\operatorname{tr}(a_1x))^m.$$

Therefore we can apply dominated convergence and write $\lim_{n\to\infty}I_\kappa(a_n)=I_\kappa(a)$ and this implies that 16 holds even for a singular a. Summing up all equalities 16 and using 10 we get easily 15. The fact that the mass of $\gamma(p,a)$ is one is obtained by doing $s=0$ in 15. $\square$

Let p be in the so called Gyndikin set Λ of order k defined by

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

and let a in $\overline{\mathcal{P}}_k$. We define the standard non central Wishart distribution $\gamma(p,a)$ on $\overline{\mathcal{P}}_k$ as the unique probability such that 15 holds. For $p>(k-1)/2$ its existence is given by Proposition 3.1. If p is the half integer $n/2$, the existence of $\gamma(p,a)(dt)$ as well a Gaussian interpretation comes from 7. Actually, these values of $p\in\Lambda$ are the only ones such that $\gamma(p,a))$ does exist. More specifically

Proposition 3.2. For $p>0$ and a in $\overline{\mathcal{P}}_k$, there exists a probability $\gamma(p,a)$ such that 15 holds if and only if p is in Λ.

Since the proof requires some notations which will be introduced in the next sections, we postpone it to Section 7. Section 4 considers the practical case $k=2$, Section 5 concentrates on the natural exponential family generated by $\gamma(p,a)$.

4 Zonal polynomials for dimension two and Legendre polynomials .

This section computes the zonal polynomials and the density of the noncentral Wishart for $k=2$. We shall express the zonal polynomials in terms of the familiar Legendre polynomials $(P_n)_{n\geq 0}$ as defined by their generating formula

$$\frac{1}{(1-2xt+t^2)^{1/2}}=\sum_{n=0}^{\infty}t^nP_n(x).$$

Our favorite textbook on the subject is Rainville *Special Functions* (1960).

Proposition 4.1. Let m be an non negative integer and let $\kappa = (m_1, m_2)$ with $m_1 + m_2 = m$ and $m_1 \geq m_2 \geq 0$. Then for a symmetric non singular matrix x of order 2 one has

$$\int_{\mathbb{SO}(2)} \Delta_\kappa(u^{-1}xu)du = (\det x)^{m/2} P_{m_1-m_2}(\frac{\operatorname{tr} x}{2(\det x)^{1/2}}).$$

For a singular non zero matrix x one has $\int_{\mathbb{SO}(2)} \Delta_\kappa(u^{-1}xu)du = 0$ if $m_2 > 0$ and

$$\int_{\mathbb{SO}(2)} \Delta_{(m,0)}(u^{-1}xu)du = \frac{(2m)!}{2^{2m}m!^2}(\operatorname{tr} x)^m.$$

Proof. A typical element of $\mathbb{SO}(2)$ is

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Suppose that $x = \operatorname{diag}(\lambda_1, \lambda_2)$. Then $\Delta_\kappa(x) = \lambda_1^{m_1-m_2}(\lambda_1\lambda_2)^{m_2}$ and

$$\begin{aligned} \int_{\mathbb{SO}(2)} \Delta_\kappa(u^{-1}xu)du &= \frac{1}{2\pi}\int_0^{2\pi} \Delta_\kappa(R(-\theta)xR(-\theta))d\theta \\ &= (\lambda_1\lambda_2)^{m_2}\frac{1}{2\pi}\int_0^{2\pi}(\lambda_1\cos^2\theta + \lambda_2\sin^2\theta)^{m_1-m_2}d\theta \end{aligned}$$

Now we compute the generating function

$$\begin{aligned} \sum_{n=0}^\infty t^n \frac{1}{2\pi}\int_0^{2\pi}(\lambda_1\cos^2\theta + \lambda_2\sin^2\theta)^n d\theta &= \frac{1}{2\pi}\int_0^{2\pi}\frac{d\theta}{1 - t(\lambda_1\cos^2\theta + \lambda_2\sin^2\theta)} \\ &= \frac{2}{\pi}\int_0^{\pi/2}\frac{d\theta}{1 - t(\lambda_1\cos^2\theta + \lambda_2\sin^2\theta)} \\ &= \frac{1}{(1-t\lambda_1)^{1/2}(1-t\lambda_2)^{1/2}} \end{aligned}$$

(use the change of variable $u = \tan\theta$). In the non singular case we have

$$\frac{1}{(1-t\lambda_1)^{1/2}(1-t\lambda_2)^{1/2}} = \sum_{n=0}^\infty t^n(\lambda_1\lambda_2)^{n/2}P_n(\frac{\lambda_1+\lambda_2}{2(\lambda_1\lambda_2)^{1/2}})$$

Thus the result is proved when x is a non singular diagonal matrix. Now we have seen that the zonal polynomial is a symmetric function of the eigenvalues. Thus the result is proved in the non singular case. The proof in the singular case is similar. □

The coefficient C_κ. In order to have a complete knowledge of the zonal polynomial $C_\kappa(x)$ for $k=2$ we explicit it from the value given in Muirhead page 237. For $\kappa=(m_1,m_2)$ with $m_1+m_2=m$ and $m_1\geq m_2\geq 0$ we have to distinguish the case $m_2=0$ and the case $m_2>0$. We get

$$\begin{aligned} C_{(m,0)} &= \frac{2^{2m}m!^2}{(2m)!} \\ C_{(m_1,m_2)} &= 2^{2m}m!m_1!\times\frac{1}{2}\times\frac{3}{2}\times\cdots\times\frac{2m_2-1}{2}\times\frac{2(m_1-m_2)+1}{(2m_1+1)!(2m_2)!} \end{aligned}$$

The density of the standard non central Wishart for dimension 2. In this case the non central distribution $\gamma(p,a)(dt)$ for $p>1/2$ and for a non singular is given by

$$\begin{aligned} &\gamma(p,a)(dt) = e^{-\operatorname{tr}(t+a)}(\det t)^{p-\frac{3}{2}} \\ &\times\left(\sum_{m=0}^{\infty}\frac{(\det a\det t)^{m/2}}{m!}\sum_{\kappa\in E_m}C_{m_1,m_2}\frac{P_{m_1-m_2}\left(\frac{\operatorname{tr}(at)}{2(\det a\det t)^{1/2}}\right)}{\Gamma(m_1+p)\Gamma(m_2+p+\frac{1}{2})}\right)\mathbf{1}_{\mathcal{P}_2}(t)dt \end{aligned}$$

If a is singular but not zero then $\gamma(p,a)(dt)$ for $p>1/2$ is

$$\gamma(p,a)(dt) = e^{-\operatorname{tr}(t+a)}(\det t)^{p-\frac{3}{2}}\left(\sum_{m=0}^{\infty}\frac{(\operatorname{tr}at)^m}{m!\Gamma(m+p)}\right)\mathbf{1}_{\mathcal{P}_2}(t)\frac{dt}{\Gamma(p+\frac{1}{2})}.$$

When a in non zero singular, then a has rank 1 and can be written $a=\lambda m\otimes m$ where $m\in\mathbb{R}^2\setminus\{0\}$ and λ is any non zero real number. With this presentation, $\operatorname{tr}(at)=\lambda\operatorname{tr}(m\otimes mt)=\lambda mtm^T$. The case where $\lambda=N\geq 2$ is an integer is specially useful:

Proposition 4.2. Let $Z_1,\ldots,Z_N$ with $N\geq 2$ be standard normal centered random variables of $\mathbb{R}^2$ and let $m\in\mathbb{R}^2\setminus\{0\}$. Then the distribution of the following random symmetric matrix of order 2

$$T=\frac{1}{2}[(Z_1+m)(Z_1+m)^T+\cdots+(Z_N+m)(Z_N+m)^T]$$

is

$$\gamma(N/2,Nm\otimes m)(dt) = e^{-\operatorname{tr}(t)-N\|m\|^2}(\det t)^{\frac{N-3}{2}}\left(\sum_{j=0}^{\infty}\frac{N^j(mtm^T)^j}{j!\Gamma(j+\frac{N}{2})}\right)\mathbf{1}_{\mathcal{P}_2}(t)\frac{dt}{\Gamma(\frac{N+1}{2})}.$$

5 The general non central Wishart

Proposition 5.1. Let b be a non singular 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 $a \in \overline{\mathcal{P}}_k$. We also write $\sigma = q^2$.

1. The image of $\gamma(p, a)$ by $t \mapsto utu^*$ is $\gamma(p, uau^*)$.
2. The image of $\gamma(p, qaq)$ by $t \mapsto qtq$ is a distribution denoted $\gamma(p, a; q^2) = \gamma(p, a; \sigma)$ such that
$$\int_{\overline{\mathcal{P}}_k} e^{-\operatorname{tr}(st)}\gamma(p, a; q^2)(dt) = \frac{1}{\det(1 + qsq)^p} e^{-\operatorname{tr}((I_k+qsq)^{-1}qsqqaq)} \tag{17}$$
holds.
3. The image of $\gamma(p, a)$ by $t \mapsto btb^*$ is $\gamma(p, q^{-1}uau^*q^{-1}; q^2)$.
4. The natural exponential family generated by $\gamma(p, a)$ is
$$F(\gamma(p, a)) = \{\gamma(p, a; q^2);\ q \in \mathcal{P}_k\} = \{\gamma(p, a; \sigma);\ \sigma \in \mathcal{P}_k\}.$$
5. If the distribution of X is $\gamma(p, a; q^2) = \gamma(p, a, \sigma)$ denote $\omega = \sigma a \sigma$. If c is a positive constant, then the distribution of cX is $\gamma(p, \frac{a}{c^2}, c\sigma)$ (note that $\omega = c\sigma\frac{a}{c^2}c\sigma$).

   If z is a line vector of $\mathbb{R}^k$ denote $\lambda_z = z\sigma z^T$. Then the distribution of the real random variable zXz^T is a non central gamma distribution of Laplace transform
$$\mathbb{E}(e^{-s_1 zXz^T}) = \frac{1}{(1 + s_1\lambda)^p} e^{-\frac{s_1}{1+s_1\lambda} z\omega z^T},$$
   thus with shape parameter p, with scale parameter λ_z and with non centrality parameter
$$a_z = \lambda^{-2} z\omega z^T = \frac{z\omega z^T}{(z\sigma z^T)^2}.$$
   In particular the mean m_z and the variance v_z of zXz^T are respectively
$$m_z = pz\sigma z^T + z\omega z^T, \quad v_z = z\sigma z^T(p\, z\sigma z^T + 2z\omega z^T). \tag{18}$$

Proof. 1. We write

$$\begin{aligned}\int_{\mathcal{P}_k} e^{-\operatorname{tr}(sutu^*)}\gamma(p,a)(dt) &= \int_{\mathcal{P}_k} e^{-\operatorname{tr}(u^*sut)}\gamma(p,a)(dt)\\ &= \frac{1}{\det(I_k+s)^p}e^{-\operatorname{tr}((I_k+u^*su)^{-1}u^*sua)}\\ &= \frac{1}{\det(I_k+s)^p}e^{-\operatorname{tr}(u(u^*u+u^*su)^{-1}u^*suau^*)}\\ &= \frac{1}{\det(I_k+s)^p}e^{-\operatorname{tr}((I_k+s)^{-1}suau^*)}.\end{aligned}$$

2. Standard.

3. Since the transformation $\varphi_b : t \mapsto btb^*$ satisfies $\varphi_b = \varphi_q \circ \varphi_u$ we observe that from part one the image of $\gamma(p,a)$ is $\gamma(p,uau^*)$. We rewrite it as $\gamma(p,uau^*) = \gamma(p,qq^{-1}uau^*q^{-1}q)$. By definition of $\gamma(p,a,q^2)$ its image by φ_q is $\gamma(p,q^{-1}uau^*q^{-1};q^2)$ as announced.

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

$$\int_{\mathcal{P}_k} e^{-\operatorname{tr}(st)}\mu(dt) = \left(\frac{\det(I_k+s_0+s)}{\det(I_k+s_0)}\right)^p e^{-\operatorname{tr}((I_k+s_0+s)^{-1}(s_0+s)a)}e^{\operatorname{tr}((I_k+s_0)^{-1}s_0a)}.$$

We rewrite the second member by introducing $q=(I_k+s_0)^{-1/2}$ or $s_0 = q^{-2}-I_k$ and we get the second member of 17. This proves the result.

5. If s_1 is a positive number, apply formula 17 to the symmetric matrix $s = s_1z^Xz$. In order to prove the formula, we choose an orthonormal basis $e=(e_1,\ldots,e_k)$ of $\mathbb{R}^k$ such that $zq=\lambda^{1/2}e_1$. With such a choice the representative matrices M and $(I_k+M)^{-1}M$ of the endomorphisms s_1qz^Tzq and $(I_k+qsq)^{-1}qsq$ in the above basis is simply, by blocks

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

Thus $\det(I_k+qsq) = 1+s_1\lambda$ and $\operatorname{tr}((I_k+qsq)^{-1}qsqqaq) = \frac{s_1\lambda}{1+s_1\lambda}\operatorname{tr}e_1^Te_1qaq$ Since

$$\operatorname{tr}e_1^Te_1qaq = \frac{1}{\lambda}\operatorname{tr}qz^Tzqqaq = \frac{1}{\lambda}z\sigma a\sigma z^T$$

we get the result. The mean and the variance are obtained from 5. $\square$

The distribution $\gamma(p,a,q^2)$ is called the general non central Wishart distribution. For $p>(k-1)/2$ its density is obtained by taking the density of $\gamma(p,qaq)$ as obtained from Proposition 2.2 and by taking its image by $t\mapsto x=qtq$. Thus the distribution $\gamma(p,a,q^2)(dx)$ for $p>(k-1)/2$ is

$$(\det q)^{-2p}e^{-\operatorname{tr}(q^{-1}xq^{-1}+qaq)}(\det x)^{p-\frac{k+1}{2}}$$
$$\left(\sum_{m=0}^{\infty}\sum_{\kappa\in E_m}\frac{C_\kappa(q^{1/2}a^{1/2}q^{-1/2}xq^{-1/2}a^{1/2}q^{1/2})}{m!(p)_\kappa}\right)\mathbf{1}_{\mathcal{P}_k}(x)\frac{dx}{\Gamma_{\mathcal{P}_k}(p)}.\quad\square$$

6 The moments and the variance function

6.1 Leibnitz formula.

First we observe that the exponential family $F(\gamma(p,a))$ can be generated by the unbounded positive measure $\mu(dt) = e^{\,\mathrm{tr}\,(a+t)}\gamma(p,a)(dt)$ whose Laplace transform is defined on $-\mathcal{P}_k$ by

$$L_\mu(\theta) = \int_{\overline{\mathcal{P}}_k} e^{\,\mathrm{tr}\,(\theta x)}\mu(dx) = \frac{1}{(-\theta)^p} e^{\,\mathrm{tr}\,(a(-\theta)^{-1})}. \tag{19}$$

We define the following two functions on $-\mathcal{P}_k$:

$$\sigma = \sigma(\theta) = (-\theta)^{-1}, \quad k_\mu(\theta) = \,\mathrm{tr}\,(a\sigma) + p\log\det\sigma. \tag{20}$$

Note that $k_\mu = \log L_\mu$ is the cumulant function of $F(\gamma(p,a)) = F(\mu)$ when μ 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 17 and

$$\int_{\overline{\mathcal{P}}_k} e^{-\,\mathrm{tr}\,st}P(\theta,\mu)(dt) = \frac{L_\mu(\theta - s)}{L_\mu(\theta)}$$

Let us now recall some general facts about the moments of a multivariate exponential family generated by a measure μ 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_μ and if $P(\theta,\mu)(dx) = \frac{e^{\langle\theta,x\rangle}}{L_\mu(\theta)}\mu(dx) \in F(\mu)$ corresponds to the parameter θ then the n th differential of L_μ in the directions $h_1,\ldots,h_n$ has the following probabilistic interpretation

$$L_\mu^{(n)}(\theta)(h_1,\cdots,h_n) = L_\mu(\theta)\int_E \langle h_1,x\rangle\ldots\langle h_n,x\rangle P(\theta,\mu)(dx) \tag{21}$$

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 $\langle\theta,x\rangle = \,\mathrm{tr}\,(\theta x)$.

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 linear space F: there exists a kind of Leibnitz formula. For $F=\mathbb{R}$ it reads

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

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 Leibnitz formula is

$$(fg)^{(n)}(\theta)(h_1,\ldots,h_n) = \sum_{T\subset\{1,\ldots,n\}} f^{(|T|)}(\theta)(h_T)g^{(|T'|)}(\theta)(h_{T'}). \tag{22}$$

6.2 Two differentials of order n.

We are going to apply the above considerations to $F = E^*$ to $f(\theta) = e^{\mathrm{tr}\, a\sigma(\theta)} = e^{\langle a,\sigma\rangle}$ and to $g(\theta) = e^{p \log\det\sigma(\theta)}$. Thus $L_\mu = fg$. The next step is the computation of the respective differentials $f^{(n)}(\theta)(h_1,\ldots,h_n)$ and $g^{(n)}(\theta)(h_1,\ldots,h_n)$. For this we need the following two differentials

$$\begin{aligned} \sigma'(\theta)(h) &= \sigma h\sigma \\ (\log\det\sigma(\theta))'(h) &= \mathrm{tr}\,(\sigma h). \end{aligned}$$

The differential $g^{(n)}$ is known, if not well known: the present authors have used it in three or four papers and we shall only recall the result. For this we denote by $\mathcal{S}_n$ the group of permutations π of $\{1,\ldots,n\}$, we denote by $C(\pi)$ the set of cycles of the permutation π, by $m(\pi)$ the number of cycles and we denote

$$r_\pi(\sigma)(h_1,\ldots,h_n) = \prod_{c\in C(\pi)} \mathrm{tr}\,(\prod_{j\in c}\sigma h_j).$$

Then the differential is

$$g^{(n)}(\theta)(h_1,\ldots,h_n) = g(\theta)\sum_{\pi\in\mathcal{S}_n} p^{m(\pi)} r_\pi(\sigma)(s_1,\ldots,s_k). \tag{23}$$

The differential of $f(\theta) = e^{\mathrm{tr}\, a\sigma}$ is newer. For each π in $\mathcal{S}_n$, we introduce a quantity close to $r_\pi(\sigma)(h_1,\ldots,h_n)$ by simply replacing formally $A = \mathrm{tr}\,(\prod_{j\in c}\sigma h_j)$ by $B = \mathrm{tr}\,(\sigma a\prod_{j\in c}\sigma h_j)$. However this hardly makes sense for the following reason. Suppose that the cycle c is $(2,5,4)$ that means the permutation changing 2 in 5, 5 in 4 and 4 in 2. Then $A = \mathrm{tr}\,(\sigma h_2\sigma h_5\sigma h_4)$. The same cycle could also have been written $(5,4,2)$ 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 $\mathrm{tr}\,(\sigma a\sigma h_2\sigma h_5\sigma h_4)$ and $\mathrm{tr}\,(\sigma a\sigma h_5\sigma h_4\sigma h_2)$ do not coincide.

For this reason, for a given integer n we introduce the following set P_n of objects. The data of an element P of P_n is the data of two things

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

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

$$\begin{aligned} T^{(1)} &= (\{1\},\{2\},\{3\}) \\ T^{(2)} &= (\{1,2\},\{3\}) \\ T^{(3)} &= (\{1\},\{2,3\}) \\ T^{(4)} &= (\{2\},\{1,3\}) \\ T^{(5)} &= (\{1,2,3\}) \end{aligned}$$

Thus $T^{(1)}, \dots, T^{(5)}$ generate respectively 1,2,2,2,6 elements of P_3 and P_3 has 13 elements. As an other 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.

Here are now the functions s_P indexed by $P \in \mathrm{P}_n$ which imitate r_π. If P is given by q and the (T_j, π_j) for $j = 1, \dots, q$ we define

$$s_P(\sigma)(h_1, \dots, h_n) = \prod_{j=1}^{q} \mathrm{tr}\left(\sigma a \prod_{i \in T_j} \sigma h_{\pi_j(i)}\right).$$

Proposition 6.1.

$$(e^{\mathrm{tr}\,(a\sigma)})^{(n)}(\theta)(h_1, \dots, h_n) = e^{\mathrm{tr}\,(a\sigma)} \sum_{P \in \mathrm{P}_n} s_P(\sigma)(h_1, \dots, h_n).$$

Proof. Induction on n. $\square$

Example 1. We compute the 3 first differentials of $\theta \mapsto f(\theta) = e^{\mathrm{tr}\,(a\sigma)}$. For simplicity, we write $a' = \sigma a$ and $h'_j = \sigma h_j$. Thus from the previous proposition we get

$$\begin{aligned}
\frac{1}{f(\theta)} f'(\theta)(h_1) &= \mathrm{tr}\,(a'h'_1) \\
\frac{1}{f(\theta)} f''(\theta)(h_1, h_2) &= \mathrm{tr}\,(a'h'_1)\,\mathrm{tr}\,(a'h'_2) + \mathrm{tr}\,(a'h'_1h'_2) + \mathrm{tr}\,(a'h'_2h'_1) \\
\frac{1}{f(\theta)} f'''(\theta)(h_1, h_2) &= \mathrm{tr}\,(a'h'_1)\,\mathrm{tr}\,(a'h'_2)\,\mathrm{tr}\,(a'h'_3) + \\
&\quad \mathrm{tr}\,(a'h'_1h'_2)\,\mathrm{tr}\,(a'h'_3) + \mathrm{tr}\,(a'h'_2h'_1)\,\mathrm{tr}\,(a'h'_3) + \\
&\quad \mathrm{tr}\,(a'h'_1h'_3)\,\mathrm{tr}\,(a'h'_2) + \mathrm{tr}\,(a'h'_3h'_1)\,\mathrm{tr}\,(a'h'_2) + \\
&\quad \mathrm{tr}\,(a'h'_3h'_2)\,\mathrm{tr}\,(a'h'_1) + \mathrm{tr}\,(a'h'_2h'_3)\,\mathrm{tr}\,(a'h'_1) + \\
&\quad \mathrm{tr}\,(a'h'_1h'_2h'_3) + \mathrm{tr}\,(a'h'_2h'_1h'_3) + \mathrm{tr}\,(a'h'_3h'_2h'_1) + \\
&\quad \mathrm{tr}\,(a'h'_1h'_3h'_2) + \mathrm{tr}\,(a'h'_2h'_3h'_1) + \mathrm{tr}\,(a'h'_3h'_1h'_2).
\end{aligned}$$

Example 2. Symmetrically we compute the three first differential of $g(\theta) = e^{p \log \det \sigma(\theta)} = \frac{1}{\det(-\theta)^p}$. We still adopt the notation $h'_j = \sigma h_j$. Thus from 23 and according to 21 we get

$$\begin{aligned}
\frac{1}{g(\theta)} g'(\theta)(h_1) &= p\,\mathrm{tr}\,(h'_1) \\
\frac{1}{g(\theta)} g''(\theta)(h_1, h_2) &= p^2\,\mathrm{tr}\,(h'_1)\,\mathrm{tr}\,(h'_2) + p\,\mathrm{tr}\,(h'_1h'_2) \\
\frac{1}{g(\theta)} g'''(\theta)(h_1, h_2) &= p^3\,\mathrm{tr}\,(h'_1)\,\mathrm{tr}\,(h'_2)\,\mathrm{tr}\,(h'_3) + p\,\mathrm{tr}\,(h'_1h'_2h'_3) + p\,\mathrm{tr}\,(h'_2h'_1h'_3) \\
&\quad + p^2\,\mathrm{tr}\,(h'_1h'_2)\,\mathrm{tr}\,(h'_3) + p^2\,\mathrm{tr}\,(h'_1h'_3)\,\mathrm{tr}\,(h'_2) + p^2\,\mathrm{tr}\,(h'_3h'_2)\,\mathrm{tr}\,(h'_1).
\end{aligned}$$

6.3 Moments of order 1,2,3.

Now we combine these two examples by the Leibnitz formula for obtaining the first two moments of the non central Wishart random variable X with distribution $P(\theta,\mu)(dx) = \frac{e^{\mathrm{tr}\,(\theta x)}}{L_\mu(\theta)}\mu(dx)$ where μ is the measure defined by 19, associated to the two parameters $p \in \Lambda_k$ (see Proposition 3.2) and $a \in \overline{\mathcal{P}}_k$. Recall that $-\theta \in \mathcal{P}_k$ is positive definite and that we denote $\sigma = (-\theta)^{-1}$ and $h'_j = \sigma h_j$ and $a' = \sigma a$. Recall that k_μ is the cumulant function defined in 20. With these notations we get the first two moments

$$\begin{aligned}
\mathbb{E}(\,\mathrm{tr}\,(Xh_1)) &= k'_\mu(\theta)(h_1) = \frac{L'_\mu(\theta)(h_1)}{L_\mu(\theta)} = p\,\mathrm{tr}\,(h'_1) + \,\mathrm{tr}\,(a'h'_1) &(24)\\
\mathbb{E}(\,\mathrm{tr}\,(Xh_1)\,\mathrm{tr}\,(Xh_2)) &= \frac{L''_\mu(\theta)(h_1,h_2)}{L_\mu(\theta)} = \,\mathrm{tr}\,(a'h'_1)\,\mathrm{tr}\,(a'h'_2) + \,\mathrm{tr}\,(a'h'_1h'_2) + \,\mathrm{tr}\,(a'h'_2h'_1) &(25)\\
&+ \; p\,\mathrm{tr}\,(h'_1)\,\mathrm{tr}\,(a'h'_2) + p\,\mathrm{tr}\,(h'_2)\,\mathrm{tr}\,(a'h'_1) + p^2\,\mathrm{tr}\,(h'_1)\,\mathrm{tr}\,(h'_2) + p\,\mathrm{tr}\,(h'_1h'_2)
\end{aligned}$$

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

$$\mathbb{E}[\,\mathrm{tr}\,((X-\mathbb{E}(X))h_1)\,\mathrm{tr}\,((X-\mathbb{E}(X))h_2)] = \,\mathrm{tr}\,(a'h'_1h'_2) + \,\mathrm{tr}\,(a'h'_2h'_1) + p\,\mathrm{tr}\,(h'_1h'_2). \quad (26)$$

We now reformulate the results 24 and 26 about the mean and the covariance. The linear space S_k of real symmetric matrices of dimension k is equipped with the Euclidean structure $\langle h_1,h_2\rangle = \,\mathrm{tr}\,(h_1h_2)$.

Proposition 6.2. Let X be a noncentral Wishart random variable with parameters p,a,θ as above, with the notation $\sigma = (-\theta)^{-1}$. Then

$$\mathbb{E}(X) = m = k'_\mu(\theta) = p\sigma + \sigma\, a\, \sigma. \quad (27)$$

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

$$h \mapsto \sigma a\sigma h\sigma + \sigma h\sigma a\sigma + p\sigma h\sigma = mh\sigma + \sigma hm - p\sigma h\sigma. \quad (28)$$

Proof. Consider the symmetric matrix $v = \mathbb{E}(X) - p\sigma - \sigma\, a\,\sigma$. From 24 for each symmetric matrix h we have $\,\mathrm{tr}\,(vh) = 0$ (recall that $\,\mathrm{tr}\,(\mathbb{E}(X)h) = \mathbb{E}(\,\mathrm{tr}\,(Xh))$). Now specialize to $h = v$. Thus $\,\mathrm{tr}\,v^2 = 0$. Since v^2 is a semi positive definite matrix, this implies that $v = 0$ and 27 is proved.

For proving the second one denotes for simplicity $\mathbb{E}((X-\mathbb{E}(X))\otimes(X-\mathbb{E}(X)))$ by c. Formula 26 says that with the above scalar product we have

$$\begin{aligned}
\langle c(h_1),h_2\rangle &= \,\mathrm{tr}\,(\sigma a\sigma h_1\sigma h_2) + \,\mathrm{tr}\,(\sigma a\sigma h_2\sigma h_1) + p\,\mathrm{tr}\,(\sigma h_1\sigma h_2)\\
&= \,\mathrm{tr}\,(\sigma a\sigma h_1\sigma h_2) + \,\mathrm{tr}\,(\sigma h_1\sigma a\sigma h_2) + p\,\mathrm{tr}\,(\sigma h_1\sigma h_2)\\
&= \langle \sigma a\sigma h_1\sigma + \sigma h_1\sigma a\sigma + p\,\sigma h_1\sigma, h_2\rangle.
\end{aligned}$$

Since this is true for all h_2 we get that $c(h_1) = \sigma a\sigma h_1\sigma + \sigma h_1\sigma a\sigma + p\,\sigma h_1\sigma$ as claimed. Replacing $\sigma\, a\, \sigma$ by $m - p\sigma$ gives the second expression of the covariance of X.

The third moment $\frac{1}{L_\mu(\theta)} L_\mu'''(\theta)(h_1, h_2, h_3)$ is a monster of 27 monomials that we write now (the fourth moment is the sum of 267 monomials):

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{3} \mathrm{tr}\,(Xh_i)\right] \quad = \quad & \mathrm{tr}\,(a'h_1')\,\mathrm{tr}\,(a'h_2')\,\mathrm{tr}\,(a'h_3') + \\
& \mathrm{tr}\,(a'h_1'h_2')\,\mathrm{tr}\,(a'h_3') + \mathrm{tr}\,(a'h_2'h_1')\,\mathrm{tr}\,(a'h_3') + \\
& \mathrm{tr}\,(a'h_1'h_3')\,\mathrm{tr}\,(a'h_2') + \mathrm{tr}\,(a'h_3'h_1')\,\mathrm{tr}\,(a'h_2') + \\
& \mathrm{tr}\,(a'h_3'h_2')\,\mathrm{tr}\,(a'h_1') + \mathrm{tr}\,(a'h_2'h_3')\,\mathrm{tr}\,(a'h_1') + \\
& \mathrm{tr}\,(a'h_1'h_2'h_3') + \mathrm{tr}\,(a'h_2'h_1'h_3') + \mathrm{tr}\,(a'h_3'h_2'h_1') + \\
& \mathrm{tr}\,(a'h_1'h_3'h_2') + \mathrm{tr}\,(a'h_2'h_3'h_1') + \mathrm{tr}\,(a'h_3'h_1'h_2') + \\
& (\,\mathrm{tr}\,(a'h_1')\,\mathrm{tr}\,(a'h_2') + \mathrm{tr}\,(a'h_1'h_2') + \mathrm{tr}\,(a'h_2'h_1'))p\,\mathrm{tr}\,h_3' + \\
& (\,\mathrm{tr}\,(a'h_3')\,\mathrm{tr}\,(a'h_2') + \mathrm{tr}\,(a'h_3'h_2') + \mathrm{tr}\,(a'h_2'h_3'))p\,\mathrm{tr}\,h_1' + \\
& (\,\mathrm{tr}\,(a'h_1')\,\mathrm{tr}\,(a'h_3') + \mathrm{tr}\,(a'h_1'h_3') + \mathrm{tr}\,(a'h_3'h_1'))p\,\mathrm{tr}\,h_2' + \\
& p^3\,\mathrm{tr}\,(h_1')\,\mathrm{tr}\,(h_2')\,\mathrm{tr}\,(h_3') + p\,\mathrm{tr}\,(h_1'h_2'h_3') + p\,\mathrm{tr}\,(h_2'h_1'h_3') + \\
& p^2\,\mathrm{tr}\,(h_1'h_2')\,\mathrm{tr}\,(h_3') + p^2\,\mathrm{tr}\,(h_1'h_3')\,\mathrm{tr}\,(h_2') + p^2\,\mathrm{tr}\,(h_3'h_2')\,\mathrm{tr}\,(h_1').
\end{aligned}
$$

Remark. If one is specially interested in the diagonal elements of X and their moments, one only needs to choose the symmetric matrices as diagonal ones. The results are not really simplified with this extra hypothesis.

6.4 The variance function.

We now calculate the variance function of the natural exponential family generated by μ attached to p and a whose Laplace transform is 19. Our aim is to find the k dimensional generalization of formula 4. The trick is to express σ as a function of m when they are related by 27: a strange second degree equation. For this we coin a lemma.

Lemma 6.3. 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 solution is

$$x = a^{-1/2}(a^{1/2}ba^{1/2})^{1/2}a^{-1/2}.$$

Proof. Existence: clearly $x = a^{-1/2}(a^{1/2}ba^{1/2})^{1/2}a^{-1/2}$ is a solution. Uniqueness: If $y \in \mathcal{P}_k$ is an other solution, then $a^{1/2}ya^{1/2} \in \mathcal{P}_k$ is a root of $a^{1/2}ba^{1/2}$. Since the 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 6.4. Let a be semi positive definite. The variance function of the natural exponential family generated by the measure μ of 19 is

$$V(m)(h) = mh\sigma + \sigma hm - p\sigma h\sigma \tag{29}$$

where σ is as follows. If a is invertible we have

$$\sigma = -\frac{p}{2}a^{-1} + a^{-1/2}(a^{1/2}ma^{1/2} + \frac{p^2}{4}I_k)^{1/2}a^{-1/2}. \tag{30}$$

If a is not invertible, orthonormal coordinates in $\mathbb{R}^k$ are chosen such that a and m and σ 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 σ_1 is obtained from a_1 and m_1 by the formula 30. Furthermore $\sigma_{12} = (\sigma_1 a_1 + pI_{k_1})^{-1}m_{12}$ and $\sigma_2 = \frac{1}{p}(m_2 - \sigma_{21}m_1\sigma_{12})$, where σ_{21} is the transposed matrix of σ_{12}.

Finally, with the notation $\omega = \sigma a \sigma$ we also write $m = \omega + p\sigma$ and

$$V(m)(h) = \omega h\sigma + \sigma h\omega + p\sigma h\sigma. \tag{31}$$

Proof. Let $m \in \mathcal{P}_k$. We compute $\sigma \in \mathcal{P}_k$ such that 27 holds. For this we write $\sigma = x - \frac{p}{2}a^{-1}$ and we get $xax = m + \frac{p^2}{4}a^{-1}$. Apply the lemma to a and $b = m + \frac{p^2}{4}a^{-1}$. We get that

$$\sigma = -\frac{p}{2}a^{-1} + x = -\frac{p}{2}a^{-1} + a^{-1/2}(a^{1/2}ma^{1/2} + \frac{p^2}{4}I_k)^{1/2}a^{-1/2}.$$

Since the variance function is the endomorphism $V(m)$ defined by $m \mapsto mh\sigma + \sigma hm - p\sigma h\sigma$ we get the result in the invertible case. If a is singular, the study of the equation $\sigma a\sigma + p\sigma = m$ is easily done when the coordinates are chosen such 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 $V(m)(h) = \frac{1}{p}mhm$. Finally 31 is easily obtained from 29.

6.5 The Alam and Mitra formula and its extension

Alam and Mitra (1990) have written a remarkable paper whose importance seems to have been unnoticed from the reviewer of Mathematical Rewiews. They prove first formula 33 below:

Proposition 6.5. Let X be a noncentral Wishart random variable $\gamma(p, a, \sigma)$ as defined by 17, with the notation $\omega = \sigma\, a\, \sigma$, with $m = p\sigma + \omega$. Then for all symmetric matrix h of order k

$$\mathbb{E}((X-m)h(X-m)) = \frac{1}{2}[wh\sigma + \sigma h\omega + p\sigma h\sigma + w\,\mathrm{tr}\,(h\sigma) + \sigma\,\mathrm{tr}\,(h\omega) + p\sigma\,\mathrm{tr}\,(h\sigma)]. \tag{32}$$

In particular doing $h = I_k$ the following Alam and Mitra formula holds

$$\begin{aligned}
\mathbb{E}((X-m)^2) &= \frac{1}{2}[w\sigma + \sigma\omega + p\sigma^2 + w\,\mathrm{tr}\,(\sigma) + \sigma\,\mathrm{tr}\,(\omega) + p\sigma\,\mathrm{tr}\,(\sigma)] && (33)\\
&= \frac{1}{2}[(\omega+\sigma)^2 + (\omega+\sigma)\,\mathrm{tr}\,(\omega+\sigma) + (p-1)(\sigma^2 + \sigma\,\mathrm{tr}\,\sigma)] \\
&= \frac{1}{2p}[m^2 + m\,\mathrm{tr}\,m - \omega^2 - \omega\,\mathrm{tr}\,\omega]. && (34)
\end{aligned}$$

Remarks. Note that for $k = 1$ formula 33 gives back 5. Formula 34 is (2.1) in the Alam and Mitra' s paper, with different notations. Let us comment on the importance of 32 and of the innocent looking 34. Up to now, we have been considering a non central 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. With the above proposition, the squares X^2 are at stake. Let us compare with the ordinary Wishart distributions: for them, the literature computes not only expressions like $\mathbb{E}(\langle h_1, X\rangle \dots \langle h_n, X\rangle)$ (recall that we write $\langle h, X\rangle = \mathrm{tr}\,(hX)$) but also expressions like

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

where $\alpha_1, \dots, \alpha_n$ are arbitrary integers (see Graczyk *et al* (2002), (2004), and Letac and Massam (2004)). The Alam and Mitra formula 33 is the first of this type (thus with $n = 1$ and $\alpha_1 = 2$) for the noncentral case.

Unfortunately, for proving 32 we need a result of linear algebra that we are not going to prove. Denote by $\mathcal{S}_k$ the space of real symmetric matrices of order k, equipped with the Euclidean structure $(x, y) \mapsto \mathrm{tr}\,(xy)$. Denote now by $L_s(\mathcal{S}_k)$ the space of symmetric endomorphisms of the Euclidean space $\mathcal{S}_k$. To each $y \in \mathcal{S}_k$ we associate the elements $y \otimes y$ and $P(y)$ of $L_s(\mathcal{S}_k)$ defined respectively by

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

They provide important examples of $L_s(\mathcal{S}_k)$. If x and y are in $\mathcal{S}_k$ one can even consider $(x \otimes y + y \otimes x)$ and $P(x, y) \in L_s(\mathcal{S}_k)$ defined by

$$\begin{aligned}
(x \otimes y + y \otimes x) &= ((x+y) \otimes (x+y) - x \otimes x - y \otimes y)(h) = x\,\mathrm{tr}\,(yh) + y\,\mathrm{tr}\,(xh) \\
P(x, y)(h) &= (P(x+y) - P(x) - P(y))(h) = xhy + yhx.
\end{aligned}$$

With this notation, 29 and 31 could even be rewritten, with $Y = X - m$

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

Finally, the result 32 that we aim to prove is

$$\mathbb{E}(P(Y)) = \frac{1}{2}[P(\omega, \sigma) + (\omega \otimes \sigma + \sigma \otimes \omega) + p(P(\sigma) + \sigma \otimes \sigma)] \qquad (35)$$

Now, $L_s(\mathcal{S}_k)$ is itself a linear space, and the result that we are going to admit as a black box is the following (see Casalis and Letac (1996) Lemma 6.1 and Letac and Massam (1997) Prop. 3.1 for a proof):

Proposition 6.6. There exists a unique endomorphism Ψ of $L_s(\mathcal{S}_k)$ such that for all $y \in \mathcal{S}_k$. one has $\Psi(y \otimes y) = P(y)$. Furthermore

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

Proof of Proposition 6.5. The proof is now very easy

$$\begin{aligned}\mathbb{E}(P(Y)) &= \mathbb{E}(\Psi(Y \otimes Y)) = \Psi(\mathbb{E}(Y \otimes Y)) \\ &= \Psi(P(\omega, \sigma)) + p\Psi(P(\sigma)) \\ &= \Psi(P(\omega + \sigma)) - \Psi(P(\omega)) - \Psi(P(\sigma)) + p\Psi(P(\sigma))\end{aligned}$$

Now applying 36 we get the result 33 under the form 35. To pass from 33 to 34 use $m = p\sigma + \omega$. $\square$

7 Proof of Proposition 3.2

First we need an other variation on the Leibnitz formula: if $\theta \mapsto f(\theta)$ and $\theta \mapsto g(\theta)$ are sufficiently differentiable real functions defined on the same open subset of $\mathbb{R}^n$ and if for $j = 1, \ldots, n$ we denote $D_j = \frac{\partial}{\partial \theta_j}$ then for $a = (a_1, \ldots, a_n) \in \mathbb{N}^n$ one has

$$D_1^{a_1} \ldots D_n^{a_n}(fg)(\theta) = \sum \begin{pmatrix} a_1 \\ i_1 \end{pmatrix} \cdots \begin{pmatrix} a_n \\ i_n \end{pmatrix} D_1^{i_1} \ldots D_n^{i_n}(f)(\theta) D_1^{a_1 - i_1} \ldots D_n^{a_n - i_n}(g)(\theta) \tag{37}$$

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. For $a = 0$ the result is due to Gyndikin. We imitate the proof of the Gyndikin's theorem due to Shanbag (1987). Let $a \in \overline{\mathcal{P}}_k$. Suppose that there exists $p > 0$ and a positive measure $\mu_p(dt)$ on $\overline{\mathcal{P}}_k$ such that for all $\theta \in -\mathcal{P}_k$ one has

$$\frac{1}{(-\theta)^p} e^{\,\mathrm{tr}\,(a(-\theta)^{-1})} = \int_{\overline{\mathcal{P}}_k} e^{\,\mathrm{tr}\,(\theta t)} \mu_p(dt). \tag{38}$$

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

$$Q(\frac{\partial}{\partial \theta}) \frac{1}{(-\theta)^p} e^{\,\mathrm{tr}\,(a(-\theta)^{-1})} = \int_{\overline{\mathcal{P}}_k} Q(t) e^{\,\mathrm{tr}\,(\theta t)} \mu_p(dt).$$

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 '

$$Q(\frac{\partial}{\partial \theta}) \frac{1}{(-\theta)^p} e^{\,\mathrm{tr}\,(a(-\theta)^{-1})} = \frac{1}{(-\theta)^{n+p}} e^{\,\mathrm{tr}\,(a(-\theta)^{-1})} P_Q(p). \tag{39}$$

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

$$\frac{1}{(-\theta)^{n+p}} e^{\operatorname{tr}(a(-\theta)^{-1})} P_Q(p) = \int_{\overline{\mathcal{P}}_k} \det(t) e^{\operatorname{tr}(\theta t)} \mu_p(dt). \tag{40}$$

Note that the right hand side of 40 is ≥ 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 the singular matrices for $p = 1/2, \ldots, (k-1)/2$ from Proposition 2.1. 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 on $p = 0, 1/2, \ldots, (k-1)/2$. Furthermore, Proposition 3.1 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 μ_p exists and that $p \notin \Lambda$. Thus $P_Q(p) > 0$ and therefore there exists 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 μ_p exists, then

$$\mu_{p+\frac{1}{2}} = \mu_p * \mu_{\frac{1}{2}}$$

does exist too, as can be seen by the Laplace transform. But now $P_Q(p + \frac{1}{2}) < 0$ which is the desired contradiction.

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

$$\int_{\overline{\mathcal{P}}_k} e^{-\operatorname{tr}(st)} \gamma(p, a)(dt) = \frac{1}{\det(I_k + s)^p} e^{-\operatorname{tr}((I_k+s)^{-1} sa)}.$$

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

8 Eigenvalues of non central Wishart

We rely first on a celebrated theorem about the distribution of the eigenvalues of a random matrix of $\mathcal{P}_k$ (see Muirhead page 104).

Proposition 8.1. Let T be a random matrix of $\mathcal{P}_k$ having density f. Denote by $\Lambda_1 > \ldots > \Lambda_k > 0$ the sequence of the eigenvalues of T. Then the density of $(\Lambda_1, \ldots, \Lambda_k)$ is

$$\frac{\pi^{k^2/2}}{\Gamma_{\mathcal{P}_k}(k/2)} \prod_{1 \leq i < j \leq k} (\lambda_i - \lambda_j) \int_{\mathbb{SO}(k)} f(u^{-1} \operatorname{diag}(\lambda_1, \ldots, \lambda_k) u) du.$$

We apply this result to the standard non central Wishart:

Proposition 8.2. Let T be a non central Wishart matrix with distribution $\gamma(p, a)(dt)$ with $p > (k-1)/2$. Denote by $\Lambda_1 > \ldots > \Lambda_k > 0$ the sequence of the eigenvalues of T. Then the density of $(\Lambda_1, \ldots, \Lambda_k)$ is

$$\frac{\pi^{k^2/2}}{\Gamma_{\mathcal{P}_k}(k/2)\Gamma_{\mathcal{P}_k}(p)} e^{-(\lambda_1+\cdots+\lambda_k)}(\lambda_1 \cdots \lambda_k)^{p-\frac{k+1}{2}} \prod_{1\leq i<j\leq k} (\lambda_i - \lambda_j) \tag{41}$$

$$\times e^{-\operatorname{tr} a} \left(\sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\kappa \in E_m} \frac{C_\kappa(a) C_\kappa(\operatorname{diag}(\lambda_1, \ldots, \lambda_k))}{C_\kappa(I_k)(p)_\kappa} \right) \tag{42}$$

Proof. Just apply Proposition 8.1 to the density defined by Proposition 3.1 and use formula 13. $\square$

Remarks. (1) Note that in the above density of the eigenvalues, line 41 is the density of the eigenvalues of a standard Wishart variable with shape parameter p. It does not depend on a. The line 42 depends on a and p and appears as a perturbation of the preceding line.

(2) Suppose that the rank of a is $r < k$. Consider $\kappa = (m_1, \ldots, m_k)$ with $m_1 \geq \ldots m_k \geq m_{k+1} = 0$. Then $\Delta_\kappa(a) = 0$ if there exists i such that $r < i \leq k$ and $m_i > m_{i+1}$ since no sub determinant of a of size $i > r$ can be non zero. The definition of $C_\kappa(a)$ implies that $C_\kappa(a) = 0$ for κ satisfying the above condition. Thus if the rank of a is $r < k$ the sum in 42 on E_m can be replaced on the set of κ's of the form $(m_1, \ldots, m_r, m_{r+1}, \ldots, m_{r+1})$.

9 The estimation of σ and a.

Suppose that we have N iid observations $X_1, \ldots, X_N$ with non central Wishart distribution $\gamma(p, a, \sigma)$ defined by 17. If a is known, 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. All the methods that we shall consider will be of the following type: some functions α and β of (a, σ) been given, one inverts them into

$$a = f(\alpha, \beta) \quad \sigma = g(\alpha, \beta).$$

Each method now chooses (α, β) and estimators $(\widehat{\alpha}, \widehat{\beta})$ as functions of the observations $X_1, \ldots, X_N$. These estimators are generally unbiased, but not always. We now plug $(\widehat{\alpha}, \widehat{\beta})$ into $f(\alpha, \beta)$ and $g(\alpha, \beta)$ in order to get the estimators

$$\widehat{a} = f(\widehat{\alpha}, \widehat{\beta}) \quad \widehat{\sigma} = g(\widehat{\alpha}, \widehat{\beta}).$$

9.1 σ unknown and a known.

Then the model is a natural exponential family. Estimation of $\sigma = (-\theta)^{-1}$ is easy since the methods of natural exponential families are available, and we find a maximum likelihood estimator. We take

$$\hat{m} = \overline{X}_N = \frac{1}{N}(X_1 + \cdots + X_N)$$

has an estimator of m and we plug this value of m into the formula 30 (if a is invertible) and we get a reasonable estimator for σ. If a is singular, we use the analogous formula after 30. For further use, observe that the distribution of $\overline{X}_N$ is $\gamma(Np, N^2 a, \sigma/N)$. Thus $\omega = \sigma a \sigma$ does not change and we have from 34

$$\mathbb{E}((\overline{X}_N - m)^2) = \frac{1}{2Np}(m^2 + m \operatorname{tr} m - \omega^2 - \omega \operatorname{tr} \omega).$$

9.2 σ known and a unknown.

The classical method, probably due to T.W. Anderson (1946) is to use $m = p\sigma + \sigma a \sigma = p\sigma + \omega$ for the estimate

$$\widehat{a} = \sigma^{-1}\overline{X}_N\sigma^{-1} - p\sigma^{-1}.$$

This is specially popular when $\sigma = I_k$ since $\widehat{a} = \overline{X}_N - pI_k$. See Leung (1994) and Neudecker (2004) for variants and properties. This estimator is not always semi positive definite.

9.3 a and σ unknown. Our method.

We suggest to use 28 to estimate a in the following way. It says that $V(m)(h) = mh\sigma + \sigma hm - p\sigma h\sigma$. Let us apply this to $h = I_k$ and denote $v = V(m)(I_k)$ for simplicity. By definition we have

$$v = \mathbb{E}((X - \mathbb{E}(X)) \otimes (X - \mathbb{E}(X)))(I_k) = \mathbb{E}((X - \mathbb{E}(X)) \times \operatorname{tr}(X - \mathbb{E}(X))).$$

Thus $v = m\sigma + \sigma m - p\sigma^2$ can be rewritten $(\sigma - \frac{1}{p}m)^2 = \frac{1}{p^2}m^2 - \frac{1}{p}v$ or

$$\sigma = \frac{1}{p}(m + (m^2 - pv)^{1/2})$$

Thus finding an estimate of v will be finding an estimate of σ. This will lead to an estimate of a via 27 since

$$a = \sigma^{-1}m\sigma^{-1} - p\sigma^{-1}.$$

We now suggest the following unbiased estimator for the matrix v

$$\hat{v} = \frac{1}{N-1}\sum_{j=1}^{N}\left[(X_j - \overline{X}_N) \times \operatorname{tr}(X_j - \overline{X}_N)\right]$$

which leads to the estimators $\hat{\sigma}$ and $\hat{a}$ for σ 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 frame mentioned above, with $(\alpha, \beta) = (m, v)$. Here the functions f and g are

$$f(\alpha, \beta) = \frac{1}{p}(\alpha + (\alpha^2 - p\beta)^{1/2}), \quad g(\alpha, \beta) = f(\alpha, \beta)^{-1}\alpha f(\alpha, \beta)^{-1} - pf(\alpha, \beta)^{-1}.$$

The conditions of existence (the symmetric matrix $\hat{m}^2 - p\hat{v}$ has to be semi positive definite and this condition is not necessarily fulfilled for all samples $X_1, \ldots, X_N$) and the properties of these estimators have still to be studied. A positive side of the proposed method is that it uses only one square root of matrices, namely $(\frac{1}{p}\hat{m}^2 - \hat{v})^{1/2}$. When a is known the classical method of estimation of parameters for a natural exponential family was leading to the estimator of σ equal to

$$-\frac{p}{2}a^{-1} + a^{-1/2}(a^{1/2}\hat{m}a^{1/2} + \frac{p^2}{4}I_k)^{1/2}a^{-1/2}$$

involving two roots in the non singular case, and with an even more complicate formula in the singular case.

9.4 a and σ unknown. The method 1 of Alam and Mitra.

This method uses $\alpha = m$ and $\beta = \omega^2 + \omega \operatorname{tr} \omega$ (with the usual notation $\omega = \sigma a \sigma$). A remarkable observation of Alam and Mitra is the fact from linear algebra that $\omega \mapsto \omega^2 + \omega \operatorname{tr} \omega$ is a bijection of $\overline{\mathcal{P}}_k$ onto itself, as a consequence of the following lemma:

Lemma 7.1. If ω is a semi positive definite matrix and if $\beta = \omega^2 + \omega \operatorname{tr} \omega$ then $\nu = \operatorname{tr} \omega$ is a function of β alone.

Proof of the lemma. We can write

$$\beta + \frac{\nu^2}{4}I_k = (\omega + \frac{\nu}{2}I_k)^2$$

which leads to

$$(\beta + \frac{\nu^2}{4}I_k)^{1/2} = \omega + \frac{\nu}{2}I_k. \tag{43}$$

Taking the trace of both sides we get that the following function on $\mathbb{R}$

$$h(x) = -x(1 + \frac{k}{2}) + \operatorname{tr}(\beta + \frac{x^2}{4}I_k)^{1/2}$$

is zero for $x = \nu$. Since $x \mapsto h(x)$ is easily proved to be convex (for β is semi positive definite) and since $\lim_{x \to -\infty} h(x) = -\infty$ then ν is the only root of $h(x) = 0$. This shows that $\nu = \nu(\beta)$ is a function of β alone. $\square$

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 43 gives ω as a function of β:

$$\omega = \omega(\beta) = -\frac{\nu(\beta)}{2}I_k + (\beta + \frac{\nu(\beta)^2}{4}I_k)^{1/2}.$$

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 $\widehat{\alpha}$ and $\widehat{\beta}$. For $\widehat{\alpha}$ we take $\overline{X}_n$. For $\widehat{\beta}$ we recall first the formula 34 that we write under the form

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

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

$$\begin{aligned} \widehat{\beta}_1 &= \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X}_N)^2 \\ \widehat{\beta}_2 &= \frac{1}{N(N-1)}\sum_{i\neq j}(X_iX_j + X_i \operatorname{tr} X_j) \end{aligned}$$

They are both unbiased. For $\widehat{\beta}_1$ write $(X_i-\overline{X}_N)^2 = (X_i - m + m - \overline{X}_N)^2$ and for $\widehat{\beta}_2$ observe that X_i and X_j are independent and that $\mathbb{E}(\operatorname{tr} X_j) = \operatorname{tr}\mathbb{E}(X_j)$ by linearity of the trace.

Thus $\widehat{\beta}_3 = \widehat{\beta}_2 - 2p\,\widehat{\beta}_1$ is an unbiased estimator of β. It is not necessarily semi positive definite. For this reason, let us write $\widehat{\beta}_3 = u\,\mathrm{diag}(c_1,\ldots,c_k)\,u^{-1}$ such that u is an orthogonal matrix and such that the eigenvalues of $\widehat{\beta}_1$ satisfy $c_1 > \ldots c_j > 0 > c_{j+1} > \ldots > c_k$. The estimator $\widehat{\beta}$ that we finally consider and that we plug into f and g is simply the semi positive definite matrix

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

What is the cost of this method? Essentially:

1. The calculation of $\widehat{\beta}$ from $\widehat{\beta}_3$: a diagonalization of a symmetric matrix.
2. The calculation of $\nu(\widehat{\beta})$ (which is the trace of the estimator of ω), for instance by Newton approximation.

The cost of the calculation of $(\widehat{\beta} + \frac{\nu(\widehat{\beta})^2}{4}I_k)^{1/2}$ from the two preceding items is negligible.

9.5 a and σ unknown. The method 2 of Alam and Mitra.

The second method is based on 18. Its aesthetic value is diminished by the fact that is is not free of coordinates as the first. For any line vector $z \in \mathbb{R}^k$ define $\alpha_z = m_z$

and $\beta_z = v_z$ with the notations of 18. One can express $z\sigma z^T$ and $z\omega z^T$ with respect to α_z and β_z since from 18 we have

$$\begin{aligned} z\sigma z^T &= \frac{1}{p}(m_z + (m_z^2 - pv_z)^{1/2}) = \frac{1}{p}(\alpha_z + (\alpha_z^2 - p\beta_z)^{1/2}), \\ z\omega z^T &= (m_z^2 - pv_z)^{1/2} = (\alpha_z^2 - p\beta_z)^{1/2}. \end{aligned}$$

Now, Alam and Mitra replace in the above formula α_z and β_z by the unbiased estimators

$$\widehat{\alpha_z} = z\overline{X}_N z^T = \operatorname{tr} \overline{X}_N z^T z, \;\; \widehat{\beta}_z = \frac{1}{N-1}\sum_{i=1}^{N}(zX_i z^T - z\overline{X}_N z^T)^2 = \frac{1}{N-1}\sum_{i=1}^{N}[\operatorname{tr}(X_i - \overline{X}_N)z^T z]^2.$$

Denote for a while

$$f(z^T z) = \frac{1}{p}(\widehat{\alpha_z} + (\widehat{\alpha_z}^2 - p\widehat{\beta}_z)^{1/2}), \;\; g(z^T z) = (\widehat{\alpha_z}^2 - p\widehat{\beta}_z)^{1/2}. \tag{44}$$

In a not too clear way, Alam and Mitra define (in their 2.9 and 2.10) the estimators $\widehat{\sigma}$ and $\widehat{\omega}$ by

$$z\widehat{\sigma}z^T = \operatorname{tr}(\widehat{\sigma}z^T z) = f(zz^T), \;\; z\widehat{\omega}z^T = \operatorname{tr}(\widehat{\omega}z^T z) = g(zz^T). \tag{45}$$

It should be specified that there are no $\widehat{\sigma}$ or $\widehat{\omega}$ which satisfy this for all $z \in \mathbb{R}^k$. For instance this would imply that the function $h \mapsto g(h)$ defined on the space $\mathcal{S}_k$ of symmetric matrices satisfying for semi positive h

$$g(h) = \left(\operatorname{tr}(\overline{X}_N h)^2 - \frac{1}{N-1}\sum_{i=1}^{N}[\operatorname{tr}(X_i - \overline{X}_N)h]^2 \right)^{1/2}$$

is a linear form on $\mathcal{S}_k$. This is clearly not true for most of the $(X_1, \ldots, X_N)$: just square both sides of the expression of $g(h)$ and use the Kronecker's law of inertia for the quadratic forms. However, if we impose (as Alam and Mitra finally do) that 45 is true only for the 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 45 defines completely $\widehat{\sigma}$ and $\widehat{\omega}$. As one can see, this method is linked to the special basis e. Its advantage is the simplicity of the formulas 44.

10 References

K. Alam, A. Mitra, (1990) 'On estimated the scale and noncentrality matrices of a Wishart distribution' *Sankhā Series B*, **52** 133-143.

M. Casalis, G. Letac, (1996) 'The Lukacs-Olkin-Rubin characterisation of the Wishart distributions on symmetric cones,' *Annals of Statistics*, **24**, 763-786.

J. Faraut, A. Koranyi (1994) *Analysis on Symmetric Cones.* Oxford University Press.

P. Graczyk, G. Letac, H. Massam, (2003) 'The complex Wishart distribution and the symmetric group,' *Annals of Statistics*, **31**, , 287-309.

G. Letac, H. Massam, (1998) 'Quadratic and inverse regressions for Wishart distributions,' *Annals of Statistics*, **26**, 573-595.

G. Letac, H. Massam, (2004) 'All invariant moments of the Wishart distribution', *Scand. J. Statist.*.

R. Muirhead, (1982) *Aspects of Multivariate Analysis.* Wiley, New York.

Rainville, (1960) *Special functions.* Chelsea, New York.