PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 35, Fasc. 1 (2015), pp. 73–90

THE FISHER INFORMATION AND EXPONENTIAL FAMILIES PARAMETRIZED BY A SEGMENT OF MEANS

BY

PIOTR GRACZYK (ANGERS) AND SALHA MAMANE* (JOHANNESBURG)

Abstract. We consider natural and general exponential families $(Q_m)_{m\in M}$ on $\mathbb{R}^d$ parametrized by the means. We study the submodels $(Q_{\theta m_1+(1-\theta)m_2})_{\theta\in[0,1]}$ parametrized by a segment in the means domain from the point of view of the Fisher information. Such a parametrization allows for a parsimonious model and is particularly useful in practical situations when hesitating between two parameters m_1 and m_2. The most interesting cases are multivariate Gaussian and Wishart models with matrix parameters.

2000 AMS Mathematics Subject Classification: Primary: 62H12; Secondary: 62H10.

Key words and phrases: Fisher information, efficient estimator, exponential family, multivariate Gaussian distribution, Wishart distribution, parsimony.

1. INTRODUCTION

The Fisher information is a key concept in mathematical statistics. Its importance stems from the Cramér–Rao inequality which says that the covariance of any unbiased estimator $T(X_1,\ldots,X_n)$ of an unknown parameter θ is bounded by the inverse of the Fisher information: $\mathrm{Var}_\theta(T)-\big(I(\theta)\big)^{-1}$ is semi-positive definite. The efficiency of an estimator is based on the equality in this inequality. For some recent applications of the Fisher information in modern statistics see [1], [6], [2], [16].

The objective of this work is to study the Fisher information for exponential families $(Q_m)_{m\in M}$ parametrized by a segment of means $[m_1,m_2]$.

Exponential families of distributions are extensively used in statistics and intensively studied, cf. [7]–[10]. They are the only models for which the Cramér–Rao bound is always attained. A parametrization of the family by a segment instead of the whole means domain allows us to obtain a parsimonious model when the

* The author acknowledges financial support (co-tutelle fellowship) from the French Ministry of Foreign Affairs and the Embassy of France in South Africa.

means domain is high-dimensional. The parametrization of the mean parameter by a segment is particularly useful in practical situations when hesitating between two equally convenient mean values m_1 and m_2. Such parametrization will also serve in sequential data collection, when an updated estimate of a parameter largely differs from the previous estimate.

In this paper we prove explicit formulas for the Fisher information of the exponential families $(Q_{\theta m_1+(1-\theta)m_2})_{\theta\in[0,1]}$ if the full model is either the multivariate Gaussian family of known mean and unknown covariance matrix or a family of Wishart distributions with unknown scale parameter.

The paper is organized as follows. In Section 2, basic definitions and results on Fisher information and exponential families are recalled. Section 3 contains new results on the Fisher information of exponential Gaussian and Wishart subfamilies parametrized by a segment of means $[m_1, m_2]$. When m_1 and m_2 are colinear, we construct efficient estimators for the segment parameter θ.

2. PRELIMINARIES

DEFINITION 2.1. Consider a σ-finite measurable space $(\Omega, \mathcal{A}, \nu)$ with a family of probability density functions f_s, $s \in S \subset \mathbb{R}^d$, dominated by ν. Let $l_s = \ln f_s$. Assume that the function $s \mapsto l_s(\omega)$ is differentiable for every $\omega \in \Omega$. Consider the gradient l'_s of the map $s \mapsto l_s$ as a random vector on the statistical model $(\Omega, \mathcal{A}, f_s d\nu)$. Suppose that it satisfies $\mathbb{E}_s(\|l'_s\|^2) < \infty$. The *Fisher information matrix* is defined by $I(s) = \mathbb{E}_s(l'_s {l'_s}^T)$.

In the sequel we restrict our attention to exponential statistical models.

DEFINITION 2.2. Let $T : \Omega \to \mathbb{R}^d$. Set

$$S = \big\{s : K(s) = \ln \textstyle\int \exp\{\langle s, T\rangle\} d\nu < \infty\big\} \subset \mathbb{R}^d.$$

We suppose that the set S has a non-empty interior S^0.

The *general exponential family* generated by the measure ν and the map T is the family

$$\big\{dP_s(T, \nu) = \exp\{\langle s, T\rangle - K(s)\} d\nu = f_s d\nu : s \in S\big\}. \tag{2.1}$$

Let μ be the image of the measure ν by T on $\mathbb{R}^d$. The *natural exponential family* associated with the above general exponential family is the family of probability distributions defined by

$$\big\{dP_s(\mu) = \exp\{\langle s, x\rangle - K(s)\} d\mu : \ s \in S\big\}. \tag{2.2}$$

Natural exponential families may be viewed as a special case of general exponential families with $\Omega \subset \mathbb{R}^d$, $T(\omega) = \omega$ and $\nu = \mu$.

As usual, $\mathbb{E}_s$ denotes the integral on the exponential model $\big(\Omega, \mathcal{A}, (P_s)_{s\in S}\big)$, where $P_s = P_s(T,\nu)$. We have $l_s = \exp\{\langle s,\, T\rangle - K(s)\}$. Theorem 2.7.1 in [8] ensures that the cumulant function K and the function $s \mapsto l_s(\omega)$ are analytic on S^0. Moreover, the derivatives with respect to s can be carried out under the integral sign in

$$1 = \int \exp\{\langle s,\, T\rangle - K(s)\} d\nu$$

as long as $s \in S^0$. This gives, by taking the derivatives and by integration by parts,

$$\begin{aligned} \mathbb{E}_s l'_s &= 0, \\ -\mathbb{E}_s l''_s &= I(s). \end{aligned} \tag{2.3}$$

Similarly, we obtain the mean and the covariance:

$$m(s) = \mathbb{E}_s(T) = K'(s), \tag{2.4}$$

$$v(s) = \mathrm{Cov}_s(T) = K''(s). \tag{2.5}$$

From (2.3) and (2.5) it follows that the Fisher information of a general exponential family $P_s(T,\nu)$ equals, for $s \in S^0$,

$$I(s) = K''(s) = v(s). \tag{2.6}$$

The following result is proved in [10].

PROPOSITION 2.1. *The map $s \mapsto m(s) = \mathbb{E}_s(T) = K'(s)$ is an analytic diffeomorphism from S^0 to the open set $M = m(S^0) \subset \mathbb{R}^d$ called the* domain of the means *of the family.*

Let $\psi : M \to S^0, m \mapsto \psi(m) = (K')^{-1}(m)$ denote the inverse of the "mean" diffeomorphism K'. The general exponential family, parametrized by the domain of the means M, is given by the family of distributions

$$Q(m,T,\nu)(d\omega) = e^{\langle \psi(m)\,,T(\omega)\rangle - K(\psi(m))}\nu(d\omega), \quad m \in M. \tag{2.7}$$

The mean of the family (2.7) is equal to m. We denote the covariance of the family (2.7) by $V(m)$ and, by (2.5), we have

$$V(m) = v\big(\psi(m)\big) = K''\big(\psi(m)\big). \tag{2.8}$$

The function $V : m \in M \to V(m)$ is called the *variance function* of the exponential family.

In order to avoid confusion, when the parameter of an exponential family is the mean m, we will denote the Fisher information by $J(m)$.

THEOREM 2.1. *The Fisher information of the exponential family* (2.7) *equals*

$$J(m) = V(m)^{-1} = \psi'(m), \tag{2.9}$$

where $V(m)$ *is the variance function of the exponential family*, *given by* (2.8).

P r o o f. By Definition 2.1 and by the chain rule,

$$J(m) = \psi'(m)^T I\big(\psi(m)\big)\psi'(m) \quad \text{on } M.$$

Since $\psi(m) = (K')^{-1}(m)$, we have $\psi'(m) = \big[K''\big(\psi(m)\big)\big]^{-1}$. Thus, using formula (2.6), we get

$$J(m) = \big[K''\big(\psi(m)\big)\big]^{-1} = V(m)^{-1}. \quad \blacksquare$$

REMARK 2.1. *Note a striking contrast in the formulas* (2.6) *and* (2.9) *for the Fisher information of an exponential family parametrized either by the canonical parameter* $s \in S^0$ *or by the mean* $m \in M$. *Indeed*, *in the first case we have the equality* $I\big(\psi(m)\big) = V(m)$, *in the second* $J(m) = V(m)^{-1}$.

Finally, consider a general exponential family parametrized by a segment of means. Let $A \neq 0, B \in \mathbb{R}^d$. Define $\Theta = \{\theta \in \mathbb{R} : \theta A + B \in M\}$. The set $\Theta \subset \mathbb{R}$ is open because M is open. Suppose that $\Theta \neq \emptyset$. The parametrization by a segment of means $I \subset \Theta$ consists in considering the submodel

$$\{Q(\theta A + B, T, \nu) : \theta \in I\}. \tag{2.10}$$

Such models contain the case $\big\{Q\big(\theta m_1 + (1-\theta)m_2, T, \nu\big) : \theta \in [0,1]\big\}$ when one hesitates between two different estimations $m_1, m_2 \in M$ of the true mean m of an exponential family (2.7).

The following corollary gives the Fisher information of a general exponential family parametrized by a segment of means. By analogy to the notation $J(m)$, we denote this information by $J(\theta)$.

COROLLARY 2.1. *The Fisher information of the model* $\{Q(\theta A + B, T, \nu) : \theta \in I\}$ *equals*

$$J(\theta) = A^T\, V(\theta A + B)^{-1}\, A. \tag{2.11}$$

P r o o f. We use Definition 2.1 and the chain rule as in the proof of Theorem 2.1 for the reparametrization $f : I \to M$, $f(\theta) = \theta A + B$ with $f'(\theta) = A$. We conclude by Theorem 2.1. $\blacksquare$

3. THE FISHER INFORMATION OF GAUSSIAN AND WISHART FAMILIES PARAMETRIZED BY A SEGMENT OF MEANS

In this section we study the Fisher information for multivariate Gaussian and Wishart exponential families. These families are parametrized by symmetric positive definite matrices. Therefore, we first adapt the presentation to suit this case. We denote by $\mathbb{R}^{k\times m}$ the space of real matrices with k rows and m columns, and by $A\otimes B$ the Kronecker product of two matrices. We use the usual notation $\langle A, B\rangle = \mathrm{Tr}(A^T B)$ for the scalar product of two matrices. The operator Vec converts a $k\times m$ matrix A into a vector $\mathrm{Vec}(A)\in\mathbb{R}^{km}$ by stacking the columns one underneath the other. The Vec operator is commonly used in applications of the matrix differential calculus in statistics, cf. [13], [15].

The following properties of the Kronecker product are used in this work (cf. [13], pp. 32 and 35). For non-singular square matrices A, B we have $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$. For all matrices A, B and C such that the product ABC is well defined we have

$$\mathrm{Vec}(A\,B\,C) = (C^T\otimes A)\,\mathrm{Vec}(B). \tag{3.1}$$

In this paper we use the following convention of the matrix differential calculus: if a function $f:\mathbb{R}^{k\times p}\to\mathbb{R}^{n\times m}$ is differentiable then its derivative is a matrix $f'(x)\in\mathbb{R}^{nm\times kp}$ such that

$$\mathrm{Vec}\left(df(x)(u)\right) = f'(x)\,\mathrm{Vec}(u),\qquad u\in\mathbb{R}^{k\times p}. \tag{3.2}$$

The only exception we will make is the derivative of a function $K:\mathbb{R}^{k\times m}\to\mathbb{R}$, for which the following convention is used: the derivative of K is not a row vector but the matrix $K'(x)\in\mathbb{R}^{k\times m}$ related to the differential of K by $dK(x)(u) = \langle K'(x), u\rangle = \mathrm{Tr}\left(K'(x)^T u\right)$ for all $u\in\mathbb{R}^{k\times m}$. This convention is needed to give sense to formula (2.4) for the mean of an exponential family.

The following lemma is useful for the derivation of an alternative formula for the Fisher information of an exponential family parametrized by a segment of means and satisfying an additional condition (3.3). We will see that this condition holds for Gaussian and Wishart models.

LEMMA 3.1. *Assume that, for all $m\in M$,*

$$\langle m,\, \psi(m)\rangle = C \tag{3.3}$$

for some constant $C\in\mathbb{R}$. Then, for all $u\in M$,

$$\langle m,\, d\psi(m)(u)\rangle = -\langle u,\, \psi(m)\rangle. \tag{3.4}$$

P r o o f. By (3.3) it follows that the differential of the function $g: M\to\mathbb{R}$, $m\mapsto\langle m,\,\psi(m)\rangle$ is zero. Therefore, $dg(m)(u) = \langle m,\, d\psi(m)(u)\rangle + \langle u,\, \psi(m)\rangle = 0$ for all $m, u\in M$ and (3.4) holds true. ■

COROLLARY 3.1. *Let* $\{Q(\theta A + B, T, \nu)(d\omega) : \theta \in I\}$ *be an exponential model parametrized by a segment of means* $M \subset \mathbb{R}^{k\times m}$ *. If the condition* (3.3) *holds then the Fisher information of the model equals*

$$J(\theta) = -\frac{d^2}{d\theta^2}\big[K\big(\psi(\theta A + B)\big)\big]. \tag{3.5}$$

Proof. Let $h(\theta) = K\big(\psi(\theta A + B)\big)$ and $f(\theta) = \theta A + B$. We want to compute $h''(\theta)$. If $\theta, u \in \mathbb{R}$,

$$\begin{aligned} dh(\theta)(u) &= dK\Big(\psi\big(f(\theta)\big)\Big)\Big(d\psi\big(f(\theta)\big)\big(df(\theta)(u)\big)\Big) \\ &= \Big\langle K'\Big(\psi\big(f(\theta)\big)\Big),\, d\psi\big(f(\theta)\big)\big(df(\theta)(u)\big)\Big\rangle \\ &= \big\langle f(\theta),\, d\psi\big(f(\theta)\big)\big(df(\theta)(u)\big)\big\rangle \\ &= -\big\langle df(\theta)(u),\, \psi\big(f(\theta)\big)\big\rangle \\ &= -u\big\langle A,\, \psi\big(f(\theta)\big)\big\rangle, \end{aligned}$$

where we used successively: the convention on K' introduced after (3.2), the equality $K' \circ \psi(m) = m$, Lemma 3.1 and the formula $df(\theta)(u) = uA$. Thus we have $h'(\theta) = -\big\langle A,\, \psi\big(f(\theta)\big)\big\rangle$. Now, starting as in the computation of $h'(\theta)$ and using (3.2), we get

$$\begin{aligned} h''(\theta) &= -\big\langle A,\, d\psi\big(f(\theta)\big)(A)\big\rangle = -\operatorname{Vec}(A)^T \operatorname{Vec}\Big(d\psi\big(f(\theta)\big)(A)\Big) \\ &= -\operatorname{Vec}(A)^T \psi'(\theta A + B)\operatorname{Vec}(A). \end{aligned}$$

We conclude using (2.9) and Corollary 2.1. ∎

3.1. Exponential families of Gaussian distributions. We denote by $\mathcal{S}_d$ the vector space of $d \times d$ symmetric matrices and by $\mathcal{S}_d^+$ the open cone of positive definite matrices.

Recall the construction of the multivariate Gaussian model $\{N(u, \Sigma); \Sigma \in \mathcal{S}_d^+\}$ as a general exponential family. We consider $\Omega = \mathbb{R}^d$ equipped with a normalized Lebesgue measure $\nu(d\omega) = d\omega/(2\pi)^{d/2}$, the vector space $\mathcal{S}_d$ and the map

$$T : \mathbb{R}^d \to \mathcal{S}_d, \quad T(\omega) = -\frac{1}{2}(\omega - u)(\omega - u)^T.$$

The image of T is contained in the opposite of the cone of semi-positive definite matrices of rank one. For $s \in \mathcal{S}_d^+$, we have

$$\int_\Omega e^{\langle s, T(\omega)\rangle}\nu(d\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\frac{1}{2}\operatorname{Tr}(s(\omega-u)(\omega-u)^T)} d\omega = (\det s)^{-1/2}$$

and the integral is infinite otherwise. Thus $S = \mathcal{S}_d^+$ and the cumulant function is

$$K(s) = -\frac{1}{2}\ln\det(s), \qquad s \in S = \mathcal{S}_d^+.$$

The general exponential family is therefore

$$P_s(T,\nu)(d\omega) = \frac{1}{(2\pi)^{d/2}} e^{\langle s\,,\,-\frac{1}{2}(\omega-u)(\omega-u)^T\rangle + \frac{1}{2}\ln\det(s)} d\omega \tag{3.6}$$
$$= \frac{(\det s)^{1/2}}{(2\pi)^{d/2}} e^{-\frac{1}{2}(\omega-u)^T s(\omega-u)} d\omega,$$

which is the family of Gaussian distributions $N(u, s^{-1})$ on $\mathbb{R}^d$ with a fixed mean $u \in \mathbb{R}^d$, parametrized by $s = \Sigma^{-1}$, the inverse of the covariance matrix Σ supposed to be invertible.

The derivative of the function $X \in \mathbb{R}^{d\times d} \to \det X$ is the cofactor matrix $X^\sharp$ which equals $(\det X)(X^{-1})^T$ when X is invertible. It follows that

$$m(s) = K'(s) = -\frac{1}{2}s^{-1}, \quad s \in \mathcal{S}_d^+.$$

The means domain is $M = -\mathcal{S}_d^+$ and the inverse mean map is $\psi(m) = -\frac{1}{2}m^{-1}$. The Gaussian general exponential family parametrized by $m \in M = -\mathcal{S}_d^+$ is therefore the family

$$Q(m,T,\nu) = N(u,-2m). \tag{3.7}$$

Up to a trivial affine change of parameter $\Sigma = -2m$, this parametrization by the covariance parameter is more natural than the parametrization of the family $\big(N(u,s^{-1})\big)_{s\in\mathcal{S}_d^+}$ by the canonical parameter s.

In order to compute the variance function, recall that $XX^{-1} = I_d$ implies that $dX^{-1} = -X^{-1}dX\,X^{-1}$ and $(X^{-1})' = -X^{-1} \otimes X^{-1}$. Thus, we obtain $K''(s) = \frac{1}{2}s^{-1} \otimes s^{-1}$ and formula (2.8) implies that

$$V(m) = 2m \otimes m. \tag{3.8}$$

The Fisher information of the family $\big(N(u,s^{-1})\big)_{s\in\mathcal{S}_d^+}$ is $I(s) = \frac{1}{2}s^{-1}\otimes s^{-1}$. By Theorem 2.1 and formula (3.8), we infer that the Fisher information of the model $\big(N(u,-2m)\big)_{m\in-\mathcal{S}_d^+}$ equals $J(m) = \frac{1}{2}m^{-1} \otimes m^{-1}$.

COROLLARY 3.2. *The Fisher information matrix of the Gaussian model* $\big(N(u,\Sigma)\big)_{\Sigma\in\mathcal{S}_d^+}$ *is*

$$J(\Sigma) = \frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1}.$$

P r o o f. Using the chain rule and a reparametrization $\Sigma = -2m$ we see that the information for the new parameter Σ is $\tilde{J}(\Sigma) = \frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1} = J(\Sigma)$. ∎

Let us now consider Gaussian models parametrized by a segment of covariances.

COROLLARY 3.3. *Let C and D be two symmetric matrices and let $I \subset \mathbb{R}$ be a non-empty segment such that $I \subset \Theta = \{\theta \in \mathbb{R} : \theta C + D \in \mathcal{S}_d^+\}$. The Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ is*

$$J(\theta) = \frac{1}{2} \operatorname{Tr}\big(C(\theta C + D)^{-1} C(\theta C + D)^{-1}\big).$$

P r o o f. We use Corollary 3.2 and the chain rule with $f(\theta) = \theta C + D$. It follows that

$$\begin{aligned} J(\theta) &= \operatorname{Vec}(C)^T J(\theta C + D) \operatorname{Vec}(C) \\ &= \frac{1}{2} \operatorname{Vec}(C)^T \big((\theta C + D)^{-1} \otimes (\theta C + D)^{-1}\big) \operatorname{Vec}(C). \end{aligned}$$

Applying (3.1) we get

$$\begin{aligned} J(\theta) &= \frac{1}{2} \operatorname{Vec}(C)^T \operatorname{Vec}\big((\theta C + D)^{-1} C(\theta C + D)^{-1}\big) \\ &= \frac{1}{2} \operatorname{Tr}\big(C(\theta C + D)^{-1} C(\theta C + D)^{-1}\big). \quad ∎ \end{aligned}$$

On the other hand, we have the following alternative formula for the information $J(\theta)$.

COROLLARY 3.4. *With the notation as above the Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ is*

$$J(\theta) = -\frac{1}{2} \frac{d^2}{d\theta^2}\big(\ln \det(\theta C + D)\big). \tag{3.9}$$

P r o o f. Observe that the condition (3.3) holds for the Gaussian exponential families $Q(m, t, \nu)$:

$$\langle m\,,\, \psi(m)\rangle = -\frac{1}{2} \operatorname{Tr}(m m^{-1}) = -\frac{d}{2}.$$

The model $N(u, \theta C + D) = N(u, -2m) = Q(m, T, \nu)$, with $m = \theta A + B \in M = -\mathcal{S}_d^+$, where $A = -C/2$ and $B = -D/2$. We apply Corollary 3.1 and the fact that

$$K\big(\psi(\theta A + B)\big) = -\frac{1}{2} \ln \det(\theta C + D).$$

Thus, formula (3.9) holds true. ∎

Now we characterize the information $J(\theta)$ in terms of the eigenvalues of the matrix $D^{-1/2}CD^{-1/2}$.

THEOREM 3.1. *Let C and D be two symmetric matrices and let $I \subset \mathbb{R}$ be a segment such that $IC + D \subset \mathcal{S}_d^+$. Let us assume that $a_1, \dots, a_d$ are the eigenvalues of the matrix $D^{-1/2}CD^{-1/2}$. Then the Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ equals*

$$J(\theta) = \frac{1}{2} \sum_{j=1}^{d} \left(\frac{a_j}{1 + a_j \theta} \right)^2. \tag{3.10}$$

P r o o f. The idea of the proof is to use formula (3.9). Let $P(\lambda)$ be the characteristic polynomial of the matrix $D^{-1/2}CD^{-1/2}$. We have

$$\begin{aligned} P(\lambda) &= \det(D^{-1/2}CD^{-1/2} - \lambda I_n) \\ &= \det(D^{-1}C - \lambda I_n) = (\det D)^{-1} \det(C - \lambda D). \end{aligned}$$

On the other hand, $P(\lambda) = \prod_{j=1}^{n} (a_j - \lambda)$. It follows that

$$\det(\theta C + D) = \det D \times \theta^d P(-1/\theta) = \det D \prod_{j=1}^{d} (\theta a_j + 1).$$

The last formula allows us to compute easily $\frac{d^2}{d\theta^2}\big(\ln \det(\theta C + D)\big)$. First we see that

$$\frac{d}{d\theta}\big(\ln \det(\theta C + D)\big) = \frac{\frac{d}{d\theta} \det(\theta C + D)}{\det(\theta C + D)} = \sum_{j=1}^{d} \frac{a_j}{\theta a_j + 1}.$$

One more derivation and formula (3.9) lead to (3.10). ■

We finish by computing the Fisher information of two Gaussian models in $\mathbb{R}^d$, parametrized by an explicitly given segment of covariances. First, let A be a circulant matrix with the first row $e_2 + e_d = (0, 1, 0, \dots, 0, 1)$. Then for a segment $I \subset \mathbb{R}$ containing 0 and $\theta \in I$ we have

$$\theta A + I_d = \begin{pmatrix} 1 & \theta & 0 & \dots & 0 & \theta \\ \theta & 1 & \theta & 0 & \dots & 0 \\ 0 & \theta & 1 & \theta & 0 & \dots \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & \theta & 1 & \theta \\ \theta & 0 & \dots & 0 & \theta & 1 \end{pmatrix} \in \mathcal{S}_d^+. \tag{3.11}$$

COROLLARY 3.5. *The Fisher information of the model $\big(N(0, \theta A + I_d)\big)_{\theta \in I}$ is given by*

$$J(\theta) = \frac{1}{2} \sum_{j=0}^{d-1} \left(\frac{2\cos\big((2\pi j)/d\big)}{1 + 2\theta \cos\big((2\pi j)/d\big)} \right)^2. \tag{3.12}$$

P r o o f. Let $\mathcal{A}$ be a circulant matrix with the first row $(r_0, r_1, \ldots, r_{d-1})$. It is well known (see, e.g., [5]) and easy to check that if ϵ is a d-th root of unity, $\epsilon^d = 1$, then $a = \sum_{l=0}^{d-1} r_l \epsilon^l$ is an eigenvalue of $\mathcal{A}$ with an eigenvector $(1, \epsilon, \epsilon^2, \ldots, \epsilon^{d-1})$.

Therefore, if $\epsilon_j = e^{(2\pi j i)/d}$, $j = 0, \ldots, d-1$, are the d distinct d-th roots of unity, then the matrix $\mathcal{A}$ has d distinct eigenvalues $a_j = \sum_{l=0}^{d-1} r_l \epsilon_j^l$. In our particular case, we have

$$a_j = e^{(2\pi j i)/d} + e^{[2(d-1)\pi j i]/d} = 2\cos\left(\frac{2\pi j}{d}\right).$$

Formula (3.12) follows from Theorem 3.1. ■

Now, let us consider a tridiagonal matrix C such that

$$\theta C + I_d = \begin{pmatrix} 1 & \theta & 0 & 0 & 0 & \ldots \\ \theta & 1 & \theta & 0 & 0 & \ldots \\ 0 & \theta & 1 & \theta & 0 & \ldots \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & \theta & 1 & \theta \\ 0 & \ldots & 0 & 0 & \theta & 1 \end{pmatrix}. \tag{3.13}$$

As in the preceding case, there exists a segment $I \subset \mathbb{R}$ such that $\theta C + I_d \in \mathcal{S}_d^+$ for $\theta \in I$.

COROLLARY 3.6. *The Fisher information of the model* $\big(N(0, \theta C + I_d)\big)_{\theta \in I}$ *is given by*

$$J(\theta) = \frac{1}{2} \sum_{j=1}^{d} \left(\frac{2\cos\big([j/(d+1)]\pi\big)}{1 + 2\theta \cos\big([j/(d+1)]\pi\big)} \right)^2. \tag{3.14}$$

P r o o f. We will apply Theorem 3.1 with C and D equal to I_d. Expanding $\psi_d(\lambda) = \det(C - \lambda I_d)$ along the first row, we get $\psi_d(\lambda) = -\lambda \psi_{d-1}(\lambda) - M^{1,2}$. Expanding the minor $M^{1,2}$ along its first column gives $M^{1,2} = \psi_{d-2}(\lambda)$ and

$$\psi_d(\lambda) = -\lambda \psi_{d-1}(\lambda) - \psi_{d-2}(\lambda), \quad d \geqslant 3.$$

We set $\varphi_d(\lambda) = (-1)^d \psi_d(2\lambda)$ and obtain

$$\varphi_d(\lambda) = 2\lambda \varphi_{d-1}(\lambda) - \varphi_{d-2}(\lambda), \quad d \geqslant 3,$$

with initial conditions $\varphi_1(\lambda) = 2\lambda$, $\varphi_2(\lambda) = 4\lambda^2 - 1$. Therefore, φ_d is a Tchebyshev polynomial of the second kind [14] and it satisfies

$$\varphi_d(\cos x) = \frac{\sin(d+1)x}{\sin x}, \quad d \geqslant 1.$$

We have, for all $\lambda \in [-2, 2]$,

$$\left[\psi_d(\lambda) = 0 \Leftrightarrow \varphi_d\left(\frac{\lambda}{2}\right) = 0\right] \implies \frac{\sin(d+1)x}{\sin x} = 0, \quad x = \arccos\frac{\lambda}{2}.$$

Therefore, $\lambda_j = 2\cos\big([j/(d+1)]\pi\big)$, $1 \leqslant j \leqslant d$, are d distinct eigenvalues of the matrix C. ■

3.2. Exponential families of Wishart distributions. Wishart distributions on the cone $\overline{\mathcal{S}_d^+}$ are defined as elements of natural exponential families generated by Riesz measures (see [4]). Recall that the Riesz measures μ_p on the cone $\overline{\mathcal{S}_d^+}$ are unbounded positive measures such that their Laplace transform equals, for $t \in \mathcal{S}_d^+$,

$$\mathcal{L}\,\mu_p(t) = \int_{\overline{\mathcal{S}_d^+}} e^{-\langle t,x\rangle} d\mu_p(x) = (\det t)^{-p}.$$

By the celebrated Gindikin theorem, such measures exist if and only if p belongs to the Gindikin set $\Lambda_d = \{1/2, \ldots, (d-1)/2\} \cup \big((d-1)/2, \infty\big)$. Their support is equal to the cone $\overline{\mathcal{S}_d^+}$ if and only if $p > (d-1)/2$ and they are absolutely continuous in that case with a density $\Gamma_d(p)^{-1}(\det x)^{p-(d+1)/2}$, $x \in \mathcal{S}_d^+$, where

$$\Gamma_d(p) = \Gamma(p)\Gamma\left(p - \frac{1}{2}\right)\ldots\Gamma\left(p - \frac{d-1}{2}\right).$$

Otherwise, when $p \in \{1/2, \ldots, (d-1)/2\}$, the measures μ_p are singular and concentrated on semi-positive symmetric matrices of rank $2p$.

The family of Wishart distributions $W(p; s)$ on the cone $\overline{\mathcal{S}_d^+}$ is defined as the natural exponential family generated by the Riesz measure μ_p. It means that $p \in \Lambda_d$, $s \in S = -\mathcal{S}_d^+$ and

$$\begin{aligned} W(p; s)(dx) &= \frac{e^{\langle s\,,\,x\rangle}}{\mathcal{L}\,\mu_p(-s)}\mu_p(dx) \\ &= e^{\langle s,\,x\rangle}\big(\det(-s)\big)^p\mu_p(dx) = e^{\langle s\,,\,x\rangle - K_p(s)}\mu_p(dx) \end{aligned}$$

with $K_p(s) = -p\ln\det(-s)$. It follows that

$$\mathcal{L}\,W(p; s)(t) = \det\big(I_d + (-s)^{-1}t\big)^{-p} \quad \text{and} \quad \mu_p(dx) = e^{\mathrm{Tr}\,x}W(p; -I_d).$$

Wishart distributions are multivariate analogs of the gamma distributions $\lambda^p\Gamma(p)^{-1}e^{-\lambda x}x^{p-1}dx$ on $\mathbb{R}^+$ $(p > 0, \lambda > 0)$, considered with a canonical parameter $s = -\lambda < 0$. As in dimension one, the Wishart distributions are often parametrized by a scale parameter $\sigma = (-s)^{-1} \in \mathcal{S}_d^+$ and then the notation $\gamma(p; \sigma) = W\big(p; (-\sigma)^{-1}\big)$ is used, cf. [11]. The study of Wishart distributions is motivated by their importance as estimators of the covariance matrix of a Gaussian model in $\mathbb{R}^d$.

Let us apply our results on the Fisher information to a natural exponential family of Wishart distributions $\{W(p;s) : \ s \in -\mathcal{S}_d^+\}$. The mean equals $m(s) = K_p'(s) = p(-s)^{-1} \in M = \mathcal{S}_d^+$ and the inverse mean map $\psi : \mathcal{S}_d^+ \to -\mathcal{S}_d^+$ is of the form $\psi(m) = -pm^{-1}$.

Thus the Wishart family $Q(m, \mu_p)$ parametrized by the domain of means is, up to a trivial reparametrization $m \to p^{-1}m$, the family parametrized by its scale parameter:

$$Q(m, \mu_p) = W(p; -pm^{-1}) = \gamma(p; p^{-1}m), \qquad m \in \mathcal{S}_d^+. \tag{3.15}$$

As $v(s) = K_p''(s) = p(s^{-1} \otimes s^{-1})$, it follows that the variance function is

$$V(m) = \frac{1}{p}(m \otimes m). \tag{3.16}$$

The Fisher information of the model $\{W(p;s) : \ s \in -\mathcal{S}_d^+\}$ is

$$I(s) = ps^{-1} \otimes s^{-1}.$$

By Theorem 2.1 the Fisher information of the model $\{Q(m, \mu_p), \ m \in M\}$ is equal to $J(m) = pm^{-1} \otimes m^{-1}$.

Consequently, using the reparametrization $m \to p^{-1}m = \sigma$ and the chain rule, we see that the Fisher information matrix of the Wishart model $\{\gamma(p;\sigma) : \sigma \in \mathcal{S}_d^+\}$ parametrized by the scale parameter σ equals $J(\sigma) = p\sigma^{-1} \otimes \sigma^{-1}$.

THEOREM 3.2. *Let $I = (a, b) \subset \mathbb{R}$ and $C, D \in \mathcal{S}_d$ be such that the relation $IC + D \subset \mathcal{S}_d^+$ is satisfied. Then the Fisher information $J(\theta)$ of the Wishart model $\{\gamma(p; \theta C + D) : \ \theta \in I\}$ satisfies the formulas*

$$J(\theta) = p \operatorname{Tr} \left(C(\theta C + D)^{-1}\right)^2, \tag{3.17}$$

$$J(\theta) = -p\frac{d^2}{d\theta^2}\big(\ln \det(\theta C + D)\big),$$

$$J(\theta) = p \sum_{j=1}^{d} \left(\frac{a_j}{1 + a_j\theta}\right)^2, \tag{3.18}$$

where $a_1, \ldots, a_d$ are the eigenvalues of the matrix $D^{-1/2}CD^{-1/2}$.

P r o o f. The proofs are similar to the proofs of the analogous results for exponential Gaussian families in the previous subsection. The condition (3.3) holds true: $\langle m\, ,\, \psi(m)\rangle = -pd$, the model $\{\gamma(p; \theta C + D) : \ \theta \in I\}$ is equal to the model $\{Q(\theta pC + pD, \mu_p) : \ \theta \in I\}$ parametrized by the means and we have

$$K_p\big(\psi(\theta pC + pD)\big) = p \log \det(\theta C + D). \quad \blacksquare$$

COROLLARY 3.7. *Let $\sigma_1, \sigma_2 \in \mathcal{S}_d^+$ and let I be the open interval containing θ such that $\sigma_\theta = \theta\sigma_1 + (1-\theta)\sigma_2 \in \mathcal{S}_d^+$. The Fisher information of the model $\{\gamma(p;\sigma_\theta) :\ \theta \in I\}$ is equal to $J(\theta) = p\,\mathrm{Tr}\left(\left((\sigma_1-\sigma_2)\,\sigma_\theta^{-1}\right)^2\right)$.*

Proof. We write $\theta\sigma_1 + (1-\theta)\sigma_2 = \theta(\sigma_1-\sigma_2)+\sigma_2$ and then we apply the formula (3.17). ∎

Using (3.18) we obtain the following corollary, analogous to Corollaries 3.5 and 3.6.

COROLLARY 3.8. 1. *Let us consider the model $\{\gamma(p;\theta A + I_d) :\ \theta \in I\}$ with $\theta A + I_d$ as in* (3.11). *Then its Fisher information equals*

$$J(\theta) = p\sum_{j=0}^{d-1}\left(\frac{2\cos\left((2\pi j)/d\right)}{1+2\theta\cos\left((2\pi j)/d\right)}\right)^2.$$

2. *Let us consider the model $\{\gamma(p;\theta C + I_d) : \theta \in I\}$ with $\theta C + I_d$ as in* (3.13). *Then its Fisher information equals*

$$J(\theta) = p\sum_{j=1}^{d}\left(\frac{2\cos\left([j/(d+1)]\pi\right)}{1+2\theta\cos\left([j/(d+1)]\pi\right)}\right)^2.$$

REMARK 3.1. *Let $P_s(\mu)$ be the natural exponential family corresponding to the Gaussian general exponential family* (3.6). *If W has the law $N(u, s^{-1})$ given by* (3.6), *then $T(W)$ has the law $P_s(\mu)$. On the other hand, it is well known that $-T(W) = \frac{1}{2}(W-u)(W-u)^T$ has the Wishart law $\gamma\left(\frac{1}{2}; 2s^{-1}\right)$. This explains why the formulas for the Fisher information are the same for the Gaussian family and for the Wishart family with $p = \frac{1}{2}$.*

3.2.1. *Exponential families of non-central Wishart distributions.* Let us complete the section on the Wishart models by considering the non-central case. The main reference is [11]. Let $p \in \Lambda_d$, $a \in \overline{\mathcal{S}_d^+}$ and $\sigma \in \mathcal{S}_d^+$. The non-central Wishart distribution $\gamma(p, a; \sigma)$ is defined by its Laplace transform

$$\begin{aligned}\mathcal{L}\,\gamma(p,a;\sigma)(t) &= \int_{\overline{\mathcal{S}_d^+}} e^{-\mathrm{Tr}(tx)}\gamma(p,a;\sigma)(dx)\\ &= \det(I_d+\sigma t)^{-p}e^{-\mathrm{Tr}(t(I_d+\sigma t)^{-1}\sigma a\sigma)}\end{aligned}$$

for all $t \in \mathcal{S}_d^+$. If $p \geqslant (d-1)/2$, then non-central Wishart laws exist for all $a \in \overline{\mathcal{S}_d^+}$; if $p \in \{1/2, \ldots, (d-2)/2\}$, then a must be of rank at most $2p$ (see [12]). If $p = n/2$, $n \in \mathbb{N}$, then the non-central Wishart distributions are constructed in the following way from n independent d-dimensional Gaussian vectors $Y_1, \ldots, Y_n$.

Let $Y_j \sim N(m_j\,, \Sigma)$ and let M be the $d \times n$ matrix $[m_1, \ldots, m_n]$. Then the $d \times d$ matrix $W = Y_1Y_1^T + \ldots + Y_nY_n^T$ has the non-central Wishart distribution

$\gamma(p,a;\sigma)$ with $p=n/2$, $\sigma=2\Sigma$ and $\sigma a\sigma = MM^T$. Such Wishart distributions are studied in [15].

The non-central Wishart distributions may be constructed as a natural exponential family $\{W(p,a;s):\ s\in -\mathcal{S}_d^+\}$ generated by the positive measure $\mu = \mu_{a,p}(dx) = e^{\mathrm{Tr}(a+x)}\gamma(p,a;I_d)(dx)$. Its moment generating function is given for $s\in -\mathcal{S}_d^+$ by

$$\int_{\mathcal{S}_d^+} e^{\mathrm{Tr}(sx)}\mu_{a,p}(dx) = \det(-s)^{-p}e^{\mathrm{Tr}(a(-s)^{-1})}.$$

We have $W(p,a;s)=\gamma\big(p,a;(-s)^{-1}\big)$. Like for central Wishart families, we have $S=-\mathcal{S}_d^+$. The cumulant function is

$$K(s) = -p\log\det(-s) + \mathrm{Tr}\left(a(-s)^{-1}\right).$$

As before, we put $\sigma=(-s)^{-1}$. We see that the mean equals

$$m(s) = K'(s) = p(-s)^{-1} + (-s)^{-1}a(-s)^{-1} = p\sigma + \sigma a\sigma \tag{3.19}$$

and the covariance is of the form

$$\begin{aligned} v(s) = K''(s) &= p\sigma\otimes\sigma + (\sigma a\sigma)\otimes\sigma + \sigma\otimes(\sigma a\sigma) \\ &= -p\sigma\otimes\sigma + m\otimes\sigma + \sigma\otimes m. \end{aligned} \tag{3.20}$$

When the matrix a is non-singular, the inverse mean map $\psi(m)=s$ is such that

$$(-s)^{-1} = \sigma = -\frac{p}{2}a^{-1} + a^{-1/2}\left(a^{1/2}ma^{1/2} + \frac{p^2}{4}I_d\right)^{1/2} a^{-1/2}. \tag{3.21}$$

For other cases see [11], Proposition 4.5. In order to write the variance function $V(m)=v\big(\psi(m)\big)$ we compose the last expression from (3.20) and the formula (3.21).

For a model $\big\{W\big(p,a;\psi(\theta A+B)\big) : \theta\in I\big\}$ parametrized by a segment of means, the Fisher information $J(\theta)$ is obtained from the expression of $V(m)$ and Theorem 2.1.

EXAMPLE 3.1. Suppose that $a=I_d$, $A=\alpha I_d$ and $B=\beta I_d$, $\alpha,\beta>0$. The Fisher information on θ is

$$J(\theta) = \alpha^2 d\left((p^2+2\theta\alpha+2\beta)\left(\theta\alpha+\beta+\frac{p^2}{4}\right)^{1/2} - 2p(\theta\alpha+\beta) - \frac{p^3}{2}\right)^{-1}.$$

3.3. Applications to estimation of the mean in exponential families parametrized by a segment of means. Consider a sample $X_1,\dots,X_n$ of a random variable X from a natural exponential family $Q(m,\mu)$ parametrized by the domain of means M, where the parameter $m=\mathbb{E}\,X$ is unknown and M is open. The following qualities of the sample mean $\bar X_n$ as an estimator of m are known; for the sake of completeness we provide a short proof of properties which are less evident.

PROPOSITION 3.1. *The sample mean $\bar{X}_n$ is an unbiased, consistent and efficient estimator of the parameter m. It is a maximum likelihood estimator of m.*

P r o o f. By Theorem 2.1 we have $\operatorname{Cov} X = V(m) = J(m)^{-1}$, so the Cramér–Rao bound is attained by X. Consequently, the sample mean $\bar{X}_n$ is an efficient estimator of m. It follows from (2.7) that the sample mean $\bar{X}_n$ is a maximum likelihood estimator of m. One can also first show by (2.2) that the maximum likelihood estimator of s is $\hat{s} = (K')^{-1}(X) = \psi(X)$ and next use the functional invariance of the maximum likelihood estimator (see [3], Theorem 7.2.10). ∎

REMARK 3.2. *For general exponential families $Q(m, T, \nu)$ parametrized by an open domain of means M, all these properties remain valid for $\hat{m} = \overline{T(X)}_n$ as an estimator of $m = \mathbb{E}\, T(X)$.*

Consider an exponential family $Q(\theta A + B, \mu)$ parametrized by a segment of means $IA + B \subset M$ with $A \neq 0, B \in E$ and $\theta \in I$, a segment in $\mathbb{R}$. We will now discuss estimators of the real parameter θ when we know that the mean is of the form $\mathbb{E}\, X = m \in IA + B$.

The segment $IA + B \subset M$ is of dimension one and has an empty interior in M. That is why the efficiency and maximum likelihood properties of the estimator $\hat{m} = \bar{X}_n$ are not automatically inherited by natural estimators of the real parameter θ. Determining a maximum likelihood estimator for θ seems impossible explicitly. This is the "price to pay" for the parsimony of the segment model parametrized by $m \in IA + B$. On the other hand, the efficiency of estimators of θ may be studied thanks to Theorem 2.1 and its corollaries.

Knowing that

$$m = \theta A + B \tag{3.22}$$

for a value $\theta \in I$, we have many possibilities of writing down a solution θ of equation (3.22). If $A \neq 0$ then the solution θ is unique ($A\theta + B = A\theta' + B$ implies $\theta = \theta'$ when $A \neq 0$). For any C such that $\langle A, C\rangle \neq 0$ we have

$$\theta = \frac{\langle m - B, C\rangle}{\langle A, C\rangle}.$$

We define an estimator $\hat{\theta}_C$ of the parameter θ by

$$\hat{\theta}_C = \frac{\langle \bar{X}_n - B, C\rangle}{\langle A, C\rangle}.$$

All the estimators $\hat{\theta}_C$ are unbiased and consistent. The natural question is whether they are efficient. The variance of $\hat{\theta}_C$ may be computed by using the variance function $V(m)$ of the exponential family:

$$\operatorname{Var} \hat{\theta}_C = \frac{1}{\langle A, C\rangle^2} \operatorname{Var}\langle \bar{X}_n, C\rangle = \frac{\operatorname{Vec}(C)^T V(\theta A + B) \operatorname{Vec}(C)}{n\langle A, C\rangle^2}. \tag{3.23}$$

On the other hand, the Cramér–Rao bound is equal, by Theorem 2.1, to

$$\frac{1}{nJ(\theta)} = \frac{1}{n\,\mathrm{Vec}(A)^T\,V(\theta A+B)^{-1}\,\mathrm{Vec}(A)}. \tag{3.24}$$

When the state space is a square matrix space $\mathbb{R}^{d\times d}$ and the matrix A is invertible, we can take $C = A^{-1}$ and consider the estimator

$$\hat{\theta}_{A^{-1}} = \frac{\langle \bar{X}_n - B, A^{-1}\rangle}{d}.$$

The following theorem shows that for Gaussian and central Wishart exponential families and for linearly dependent A and B the estimator $\hat{\theta}_{A^{-1}}$ is efficient as an estimator of the mean m (with X_i replaced by $T(X_i) = -\frac{1}{2}(X_i-u)(X_i-u)^T$ in the Gaussian case). In conclusion, we obtain efficient estimators for Gaussian models parametrized by a covariance segment parameter and for Wishart models parametrized by a scale segment parameter.

THEOREM 3.3. 1. *Let $I \subset \mathbb{R}^+$ be a non-empty segment. Let $c \geqslant 0$, $A \in \mathcal{S}_d^+$ and $B = cA$.*

(1a) *Consider an n-sample $(X_1,\ldots,X_n)$ from a Gaussian family $Q(m,T,\nu)$ defined by* (3.7), *where $m = \theta A + B$, $\theta \in I$. Then*

$$\hat{\theta}_{A^{-1}} = \frac{\langle \overline{T(X)}_n - B, A^{-1}\rangle}{d}$$

is an unbiased efficient estimator of the parameter θ.

(1b) *Consider an n-sample $(X_1,\ldots,X_n)$ from a Wishart model $Q(m,\mu_p)$ defined by* (3.15), *where $m = \theta A + B$, $\theta \in I$. Then*

$$\hat{\theta}_{A^{-1}} = \frac{\langle \bar{X}_n - B, A^{-1}\rangle}{d}$$

is an unbiased efficient estimator of the parameter θ.

2. *Let $c \geqslant 0$, $C \in \mathcal{S}_d^+$ and $D = cC$.*

(2a) *Let us consider an n-sample $(X_1,\ldots,X_n)$ from a Gaussian model $\{N(u,\theta C + D), \theta \in I\}$ parametrized by a segment of covariances. An unbiased efficient estimator of θ is given by*

$$\hat{\theta} = \frac{1}{d}\left\langle \frac{1}{n}\sum_{i=1}^{n}(X_i-u)(X_i-u)^T - D, C^{-1}\right\rangle.$$

(2b) *Let us consider an n-sample $(X_1,\ldots,X_n)$ from a Wishart model $\{\gamma(p,\theta C + D), \theta \in I\}$ parametrized by a segment of scale parameters. Un unbiased efficient estimator of θ is given by*

$$\hat{\theta} = \frac{\langle p^{-1}\bar{X}_n - D, C^{-1}\rangle}{d}.$$

P r o o f. For the first part of the theorem, we give the proof in the Wishart case. The proof in the Gaussian case is identical, with $p = \frac{1}{2}$, cf. Remark 3.1. By formulas (3.23) and (3.16) we have

$$\operatorname{Var} \hat{\theta}_{A^{-1}} = \frac{1}{pd^2 n} \operatorname{Tr}\big((A\theta + B)A^{-1}(A\theta + B)A^{-1}\big) = \frac{(\theta + c)^2}{pdn}.$$

On the other hand, by (3.24) and (3.16), we obtain

$$\frac{1}{nJ(\theta)} = \frac{1}{np \operatorname{Tr}\big(A(A\theta + B)^{-1}A(A\theta + B)^{-1}\big)} = \frac{1}{np(\theta + c)^{-2}d}.$$

Thus $\operatorname{Var} \hat{\theta} = 1/\big(nJ(\theta)\big)$ and the estimator $\hat{\theta}_{A^{-1}}$ is efficient.

The second part of the theorem follows by necessary reparametrizations. For (2a), using (3.7), we write $\theta C + D = -2m$ with $m = \theta A + B$, where $A = -C/2$ and $B = -D/2$. The part (2b) follows similarly from (3.15). ∎

REMARK 3.3. *It is an open question whether $\hat{\theta}_{A^{-1}}$ may be efficient for independent A and B. Let $n = 1$. The equality $\operatorname{Var} \hat{\theta} = 1/J(\theta)$ holds if and only if, writing $D_\theta = (A\theta + B)A^{-1}(A\theta + B)A^{-1}$, the equality $d^{-2} \operatorname{Tr}(D_\theta) = 1/\operatorname{Tr}(D_\theta^{-1})$ holds for all $\theta \in I$.*

Acknowledgments. The authors would like to thank Gérard Letac for his insightful suggestions and the anonymous referee for numerous comments that strongly improved the introductory part of the paper.

REFERENCES

[1] T. Andersson and P. Händel, *IEEE Standard 1057. Cramér–Rao bound and the parsimony principle*, IEEE Transactions on Instrumentation and Measurement 55 (1) (2006), pp. 44–53.

[2] J.-F. Bercher and C. Vignat, *On minimum Fisher information distributions with restricted support and fixed variance*, Inform. Sci. 179 (2009), pp. 3832–3842.

[3] G. Casella and R.L. Berger, *Statistical Inference*, Duxbury Advances Series, 2002.

[4] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford Math. Monogr., Oxford University Press, New York 1994.

[5] R. M. Gray, *Toeplitz and Circulant Matrices: A review*, Foundations and Trends in Communications and Information Theory, Vol. 2, Issue 3, NOW Publishers, 2006, pp. 155–239.

[6] R. D. Gupta and D. Kundu, *On the comparison of Fisher information of the Weibull and GE distributions*, J. Statist. Plann. Inference 136 (2006), pp. 3130–3144.

[7] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Texts Statist., Springer, New York 1998.

[8] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer Texts Statist., Springer, New York 2005.

[9] G. Letac, *Lectures on Natural Exponential Families and Their Variance Functions*, Monogr. Mat., Vol. 50, IMPA, Rio de Janeiro 1992.

[10] G. Letac and M. Casalis, *Natural exponential families*, in: *Continuous Multivariate Distributions. Volume 1: Models and Applications*, second edition, S. Klotz, N. Balakrishnan, and N. L. Johnson (Eds.), Wiley Ser. Probab. Stat., 2000.

[11] G. Letac and H. Massam, *The noncentral Wishart as an exponential family, and its moments*, J. Multivariate Anal. 99 (7) (2008), pp. 1393–1417.

[12] G. Letac and H. Massam, *Existence and non-existence of the noncentral Wishart distributions*, preprint, 2011.

[13] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, 2007.

[14] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*, Chapman & Hall/CRC, 2003.

[15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, Hoboken, New Jersey, 2005.

[16] V. Rico-Ramirez, P. A. Quintana-Hernandez, S. Hernandez-Castro, and U. M. Diwekar, *Fisher information as a novel tool for process control applications*, in: *20th European Symposium on Computer Aided Process Engineering – ESCAPE20*, S. Pierucci and G. Buzzi Ferraris (Eds.), Elsevier B.V., 2010.

Piotr Graczyk
LAREMA, Université d'Angers
2 Bd Lavoisier, 49045 Angers Cedex 01, France
E-mail: piotr.graczyk@univ-angers.fr

Salha Mamane
School of Statistics and Actuarial Science
University of the Witwatersrand
and LAREMA, Université d'Angers
Private Bag x3, Wits 2050, South Africa
E-mail: Salha.Mamane@wits.ac.za

Received on 6.2.2014;
revised version on 7.5.2014