A CRAMÉR–RAO TYPE INEQUALITY FOR RANDOM VARIABLES IN EUCLIDEAN MANIFOLDS

by

H. Hendriks *and* F.H. Ruymgaart[*)]

Katholieke Universiteit Nijmegen

[*)] Part of this work was done while the author was visiting the Department of Statistics, UNC, Chapel Hill.

ABSTRACT

In this note the Cramér–Rao type inequality for estimators with values in an abstract manifold in Hendriks (1987) in specialized to manifolds in $\mathbb{R}^k$. In passing some useful concepts regarding location and dispersion and the potential use of the exponential mapping are discussed. The examples focus on the special orthogonal group that plays a role in Chang (1986), with the circle (see Mardia (1975)) and torus as special cases.

AMS 1980 subject classification: primary 62H12, secondary 22E99.

Key words and phrases: Cramér–Rao inequality, manifold, special orthogonal group, circle, torus.

1. INTRODUCTION: SOME BASIC CONCEPTS

In Hendriks (1987) a Cramér–Rao type lower bound has been established for estimators with values in an abstract manifold. Here we will give some more examples by first specializing to manifolds that are subsets of $\mathbb{R}^k$, $k \in \mathbb{N}$, referred to as Euclidean manifolds in the title. This restriction entails that the Euclidean metric and ordinary (componentwise) differentiation of (vector valued) functions can be used, so that an elementary formulation and proof of the inequality can be given; see Section 2. On the other hand this special case is still general enough to cover most situations that are of practical importance. In Section 3 we will in particular pay attention to the special orthogonal group, that plays a role in Chang (1986). A special case is the circle (Mardia (1975)) and the m–tori are special subgroups.

For the formulation of the inequality we need to briefly review some concepts in Hendriks (1987) in the present setting and using a different terminology. In Subsections 1.2 and 1.3 we introduce the mean location and divergence respectively of a random variable with values in a manifold in $\mathbb{R}^k$. These concepts are of general importance and they coincide with the expectation and variance respectively if the manifold is a convex set. The mean location, moreover, specializes to the mean direction as defined for the circle and the sphere (Mardia (1975), Watson (1983)). The divergence is related to Mardia's (1975) concept of divergence for the circle, as will be shown below. In Subsection 1.1 we will briefly discuss the exponential mapping as a tool to construct examples of probability measures and to deal with weak convergence on certain classes of manifolds, like the special orthogonal group in particular.

1.1. EXPONENTIAL MAP AND WRAPPING. Given the tangent space at a certain point of the manifold, the exponential mapping maps a neighborhood in the tangent space onto a neighborhood of the point on the manifold, such that the mapping is one–to–one and distances along geodesics are preserved. Although the mapping is defined only locally, in some cases it is defined on the entire tangent space and maps onto the entire manifold, where the one–to–one and distance preserving character is usually lost in the global sense. In such cases probability measures on the linear tangent space are naturally transformed into probabilities on the manifold. Moreover, such linear operations on elements in the tangent space as are often needed for asymptotic normality, result in an element that can be mapped in the manifold. In such a way weak convergence on the manifold might be defined. Here we will not be concerned with asymptotic results, however.

EXAMPLE 1.1 Let $\mathcal{M} = \mathcal{S}^1 = \{(x_1,x_2) \in \mathbb{R}^2;\ x_1^2 + x_2^2 = 1\}$. An example of such a global exponential mapping is

$$h \mapsto \mathrm{Exp}(h) = (\cos h, \sin h),\ h \in \mathbb{R}, \tag{1.1}$$

which might of course be identified with exp(i h), i imaginary unit. Given a random variable $U \overset{d}{=} \mathcal{N}(\mu,\sigma^2)$, Exp(U) defines a probability measure which is called the wrapped normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in (0,\infty)$. In Mardia (1975) the definition is essentially the same, but the relation with the exponential mapping is not mentioned. It is clear that in polar coordinates $(\cos \varphi, \sin \varphi)$, $\varphi \in (0,2\pi]$, the density of this wrapped normal distribution is given by

$$(1.2) \qquad \tilde{f}_{\mu,\sigma}(\varphi) = (2\pi\,\sigma^2)^{-\frac{1}{2}} \,\Sigma_{k=-\infty}^{\infty} \exp\left[-\tfrac{1}{2}\left[\frac{\varphi-\mu+2\,\pi\,k}{\sigma}\right]^2\right] =$$

$$= (2\pi)^{-1}\left[1 + 2\,\Sigma_{k=1}^{\infty} \exp(-\tfrac{1}{2}(k\sigma)^2)\cos k\,\varphi\right],\ 0 < \varphi \le 2\pi\,.$$

For this useful equality see Mardia (1975, formula (3.4.30)).

Now suppose that $\Phi_1,\ldots,\Phi_n$ are i.i.d. real valued random variables with finite expectation μ and finite non–zero variance σ^2. Hence $\mathrm{Exp}(\Phi_i) = (\cos\Phi_i, \sin\Phi_i)$ are i.i.d. random variables in S^1. To determine a suitable sum with a limiting law on S^1 it seems natural to first take the usual partial sum of the Φ_i in the tangent space, and then map back onto S^1 using Exp. Hence we consider

$$(1.3) \qquad \mathrm{Exp}\left[\frac{\Sigma_{i=1}^n(\Phi_i-\mu)}{n^{\frac{1}{2}}\,\sigma}\right] = \left[\cos\left[\frac{\Sigma_{i=1}^n(\Phi_i-\mu)}{n^{\frac{1}{2}}\,\sigma}\right], \sin\left[\frac{\Sigma_{i=1}^n(\Phi_i-\mu)}{n^{\frac{1}{2}}\,\sigma}\right]\right],$$

which is equivalent to the random variable considered in Mardia (1975, Section 4.3.2b). It is clear that the random variables in (1.3) converge in distribution to the wrapped normal distribution with $\mu = 0$ and $\sigma = 1$.

More generally the special orthogonal group admits an onto exponential mapping. In principle for such manifolds probability distributions and weak convergence may be defined with the aid of the usual theory for random vectors in linear spaces.

1.2. MEAN LOCATION. Let $\mathcal{M} \subset \mathbb{R}^k$ be a smooth manifold of dimension $1 \le m \le k$, let $(\mathcal{X},\mathcal{A},P)$ be a probability space and $S : \mathcal{X} \to \mathcal{M}$ a random variable in the manifold. The *mean location* of S is defined to be the point $M(S) \in \mathcal{M}$, unique by assumption, satisfying

$$(1.4) \qquad E\|S - M(S)\|^2 = \inf_{p \in \mathcal{M}} E\|S-p\|^2\,.$$

Uniqueness is typically satisfied when the law of S on $\mathcal{M}$ is unimodal.

The expectation of $S = (S_1,\dots,S_k)$ is defined in the usual way by $E(S) = (E(S_1),\dots,E(S_k))$ and hence satisfies

(1.5) $$E\|S-E(S)\|^2 = \inf_{x\in\mathbb{R}^k} E\|S-x\|^2 .$$

It is clear that $M(S) = E(S)$ when $\mathscr{M}$ is a convex set.

EXAMPLE 1.2. Let again $\mathscr{M} = \mathscr{S}^1 \subset \mathbb{R}^2$ and $S : \mathscr{X} \to \mathscr{S}^1$ be given by $S = (\cos\Phi, \sin\Phi)$ for $\Phi : \mathscr{X} \to \mathbb{R}$. Then Mardia (1975, formula (3.37)) defines the mean direction as μ, where $\mu \in (0,2\pi]$ and $r \geq 0$ are given by $(E\cos\Phi, E\sin\Phi) = r(\cos\mu, \sin\mu)$. This yields

(1.6) $$((E\cos\Phi)^2 + (E\sin\Phi)^2)^{-\frac{1}{2}}(E\cos\Phi, E\sin\Phi) \in \mathscr{S}^1,$$

as the corresponding point on the circle (assuming $r > 0$). It is easily seen that (1.6) minimizes (1.4), so that the point coincides with the mean location.

Let us introduce the modified Bessel function

(1.7) $$I_m(\kappa) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\kappa)^{2k+m}}{\Gamma(m+k+1)\Gamma(k+1)}, \quad \kappa \in \mathbb{R}, m \in \mathbb{Z},$$

where Γ denotes the gamma–function and $\mathbb{Z}$ the set of all integers. It is shown in Mardia (1975) that unimodal symmetric distributions on the circle have the symmetry point as mean direction. Hence it follows that for both the wrapped normal distribution in (1.2) and the von Mises distribution with density

(1.8) $$f_{\mu,\kappa}(\varphi) = \frac{\exp(\kappa\cos(\varphi-\mu))}{2\pi I_0(\kappa)}, \quad 0 < \varphi \leq 2\pi,$$

where $\kappa > 0$, $0 < \mu \leq 2\pi$, the mean direction is μ and hence $(\cos\mu, \sin\mu)$ the mean location.

1.3. DIVERGENCE. Let $\mathcal{T}_{M(S)}(\mathcal{M})$ be the tangent space to $\mathcal{M}$ at M(S). We may and will identify this tangent space with an m–dimensional subspace of $\mathbb{R}^k$. For arbitrary non–zero $h \in \mathcal{T}_{M(S)}(\mathcal{M})$ we define the *divergence* of S in the direction h by

$$(1.9) \qquad D_h(S) = E < S - M(S),\ h >^2 ,$$

where $<\cdot,\cdot>$ denotes inner product in $\mathbb{R}^d$ for any dimension $d \in \mathbb{N}$.

When $\mathcal{M}$ is a convex set this concept coincides with the ordinary variance in a multivariate setting, since M(S) = E(S) so that

$$(1.10) \qquad D_h(S) = \text{Var } \Sigma_{j=1}^k h_j S_j .$$

Values of S for which S–M(S) is almost perpendicular to h don't contribute very much to the value of the divergence. Such points may be found on closed compact manifolds without boundary, like $\mathcal{S}^{k-1}$, in a neighborhood around the point across from M(S). But even then it is a reasonable measure for dispersion when the law of S is concentrated about M(S).

EXAMPLE 1.3. In the notation of Example 1.2 it follows by elementary geometry that (h=1)

$$(1.11) \qquad D_1(S) = E \sin^2(\Phi-\mu) = 1 - E \cos^2(\Phi-\mu),$$

where $\mu \in (0,2\pi]$ is determined by $M(S) = (\cos \mu, \sin \mu)$. In Mardia (1975, formula (3.3.9)) the circular variance is defined by

$$(1.12) \qquad V_0(S) = 1 - E \cos (\Phi-\mu),$$

and, apart from conditioning, the divergence (Madia (1975, formula (5.1.6))) as

(1.13) $$\operatorname{div}(S,\mu) = \frac{E\sin^2(\Phi-\mu)}{(E\cos(\Phi-\mu))^2}\,.$$

Let us now assume that S has the distribution with density in (1.7). It follows from Mardia (1975, formulas (5.1.9) and (3.4.47)) that

(1.14) $$D_1(S) = I_1(\kappa)/(\kappa\, I_0(\kappa))\ ,$$

and $V_0(S) = 1 - I_1(\kappa)/I_0(\kappa)$ respectively.

When S has the wrapped normal distribution with density (1.2) we have, writing $\exp(-\frac{1}{2}\sigma^2) = \rho$,

(1.15) $$\begin{aligned} D_1(S) &= 1 - \int_0^{2\pi} \cos^2(\varphi-\mu)\ \tilde{f}_{\mu,\sigma}(\varphi)d\varphi = \\ &= 1 - \int_0^{2\pi} \frac{\cos^2\varphi}{2\pi} \left[1 + 2\,\Sigma_{k=1}^{\infty}\ \rho^{k^2}\ \cos k\ \varphi\right] d\varphi = \\ &= 1 - \tfrac{1}{2} - (1/\pi)\rho^4\ (1/2)\pi = \tfrac{1}{2}(1-\rho^4)\ . \end{aligned}$$

According to Mardia (1975, formula (3.4.32)) the circular variance equals $V_0(S) = 1-\rho$.

2. THE INEQUALITY

Let $(\mathscr{X},\mathscr{A}, \{P_\theta, \theta \in \Theta\})$ be a statistical model with $\{P_\theta, \theta \in \Theta\} << \nu$ for some σ–finite measure ν on $(\mathscr{X},\mathscr{A})$ and $\Theta \subset \mathbb{R}^p$ open $(p \in \mathbb{N})$. Let us introduce

(2.1) $$f_\theta = d\,P_\theta/d\nu,\ \ \ell_\theta = \log f_\theta\ ;\ \ \theta \in \Theta.$$

Throughout $t \in \Theta$ will be a fixed parameter point. We may think of Θ as an open neighborhood of t. The function $\theta \mapsto \ell_\theta$, $\theta \in \Theta$, is supposed to have continuous partial derivatives, and the directional derivative at $\theta \in \Theta$ in the direction $\xi \in \mathbb{R}^p$, will be written $\nabla_\xi\, \ell_\theta$. It will be assumed that

(2.1) $\quad I_{\xi,\xi}(t) = E_t(\nabla_\xi \ell_t)^2 > 0.$

For a smooth m–dimensional manifold $\mathcal{M}$ $(1 \le m \le k)$ we are given a statistic $S : \mathcal{X} \to \mathcal{M}$, for which the expectation $E_\theta(S)$, the covariance matrix $\Sigma_\theta(S)$, and the mean location $M_\theta(S)$ exist for all $\theta \in \Theta$. The components of the vector–function $\theta \mapsto E_\theta(S)$ and $\theta \mapsto M_\theta(S)$, $\theta \in \Theta$, are supposed to have continuous partial derivatives and the k–vectors of directional derivatives at θ in the direction ξ will be written $\nabla_\xi E_\theta(S)$ respectively $\nabla_\xi M_\theta(S)$.

For each $\theta \in \Theta$ let $h_\theta \in \mathcal{T}_{M_\theta(S)}(\mathcal{M})$ be non–zero and such that the components have continuous partial derivatives on Θ. The m–vector of directional derivatives at θ in the direction ξ will be denoted by $\nabla_\xi h_\theta$.

THEOREM 2.1. *Under the conditions mentioned above we have*

$$(2.2)\quad D_{t,h_t}(S) \ge \frac{\{<\nabla_\xi M_t(S), h_t> + <M_t(S)-E_t(S), \nabla_\xi h_t>\}^2}{I_{\xi,\xi}(t)}.$$

PROOF. Sincwe $h_\theta \in \mathcal{T}_{M_\theta(S)}(\mathcal{M})$ there exists a smooth curve $\gamma_\theta : (-\epsilon,\epsilon) \to \mathcal{M}$, $\epsilon > 0$, with $\gamma_\theta(0) = M_\theta(S)$ and $\gamma_\theta'(0) = h_\theta$. The definition of mean location entails, moreover, that

$$(2.3)\quad 0 = \frac{d}{du} E_\theta \|S - \gamma_\theta(u)\|^2 \Big|_{u=0} = -2\, E_\theta < S-M_\theta(S), h_\theta>, \ \textit{for all } \theta \in \Theta.$$

Let us next choose an arbitrary direction $\xi \in \mathbb{R}^p$ at the parameter point $t \in \Theta$. It follows from (2.3) that

(2.4) $$0 = \nabla_\xi \int_{\mathscr{X}} <S-M_t(S), h_t> f_t \, d\mu =$$

$$= E_t < S-M_t(S), \nabla_\xi h_t> - <\nabla_\xi M_t(S), h_t> + E_t (<S-M_\theta(S), h_t> \cdot \nabla_\xi \ell_t) .$$

Rewriting this inequality yields

(2.5) $$\{E_t(<S-M_t(S), h_t> \cdot \nabla_\xi \ell_t)\}^2 = \{<\nabla_\xi M_t(S), h_t> + <M_t(S) - E_t(S), \nabla_\xi h_t>\}^2 .$$

The statement of the theorem is immediate from application of Schwarz's inequality to the expression on the left in (2.5). Q.E.D.

Next let us introduce for each pair of directions $\xi, \eta \in \mathbb{R}^p$

(2.6) $$\nabla_{\xi,\eta}(t) = \{<\nabla_\xi M_t(S), h_t> + <M_t(S) - E_t(S), \nabla_\xi h_t>\}$$

$$\times \{<\nabla_\eta M_t(S), h_t> + <M_t(S) - E_t(S), \nabla_\eta h_t>\} ,$$

(2.7) $$I_{\xi,\eta}(t) = E_t(\nabla_\xi \ell_t)(\nabla_\eta \ell_t) .$$

The definition of directional derivative entails the linearity in each of the components ξ and η separately, so that both are bilinear forms that, moreover, are easily seen to be symmetric and semi–definite positive. Consequently there exist symmetric semi–definite positive matrices $\mathscr{D}_t$ and $\mathscr{I}_t$ (the Fisher information matrix) such that

(2.8) $$<\mathscr{D}_t \xi,\eta> = \Delta_{\xi,\eta}(t), \;\; <\mathscr{I}_t \, \xi,\eta> = I_{\xi,\eta}(t); \; \xi,\eta \in \mathbb{R}^p .$$

THEOREM 2.2. *Let in addition to the assumptions of the previous theorem the matrix* $\mathscr{I}_t$ *be definite positive, and let us denote the largest eigenvalue of the matrix* $\mathscr{D}_t \mathscr{I}_t^{-1}$ *or, equaivalently, of the matrix* $\mathscr{I}_t^{-\frac{1}{2}} \mathscr{D}_t \, \mathscr{I}_t^{-\frac{1}{2}}$ *by* $\lambda_{max}(\mathscr{D}_t,\mathscr{I}_t)$. *Then we have*

(2.9) $$D_{t,h_t}(S) \geq \lambda_{max}(\mathscr{D}_t,\mathscr{I}_t) .$$

PROOF. It suffices to remark that under the present conditions (2.2) is true for each $\xi \in \mathbb{R}^p$, so that the right hand side of (2.2) may be even replaced by $\max_{\xi \in \mathbb{R}^p} \langle \mathscr{D}_t\, \xi,\xi \rangle \,/\, \langle \mathscr{I}_t\, \xi,\xi \rangle$. The relation with the maximal eigenvalue is well–known from matrix–theory, see e.g. Rao (1973). Q.E.D.

Our main examples, including the circle, will be postponed to the next section. By way of a discussion of the result, however, it might be interesting to see what happens when $\mathscr{M} = \mathbb{R}^k$ (so that m=k). We have seen in Subsection 1.2 that $M_\theta(S) = E_\theta(S)$ in this case. The tangent space at any $E_\theta(S) \in \mathscr{M}$ may now be associated with $\mathbb{R}^k$ itself and hence does not depend on $\theta \in \Theta$. Accordingly, for the smooth vector field we may simply take the constant field $h_\theta = h$, $\theta \in \Theta$, for some fixed $h \in \mathbb{R}^k$, so that $\nabla_\xi h_t = 0$. Hence (2.9) simplifies to

$$D_{t,h}(S) = \mathrm{Var}_t \langle S,h \rangle \geq \max_{\xi \in \mathbb{R}^p} \frac{\langle \nabla_\xi E_t(S),h \rangle^2}{I_{\xi,\xi}(t)} , \tag{2.10}$$

which is the ordinary Cramér–Rao lower bound. By observing that $\nabla_\xi E_t(S) = \Sigma_{\alpha=1}^p (\xi_\alpha\, \partial E_t(S_1)/\partial t_\alpha, \dots, \xi_\alpha\, \partial E_t(S_k)/\partial t_\alpha)$, it easily follows that

$$\mathscr{D}_t = \{\Sigma_{q=1}^k \Sigma_{r=1}^k\, h_q h_r (\partial E_t(S_q)/\partial t_\alpha)(\partial E_t(S_r)/\partial t_\beta\}_{\alpha,\beta \in \{1,\dots,p\}} . \tag{2.11}$$

Relations (2.2) and (2.9) also suggest that when the distribution of S becomes more concentrated about the mean location, the expectation will more and more approximate this mean location so that the lower bound will approach the ordinary lower bound.

3. THE SPECIAL ORTHOGONAL GROUP

The k × k–matrices $M = \{M_{qr}\}_{q,r \in \{1,\dots,k\}}$ are considered as elements of $\mathbb{R}^{k \times k}$; the class $\mathscr{M}_k$ of all these matrices is therefore identical to $\mathbb{R}^{k \times k}$. The transpose of

$M = \{M_{qr}\}$ is defined as $M^* = \{M^*_{qr}\} = \{M_{rq}\}$, and the trace of M by $\mathrm{tr}\, M = \Sigma^k_{q=1} M_{qq}$. Given any two matrices $M_q, M_2 \in \mathscr{M}_k$ we have

(3.1) $$\mathrm{tr}(M_1+M_2) = \mathrm{tr}\, M_1 + \mathrm{tr}\, M_2,\ \mathrm{tr}\, M_1\, M_2 = \mathrm{tr}\, M_2\, M_1,$$

and the important relation with the inner product

(3.2) $$<M_1,M_2> = \mathrm{tr}\, M_1^*\, M_2\ .$$

For any smooth manifold $\mathscr{M} \subset \mathscr{M}_k$ the metric will always be derived from (3.2). Let I denote the identity matrix.

The smooth manifold that we will be concerned with has, moreover, a group structure and is called the *special orthogonal group*

(3.3) $$\mathscr{SO}_k = \{O \in \mathscr{M}_k : O^*O = I,\ \textit{and}\ \ \det O=1\},$$

consisting of all orthonormal matrices with determinant +1. Let us also introduce the *linear subspace* of all *skew–symmetric* matrices

(3.4) $$\mathscr{H}_k = \{H \in \mathscr{M}_k : H = -H^*\}\ .$$

The dimension of this subspace is $\frac{1}{2}(k-1)k$ and we have the useful relation

(3.5) $$\mathrm{tr}\, H=0,\ \ H \in \mathscr{H}_k\ .$$

We will make use of the well–known fact (see e.g. Curtis (1985)) that

(3.6) $$\mathscr{T}_O(\mathscr{SO}_k) = \{OH : H \in \mathscr{H}_k\} = \{HO : H \in \mathscr{H}_k\},\ \ O \in \mathscr{SO}_k\ .$$

It will be tactily understood that the conditions mentioned in the beginning of Section 2 are satisfied and that (2.1) holds for all $\xi \in \mathbb{R}^p$ so that $\det \mathscr{T}_t > 0$. Hence the expression on the right in (2.2) may be replaced by its maximum over all $\xi \in \mathbb{R}^p$.

When $\mathscr{M} = \mathscr{SO}_k$ a suitable choice for the vector field is $h_\theta = M_\theta(S)\, H$, $\theta \in \Theta$, for a fixed $H \in \mathscr{H}_k$. It follows that the left–hand side of (2.2) reduces to

$$E_t\{\mathrm{tr}(S - M_t(S))^* M_t(S)H\}^2 = E_t\{\mathrm{tr}\, S^* M_t(S)H\}^2 , \tag{3.7}$$

because $\mathrm{tr}\, M_t(S)^* M_t(S)H = \mathrm{tr}\, H = 0$, due to the orthonormality of $M_t(S)$ and the skew–symmetry of H. To compute the right–hand side let us note that

$$\begin{aligned} &\mathrm{tr}\{(\nabla_\xi M_t(S))^* M_t(S)H + (M_t(S) - E_t(S))^* \nabla_\xi M_t(S)H\} = \\ &= \mathrm{tr}\{(\nabla_\xi M_t(S))^* M_t(S)(H+H^*) - E_t(S)^* \nabla_\xi M_t(S)H\} = \\ &= -\mathrm{tr}\, E_t(S)^* \nabla_\xi M_t(S)H . \end{aligned} \tag{3.8}$$

It follows that (2.2) now assumes the form

$$E_t\{\mathrm{tr}\, S^* M_t(S)H\}^2 \geq \max_{\xi \in \mathbb{R}^p} \frac{\{\mathrm{tr}\, E_t(S)^* \nabla_\xi M_t(S)H\}^2}{I_{\xi,\xi}(t)} . \tag{3.9}$$

EXAMPLE 3.1. The circle may be identified with $\mathscr{SO}_2$. Let us take $\Theta \subset \mathbb{R}$ (i.e. p=1), and let us represent $S : \mathscr{X} \to \mathscr{SO}_2$ by means of a real valued statistic $\Phi : \mathscr{X} \to \mathbb{R}$ according to

$$S = \begin{bmatrix} \cos\Phi & \sin\Phi \\ -\sin\Phi & \cos\Phi \end{bmatrix} . \tag{3.10}$$

In this case we can find an explicit expression for $M_t(S)$ by noticing that minimizing $E_t\|S-O\|^2 = E_t\|S\|^2 + \|O\|^2 - 2\, E_t <S,O> = 4 - E_t<S,O>$ over $O \in \mathscr{SO}_2$ is equivalent with maximizing $<E_t(S),O>$ over $O \in \mathscr{SO}_2$. Let us write for brevity

$$c(\theta) = E_\theta \cos\Phi,\ s(\theta) = E_\theta \sin\Phi,\ \ \theta \in \Theta, \tag{3.11}$$

$$r^2(\theta) = c^2(\theta) + s^2(\theta),\ \ \theta \in \Theta, \tag{3.12}$$

and assume that $r(t) > 0$. It is clear that $(1/r(t))E_t(S) \in \mathscr{SO}_2$, so that obviously this maximum is attained for (see also (1.6))

$$(3.13) \qquad M_t(S) = (1/r(t))\, E_t(S) = \begin{bmatrix} c(t)/r(t) & s(t)/r(t) \\ -s(t)/r(t) & c(t)/r(t) \end{bmatrix} .$$

Under the present assumptions, moreover, we may replace $I_{\xi,\xi}(t)$ by $\mathscr{I}_t$, simply write $M_t'(S)$ rather than $\nabla_\xi\, M_t(S)$. Note that any non–zero $H \in \mathscr{H}_2$ is of the form

$$(3.15) \qquad H = \begin{bmatrix} 0 & h \\ -h & 0 \end{bmatrix}, \ h \in \mathbb{R}\backslash\{0\} .$$

Substitution in (3.9) yields

$$(3.16) \qquad \mathrm{tr}\, S^*\, M_t(S)H = 2h\left[\frac{c(t)}{r(t)}\, \sin \Phi - \frac{s(t)}{r(t)}\, \cos \Phi\right] ,$$

$$(3.17) \qquad \mathrm{tr}\, E_t(S)^*\, M_t'(S)H = 2h\left[\left[\frac{c(t)}{r(t)}\right]'\, \sin \Phi - \left[\frac{s(t)}{r(t)}\right]'\, \cos \Phi\right] .$$

It is clear from (3.16) that

$$(3.18) \qquad E_t\, \mathrm{tr}\, S^*\, M_t(S)H = 0 .$$

Combining (3.16)–(3.18) it follows that (2.9) reduces to

$$(3.19) \quad \mathrm{Var}_t\left[\frac{c(t)}{r(t)}\, \sin \Phi - \frac{s(t)}{r(t)}\, \cos \Phi\right] \geq \frac{\left[\left[\frac{c(t)}{r(t)}\right]' s(t) - \left[\frac{s(t)}{r(t)}\right]' c(t)\right]^2}{\mathscr{I}_t} .$$

To see that this is the same inequality as in Mardia (1975, formula (5.1.5)), apart from the conditioning, we need only realize that there exists a smooth function $\mu : \Theta \to (0,2\pi]$ such that

(3.20) $\frac{c(\theta)}{r(\theta)} = \cos \mu(\theta), \frac{s(\theta)}{r(\theta)} \sin \mu(\theta), \ \theta \in \Theta$.

Substitution in (3.19) yields (see also (1.11) and (1.13))

(3.21) $$E_t \sin^2(\Phi-\mu(t)) \geq \frac{\{\mu'(t)E_t \cos(\Phi-\mu(t))\}^2}{\mathcal{I}_t} .$$

Note that in Maridia (1975) the special case $\mu(t) = t$ is considered.

For the manifold $\mathcal{M}$ we might also take an m–dimensional subgroup $\mathcal{G} \subset \mathcal{SO}_k$. In this case we have to replace the tangent space in (3.6) by

(3.22) $$\mathcal{T}_O(\mathcal{G}) = \{OH : H \in \mathcal{H}\} = \{HO : H \in \mathcal{H}\} ,$$

where $\mathcal{H}$ is a uniquely determined m–dimensional subspace of $\mathcal{H}_k$. Important examples of subgroups of $\mathcal{SO}_k$ are the m–tori, $m \in \mathbb{N}$ with $2m \leq k$, where a 1–torus is also called a circle subgroup.

EXAMPLE 3.2. Let $\mathcal{G} \subset \mathcal{SO}_4$ be a 2–torus, and let as before $\Theta \subset \mathbb{R}$ (i.e. p=1). We may now represent $S : \mathcal{X} \to \mathcal{G}$ by means of a random vector $(\Phi,\tilde{\Phi}) : \mathcal{X} \to \mathbb{R}^2$ by defining S to be the random 4 × 4–matrix

(3.23) $$S = \text{block diagonal}\begin{bmatrix} \cos \Phi & \sin \Phi \\ -\sin \Phi & \cos \Phi \end{bmatrix}, \begin{bmatrix} \cos \tilde{\Phi} & \sin \tilde{\Phi} \\ -\sin \tilde{\Phi} & \cos \tilde{\Phi} \end{bmatrix} .$$

It is easy to see that the mean location is the 4 × 4–matrix

(3.24) $$M_t(S) = \text{block diagonal}\begin{bmatrix} c(t)/r(t) & s(t)/r(t) \\ -s(t)/r(t) & c(t)/r(t) \end{bmatrix}, \begin{bmatrix} \tilde{c}(t)/\tilde{r}(t) & \tilde{s}(t)/\tilde{r}(t) \\ -\tilde{s}(t)/\tilde{r}(t) & \tilde{c}(t)/\tilde{r}(t) \end{bmatrix} ,$$

where $\tilde{c}$, $\tilde{s}$ and $\tilde{r}$ are defined as in (3.11) and (3.12) with Φ replaced by $\tilde{\Phi}$. The tangent space, finally, is given by (3.22) with $\mathcal{H}$ the 2–dimensional subspace of $\mathcal{H}_4$ determined by the 4 × 4–matrices

(3.25) $$H = \text{block diagonal} \begin{bmatrix} 0 & h \\ -h & 0 \end{bmatrix}, \begin{bmatrix} 0 & \tilde{h} \\ -\tilde{h} & 0 \end{bmatrix},$$

where $(h,\tilde{h}) \in \mathbb{R}^2$.

REFERENCES

[1] CHANG, T. (1986). Spherical regression. *Ann. Statist. 14*, 907–924.

[2] CURTIS, M.L. (1985). *Matrix Groups*, 2nd ed., Springer, New York.

[3] HENDRIKS, H. (1987). A Cramér–Rao type lower bound for estimators with values in a certain manifold. Report Math. Inst., Kath. Un. Nijmegen.

[4] MARDIA, K.V. (1972). *Statistics of Directional Data.* Ac. Press, New York.

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

[6] WATSON, G.S. (1983). *Statistics on Spheres.* Wiley, New York.