Annals of Statistics
1984, Vol. 12, No. 3, 1085–1089

ON THE INFORMATION MATRIX FOR SYMMETRIC DISTRIBUTIONS ON THE HYPERSPHERE[1]

BY LOUIS-PAUL RIVEST

Laval University

A general definition of symmetry for a directional model is given. Then the information matrix for the parameters indexing a symmetric density is shown to be block-diagonal: one block for the location and one block for the shape. This result is used to construct an algorithm for the efficient estimation of the parameters. Examples are given; a new symmetric distribution, the FB_6 is introduced; it generalizes the classical distributions on the hypersphere.

1. Introduction. The Fisher information matrix plays an important role in the asymptotic theory of any statistical model. In this work the Fisher information matrix for the parameters indexing a symmetric distribution on the hypersphere is shown to be block diagonal: one block for its generalized location and one block for its shape. For instance the information matrix for $(\boldsymbol{\mu}, \kappa)$ indexing the Fisher-von Mises density:

$$f(\mathbf{u}) = c(\kappa)\exp(\kappa \mathbf{u}'\boldsymbol{\mu}) \quad \boldsymbol{\mu} \in S_k,\ \kappa > 0$$

where S_k is the unit sphere in R^k is made of two blocks: $a(k-1) \times (k-1)$ matrix for the parameters indexing $\boldsymbol{\mu}$ and a 1×1 matrix for κ.

In Section 2, a general definition of symmetry is given. Section 3 contains a proof of the main result; an algorithm for the estimation of the parameters of a symmetric density is also suggested. Section 4 presents some symmetric models.

2. Symmetric models.

NOTATION. $.O(k)$ is the group of orthogonal transformations on R^k and $\mathscr{A}$ is a subgroup of $O(k)$

$.f(\mathbf{u})$ is a density with respect to the Lebesgue measure on S_k.

$.r_1, r_2, \cdots, r_k$ are the components of a vector $\mathbf{r}$ in R^k.

DEFINITION 1. *$\mathscr{A}$-symmetric models.* A density $f(\mathbf{u})$ is said to be $\mathscr{A}$-symmetric if there exists a rotation $\mathbf{P}$ in $O(k)$ such that the density g of $\mathbf{r} = \mathbf{P}'\mathbf{u}$ satisfies:

$$(1)\quad \begin{aligned} &\text{i) } E(r_1) \geq 0 \\ &\text{ii) } E(r_2^2) \geq E(r_3^2) \geq \cdots \geq E(r_k^2) \\ &\text{iii) } g(\mathbf{r}) = g(\mathbf{Hr}) \end{aligned}$$

for all $\mathbf{H}$ in $\mathscr{A}$.

Received March 1983; revised October 1983.

[1] This research is supported by the NSERC grant #A5244 and the FCAC grant #EQ1023.

AMS 1980 *subject classification.* Primary 62F12; secondary 62E20.

Key words and phrases. Bingham distribution, Fisher distribution, maximum likelihood estimation, von Mises distribution.

REMARK 1. Since the determinant of $\mathbf{H}$ is ± 1 (1) iii) implies that $\mathbf{r}$ and $\mathbf{Hr}$ have the same density for any $\mathbf{H}$ in $\mathscr{A}$.

REMARK 2. $\mathscr{A}$-symmetry is not uniquely defined: for any rotation $\mathbf{Q}$ in $O(k)$ let $\mathscr{A}_1 = \{\mathbf{Q}'\mathbf{PHP}'\mathbf{Q}\colon \mathbf{H} \in \mathscr{A}\}$; then if f is $\mathscr{A}$-symmetric, for any $\mathbf{H}$ in $\mathscr{A}_1$, the density $g_1(\mathbf{s})$ of $\mathbf{s} = \mathbf{Q}'\mathbf{u}$ satisfies $g_1(\mathbf{s}) = g_1(\mathbf{Hs})$; hence if $\mathbf{s}$ satisfies (i) and (ii), f is also $\mathscr{A}_1$-symmetric. $\mathscr{A}$-symmetry and $\mathscr{A}_1$-symmetry are equivalent for f.

A way out of this problem is to define $\mathbf{P}$ in terms of the moments of $\mathbf{u}$. For instance, as will be shown in Proposition 1, the symmetries in Examples 2 and 3 are defined using subgroups $\mathscr{A}$ for which $\mathbf{P}$ is the eigenvector matrix of $E(\mathbf{u}\,\mathbf{u}')$.

EXAMPLES.

1) If $\mathscr{A} = O(k)$ then f is $\mathscr{A}$-symmetric if and only if it is equal to the uniform density on S_k.

2) If

$$\mathscr{A} = \left\{\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H} \end{pmatrix} : \mathbf{H} \in O(k-1)\right\}$$

$\mathscr{A}$-symmetry is equivalent to rotational symmetry.

3) If $\mathscr{A} = \{\operatorname{diag}(\pm 1)\}$ then it can be named strong antipodal symmetry. The Bingham density (Mardia, 1972) is strongly antipodal.

The key concept needed to establish the main result is a symmetry weaker than the three types aforementioned:

DEFINITION 2. *$\mathscr{Q}$-symmetry.* A density f is said to be $\mathscr{Q}$-symmetric if it is $\mathscr{A}$-symmetric where

$$\mathscr{A} = \left\{\begin{pmatrix} 1 & 0 \\ 0 & \operatorname{diag}(\pm 1) \end{pmatrix}\right\}.$$

Let $\{\mathbf{p}_i\}_{i=1}^k$ denote the columns of $\mathbf{P}$.

PROPOSITION 1. *If f is $\mathscr{Q}$-symmetric*

$$E(\mathbf{u}) = \mathbf{p}_1 E(r_1), \quad E(\mathbf{uu}') = \mathbf{P}\operatorname{diag}\{E(r_i^2)\}\mathbf{P}'.$$

PROOF. Let $\mathbf{H}_i$ be a diagonal matrix of 1 except for its (i, i) entry which is equal to -1 $(i > 1)$. Remark 1 implies that $\mathbf{H}_i\mathbf{r}$ and $\mathbf{r}$ have the same distribution; therefore

$$E(r_i) = E(\mathbf{p}_i'\mathbf{u}) = 0$$

$$E(r_i r_j) = E(\mathbf{p}_i'\mathbf{u}\,\mathbf{u}'\mathbf{p}_j) = 0 \quad \text{for} \quad j \neq i. \quad \square$$

3. The information matrix. Let $\{\mathbf{u}_i\}_{i=1}^n$ be a sample from a $\mathscr{Q}$-symmetric

density $f(\mathbf{u})$ which can be written as

$$f(\mathbf{u}) = g(\mathbf{P}'\mathbf{u}; \boldsymbol{\phi}); \tag{2}$$

this section considers the problem of estimating $\boldsymbol{\phi}$, a vector valued shape parameter and $\mathbf{P}$. Let $\boldsymbol{\psi}$ be a vector valued parameter indexing $\mathbf{P}$. Note that the dimension of $\boldsymbol{\psi}$ depends on the symmetry of the model. If f is rotationally symmetric, two rotations whose first columns are equal will yield the same parameterization for (2). In this case only $k - 1$ parameters will parametrize $\mathbf{P}$ while in general $k(k-1)/2$ are needed.

NOTATION. $\mathcal{i} = \mathcal{i}(\boldsymbol{\phi}, \boldsymbol{\psi})$ denotes the Fisher information for one observation. It can be written as

$$\mathcal{i} = \begin{pmatrix} \mathcal{i}_{\boldsymbol{\phi}\boldsymbol{\phi}} & \mathcal{i}_{\boldsymbol{\phi}\boldsymbol{\psi}} \\ \mathcal{i}_{\boldsymbol{\psi}\boldsymbol{\phi}} & \mathcal{i}_{\boldsymbol{\psi}\boldsymbol{\psi}} \end{pmatrix}$$

where

$$\mathcal{i}_{\boldsymbol{\phi}\boldsymbol{\phi}} = E(\mathbf{U}_{\boldsymbol{\phi}}\mathbf{U}_{\boldsymbol{\phi}}'), \ \mathcal{i}_{\boldsymbol{\psi}\boldsymbol{\psi}} = E(\mathbf{U}_{\boldsymbol{\psi}}\mathbf{U}_{\boldsymbol{\psi}}') \ \mathcal{i}_{\boldsymbol{\psi}\boldsymbol{\phi}} = \mathcal{i}_{\boldsymbol{\phi}\boldsymbol{\psi}}' = E(\mathbf{U}_{\boldsymbol{\psi}}\mathbf{U}_{\boldsymbol{\phi}}')$$

and $\mathbf{U}_{\boldsymbol{\phi}}$, $\mathbf{U}_{\boldsymbol{\psi}}$ denote the score vectors of partial derivatives of $\ln f(\mathbf{u})$ with respect to the elements of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ respectively.

PROPOSITION 2. *If $g(\mathbf{r}; \boldsymbol{\phi})$ is a differentiable function of $\boldsymbol{\phi}$ which as a function of $\mathbf{r}$ is differentiable and satisfies* (1) iii) *in S_k*

$$\mathcal{i}_{\boldsymbol{\phi}\boldsymbol{\psi}} = \mathbf{0}.$$

PROOF. Let $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_j$ be components of $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ respectively. Let

$$g_{\phi_i}(\mathbf{r}) = \frac{\partial}{\partial \phi_i} g(\mathbf{r}; \boldsymbol{\phi}), \quad g^{(\ell)}(\mathbf{r}; \boldsymbol{\phi}) = \frac{\partial}{\partial r_\ell} g(\mathbf{r}; \boldsymbol{\phi})$$

and $\mathbf{g}^{(\cdot)}(\mathbf{r})$ the vector of $\{g^{(\ell)}(\mathbf{r}; \boldsymbol{\phi})\}$. For $m > 1$ let $\mathbf{H}_m$ be a diagonal matrix of 1 except for its (m, m) entry which is -1. The assumptions on g imply:

$$g_{\phi_i}(\mathbf{H}_m\mathbf{r}) = g_{\phi_i}(\mathbf{r}), \quad \mathbf{g}^{(\cdot)}(\mathbf{H}_m\mathbf{r}) = \mathbf{H}_m\mathbf{g}^{(\cdot)}(\mathbf{r}). \tag{3}$$

If $\mathbf{r} = \mathbf{P}'\mathbf{u}$ one can write

$$U_{\phi_i} = \frac{\partial}{\partial \phi_i} \ln f(\mathbf{u}) = g_{\phi_i}(\mathbf{r})/g(\mathbf{r}; \boldsymbol{\phi})$$

$$U_{\psi_j} = \frac{\partial}{\partial \psi_j} \ln f(\mathbf{u}) = \{\mathbf{g}^{(\cdot)}(\mathbf{r})\}'\mathbf{A}_j\mathbf{r}/g(\mathbf{r}; \boldsymbol{\phi}) = \operatorname{tr}[\mathbf{A}_j\mathbf{r}\{\mathbf{g}^{(\cdot)}(\mathbf{r})\}']/g(\mathbf{r}; \boldsymbol{\phi})$$

where $\mathbf{A}_j = (\partial/\partial\psi_j\mathbf{P}')\mathbf{P}$. Since $\mathbf{P}'\mathbf{P} = \mathbf{I}$ $(\partial/\partial\psi_j\mathbf{P}')\mathbf{P} + \mathbf{P}'(\partial/\partial\psi_j\mathbf{P}) = 0$ and $\mathbf{A}_j = -\mathbf{A}_j'$. Thus

$$\begin{aligned} E(U_{\phi_i}U_{\psi_j}) &= E(g^{-2}(\mathbf{r}; \boldsymbol{\phi})g_{\phi_i}(\mathbf{r})\operatorname{tr}[\mathbf{A}_j\mathbf{r}\{\mathbf{g}^{(\cdot)}(\mathbf{r})\}']) \\ &= \operatorname{tr} \mathbf{A}_j\mathbf{B}_i \end{aligned}$$

where $\mathbf{B}_i = E[g^{-2}(\mathbf{r};\ \phi)g_{\phi_i}(\mathbf{r})\mathbf{r}\{\mathbf{g}^{(\cdot)}(\mathbf{r})\}']$. Since $\mathbf{r}$ and $\mathbf{H}_m\mathbf{r}$ have the same density,

$$\mathbf{B}_i = E[g^{-2}(\mathbf{H}_m\mathbf{r};\ \phi)g_{\phi_i}(\mathbf{H}_m\mathbf{r})\mathbf{H}_m\mathbf{r}\{\mathbf{g}^{(\cdot)}(\mathbf{H}_m\mathbf{r})\}']$$

which on using (3) is seen to be equal to

$$\mathbf{H}_m E[g^{-2}(\mathbf{r};\ \phi)g_{\phi_i}(\mathbf{r})\mathbf{r}\{\mathbf{g}^{(\cdot)}(\mathbf{r})\}']\mathbf{H}_m.$$

Right and left multiplication by $\mathbf{H}_m$ changes the sign of all the elements of the mth line and the mth column except for the (m, m) entry. Since $\mathbf{B}_i = \mathbf{H}_m\mathbf{B}_i\mathbf{H}_m$ is true for any $m > 1$, $\mathbf{B}_i$ is a diagonal matrix; furthermore

$$\mathrm{tr}\mathbf{A}_j\mathbf{B}_i = \mathrm{tr}\mathbf{B}_i'\mathbf{A}_j' = -\mathrm{tr}\mathbf{A}_j\mathbf{B}_i = 0. \qquad \square$$

For circular variables, Proposition 2 is equivalent to the classical result that the maximum likelihood estimate of location is asymptotically independent of the maximum likelihood estimate of the scale or shape parameter for symmetric densities.

This result suggests a two-step algorithm to estimate $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ (or $\mathbf{P}$). First given an $O_p(n^{-1/2})$ estimate $\hat{\hat{\mathbf{P}}}$ of $\mathbf{P}$, maximize

$$\Pi g(\hat{\hat{\mathbf{P}}}'\mathbf{u}_i;\ \phi)$$

to get $\hat{\boldsymbol{\phi}}$. Then maximize

$$\Pi g(\mathbf{P}'\mathbf{u}_i;\ \hat{\boldsymbol{\phi}})$$

to obtain $\hat{\mathbf{P}}$. The standard asymptotic theory of maximum likelihood estimation (Cox and Hinkley, 1974, Chapter 9) shows that $\hat{\boldsymbol{\phi}}$ and $\hat{\mathbf{P}}$ are efficient estimates. An iteration of this procedure will lead to the maximun likelihood estimates.

For a rotationally symmetric model, any matrix $\hat{\hat{\mathbf{P}}}$ with its first column proportional to $\sum \mathbf{u}_i$ provides an $O_p(n^{-1/2})$ estimate of $\boldsymbol{\psi}$ if $E(r_1) > 0$. When all the population eigenvalues of $E(\mathbf{u}\,\mathbf{u}')$ are different, the matrix of the eigenvectors of $\sum \mathbf{u}_i\mathbf{u}_i'$ is an $O_p(n^{-1/2})$ estimate of $\mathbf{P}$ (Tyler, 1981).

For the algorithm to be valid, $g(\mathbf{r};\ \phi)$ has to be parametrized in such a way that the matrix $\mathbf{P}$ of Definition 1 is the same for all the values of $\boldsymbol{\phi}$.

For instance Bingham and Mardia (1978) model on S_3:

$$f(\mathbf{u}) = \{F(\tau,\ \nu)\}^{-1}\exp\{-\tau(\boldsymbol{\mu}'\mathbf{u} - \nu)^2\}, \quad \tau \in R, \quad \nu > 0$$

has a mean direction equal to $\boldsymbol{\mu}$ when $\tau > 0$ and $-\boldsymbol{\mu}$ when $\tau < 0$. If the sign of τ is not known a priori, a reparametrization is needed in order to use the algorithm of this section.

4. Examples. The following examples discuss distributions on S_3.

1) A reparametrization of Bingham and Mardia (1978) model:

$$f(\mathbf{u}) = c(\kappa,\ \gamma)\exp\{\kappa\mathbf{p}_1'\mathbf{u} + \gamma(\mathbf{p}_1'\mathbf{u})^2\}, \quad \kappa > 0,\ \gamma \in R.$$

When $\gamma < 0$ and $\kappa < -2\gamma$, f has a maximum on a small circle in a plane orthogonal to $\mathbf{p}_1$. This model is rotationally symmetric. Bingham and Mardia proved that the information matrix is block diagonal; the two-step algorithm they used to

maximize the likelihood is similar to the one of Section 3.

2) The FB_5 distribution:

$$f(\mathbf{u}) = c(\kappa, \beta)^{-1}\exp\{\kappa\mathbf{p}_1'\mathbf{u} + \beta\{(\mathbf{p}_2'\mathbf{u})^2 - (\mathbf{p}_3'\mathbf{u})^2\}\}$$

where $\kappa \geq 0$, $\beta \geq 0$.

Kent (1982) pointed out that when $\kappa > 0$, $\mathbf{p}_1$ is the mean direction of $\mathbf{u}$, $\mathbf{p}_2$ is the major axis and $\mathbf{p}_3$ is the minor axis, therefore the matrix $\mathbf{P}$ satisfies the requirements of Definition 1. The density of $\mathbf{r} = \mathbf{P}'\mathbf{u}$ is equal to:

$$g(\mathbf{r}) = c(\kappa, \beta)^{-1}\exp\{\kappa r_1 + \beta(r_2^2 - r_3^2)\}$$

so that $f(\mathbf{u})$ is $\mathscr{Q}$-symmetric.

When $\kappa > 0$, estimating $\mathbf{p}_1$ by $\bar{\mathbf{p}}_1 = \sum \mathbf{u}_i/\|\sum \mathbf{u}_i\|$ and $\mathbf{p}_2$ and $\mathbf{p}_3$ by the eigenvectors of the nonzero eigenvalues of $(\mathbf{I} - \bar{\mathbf{p}}_1\mathbf{p}_1')\sum \mathbf{u}_i\mathbf{u}_i'(\mathbf{I} - \bar{\mathbf{p}}_1\bar{\mathbf{p}}_1')$ yields $\hat{\hat{\mathbf{P}}}$, the matrix of constrained eigenvectors (Kent 1982, page 74). This matrix is an $O_p(n^{-1/2})$ consistent estimate of $\mathbf{P}$. Thus Kent's moment estimates for κ and β correspond to the ones obtained at step 1 of the algorithm. They are therefore efficient.

3) As a generalization of the previous distribution, the FB_6 density is suggested:

$$f(\mathbf{u}) = c(\kappa, \beta, \gamma)^{-1}\exp\{\kappa\mathbf{p}_1'\mathbf{u} + \gamma(\mathbf{p}_1'\mathbf{u})^2 + \beta((\mathbf{p}_2'\mathbf{u})^2 - (\mathbf{p}_3'\mathbf{u})^2)\},$$

where $\kappa, \beta > 0$ and $\gamma \in R$. As for the FB_5 distribution $\mathbf{P}$ satisfies the requirement of Definition 1. The density of $\mathbf{r} = \mathbf{P}'\mathbf{u}$ is given by

$$g(\mathbf{r}) = c(\kappa, \beta, \gamma)^{-1}\exp\{\kappa r_1 + \gamma r_1^2 + \beta(r_2^2 - r_3^2)\};$$

it is $\mathscr{Q}$-symmetric. Besides being a generalization of the previous two examples, this model contains as a special case the Bingham distribution (if $\kappa = 0$). Kent's matrix of constrained eigenvectors will yield an $O_p(n^{-1/2})$ consistent estimate of $\mathbf{P}$ when $\kappa > 0$.

REFERENCES

BINGHAM, C. and MARDIA, K. V. (1978). A small circle distribution on the sphere. *Biometrika* **65** 379–389.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

KENT, J. T. (1982). The Fisher-Bingham distribution on the sphere. *J. Roy. Statist. Soc. B* **44** 71–80.

MARDIA, K. V., (1972). *Statistics of Directional Data.* Academic, London.

TYLER, D. E., (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ LAVAL
QUÉBEC, P.Q.,
CANADA G1K 7P4