The Annals of Statistics
1979, Vol. 7, No. 3, 599–606

MAXIMUM LIKELIHOOD ESTIMATORS FOR THE MATRIX VON MISES-FISHER AND BINGHAM DISTRIBUTIONS

BY P. E. JUPP[1] AND K. V. MARDIA

University of Leeds

It has been conjectured by Khatri and Mardia that with probability one MLEs for the parameters of the von Mises-Fisher matrix distribution exist and are unique. We prove that, except for small sample sizes, this conjecture is true, both in the case where the parameter matrix has known rank and in the unrestricted case. The corresponding result for the matrix Bingham distribution is proven also.

1. Introduction and summary. The Stiefel manifold $O(n, p)$ of $n \times p$ matrices $\mathbf{X}$ such that $\mathbf{XX}' = \mathbf{I}_n$ admits a distinguished probability distribution, the uniform distribution. This is induced from the normalized Haar measure on the orthogonal group $O(p)$. Background on Haar measure and uniform measures on compact homogeneous spaces like the Stiefel and Grassman manifolds can be found in Nachbin (1965) (especially pages 130–131) and in Farrell (1976). The von Mises-Fisher distribution $M(p, n, \mathbf{F})$ on $O(n, p)$ was defined by Downs (1972) to have density

$$a(\mathbf{F}) \exp \operatorname{tr}(\mathbf{FX}'), \qquad \mathbf{X} \in O(n, p) \tag{1.1}$$

with respect to the uniform distribution. Here $\mathbf{F}$ is an $n \times p$ matrix. The Bingham distribution $B(p, n, \mathbf{A})$ on $O(n, p)$ has density

$$\alpha(\mathbf{A}) \exp \operatorname{tr}(\mathbf{XAX}'), \qquad \mathbf{X} \in O(n, p) \tag{1.2}$$

with respect to the uniform distribution. Here $\mathbf{A}$ is a symmetric $p \times p$ matrix. We assume without loss of generality that $\operatorname{tr} \mathbf{A} = 0$. The Bingham distribution $B(p, 1, \mathbf{A})$ was introduced by Bingham (1974, 1976) and $B(p, n, \mathbf{A})$ is a particular case of a distribution introduced by Khatri and Mardia (1977).

Khatri and Mardia (1977) derive expressions for the maximum likelihood estimator (MLE) of $\mathbf{F}$ in the von Mises-Fisher family (and the subfamilies where rank $\mathbf{F}$ is known) and show that in many cases the MLE exists and is unique. They suggest that, as the von Mises-Fisher family is an exponential family, this is true in general. Here we show that this is indeed the case; for all except the smallest sample sizes, with probability one, the MLE exists and is unique. We prove also the analogous result for the Bingham family extending results of Bingham (1974, 1976). Our results can be contrasted with those of Berk (1972) which yield almost sure eventual existence of MLE's without giving any information on the required sample size.

Received October 1977; revised February 1978.

[1]Research carried out while this author was on a Research Fellowship funded by the Science Research Council.

AMS 1970 *subject classifications.* Primary 62F10; secondary 62F05.

Key words and phrases. Maximum likelihood estimator, von Mises-Fisher matrix distribution, Bingham matrix distribution, exponential family.

2. Existence and uniqueness of M.L.E.'s. On the vector space $\mathbb{R}(n, p)$ of $n \times p$ matrices there is an inner product defined by $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr} \mathbf{AB}'$. Thus we can rewrite the density of $M(p, n, \mathbf{F})$ as

$$(2.1) \qquad \exp\{\langle \mathbf{F}, \mathbf{X} \rangle - \kappa(\mathbf{F})\}, \qquad \mathbf{X} \in O(n, p)$$

where $\kappa(\mathbf{F}) = -\log a(\mathbf{F}) = \log {}_0F_1(p/2, \frac{1}{4}\mathbf{FF}')$ and ${}_0F_1$ is a hypergeometric function of matrix argument. Similarly, we can rewrite the density of $B(p, n, \mathbf{A})$ as

$$(2.2) \qquad \exp\{\langle \mathbf{A}, \mathbf{X}'\mathbf{X} \rangle - \kappa_B(\mathbf{A})\}, \qquad \mathbf{X} \in O(n, p)$$

where $\kappa_B(\mathbf{A}) = -\log \alpha(\mathbf{A}) = \log {}_1F_1(n/2, p/2; \mathbf{A})$ and ${}_1F_1$ is a confluent hypergeometric function of matrix argument (see Equation (3.8) of Bingham, 1976). In some ways it is more natural to think of $B(p, n, \mathbf{A})$ as being on the Grassmann manifold $G_n(\mathbb{R}^p)$ rather than on $O(n, p)$. $G_n(\mathbb{R}^p)$ is the set of $p \times p$ matrices $\mathbf{Y}$ such that $\mathbf{Y} = \mathbf{Y}' = \mathbf{Y}^2$ and $\operatorname{tr} \mathbf{Y} = n$. There is a function from $O(n, p)$ to $G_n(\mathbb{R}^p)$ defined by $\mathbf{Y} = \mathbf{X}'\mathbf{X}$. On $G_n(\mathbb{R}^p)$ there is a distinguished probability distribution, the uniform distribution, which may be defined as the distribution of $\mathbf{Y}$ when $\mathbf{X}$ has the uniform distribution on $O(n, p)$. If $\mathbf{X}$ is distributed as $B(p, n, \mathbf{A})$ then the distribution of $\mathbf{Y}$ has density

$$(2.3) \qquad \exp\{\langle \mathbf{A}, \mathbf{Y} \rangle - \kappa_B(\mathbf{A})\}, \qquad \mathbf{Y} \in G_n(\mathbb{R}^p)$$

with respect to the uniform distribution.

We shall usually use $B(p, n, \mathbf{A})$ to refer to this distribution on $G_n(\mathbb{R}^p)$. Let $S_0(p)$ denote the vector space of symmetric $p \times p$ matrices with zero trace and define $t(\mathbf{Y}) \in S_0(p)$ by $t(\mathbf{Y}) = \mathbf{Y} - (n/p)\mathbf{I}_p$. We can rewrite the density of $B(p, n, \mathbf{A})$ as

$$(2.4) \qquad \exp\{\langle \mathbf{A}, t(\mathbf{Y}) \rangle - \kappa_B(\mathbf{A})\}, \qquad \mathbf{Y} \in G_n(\mathbb{R}^p).$$

It is now clear that in the terminology of Barndorff-Nielson (1973, Section 5) (2.1) and (2.4) are minimal standard representations of the exponential families $M(p, n, \cdot)$ and $B(p, n, \cdot)$. (That the representations are minimal, i.e., the dimensions of the parameter spaces cannot be reduced, follows, for example, from the existence of densities in Proposition 2.) Further, as $O(n, p)$ and so $G_n(\mathbb{R}^p)$ are compact, $M(p, n, \mathbf{F})$ and $B(p, n, \mathbf{A})$ are defined for all $\mathbf{F}$ in $\mathbb{R}(n, p)$ and all $\mathbf{A}$ in $S_0(p)$. Thus the maximal sets on which κ and κ_B are defined are open, i.e., the families are regular. Similarly, if we consider the sample means $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ of samples of size N from $M(p, n, \mathbf{F})$ and $B(p, n, \mathbf{A})$ we again get minimal standard representations of regular families. The densities are

$$(2.5) \qquad \exp\{\langle \mathbf{F}, \overline{\mathbf{X}} \rangle - \kappa(\mathbf{F})\}, \qquad \overline{\mathbf{X}} \in \mathbb{R}(n, p)$$

$$(2.6) \qquad \exp\{\langle \mathbf{A}, t(\overline{\mathbf{Y}}) \rangle - \kappa_B(\mathbf{A})\}, \qquad t(\overline{\mathbf{Y}}) \in S_0(p)$$

with respect to the probability distributions on $\mathbb{R}(n, p)$ and $S_0(p)$ corresponding to the uniform case $\mathbf{F} = \mathbf{0}$, $\mathbf{A} = \mathbf{0}$. Now we can apply the general theory of MLEs for exponential families.

The principal result we need is the following direct consequence of Theorem 7.1 of Barndorff-Nielsen (1973).

PROPOSITION 1. *Let C be the convex hull of $O(n,p)$ in $\mathbb{R}(n,p)$ and C_B be the convex hull of $t(G_n(\mathbb{R}^p))$ in $S_0(p)$. For $\mathbf{F} \in \mathbb{R}(n,p)$, $\mathbf{A} \in S_0(p)$, denote the expectations $E_{M(p,n,\mathbf{F})}(\mathbf{X})$ and $E_{B(p,n,\mathbf{A})}(t(\mathbf{Y}))$ by $\tau(\mathbf{F})$ and $\tau_B(\mathbf{A})$ respectively. Then*

(a) *there is a* MLE $\hat{\mathbf{F}}$ *of* $\mathbf{F}$ *if and only if* $\overline{\mathbf{X}} \in \operatorname{int} C$.

If $\hat{\mathbf{F}}$ *exists it is given uniquely by*

$$\overline{\mathbf{X}} = \tau(\hat{\mathbf{F}}) = (\partial\kappa/\partial\mathbf{F})_{\mathbf{F}=\hat{\mathbf{F}}} \tag{2.7}$$

(b) *there is a* MLE $\hat{\mathbf{A}}$ *of* $\mathbf{A}$ *if and only if* $t(\overline{\mathbf{Y}}) \in \operatorname{int} C_B$.

If $\hat{\mathbf{A}}$ *exists it is given uniquely by*

$$t(\overline{\mathbf{Y}}) = \tau_B(\hat{\mathbf{A}}) = (\partial\kappa_B/\partial\mathbf{A})_{\mathbf{A}=\hat{\mathbf{A}}}. \tag{2.8}$$

To use this proposition we do not need to calculate C or C_B once we know that $\overline{\mathbf{X}}$, $t(\overline{\mathbf{Y}})$ have densities with respect to Lebesgue measure. From now on $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ will denote the sample means from any distributions on $O(n,p)$ and $G_n(\mathbb{R}^p)$ which are absolutely continuous with respect to the uniform distribution. The next result follows from Theorems 2 and 3 of Jupp and Mardia (1977).

PROPOSITION 2. (a) *In* $\mathbb{R}(n,p)$, $\overline{\mathbf{X}}$ *has a density with respect to Lebesgue measure if and only if either*

(i) $N \geqslant 2$ *and* $n < p$; *or*

(ii) $N \geqslant 3$ *and* $n = p \geqslant 3$.

(b) *In* $S_0(p)$, $t(\overline{\mathbf{Y}})$ *has a density with respect to Lebesgue measure if* $N \geqslant p$.

COROLLARY. *Under the conditions of Proposition* 2, *with probability* 1, $\overline{\mathbf{X}} \in \operatorname{int} C$ *and* $t(\overline{\mathbf{Y}}) \in \operatorname{int} C_B$.

PROOF. Certainly $\overline{\mathbf{X}} \in C$ and $t(\overline{\mathbf{Y}}) \in C_B$. Also $\partial C = C \setminus \operatorname{int} C$, ∂C_B have Lebesgue measure zero in $\mathbb{R}(n,p)$ and $S_0(p)$ from Corollary 8.6.2 of Farrell (1976).

Combining the above we get the following main result on MLEs for the von Mises-Fisher and Bingham families.

THEOREM 1.

(a) *For the von Mises-Fisher family on* $O(n,p)$: *if* $N = 1$, *there is no* MLE *of* $\mathbf{F}$; *if* $N = 2$ *and* $n = p$, *then with positive probability there is no* MLE *of* $\mathbf{F}$; *if either* (i) $N \geqslant 2$ *and* $n < p$ *or* (ii) $N \geqslant 3$ *and* $n = p$, *then with probability* 1 *there is a unique* MLE $\hat{\mathbf{F}}$ *of* $\mathbf{F}$ *and*

$$\overline{\mathbf{X}} = \tau(\hat{\mathbf{F}}) = (\partial\kappa/\partial\mathbf{F})_{\mathbf{F}=\hat{\mathbf{F}}}.$$

(b) *For the Bingham family on* $G_n(\mathbb{R}^p)$, *if* $N \geqslant p$ *then with probability* 1 *there is a unique* MLE $\hat{\mathbf{A}}$ *of* $\mathbf{A}$ *and*

$$t(\overline{\mathbf{Y}}) = \tau(\hat{\mathbf{A}}) = (\partial\kappa_B/\partial\mathbf{A})_{\mathbf{A}=\hat{\mathbf{A}}}.$$

PROOF. The existence statements follow immediately from the results above. Nonexistence in (a) for $N = 1$ follows from the fact that $O(n,p) \subset \partial C$. (To see this, note that for $\mathbf{X} \in O(n,p)$ $\langle \mathbf{X}, \mathbf{X} \rangle = n$ and argue as for the sphere S^{p-1} in $\mathbb{R}^p$.)

In the case $N = 2$ define $\rho(\mathbf{X})$ to be the maximum modulus of an eigenvalue of $\mathbf{X}$. Then for $\mathbf{X} \in O(n, p)$, $\rho(\mathbf{X}) = 1$. Also ρ is convex, so

$$\operatorname{int} C \subset \{\mathbf{X} \in \mathbb{R}(n, p) | \rho(\mathbf{X}) < 1\}.$$

Now for a sample $(\mathbf{X}_1, \mathbf{X}_2)$ from $M(p, p, \mathbf{F})$, with positive probability, 1 is an eigenvalue of $\mathbf{X}_1^{-1}\mathbf{X}_2$ (consider the cases p odd and $\det X_1 = \det X_2$, p even and $\det \mathbf{X}_1 = -\det \mathbf{X}_2$), so $\rho(\overline{\mathbf{X}}) = \rho\{\frac{1}{2}(\mathbf{I}_p + \mathbf{X}_1^{-1}\mathbf{X}_2)\} = 1$. Thus $\overline{\mathbf{X}} \in \partial C$ and there is no MLE.

3. The von Mises-Fisher family—singular value version. We can reformulate the above results for the von Mises-Fisher family using singular value decompositions. Recall the singular value decomposition $\mathbf{F} = \mathbf{\Delta D}_{\phi}\mathbf{\Gamma}$ where $\mathbf{\Delta}'\mathbf{\Delta} = \mathbf{I}_r = \mathbf{\Gamma\Gamma}'$, $\mathbf{D}_{\phi} = \operatorname{diag}(\phi_1, \ldots, \phi_r)$ with $\phi_1 \geqslant \phi_2 \geqslant \cdots \geqslant \phi_r > 0$ (see, for example, Rao, 1973, page 42, or Farrell, 1976, page 120). Such a decomposition always exists and $\mathbf{D}_{\phi}$ is unique. If we insist that the first nonzero element in each row of $\mathbf{\Gamma}$ is positive, then if the ϕ_i are distinct, $\mathbf{\Delta}$ and $\mathbf{\Gamma}$ are unique. Let $\overline{\mathbf{X}} = \overline{\mathbf{\Delta}}\mathbf{D}_{g}\overline{\mathbf{\Gamma}}$, $\mathbf{F} = \hat{\mathbf{\Delta}}\mathbf{D}_{\hat{\phi}}\hat{\mathbf{\Gamma}}$ be singular value decompositions of the sample mean and MLE.

THEOREM 2. *For the von Mises-Fisher family on $O(n, p)$ if either* (i) $\geqslant 2$ *and* $n < p$ *or* (ii) $N \geqslant 3$ *and* $n = p \geqslant 3$, *then with probability* 1, $\hat{\mathbf{\Delta}}$, $\mathbf{D}_{\hat{\phi}}$, $\hat{\mathbf{\Gamma}}$ *exist and are given uniquely by*

$$\hat{\mathbf{\Delta}} = \overline{\mathbf{\Delta}}, \hat{\mathbf{\Gamma}} = \overline{\mathbf{\Gamma}}, g_i = u_i(\hat{\boldsymbol{\phi}}), \qquad 1 \leqslant i \leqslant r \tag{3.1}$$

where $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_r)$ *and* $u_i(\boldsymbol{\phi}) = (\partial/\partial\phi_i)\, {}_0F_1(p/2, \frac{1}{4}\mathbf{D}_{\phi}^2)$.

PROOF. First suppose that $n = p = r$. Then a calculation using invariance of Haar measure on $O(p)$ gives

$$\tau(\mathbf{\Delta D}_{\phi}\mathbf{\Gamma}) = \mathbf{\Delta}\tau(\mathbf{D}_{\phi})\mathbf{\Gamma}. \tag{3.2}$$

Also, as a $p \times p$ matrix is diagonal if and only if it commutes with all orthogonal diagonal matrices, $\tau(\mathbf{D}_{\phi})$ is diagonal.

In the general case choose $\mathbf{\Delta}_0 \in O(n)$, $\mathbf{\Gamma}_0 \in O(p)$ such that $\mathbf{\Delta}_0 = (\mathbf{\Delta}, \mathbf{\Delta}_1)$, $\mathbf{\Gamma}_0' = (\mathbf{\Gamma}', \mathbf{\Gamma}_1')$ and put $\mathbf{\Delta}_1 = \text{block diag}[\mathbf{\Delta}_0, \mathbf{I}_{p-n}]$. Then $\tau(\mathbf{\Delta}_2\mathbf{D}_{(\phi,0)}\mathbf{\Gamma}_0) = \mathbf{\Delta}_1\tau(\mathbf{D}_{(\phi,0)})\mathbf{\Gamma}_0$, so (3.2) holds in this case also and $\tau(\mathbf{D}_{\phi})$ is diagonal. As $\overline{\mathbf{X}}$ has a density, the g_i are distinct with probability 1. Now we can use (3.2), Theorem 1(a) and uniqueness of polar decomposition to get (3.1).

REMARK. The maximum likelihood equations (3.1) were obtained by Khatri and Mardia (1977).

4. The von Mises-Fisher family—F of given rank. In the von Mises-Fisher family on $O(n, p)$, by restricting $\mathbf{F}$ to lie in $\mathbb{R}(n, p; r)$, the set of $n \times p$ matrices of rank r, we obtain a curved exponential subfamily. Again, except for small sample size N, with probability 1 the MLE of $\mathbf{F}$ exists and is unique.

THEOREM 3. *For the von Mises-Fisher family on $O(n, p)$ restricted by rank $\mathbf{F} = r$, if either* (i) $N \geqslant 2$ *and* $n < p$ *or* (ii) $N \geqslant 3$ *and* $n = p \geqslant 3$, *then with probability* 1 *there is a unique* MLE $\hat{\mathbf{F}}$ *of* $\mathbf{F}$ *and*

$$\hat{\mathbf{F}} = \bar{\boldsymbol{\Delta}}\mathbf{D}_{(\tilde{\boldsymbol{\phi}}, \mathbf{0})}\bar{\boldsymbol{\Gamma}}, \qquad g_i = u_i(\tilde{\boldsymbol{\phi}}), \qquad 1 \leqslant i \leqslant r, \tag{4.1}$$

where

$$\bar{\mathbf{X}} = \bar{\boldsymbol{\Delta}}\mathbf{D}_{\mathbf{g}}\bar{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \cdots, \hat{\phi}_r) \text{ and } u_i(\boldsymbol{\phi}) = \frac{\partial}{\partial \phi_i}\, {}_0F_1\left(\frac{p}{2}, \frac{1}{4}\mathbf{D}_{\boldsymbol{\phi}}^2\right).$$

PROOF. We have rank $\bar{\mathbf{X}} < n$ precisely when at least one of the $n \times n$ minors of $\bar{\mathbf{X}}$ has determinant zero. Thus the set of $\bar{\mathbf{X}}$ in $\mathbb{R}(n,p)$ with rank $\bar{\mathbf{X}} < n$ is an algebraic variety of codimension one (the set of zeros of a polynomial) and so has Lebesgue measure zero. As $\bar{\mathbf{X}}$ has a density, with probability 1 rank $\bar{\mathbf{X}} = n$. Similarly, with probability 1 the eigenvalues of $\bar{\mathbf{X}}\bar{\mathbf{X}}'$ are distinct. Thus, with probability 1, $\bar{\mathbf{X}} \in \operatorname{int} C$, rank $\bar{\mathbf{X}} = n$, and the eigenvalues of $\bar{\mathbf{X}}\bar{\mathbf{X}}'$ are distinct. We shall assume that $\bar{\mathbf{X}}$ satisfies these conditions.

Existence. For $\boldsymbol{\Delta} \in O(n)$, $\boldsymbol{\Gamma} \in O(n,p)$, and $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_r)$, let $l(\boldsymbol{\phi})$ be the log-likelihood of $\bar{\mathbf{X}}$ at $\boldsymbol{\Delta}\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}\boldsymbol{\Gamma}$,

$$\begin{aligned} l(\boldsymbol{\phi}) &= \langle \boldsymbol{\Delta}\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}\boldsymbol{\Gamma}, \bar{\mathbf{X}} \rangle - \kappa(\boldsymbol{\Delta}\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}\boldsymbol{\Gamma}) \\ &= \langle \mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}, \boldsymbol{\Delta}'\bar{\mathbf{X}}\boldsymbol{\Gamma}' \rangle - \kappa(\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}). \end{aligned} \tag{4.2}$$

Now $\boldsymbol{\Delta}'\bar{\mathbf{X}}\boldsymbol{\Gamma}' \in \operatorname{int} C$ so $(\mathbf{I}_r, \mathbf{0})\bar{\mathbf{X}}\boldsymbol{\Gamma}' \in \operatorname{int} C_r$ where C_r is the convex hull of $O(r,p)$ in $\mathbb{R}(r,p)$, and $\partial l/\partial \phi_i = (\boldsymbol{\Delta}\bar{\mathbf{X}}\boldsymbol{\Gamma}')_{ii} - \partial/(\partial\phi_1)\kappa(\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})})$. By Proposition 1(a) and (3.1) for $O(r,p)$, there is a unique solution $\bar{\boldsymbol{\phi}}$ of $\partial l/\partial \phi_i = 0$, $1 \leqslant i \leqslant r$, and this gives a maximum of l. Now put $g(\boldsymbol{\Delta}, \boldsymbol{\Gamma}) = \max_{\boldsymbol{\phi}}\{\langle \boldsymbol{\Delta}\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}\boldsymbol{\Gamma}, \bar{\mathbf{X}} \rangle - \kappa(\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})})\}$. We wish to show that g has a maximum. As g is defined on the compact set $O(n) \times O(n,p)$ we need only show that g is continuous. As the matrix $((\partial^2/\partial\phi_i\partial\phi_j)\kappa(\mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}))$ is nonsingular (see Theorem 5.4 of Barndorff-Nielsen, 1973), by the inverse function theorem $\bar{\boldsymbol{\phi}}$ is a differentiable (so continuous) function of the $(\boldsymbol{\Delta}'\bar{\mathbf{X}}\boldsymbol{\Gamma}')_{ii}$, $1 \leqslant i \leqslant r$. Thus g is continuous as required.

Uniqueness. $\mathbb{R}(n,p;r)$ is a submanifold of $\mathbb{R}(n,p)$ of codimension $(n-r)(p-r)$. Indeed, if $\mathbf{F} \in \mathbb{R}(n,p;r)$, there are $\boldsymbol{\Delta}_0 \in O(n)$, $\boldsymbol{\Gamma}_0 \in O(p)$ such that $\boldsymbol{\Delta}_0'\mathbf{F}\boldsymbol{\Gamma}_0' = \mathbf{D}_{(\boldsymbol{\phi}, \mathbf{0})}$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_r)$. If $\mathbf{G} = \boldsymbol{\Delta}_0\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}\boldsymbol{\Gamma}_0$ is near enough to $\mathbf{F}$, then rank $\mathbf{A} = r$ and

$$\operatorname{rank} \mathbf{G} = \operatorname{rank}\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I}_{n-r} \end{pmatrix}\boldsymbol{\Delta}_0'\mathbf{G}\boldsymbol{\Gamma}_0 = \operatorname{rank}\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix}.$$

Thus $\mathbf{G} \in \mathbb{R}(n,p;r)$ if and only if $\mathbf{D} = \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. Therefore $\mathbb{R}(n,p;r)$ is a submanifold of codimension $(n-r)(p-r)$ and the tangent space to $\mathbb{R}(n,p;r)$ at $\mathbf{F}$ is the space of matrices $\boldsymbol{\Delta}_0\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{pmatrix}\boldsymbol{\Gamma}_0$.

The derivative at $\mathbf{F}$ of the log-likelihood function at $\bar{\mathbf{X}}$ is $\bar{\mathbf{X}} - \tau(\mathbf{F})$, so at a MLE $\hat{\mathbf{F}}$ in $\mathbb{R}(n,p;r)$ $\bar{\mathbf{X}} - \tau(\hat{\mathbf{F}})$ is normal to the tangent space to $\mathbb{R}(n,p;r)$ at $\hat{\mathbf{F}}$. If $\mathbf{F} = \hat{\boldsymbol{\Delta}}\mathbf{D}_{\boldsymbol{\phi}}\hat{\boldsymbol{\Gamma}}$, $\mathbf{X} = \bar{\boldsymbol{\Delta}}\mathbf{D}_{\mathbf{g}}\bar{\boldsymbol{\Gamma}}$ are the singular value decompositions, a calculation shows that a necessary condition for $\hat{\mathbf{F}}$ to be a MLE is

$$(4.3) \qquad \hat{\boldsymbol{\Delta}}_0'\bar{\boldsymbol{\Delta}}\mathbf{D_g}\bar{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}_0' = \text{block diag}[\mathbf{D}_{(\psi,\mathbf{0})}, \mathbf{E}]$$

where $\hat{\boldsymbol{\Delta}}_0 = (\hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\Delta}}_1) \in O(n)$, $\hat{\boldsymbol{\Gamma}}_0' = (\hat{\boldsymbol{\Gamma}}'\hat{\boldsymbol{\Gamma}}_1') \in O(p)$, $\psi = (\psi_1, \cdots, \psi_r)$ and satisfies $\mathbf{D}_{(\psi,\mathbf{0})} = \tau(\mathbf{D}_{(\phi,\mathbf{0})})$, and $\mathbf{E} \in \mathbb{R}(n-r, p-r)$. If $\mathbf{E} = \mathbf{U}\mathbf{D_e}\mathbf{V}$ is a singular value decomposition, then

$$(4.4) \qquad \hat{\boldsymbol{\Delta}}_0'\bar{\boldsymbol{\Delta}}\mathbf{D_g}\bar{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}_0' = \text{block diag}[\mathbf{I}_r, \mathbf{U}] \times \mathbf{D}_{(\psi,\mathbf{e})} \times \text{block diag}[\mathbf{I}_r, \mathbf{V}].$$

As the eigenvalues of $\bar{\mathbf{X}}\bar{\mathbf{X}}'$ are distinct, uniqueness of singular value decomposition (see Theorem 8.5.1 of Farrell, 1976) gives $\psi_i = g_i$ for $1 \leqslant i \leqslant r$ and the existence of a diagonal orthogonal matrix $\mathbf{H}$ such that $\hat{\boldsymbol{\Delta}}_0'\bar{\boldsymbol{\Delta}}\mathbf{H} = \text{block diag}\,[\mathbf{I}_r, \mathbf{U}]$ and $\mathbf{H}'\bar{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Gamma}}_0' = \text{block diag}\,[\mathbf{I}_r, \mathbf{V}]$. Then $\hat{\mathbf{F}} = \hat{\boldsymbol{\Delta}}_0\mathbf{D}_{(\hat{\phi},\mathbf{0})}\hat{\boldsymbol{\Gamma}}_0 = \bar{\boldsymbol{\Delta}}\mathbf{D}_{(\hat{\phi},\mathbf{0})}\bar{\boldsymbol{\Gamma}}$ and $g_i = u_i(\hat{\phi})$, $1 \leqslant i \leqslant r$.

5. The Bingham family—spectral version. Recall that if $\mathbf{A}$ is a symmetric $p \times p$ matrix it has a spectral decomposition $\mathbf{A} = \mathbf{M}\mathbf{D}_\zeta\mathbf{M}'$ with $\mathbf{M} \in O(p)$, $\mathbf{D}_\zeta = \text{diag}(\zeta_1, \cdots, \zeta_p)$, $\zeta_1 \leqslant \zeta_2 \leqslant \cdots \leqslant \zeta_p$. $\mathbf{D}_\zeta$ is unique and $\mathbf{M}$ is unique up to post-multiplication by $\mathbf{H} \in O(p)$ satisfying $\mathbf{H}\mathbf{D}_\zeta\mathbf{H}' = \mathbf{D}_\zeta$. If the ζ_i are distinct (as happens outside a set of Lebesgue measure zero) uniqueness of $\mathbf{M}$ can be ensured by insisting that the first nonzero element in each row of $\mathbf{M}$ be positive. We can reformulate Theorem 1(b) in spectral language, generalising results given by Bingham (1974, 1976) for $n = 1$.

THEOREM 4. *For the Bingham family on $G_n(\mathbb{R}^p)$, for $N \geqslant p$, with probability* 1 *there is a unique* MLE $\hat{\mathbf{A}}$ *of* $\mathbf{A}$. *If $t(\bar{\mathbf{Y}}) = \bar{\mathbf{M}}\mathbf{D_g}\bar{\mathbf{M}}'$, $\hat{\mathbf{A}} = \hat{\mathbf{M}}\mathbf{D}_{\hat{\zeta}}\hat{\mathbf{M}}'$ are the spectral decompositions, then*

$$(5.1) \qquad \hat{\mathbf{M}} = \bar{\mathbf{M}},\ g_i = v_i(\hat{\zeta}), \qquad 1 \leqslant i \leqslant p$$

where $\hat{\zeta} = (\hat{\zeta}_1, \cdots, \hat{\zeta}_p)$ and $v_i(\zeta) = \partial/\partial\zeta_i \log {}_1F_1(n/2; p/2; \mathbf{D}_\zeta)$.

PROOF. As in the proof of (3.2) one can readily show that $\tau_B(\boldsymbol{\Delta}\mathbf{A}\boldsymbol{\Delta}') = \boldsymbol{\Delta}\tau_B(\mathbf{A})\boldsymbol{\Delta}'$ for $\boldsymbol{\Delta} \in O(p)$ and that if $\mathbf{A}$ is diagonal then $\tau_B(\mathbf{A})$ is diagonal. For almost all $\bar{\mathbf{Y}}$, the eigenvalues of $t(\bar{\mathbf{Y}})$ are distinct, and by Theorem 1(b) for almost all $\bar{\mathbf{Y}}$, $t(\bar{\mathbf{Y}}) = \tau_B(\hat{\mathbf{A}}) = \hat{\mathbf{M}}\tau_B(\mathbf{D}_{\hat{\zeta}})\hat{\mathbf{M}}'$. The result follows from uniqueness of spectral decomposition. There is also a version for restricted MLEs given by Bingham (1974) in the case $n = 1, p = 3$.

THEOREM 5.

(i) *If $\mathbf{D}_\zeta$ is known, then $\bar{\mathbf{M}}$ is a* MLE *of* $\mathbf{M}$. *If the ζ_i are distinct then the* MLE *of* $\mathbf{M}$ *is unique.*

(ii) *If $\mathbf{M}$ is known and $N \geqslant p$, then, with probability* 1, *$\mathbf{D}_{\hat{\zeta}}$ exists and is unique. Put $\gamma_i = (\hat{\mathbf{M}}t(\bar{\mathbf{Y}})\hat{\mathbf{M}}')_{ii}$ and define $i_0 = 0$, $i_r = \max\{i : \gamma_i = \min_{j > i_{r-1}}\gamma_j\}$. Then $\hat{\zeta}_j = \hat{\zeta}_{i_r}$, $i_{r-1} < j \leqslant i_r$ and $(i_r - i_{r-1})^{-1}\Sigma_{k=i_{r-1}+1}^{i_r}\gamma_k = v_j(\hat{\zeta})$, $i_{r-1} < j \leqslant i_r$.*

PROOF. (i) The log-likelihood of $\mathbf{MD}_\zeta\mathbf{M}'$ at $\bar{\mathbf{Y}}$ is

$$\langle \mathbf{MD}_\zeta\mathbf{M}', t(\bar{\mathbf{Y}})\rangle - \kappa_B(\mathbf{D}_\zeta) = \langle \mathbf{D}_\zeta, \mathbf{M}'\overline{\mathbf{M}}\mathbf{D}_g\overline{\mathbf{M}}'\mathbf{M}\rangle - \kappa_B(\mathbf{D}_\zeta)$$

and this is maximised when $\mathbf{M}'\mathbf{M} = \mathbf{I}_p$, as can be seen by an argument using Lagrange multipliers. (See also Theobald, 1975).

(ii) This follows from Theorem 1(b) and a consideration of maxima on the boundary of the closed convex region $\hat{\zeta}_1 \leqslant \hat{\zeta}_2 \leqslant \cdots \leqslant \hat{\zeta}_p$. The important point is that τ_B is equivariant and order-preserving, i.e., if $\tau_B(\mathrm{diag}(\theta_1, \cdots, \theta_p)) = \mathrm{diag}(h_1, \cdots, h_p)$ then permutation of the θ_i induces the same permutation of the h_i and if $\theta_i < \theta_j$ then $h_i < h_j$.

REMARK. The criterion introduced on an intuitive basis by Mardia and Khatri (1977) for testing for uniformity on $G_n(\mathbb{R}^p)$ (and so on $O\cdot(n,p)$) is approximately the likelihood ratio test for uniformity among the Bingham family. More precisely, we have

$$S_{\mathrm{U}} = -2\log\lambda + 0\big(N\|t(\bar{\mathbf{Y}})\|^3\big)$$

where $-2\log\lambda = N\{\langle\hat{\mathbf{A}}, t(\bar{\mathbf{Y}})\rangle - \kappa_B(\hat{\mathbf{A}})\}$ is the likelihood ratio test statistic, $S_{\mathrm{U}} = Na^{-1}\|t(\bar{\mathbf{Y}})\|^2 = Na^{-1}(\mathrm{tr}(\bar{\mathbf{Y}}^2) - n^2/p)$ is the statistic of Mardia and Khatri and $a = 2n(p-n)/\{p(p-1)(p+2)\}$, and $\|t(\bar{\mathbf{Y}})\| = \langle t(\bar{\mathbf{Y}}), t(\bar{\mathbf{Y}})\rangle^{\frac{1}{2}}$. This can be seen by using expansions of $\kappa_B(\mathbf{A})$ in terms of zonal polynomials analogous to those of Section 4 of Mardia and Khatri's paper and the fact that

$$t(\bar{\mathbf{Y}}) = \frac{\partial}{\partial \mathbf{A}}\kappa_B(\mathbf{A})|_{\mathbf{A}=\hat{\mathbf{A}}}.$$

6. An example. Examples of data on Stiefel manifolds arise in astronomy, vectorcardiography, etc. One such example from astronomy is given by the orbits of comets. The orientation of a comet's orbit can be specified by the celestial longitude (L) and latitude (θ) of its perihelion and the longitude (ν) of its ascending node (thus specifying the sense of rotation). The direction of the perihelion is $\mathbf{x}_1 = (\cos\theta\cos L, \cos\theta\sin L, \sin\theta)$ and the directed unit normal to the orbit given by the right hand rule is

$$\mathbf{x}_2 = (\sin\theta\sin\nu, -\sin\theta\cos\nu, -\cos\theta\sin(\nu - L))/r$$

where $r^2 = \sin^2\theta + \cos^2\theta\sin^2(\nu - L)$. The orientation of the orbit can thus be specified by the matrix $\mathbf{X}$ in $O(2, 3)$ given by $\mathbf{X}' = (\mathbf{x}_1', \mathbf{x}_2')$. As there is, a priori, no symmetry, an appropriate model for the distribution of these matrices is the matrix von Mises-Fisher family. The likelihood ratio test for uniformity on $O(n, p)$ against the alternative of a von Mises-Fisher distribution uses the generalized Rayleigh statistic $\bar{R}$ where $\bar{R}^2 = \mathrm{tr}(\bar{\mathbf{X}}'\bar{\mathbf{X}})$. Under the hypothesis of uniformity $pN\bar{R}^2$ is distributed asymptotically as χ^2_{np}. This test was applied in an analysis of the orbits of $N = 240$ long-period comets in Marsden's (1972) catalogue (pages 62–63).

It was found that $3N\bar{R}^2 = 215$. As $P(\chi_6^2 > 215) < 10^{-4}$, the hypothesis of uniformity is rejected. Following astronomers (see Tyror, 1957), Mardia (1975) analysed the accretion theory of Lyttleton by using just the perihelion direction. However, the accretion theory should be tested using the above formulation since the normal to the orbit of each comet contains additional information relevant to the problem. For a practical example in vectorcardiography, leading to observations on $O(2, 3)$, see Downs (1972) and Mardia and Khatri (1977).

REFERENCES

[1] Barndorff-Nielsen, O. (1973). Exponential families and conditioning. Sc.D. Thesis, Univ. Copenhagen.

[2] Berk, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Statist.* **43** 193–204.

[3] Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.* **2** 1201–1225.

[4] Bingham, C. (1976). Expansions related to a hypergeometric function arising in an antipodally symmetric distribution on the sphere. Directional Data Project, Research Report 11, Depart. Statist., Leeds Univ.

[5] Downs, T. D. (1972). Orientation statistics. *Biometrika* **59** 665–676.

[6] Farrell, R. H. (1976). Techniques of multivariate calculation. *Lecture Notes in Mathematics* **520**. Springer-Verlag, Berlin.

[7] Jupp, P. E. and Mardia, K. V. (1977). Density-preserving statistics and densities for sample means. *Ann. Probability* **6** 688–694.

[8] Khatri, C. G. and Mardia, K. V. (1977). The von Mises-Fisher matrix distribution in orientation statistics. *J. Roy. Statist. Soc. Ser. B* **39** 95–106.

[9] Mardia, K. V. (1975). Statistics of directional data (with discussion). *J. Roy. Statist. Soc. Ser. B* **37** 349–393.

[10] Mardia, K. V. and Khatri, C. G. (1977). Uniform distribution on a Stiefel manifold. *J. Mult. Analysis* **7** 468–473.

[11] Marsden, B. G. (1972). *Catalogue of Cometary Orbits*. Smithsonian Astrophysical Laboratory.

[12] Nachbin, L. (1965). *The Haar Integral*. van Nostrand, Princeton.

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

[14] Theobald, C. M. (1975). An inequality for the trace of the product of two symmetric matrices. *Proc. Cambridge Philos. Soc.* **77** 265–267.

[15] Tyror, J. G. (1957). The distribution of the directions of perihelion of long-period comets. *Mon. Not. R. Astr. Soc.* **117** 369–379.

Department of Statistics
University of Leeds
Leeds LS2 9JT
England