Sankhyā : The Indian Journal of Statistics
1992, Volume 54, Series A, Pt. 1, pp. 123–127.

MAXIMUM LIKELIHOOD CHARACTERIZATION OF THE VON MISES-FISHER MATRIX DISTRIBUTION

By SUMITRA PURKAYASTHA

Indian Statistical Institute

and

RAHUL MUKERJEE*

Indian Institute of Management

SUMMARY. A characterization of the von Mises-Fisher matrix distribution, extending a result of Bingham and Mardia (1975) for distributions on sphere to distributions on Stiefel manifold, is obtained.

1. Introduction and main result

Bingham and Mardia (1975)—hereafter, abbreviated to BM—proved that under mild conditions a rotationally symmetric family of distributions on the sphere must be the von Mises-Fisher family if the mean direction is a maximum likelihood estimator (MLE) of the location parameter. In view of Downs' (1972) extension of the von Mises-Fisher distribution to a Stiefel mainfold (for further references, see Jupp and Mardia (1979)), it has been attempted here to extend the result in BM in the direction of Downs' work.

Let S_{np} be the class of $n \times p$ $(n \leqslant p)$ matrices $\boldsymbol{M}$ satisfying $\boldsymbol{M}\boldsymbol{M}' = \boldsymbol{I}_n$. For $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \in S_{np}$ with $\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{X}_i$ having full row rank, define the polar component of $\boldsymbol{X}$ as the matrix $(\boldsymbol{X}\boldsymbol{X}')^{-\frac{1}{2}}\boldsymbol{X}$(cf. Downs, 1972). Then the following result, proved in the next section, holds.

Theorem. *Let $\mathcal{F} = \{p(\boldsymbol{X}; \boldsymbol{A}) = f[tr(\boldsymbol{A}\boldsymbol{X}')] \mid \boldsymbol{A} \in S_{np}\}$ be a class of non-uniform densities on S_{np}. Assume that f is lower semi-continuous at the point n. Furthermore, suppose that for every positive integral N and for all random samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$, with $\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{X}_i$ of full row rank, the polar component of $\boldsymbol{X}$ is a MLE of $\boldsymbol{A}$. Then*

$$p(\boldsymbol{X}\,;\boldsymbol{A}) = K\,exp\{\lambda tr(\boldsymbol{A}\boldsymbol{X}')\},\ \boldsymbol{X} \in S_{np}, \qquad \ldots \quad (1.1)$$

for some constants λ and K, both positive.

Paper received. June 1989 ; revised May 1990.

AMS (1980) *subject classification.* 62E10, 62H05.

Key words and phrases. Maximum likelihood characterization, orientation statistics, von Mises-Fisher matrix distribution.

*On leave from Indian Statistical Institute, Calcutta, India.

Remark 1. The class $\mathscr{F}$ considered above has the following property. $p(\boldsymbol{X}; \boldsymbol{A}) = p(\boldsymbol{XB}; \boldsymbol{A})$ for all $p \times p$ orthogonal matrix $\boldsymbol{B}$ with det $(\boldsymbol{B}) = 1$ that satistics $\boldsymbol{AB} = \boldsymbol{A}$. Because of this geometric consideration the matrix $\boldsymbol{A}$ can be thought of as a location parameter for the class $\mathscr{F}$. Thus $\mathscr{F}$ is a natural extension of the class considered in BM.

Remark 2. The converse of the theorem is also true, i.e, if $\boldsymbol{X}$ has the density (1.1), then for i.i.d. observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ from $p(\boldsymbol{X}; \boldsymbol{A})$ the polar component of $\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{X}_i$ is the MLE of $\boldsymbol{A}$ (cf. Downs (1972)).

2. Proof of the theorem

For $n = 1$, our theorem follows from Theorem 2 in BM. Throughout this section, we therefore consider the case $n \geqslant 2$, and it appears that this generalization is non-trivial especially for odd n. Observe that the condition regarding the MLE of $\boldsymbol{A}$ is equivalent to the following : for every positive integral N and every choice of matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$, $\boldsymbol{A} \in S_{np}$ with $\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{X}_i$ of full row rank, the relation

$$\prod_{i=1}^{N} f[\mathrm{tr}(\hat{\boldsymbol{A}}\boldsymbol{X}_i')] \geqslant \prod_{i=1}^{N} f[\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}_i')] \quad \ldots \quad (2.1)$$

holds, where $\hat{\boldsymbol{A}} = (\boldsymbol{XX}')^{-\frac{1}{2}}\boldsymbol{X}$. The following lemmas will be helpful.

Lemma 1. *For every positive integral N and every choice of matrices $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N$, $U \in S_{nn}$ with $\boldsymbol{C} = \sum_{i=1}^{N} \boldsymbol{C}_i$ positive definite, the relation*

$$\prod_{i=1}^{N} f[tr(\boldsymbol{C}_i)] \geqslant \prod_{i=1}^{N} f[tr(\boldsymbol{U}\boldsymbol{C}_i)] \quad \ldots \quad (2.2)$$

holds.

Proof. Let $\boldsymbol{L} = (\boldsymbol{I}_n, \boldsymbol{0}) \in S_{np}$. Then the lemma follows from (2.1) taking $\boldsymbol{X}_i = \boldsymbol{C}_i'\boldsymbol{L}$, $1 \leqslant i \leqslant N$, and $\boldsymbol{A} = (\boldsymbol{U}, \boldsymbol{0}) \in S_{np}$.

Lemma 2. *For each $x \in [-n, n]$, $f(n) \geqslant f(x)$.*

Proof. Follows taking $N = 1$, $\boldsymbol{C}_1 = \boldsymbol{I}_n$ in (2.2) and observing that for each $u \in [-n, n]$, there exists $\boldsymbol{U} \in S_{nn}$ satisfying $\mathrm{tr}(U) = u$.

Lemma 3. *For each $x \in [-n, n]$, $f(x) < \infty$.*

Proof. In consideration of Lemma 2, it is enough to show that

$$f(n) < \infty. \quad \ldots \quad (2.3)$$

Taking $N = 2$, $U = C_1'$ in (2.2), we get $f[\mathrm{tr}(C_1)]f(\mathrm{tr}(C_2)] \geqslant f(n)f[\mathrm{tr}(C_1'C_2)]$, for every $C_1, C_2 \in S_{nn}$ such that C_1+C_2 is positive definite. Hence if (2.3) does not hold then $f(n) = \infty$, and for every $C_1, C_2 \in S_{nn}$ such that C_1+C_2 is positive definite, one must have either (a) $f[\mathrm{tr}(C_1'C_2)] = 0$, or (b) $f[\mathrm{tr}(C_1)]$ $f[\mathrm{tr}(C_2)] = \infty$.

For real α, u and positive integral m, define the matrices

$$H_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \quad Q_{m\alpha} = I_m \otimes H_\alpha, \quad Q^*_{m\alpha}(u) = \begin{pmatrix} Q_{m\alpha} & 0 \\ 0' & u \end{pmatrix},$$

Consider first the case of odd n. If $n = 2m+1 (m \geqslant 1)$ and (2.3) does not hold, then taking $C_1 = Q^*_{m\alpha}(1)$, $C_2 = Q^*_{m(-\alpha)}(1)$, $-\pi/2 < \alpha < \pi/2$ (note that then $C_1, C_2 \in S_{nn}$ and C_1+C_2 is positive definite), it follows from the discussion in the last paragraph that for each $\alpha \in (-\pi/2), \pi/2)$, either (a) $f(1+2m\cos 2\alpha) = 0$, or (b) $f(1+2m \cos\alpha) = \infty$. The condition (b) cannot hold over a set of positive Lebesgue measure. Hence (a) must hold almost everywhere (a.e.) over $\alpha \in (-\pi/2, \pi/2)$, i.e., $f(x) = 0$ a.e. over $x \in (-(2m-1), (2m+1))$ and a contradiction is reached in consideration of lower semicontinuity of f at the point $n(= 2m+1)$ (cf. (2.4) below). Similarly, for even $n(= 2m, m \geqslant 1)$, if (2.3) does not hold, then taking $C_1 = Q_{m\alpha}$, $C_2 = Q_{m(-\alpha)}$, $-\pi/2 < \alpha < \pi/2$, it follows as before that for each $\alpha \in (-\pi/2, \pi/2)$, either (a) $f(n\cos 2\alpha) = 0$, or (b) $f(n\cos\alpha) = \infty$, and a contradiction is reached again by the lower semicontinuity of f at n.

Lemma 4. *For each* $x \in [-n, n], f(x) > 0$.

Proof. First note that

$$f(n) > 0, \qquad \ldots \quad (2.4)$$

for otherwise by Lemma 2, $f(x) = 0$ for each $x \in [-n, n]$, which is impossible as f is a density. Also, observe that for any given $\theta \in [0, \pi]$, there exists η satisfying (cf. BM)

$$\text{(i) } -\tfrac{1}{2}\theta \leqslant \eta \leqslant 0, \text{ (ii) } \cos\theta + 2\cos\eta > 0, \text{ (iii) } \sin\theta + 2\sin\eta = 0. \ldots \quad (2.5)$$

Consider first the case of odd n. For $n = 2m+1 (m \geqslant 1)$, define

$$\mathscr{B} = \{\theta : \theta \in [0, \pi], f(1+2m\cos\theta) = 0\}.$$

If $\mathscr{B}$ is non-empty, then for each $\theta \in \mathscr{B}$, one can choose η satisfying (2.5) and then employ (2. 2) with $N = 3$, $C_1 = Q^*_{m\theta}(1)$, $C_2 = C_3 = Q^*_{m\eta}(1)$, $U = Q^*_{m\alpha}(1)$, where $\alpha = -(\theta+\eta)/2$, to obtain $f[1+2m\cos(\tfrac{1}{2}(\theta-\eta))] = 0$; but as in Lemma

2 in BM, because of (2.4) and lower semi-continuity of f at n, this leads to a contradiction. Hence $\mathscr{S}$ is empty and

$$f(x) > 0 \text{ for all } x \in [-(2m-1), (2m+1)]. \qquad \ldots \quad (2.6)$$

We shall now show that $f(x) > 0$ also for $x \in [-(2m+1), -(2m-1))$. If possible, let there exist $x_0 \in [-(2m+1), -(2m-1))$ such that $f(x_0) = 0$. Let $\theta(\epsilon[0, \pi])$ be such that $\cos\theta = (x_0+1)/(2m)$, and corresponding to this θ, find η satisfying (2.5). Taking $N = 3$, $C_1 = Q^*_{m\theta}(-1)$, $C_2 = C_3 = Q^*_{m\eta}(1)$, $U = Q^*_{m(-\theta)}(1)$ in (2.2), and using Lemma 3, one then gets $f(2m-1)\{f[1+2m\cos(\eta-\theta)]\}^2 = 0$, which is impossible by (2.6). This proves the lemma for odd n. The proof for even n is similar.

Lemma 5. *For every positive integral N' and every choice of matrices $C_1, \ldots, C_N$, $U \in S_{nn}$ with $\sum_{i=1}^{N'} C_i$ non-negative definite, the relation*

$$\prod_{i=1}^{N'} f[tr(C_i)] \geqslant \prod_{i=1}^{N'} f[(tr(UC_i)]$$

holds.

Proof. In view of Lemma 1, it is enough to consider the case when $C = \sum_{i=1}^{N'} C_i$ is positive semidefinite. Obviously, then $I+\nu C$ is positive definite for every positive integral ν. In Lemma 1, now take $N = 1+\nu N'$, and choose the C_i's such that one of them equals I and the rest are given by ν copies of each of $C_1 \ldots, C_N$. The rest of the proof follows using agruments similar to those in Lemma 3 in BM.

We now proceed to the final step of our proof. For $n = 2m+1$ $(m \geqslant 1)$, in Lemma 5 taking $N' = N$, $C_i = Q^*_{m\theta_i}(1)$ $(1 \leqslant i \leqslant N)$, $U = Q^*_{m(-\alpha)}(1)$, where

$$\sum_{i=1}^{N} \cos\theta_i \geqslant 0, \ \sum_{i=1}^{N} \sin\theta_i = 0, \qquad \ldots \quad (2.7)$$

it follows that for every positive integral N and for every α, $\prod_{i=1}^{N} f(1+2m\cos\theta_i) \geqslant \prod_{i=1}^{N} f(1+2m\cos(\theta_i-\alpha))$, whenever the θ_i's satisfy (2.7). Writing $h(\theta) = \log f(1+2m\cos\theta)$, which is well-defined by Lemmas 3.4, it follows that for each positive integral N and each α,

$$\sum_{i=1}^{N} h(\theta_i) \geqslant \sum_{i=1}^{N} h(\theta_i-\alpha), \qquad \ldots \quad (2.8)$$

whenever the θ_i's satisfy (2.7). The relation (2.8) is equivalent to the relation (4) in BM and hence as in BM, $h(\theta) = a\cos\theta + b$, for every θ, where $a(\geqslant 0)$ and b are some constants. By the definition of $h(\theta)$, one obtains

$$f(x) = K\exp(\lambda x),\ \text{for } x \in [-(2m-1), (2m+1)] \qquad \ldots \quad (2.9)$$

where $K(>0)$ and $\lambda(\geqslant 0)$ are constants. By Lemma 5, for every $C, U \in S_{nn}$, $f[\text{tr}(C)]f[-\text{tr}(C)] \geqslant f[\text{tr}(UC)]f[-\text{tr}(UC)]$, so that $f(x)f(-x)$ remains constant over $x \in [-n, n]$. This, together with (2.9), implies that $f(x) = K\exp(\lambda x)$, for each $x \in [-n, n]$, where K, λ are constants, both positive, the positiveness of λ being a consequence of the stipulated non-uniformity of f. This proves the theorem for odd n. The proof for even n is similar.

Acknowledgement. The authors are thankful to a referee for very constructive suggestions.

REFERENCES

BINGHAM, M. S. and MARDIA, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. In : *Statistical Distributions in Scientific Work*, vol. 3 (G. P. Patil *et al.* eds.), Reidel, Dordrecht-Holland, 387-398.

DOWNS, T. D. (1972). Orientation statistics. *Biometrika*, 59, 665-676.

JUPP, P. E. and MARDIA, K. V. (1979). Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions. *Ann. Statist*, 7, 599-606.

STAT-MATH DIVISION
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
CALCUTTA 700 035
INDIA.

INDIAN INSTITUTE OF MANAGEMENT, CALCUTTA
JOKA, DIAMOND HARBAR ROAD
POST BOX NO 16757
CALCUTTA 700027
INDIA.