Sankhyā : The Indian Journal of Statistics
1991, Volume 53, Series A, Pt. 1, pp. 70-83.

A ROTATIONALLY SYMMETRIC DIRECTIONAL DISTRIBUTION : OBTAINED THROUGH MAXIMUM LIKELIHOOD CHARACTERIZATION

SUMITRA PURKAYASTHA

Indian Statistical Institute

SUMMARY. A circularly symmetric directional distribution is obtained by showing that in the class of circularly symmetric distributions on circle it is the only distribution for which the circular median is a maximum likelihood estimate of the location parameter. Subsequently, quently, this result is extended to the spherical case.

1. Introduction

Teicher (1961) proved that under very mild conditions a translation parameter family of distributions on the real line must be normal if the sample mean is a maximum likelihood estimate of the translation parameter. Later, Ghosh and Rao (1971) solved the same problem with 'sample mean' replaced by 'sample median' and obtained a characterization of the Laplace distribution. A proof of the latter result may be found in Kagan *et al.* (1973, 413–414).

The above two results in linear data were followed by a result of Bingham and Mardia (1975) in directional data which states that under mild conditions a rotationally symmetric family of densities on the sphere must be the von Mises—Fisher family if the mean direction is a maximum likelihood estimate of the location parameter.

Wth the above mentioned results in mind, our aim in this paper is to characterize that rotationally symmetric directional distribution for which the median direction is a maximum likelihood estimate of the location parameter. In Section 2, we settle this problem for distributions on circle (Theorem 2.1). This result is extended to higher dimensional spheres in Section 3 (Theorem 3.1). Finally, some general remarks in this context appear in Section 4.

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

Key words and phrases. Circular symmetry, maximum likelihood characterization, median direction, rotational symmetry.

2. The circular case

Definition 2.1. Let $x_1, \ldots, x_n \in S^1$. Then any point $x_0 \in S^1$ is called a circular median of $x_1, \ldots, x_n$ if

$$\sum_{i=1}^{n} \cos^{-1}(x_i' x_0) = \min_{\xi \in S^1} \sum_{i=1}^{n} \cos^{-1}(x_i' \xi). \quad \ldots \quad (2.1)$$

[For $-1 \leqslant x \leqslant 1$, $\cos^{-1}(x)$ is the unique angle $\theta \in [0, \pi]$ such that $\cos \theta = x$].

Remark 2.1. The sum appearing in the right hand side of (2.1) is a continuous function in ξ so that it makes sense to talk of its minimum and define x_0 accordingly. Observe, however, that x_0 may not be unique.

Remark 2.2. The Definition 2.1 is actually the circular analogue of the spherical median given in Fisher (1985). It is also related to the one given in Mardia (1972, 28–33).

Remark 2.3. Because circular median may not be unique, we adopt the following conventions about the choice of median direction for sample sizes $n = 2$, 3 and 4 respectively. This choice is motivated by the natural requirements of a measure of central tendency.

Notation. For two points $a, b \in S^1$; $[a, b]$ denotes the arc of S^1 with initial point a, end point b and taken in clockwise sense.

A. *Sample size* $n = 2$. Assume, without loss of generality, that length of $[x_1, x_2] \leqslant$ length of $[x_2, x_1]$. Then in this case the sum appearing in the right hand side of (2.1) remains constant for $\xi \in [x_1, x_2]$, and moreover $\xi \in [x_1, x_2]$, $\mu \in S_1 - [x_1, x_2]$ implies

$$\sum_{i=1}^{2} \cos^{-1}(x_i' \xi) \leqslant \sum_{i=1}^{2} \cos^{-1}(x_i' \mu)$$

Therefore, we agree to take the mid-point of $[x_1, x_2]$, which is the same as the mean direction, as the median direction.

B. *Sample size* $n = 3$. Write $\cos^{-1}(x_1' x_2) = \alpha$ and $\cos^{-1}(x_2' x_3) = \beta$. Then, $0 \leqslant \alpha \leqslant \pi$ and $0 \leqslant \beta \leqslant \pi$. Assume, without loss of generality, that either (a) $0 \leqslant \alpha + \beta \leqslant \pi$ or (b) $\alpha \leqslant \beta$, $\alpha + \beta > \pi$ and $\alpha + 2\beta \leqslant 2\pi$.

The two cases are illustrated in the following figures :

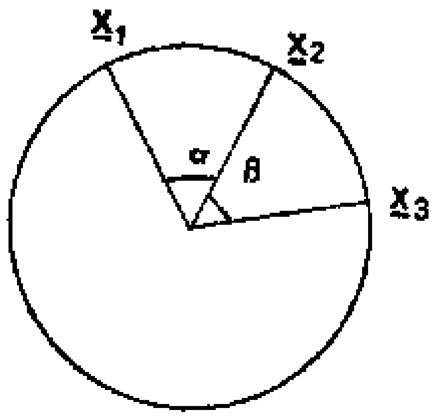

Fig. 2.1 (a)

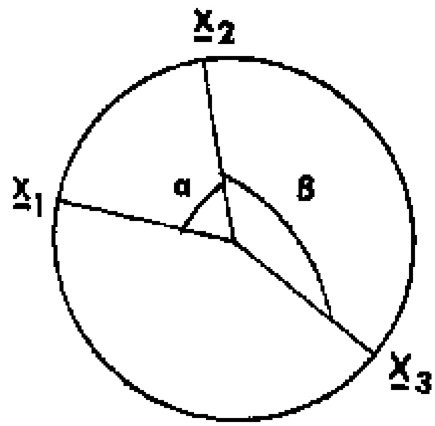

Fig. 2.1 (b)

It is now easy to see that in each of the cases

$$\left\{\mu \in S^1 : \sum_{i=1}^{3} \cos^{-1}(x_i' \mu) = \min_{\xi \in S^1} \sum_{i=1}^{3} \cos^{-1}(x_i' \xi)\right\} = \{x_2\}.$$

Hence, in both the cases, we take x_2 as the median directiim.

C. *Sample size* $n = 4$. Write $\cos^{-1}(x_i' x_{i+1}) = \alpha_i$, $1 \leqslant i \leqslant 3$. Then, $0 \leqslant \alpha_i \leqslant \pi$ for every i. Assume, without loss of generality, that either (a) $0 \leqslant \alpha_1+\alpha_2+\alpha_3 \leqslant \pi$ or (b) $\pi < \alpha_1+\alpha_2+\alpha_3 \leqslant 2\pi$, $\alpha_1+\alpha_2 \leqslant \pi$ and $\alpha_2+\alpha_3 \leqslant \pi$. The two cases are illustrated in the following figures :

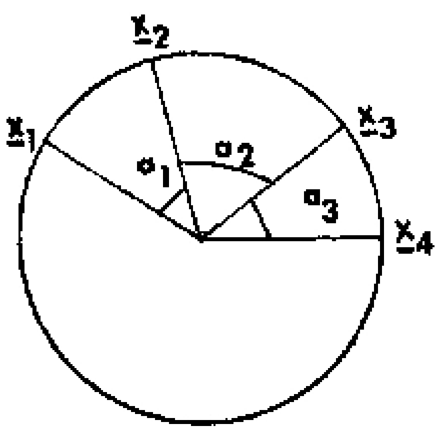

Fig. 2.2 (a)

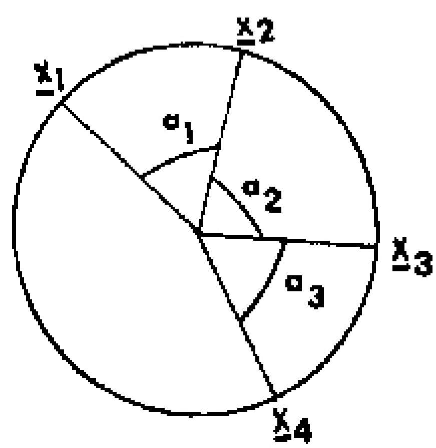

Fig. 2.2 (b)

It turns out that in each of the cases

$$\left\{\mu \in S^1 : \sum_{i=1}^{4} \cos^{-1}(x_i' \mu) = \min_{\xi \in S^1} \sum_{i=1}^{4} \cos^{-1}(x_i' \xi)\right\} = [x_2, x_3]$$

We agree to take the mid-point of $[x_2, x_3]$ as the median direction.

Theorem 2.1. *Let $\{p(x\,;\,\theta) = f(x'\theta) \mid \theta \in S^1\}$ be a class of circularly symmetric non-uniform densities on S^1. Suppose $f(t) > 0$ for every $t \in (-1,\ 1)$ and moreover $f(t)$ is right-continuous at $t = -1$. If the median direction be a maximum likelihood estimate of θ for $n = 4$ samples, then*

$$p(x\,;\,\theta) = \frac{a}{2(1-e^{-a\pi})}\, e^{-a\cos^{-1}(x'\theta)},\ x \in S^1,\ a > 0. \qquad \ldots \quad (2.2)$$

Proof. The fact that the median direction is a maximum likelihood estimate of θ implies

$$\prod_{i=1}^{4} f(x_i'\, x_0) \geqslant \prod_{i=1}^{4} f(x_i'\,\theta)\ \forall\ \theta \in S^1, \qquad \ldots \quad (2.3)$$

and for all samples $(x_1, \ldots, x_4)$ of size $n = 4$; x_0 being the median direction.

Write $\cos^{-1}(x_i' x_{i+1}) = \alpha_i$ for $i = 1,\ 2,\ 3$. Define $g(t) = f(\text{Cos}\ t)$, $0 \leqslant t \leqslant \pi$. Then, our choice of the median direction (described in part C of Remark 2.3) and (2.3) above, applied to several choices of θ, imply the following :

for every $0 \leqslant \alpha_1, \alpha_2, \alpha_3 \leqslant \pi$ with $0 \leqslant \alpha_1+\alpha_2+\alpha_3 \leqslant \pi$,

$$g\left(\alpha_1+\frac{\alpha_2}{2}\right) g^2\left(\frac{\alpha_2}{2}\right) g\left(\frac{\alpha_2}{2}+\alpha_3\right)$$

$$\geqslant \begin{cases} g\left(\alpha_1+\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}-x\right) g\left(\frac{\alpha_2}{2}+\alpha_3-x\right), \\ \qquad\qquad 0 \leqslant x \leqslant \frac{\alpha_2}{2} \qquad \ldots \quad (2.4.1) \\ g\left(\alpha_1+\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}+x\right) g\left(x-\frac{\alpha_2}{2}\right) g\left(\frac{\alpha_2}{2}+\alpha_3-x\right), \\ \qquad\qquad \frac{\alpha_2}{2} \leqslant x \leqslant \frac{\alpha_2}{2}+\alpha_3 \qquad \ldots \quad (2.4.2) \\ g\left(\alpha_1+\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}+x\right) g\left(x-\frac{\alpha_2}{2}\right) g\left(x-\frac{\alpha_2}{2}-\alpha_3\right), \\ \qquad\qquad \frac{\alpha_2}{2}+\alpha_3 \leqslant x \leqslant \pi-\left(\alpha_1+\frac{\alpha_2}{2}\right) \qquad \ldots \quad (2.4.3) \end{cases}$$

for every $0 \leqslant \alpha_1, \alpha_2, \alpha_3 \leqslant \pi$ with $\pi < \alpha_1+\alpha_2+\alpha_3 \leqslant 2\pi$, $\alpha_1+\alpha_2 \leqslant \pi$ and $\alpha_2+\alpha_3 \leqslant \pi$,

$$g\left(\alpha_1+\frac{\alpha_2}{2}\right) g^2\left(\frac{\alpha_2}{2}\right) g\left(\frac{\alpha_2}{2}+\alpha_3\right)$$

$$\geqslant g\left(\alpha_1+\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}+x\right) g\left(\frac{\alpha_2}{2}-x\right) g\left(\frac{\alpha_2}{2}+\alpha_3-x\right), 0 \leqslant x \leqslant \frac{\alpha_2}{2}. \quad \ldots \quad (2.5)$$

Observe that if we choose $\alpha_1 = \alpha_3 = 0$ in (2.4.1), we obtain

$$g^4\left(\frac{\alpha_2}{2}\right) \geqslant g^2\left(\frac{\alpha_2}{2}+x\right) g^2\left(\frac{\alpha_2}{2}-x\right) \text{ for } 0 \leqslant x \leqslant \frac{\alpha_2}{2}$$

$$\Rightarrow g^2\left(\frac{\alpha_2}{2}\right) \geqslant g(y)\, g(\alpha_2-y) \qquad \text{for } 0 \leqslant y \leqslant \alpha_2 \quad \ldots \quad (2.6)$$

Therefore, if $g(y) = \infty$ for some $0 \leqslant y < \pi$, then (2.6) implies that for every $\alpha_2 \in (y, \pi)$ either $g(\alpha_2-y) = 0$ or $g\left(\frac{\alpha_2}{2}\right) = \infty$. The former condition cannot be satisfied because of the restriction on f put forth in the statement of the theorem and the latter condition cannot be satisfied on a set of positive lebesgue measure. Therefore,

$$g(t) < \infty \text{ for every } 0 \leqslant t < \pi \quad \ldots \quad (2.7)$$

We also have $g(t) < 0$ for every $0 < t < \pi$. Observe further that if we choose $\alpha_1 = \alpha_2 = \alpha_3 =$ in (2.4.3), we obtain

$$g^4(0) \geqslant g^4(x), \quad \text{for every } 0 \leqslant x \leqslant \pi,$$

which implies

$$g(0) > 0,$$

oterwise $g(x) = 0$ for all $0 \leqslant x \leqslant \pi$, a contradiction to the fact that f is a density on S^1. Therefore, we can define

$$h(t) = \log g(t),\ 0 \leqslant t < \pi. \quad \ldots \quad (2.8)$$

With little modification, the conditions (2.4.1), (2.4.2) and (2.5), alongwith (2.8), now imply the following :

for every $0 \leqslant \alpha_1, \alpha_3 < \pi$, $0 \leqslant \alpha_2 < \pi$ with $0 \leqslant \alpha_1+\alpha_2+\alpha_3 < \pi$,

$$h\left(\alpha_1+\frac{\alpha_2}{2}\right)+2h\left(\frac{\alpha_2}{2}\right)+h\left(\frac{\alpha_2}{2}+\alpha_3\right)$$

$$\geqslant \begin{cases} h\left(\alpha_1+\frac{\alpha_2}{2}+x\right)+h\left(\frac{\alpha_2}{2}+x\right)+h\left(\frac{\alpha_2}{2}-x\right)+h\left(\frac{\alpha_2}{2}+\alpha_3-x\right), \\ \qquad\qquad 0 \leqslant x \leqslant \frac{\alpha_2}{2} \quad \ldots \quad (2.9.1) \\ h\left(\alpha_1+\frac{\alpha_2}{2}+x\right)+h\left(\frac{\alpha_2}{2}+x\right)+h\left(x-\frac{\alpha_2}{2}\right)+h\left(\frac{\alpha_2}{2}+\alpha_3-x\right), \\ \qquad\qquad \frac{\alpha_2}{2} \leqslant x \leqslant \frac{\alpha_2}{2}+\alpha_3 \quad \ldots \quad (2.9.2) \end{cases}$$

for every $0 < \alpha_1, \alpha_2, \alpha_3 < \pi$ with $\pi < \alpha_1+\alpha_2+\alpha_3 < 2\pi$, $\alpha_1+\alpha_2 < \pi$ and $\alpha_2+\alpha_3 < \pi$,

$$h\left(\alpha_1+\frac{\alpha_2}{2}\right)+2h\left(\frac{\alpha_2}{2}\right)+h\left(\frac{\alpha_2}{2}+\alpha_3\right)$$

$$\geqslant h\left(\alpha_1+\frac{\alpha_2}{2}+x\right)+h\left(\frac{\alpha_2}{2}+x\right)+h\left(\frac{\alpha_2}{2}-x\right)+h\left(\frac{\alpha_2}{2}+\alpha_3-x\right), 0 \leqslant x \leqslant \frac{\alpha_2}{2} \quad \ldots \quad (2.10)$$

With the previous steps in mind, we now proceed to the main steps of our proof.

At first we prove that

$$h \text{ is concave on } \left[0, \frac{\pi}{2}\right]. \quad \ldots \quad (2.11)$$

To see this, choose and fix $t_1, t_2 \in \left[|0, \frac{\pi}{2}\right]$ with $t_1 < t_2$. Now use (2.9.1) with $\alpha_1 = \alpha_3 = 0$, $\alpha_2 = t_1+t_2$ and $x = \frac{t_2-t_1}{2}$ to obtain

$$h\left(\frac{t_1+t_2}{2}\right) \geqslant \frac{1}{2}\{h(t_1)+h(t_2)\},$$

establishing (2.11).

Next we prove that

$$h(t) = -at+b \text{ for } 0 \leqslant t < \frac{\pi}{2}, \quad \ldots \quad (2.12)$$

for two constants a and b.

As a consequence of (2.11), we know that h is differentiable on $\left(0, \frac{\pi}{2}\right)$ except (possibly) on a subset $\mathscr{B}$ of $\left(0, \frac{\pi}{2}\right)$, which is at most countable. Write $\mathscr{A} = \left(0, \frac{\pi}{2}\right) - \mathscr{B}$. Choose and fix $t_1, t_2 \in \mathscr{A}$ with $t_1 < t_2$. Choose moreover $t_3 \in \mathscr{A}$ with $t_3 < t_1$. In (2.9.1), take now $\alpha_1 = t_1 - t_3$, $\alpha_2 = 2t_3$ and $\alpha_3 = t_1 - t_3$ to obtain

$$h(t_1) + 2h(t_3) + h(t_2)$$
$$\geqslant h(x+t_1) + h(t_3+x) + h(t_3-x) + h(t_2-x),\ 0 \leqslant x \leqslant t_3.$$

Thus the function $h^* : [0, t_3] \to \mathscr{R}$ defined by

$$h^*(x) = h(x+t_1) + h(t_3+x) + h(t_3-x) + h(t_2-x)$$

is maximized at $x = 0$. Moreover, h is differentiable at each of t_1, t_2 and t_3. Hence,

$$\lim_{x \to 0+} \frac{h^*(x) - h^*(0)}{x} \leqslant 0$$
$$\Longrightarrow h'(t_1) \leqslant h'(t_2).$$

Interchanging the role of t_1 and t_2 in the argument above, we obtain

$$h'(t_2) \leqslant h'(t_1),$$

and consequently, for every $t_1, t_2 \in \mathscr{A}$ we obtain

$$h'(t_1) = h'(t_2).$$

This implies, in view of the concavity of h on $\left(0, \frac{\pi}{2}\right)$ and the definition of $\mathscr{A}$, that h is differentiable everywhere on $\left(0, \frac{\pi}{2}\right)$ with a constant derivative. Thus

$$h(t) = -at + b,\ 0 < t < \frac{\pi}{2}. \qquad \ldots \quad (2.13)$$

To complete the proof of (2.12), we how prove that $h(0) = b$. To see this, first choose three small positive numbers $\alpha_1, \alpha_2, \alpha_3$, and for this choice of α_i's, use (2.9.1) with $x = \frac{\alpha_2}{2}$ to obtain

$$h\left(\alpha_1 + \frac{\alpha_2}{2}\right) + 2h\left(\frac{\alpha_2}{2}\right) + h\left(\frac{\alpha_2}{2} + \alpha_3\right) \geqslant h(\alpha_1+\alpha_2) + h(\alpha_2) + h(0) + h(\alpha_3),$$

which implies, in view of (2.13),

$$b \geqslant h(0).$$

Similarly, if we choose two small positive numbers α_1, α_3, and $\alpha_2 = 0$, and for this choice of α_i's, use (2.9.2) with $x = \frac{\alpha_3}{2}$ to obtain

$$h(0) \geqslant b.$$

Thus (2.12) follows.

Next we prove that

$$h \text{ is concave on } \left(\frac{\pi}{2}, \pi\right) \qquad \dots \quad (2.14)$$

Choose and fix $t_1, t_2 \in \left(\frac{\pi}{2}, \pi\right)$ such that $t_1 < t_2$. Now use (2.10) with $\alpha_1 = \alpha_3 = t_1$, $\alpha_2 = t_2 - t_1$, and $x = \frac{t_2 - t_1}{2}$ to obtain

$$2h\left(\frac{t_1+t_2}{2}\right) + 2h\left(\frac{t_2-t_1}{2}\right) \geqslant h(t_2)+h(t_2-t_1)+h(0)+h(t_1),$$

which implies (2.14), since from (2.12) we have

$$2h\left(\frac{t_2-t_1}{2}\right) = h(t_2-t_1)+h(0).$$

The next assertion is analogous to (2.12). We prove that

$$h(t) = -ct+d \text{ for } \frac{\pi}{2} < t < \pi, \qquad \dots \quad (2.15)$$

for two constants c and d. This is an immediate consequence of (2.14) and of arguments similar to those required to establish (2.12).

Now we prove that

$$h \text{ is convex on } \left(\frac{\pi}{2}-\delta, \frac{\pi}{2}+\delta\right), \qquad \dots \quad (2.16)$$

if we choose δ to be a sufficienty small positive number, say $\delta = \frac{\pi}{8}$. So see this choose and fix $t_1, t_2 \in \left(\frac{\pi}{2}-\delta, \frac{\pi}{2}+\delta\right)$ with $t_1 < t_2$. Choose now $\alpha_1, \alpha_2, \alpha_3 \in (0, \pi)$ such that $t_1 = \alpha_1 + \frac{\alpha_2}{2}$, $t_2 = \alpha_3 + \frac{\alpha_2}{2}$, $\alpha_3 - \alpha_1 < \alpha_2$, $\alpha_1 + \alpha_2 < \pi$, and $\alpha_2 + \alpha_3 < \pi$; such a choice is possible since δ is assumed to ho a small positive number. Now with these quantities $\alpha_1, \alpha_2, \alpha_3$ and $x = \frac{\alpha_3 - \alpha_1}{2}$ use (2.9.1) if $t_1 + t_2 \leqslant \pi$, (2.10) if $t_1 + t_2 > \pi$ to obtain

$$h(t_1)+2h\left(\frac{\alpha_2}{2}\right)+h(t_2) \geqslant 2h\left(\frac{t_1+t_2}{2}\right)+h\left(\frac{\alpha_2}{2}+\frac{\alpha_3-\alpha_1}{2}\right)+h\left(\frac{\alpha_2}{2}-\frac{\alpha_3-\alpha_1}{2}\right). \qquad \dots \quad (2.17.1)$$

However, we may choose α_i's in a way so tuat both $\frac{\alpha_2}{2}-\frac{\alpha_3-\alpha_1}{2}$ and $\frac{\alpha_2}{2}+\frac{\alpha_3-\alpha_1}{2}$ are in $(0, \frac{\pi}{2})$, which implies, in view of (2.12),

$$2h\left(\frac{\alpha_2}{2}\right) = h\left(\frac{\alpha_2}{2}+\frac{\alpha_3-\alpha_1}{2}\right)+h\left(\frac{\alpha_2}{2}-\frac{\alpha_3-\alpha_1}{2}\right). \quad \ldots \quad (2.1.72)$$

The assrtion (2.16) is now an immediate consequence of (2.17.1) and (2.17.2)

Employing arguments simitar to those required to establish (2,12), we now obtain from (2.9.1), (2.10) and (2.16) that the graph of h on $\left(\frac{\pi}{2}-\delta, \frac{\pi}{2}+\delta\right)$ is a straight line. In view of (2.12) and (2.15), this implies

$$a = c, \quad b = d,$$

where a, b, c, d are the constants obtained in (2.12) and (2·15).

We have thus proved that

$$h(t) = -at+b \text{ for } 0 \leqslant t < \pi$$

i.e.

$$g(t) = \exp(-at+b) \text{ for } 0 \leqslant t < \pi,$$

for two constants a and b. Morover, right continuity of $f(t)$ at $t = -1$ inplies left-continuity of $g(t)$ at $t = \pi$. Hence,

$$g(t) = \exp(-at+b) \text{ for } 0 \leqslant t \leqslant \pi.$$

Observe now that in view of the fact that $g(0) \geqslant g(x)$ for every $x \in [0, \pi]$ and stipulated non-uniformity of f, we obtain.

$$a > 0$$

From what we have done so far it is now clear that

$$p(x\,;\theta) = e^b\, e^{-a \cos^{-1}(x'\theta)},\ x \in S^1,\ a > 0.$$

The fact that $e^b = \frac{a}{2(1-e^{-a\pi})}$ is an easy exercise in integration and so we omit it. This completes the proof of the theorem.

Remark 2.4. The theorem is false if we require the median direction to be a maximum likelihood estimate of θ for $n = 0$ samples. To see this, consider the following class $\mathcal{F}_1$ of circularly symmetric non-uniform densities on S^1 :

$$\mathcal{F}_1 = \{p(x\,;\theta) = f(x'\theta) \mid \theta \in S^1\},$$

where

$$f(t) = K \exp\{h(\text{Cos}^{-1} t)\},\ -1 \leqslant t \leqslant 1$$

with

$$h : [0, \pi] \to \mathcal{R} \text{ being defined as}$$

$$h(u) = -au+b,\ 0 \leqslant u \leqslant \frac{\pi}{2},$$

$$= -cu+d,\ \frac{\pi}{2} \leqslant u \leqslant \pi,$$

where $c > a > 0$, $\frac{\pi}{2}\ (c-a) = d-b$ and

$$\frac{1}{K} = \frac{2e^{b}\left(1-e^{-\frac{a\pi}{2}}\right)}{a} + \frac{2e^{d}\left(e^{-\frac{c\pi}{2}} - e^{-c\pi}\right)}{c}.$$

Obviously f is not of the form as described in (2.2) and moreover it is easy to check that with this choice of f, the median direction is indeed a maximum likelihood estimate of θ.

Remark 2.5. The theorem is false if we require the median direction to be a maximum likelihood estimate of θ for $n = 3$ samples. To see this, consider the following class $\mathcal{F}_2$ of densities on S^1 :

$$\mathcal{F}_2 = \{p(x\,;\,\theta) = f(x'\,\theta) \mid \theta \in S^1\}$$

where

$$f(t) = K \exp \{h(\mathrm{Cos}^{-1}\, t)\},\ -1 \leqslant t \leqslant 1$$

with

$$h : [0, \pi] \to \mathcal{R} \text{ being defined as}$$

$$h(u) = au^2+bu+c,$$

where $a > 0$, $2a\pi+b < 0$ and

$$\frac{1}{K} = 2 \int_0^{\pi} \exp\,(au^2+bu+c)\, du.$$

Routine algebraic computation now leads to the fact that $\mathcal{F}_2$ serves as a counter-example to the assertion of Theorem 2.1.

Remark 2.6. In Remark 2.3, we have seen how to get rid of the non-uniqueness of median direction by choosing the median direction in a meaningful way. It should be pointed out that for the assertion of Theorem 2.1 to hold this choice is crucial. In fact, if we require any median direction (i.e. any point on S^1 satisfying Definition 2.1) to be a maximum likelihood estimate of θ, then even with $n = 2$ samples Theorem 2.1 holds true so that Theorem 2.1 with $n = 4$ samples follows immediately and the counter-example described in Remark 2.4 ceases to be one.

In this context it is also worthwhile to mention that the theorem of Ghosh and Rao (vide Kagan *et al.* (1973), 413-414)) depends crucially on their choice of median and if this special care is not taken then their counter-example (vide Ghosh and Rao (1971)) ceases to be one. This fact also serves as a motivation for our choice of median direction described in Remark 2.3.

3. The spherical case

We shall discuss only the case with S^2. For S^p with $p > 2$, the discussion is essentially same.

Definition 3.1. Let $x_1, \ldots, x_n \in S^2$. Then any point $x_0 \in S^2$ is called a spherical median of $x_1, \ldots, x_n$ if

$$\sum_{i=1}^{n} \cos^{-1}(x_i' x_0) = \min_{\xi \in S^2} \sum_{i=1}^{n} \cos^{-1}(x_i' \xi) \qquad \ldots \quad (3.1)$$

Remark 3.1. The Definition 3.1 is due to Fisher (1985).

The extension of Theorem 2.1 to S^2 poses some special problem since the location of one possible median direction for every sample of size $n = 4$ becomes difficult. However, it turns out that in order to prove a theorem analogous to Theorem 2.1 for S^2 it is enough to consider all possible sample of size $n = 4$ lying on some great circle. The following remark regarding the convention about choice of median direction for samples from S^2 is worth mentioning.

Remark 3.2. A. *Sample size* $n = 2$. Suppose $x_1, x_2 \in S^2$. Denote by C, the great circle passing through x_1 and x_2. Moreover, $[x_1, x_2]$ denotes the arc connecting x_1, x_2 and taken along C. Suppose the length of $[x_1, x_2] \leqslant$ length of $C - [x_1, x_2]$. Then in this case the sum appearing in the right-hand side of (3.1) remains constant for $\xi \in [x_1, x_2]$, and moreover $\xi \in [x_1, x_2]$, $\mu \in S^2 - [x_1, x_2]$ implies

$$\sum_{i=1}^{2} \cos^{-1}(x_i'\xi) \leqslant \sum_{i=1}^{2} \cos^{-1}(x_i'\mu).$$

Therefore, we agree to take the mid-point of $[x_1, x_2]$, which is the same as the mean direction, as the median direction.

B. *Sample size* $n = 4$. Suppose $x_1, \ldots, x_4 \in S^2$ are such that $x_1, \ldots, x_4$ lie on a great circle, say C. Denote the circular median of $x_1, \ldots, x_4$ by x_0. Then, by an argument similar to that in Part A above it makes sense to choose x_0 as the spherical median of $x_1, \ldots, x_4$.

In view of our statement and proof of Theorem 2.1 and Remark 3.2 above, the following theorem is now immediate.

Theorem 3.1. *Let* $\{p(x\,;\,\theta) = f(x'\,\theta) \mid \theta \in S^2\}$ *be a class of rotationally symmetric non-uniform densities on* S^2. *Suppose* $f(t) > 0$ *for every* $t \in (-1, 1)$ *and moreover* $f(t)$ *is right continuous at* $t = -1$. *If the median direction be a maximum likelihood estimate of* θ *for* $n = 4$ *samples, then*

$$p(x\,;\,\theta) = \frac{a^2+1}{2\pi(1+e^{-a\pi})}\, e^{-a\cos^{-1}(x'\theta)}\,;\; x \in S^2,\, a > 0. \qquad \ldots \quad (3.2)$$

Remark 3.3. For S^p with $p > 2$, the density obtained in (3.2) is as follows :

$$p(x\,;\,\theta) = \frac{\Gamma\left(\frac{p}{2}\right) \cdot I_{p-1}(a)}{\pi^{p/2}} \cdot e^{-a\cos^{-1}(x'\theta)}\,;\; x \in S^p,\, \theta \in S^p,\, a > 0, \qquad \ldots \quad (3.3)$$

where

$$I_n(a) = \frac{(a^2+2^2)(a^2+4^2)\ldots(a^2+n^2)a}{n\,!\,(1-e^{-a\pi})} \text{ if } n \text{ is even}$$

$$= \frac{(a^2+1^2)\ldots(a^2+n^2)}{n\,!\,(1+e^{-a\pi})} \qquad \text{if } n \text{ is odd.} \qquad \ldots \quad (3.4)$$

Remark 3.4. Theorem 3.1 is false if we require the median direction to be a maximum likelihood estimate of θ for $n = 2$ samples. To see this, consider the following class $\mathcal{F}$ of rotationally symmetric non-uniform densities on S^2 :

$$\mathcal{F} = \{p(x\,;\,\theta) = f(x'\,\theta) \mid \theta \in S^2\}$$

where

$$f(t) = K \exp\{h(\mathrm{Cos}^{-1}\, t)\},\; -1 \leqslant t \leqslant 1$$

with the same h as in Remark 2.4 and

$$\frac{1}{K} = 2\pi \left\{ \frac{e^b\left(1-ae^{-\frac{a\pi}{2}}\right)}{a^2+1} + \frac{e^d\left(c\,e^{-\frac{c\pi}{2}}+e^{-c\pi}\right)}{c^2+1} \right\}.$$

In order now to verify that $\mathcal{F}$ indeed serves as a counter-example to the assertion of Theorem 3.1, take $x_1, x_2 \in S^2$. Suppose the length of $[x_1, x_2]$ $\leqslant$ length of $C-[x_1, x_2]$. Then, it is easy to see that for every $\theta \in S^2 - C$ (for the definition of C see part A of Remark 3.2), $\exists\, \theta^* \in [x_1, x_2]$ such that

$$f(x_1'\,\theta^*)\, f(x_2'\,\theta^*) \geqslant f(x_1'\,\theta)\, f(x_2'\,\theta). \qquad \ldots \quad (3.5)$$

with (3.5) in mind, the rest of the verification consists in routine algebraic computation and so we omit it.

Remark 3.5. We have not discussed about the location of spherical median for $n = 3$ samples in Remark 3.2. Therefore, the question of validity of Theorem 3.1 for $n = 3$ samples remains open.

Remark 3.6. We have assumed both in Theorem 2.1 and Theorem 3.1 that $f(t) > 0$ for every $t \in (-1, 1)$. However, none of the results mentioned in Section 1 puts any such restriction on the density to be characterized. It is, therefore, of some interest to see if this assumption can be relaxed.

4. Some general remarks

Remark 4.1. The three results mentioned in the introduction has the following common feature : the specific form of the maximum likelihood estimate of the location parameter can be thought of as that x_0 which minimizes

$$\sum_{i=1}^{n} d(x_i, x) \qquad \ldots \quad (4.1)$$

over $x \in \mathscr{X}$, where $\mathscr{X}$ is the sample space under consideration and the sum in (4.1) is a measure of distance between $\{x_1, \ldots, x_n\}$ and x for some $d : \mathscr{X} \times \mathscr{X} \to \mathcal{R}_+$, the set of non-negative numbers. For example, in Teicher (1961) $\mathscr{X} = \mathcal{R}^1$ and $d(x, y) = (x-y)^2$, in Ghosh and Rao (1971) $\mathscr{X} = \mathcal{R}^1$ and $d(x, y) = |x-y|$ and in Bingham and Mardia (1975) $\mathscr{X} = S^p$ and $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_2^2$ = square of the l_2-norm of $\boldsymbol{x}-\boldsymbol{y}$ The density characterized then turns out to be of the form

$$Ae^{-ad(x, \theta)}, \ \theta \in \mathscr{X}$$

where $a > 0$ and $A > 0$ is a constant depending on a [The von Mises—Fisher density $Ae^{b(\boldsymbol{x}'\boldsymbol{\theta})}$ is easily seen to have the alternative representation $Ae^{-a\|\boldsymbol{x}-\boldsymbol{\theta}\|^2}$ since $\|\boldsymbol{x}\| = \|\boldsymbol{\theta}\| = 1$]. Thus the way mean, median or mean direction is defined is captured in the form of the density characterized.

In the problem considered in this paper, we have $\mathscr{X} = S^p$, $d(\boldsymbol{x}, \boldsymbol{y}) = \cos^{-1} (\boldsymbol{x}' \boldsymbol{y})$ = the geodesic distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. In view of the observation mentioned in the last paragraph it is, therefore, expected that the density characterized should have the form as in (2.2) and (3.2).

Remark 4.2. In both Theorem 2.1 and Theorem 3.1, we have assume that the median direction is a maximum likelihood estimate of $\boldsymbol{\theta}$ for $n = 4$ samples. The same result, therefore, holds if n is assumed to vary over a set of natural numbers containing some multiple of 4. However, the question of validity of the assertion remains open if n is assumed to very over a set (finite or infinite) containing no multiple of 4. Similar remarks hold for the result of Teicher and that of Ghosh and Rao.

Acknowledgement. The author is grateful to Prof. J. K. Ghosh for several helpful discussions.

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, Characterizations and Applications. G. P. Patil, S. Kotz and J. K. Ord (ed.). Reidel, Dordrecht and Boston 387-398.

Fisher, N. I. (1985). Spherical medians. *J. R. Statist. Soc.*, B, **47**, 342-348.

Ghosh, J. K. and Rao, C. R. (1971). A note on some translation parameter families of densities for which the median is an m.l.e. *Sankhyā*, A, **33**, 91-95.

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, Wiley, New York.

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

Teicher, H. (1961). Maximum likelihood characterization of distributions. *Ann. Math. Statist.*, **32**, 1214-1222.

Division of Theoretical Statistics and Mathematics
Indian Statistical Institute
203 B. T. Road
Calcutta 700 035
India.