SOME CONTRIBUTIONS TO DIRECTIONAL STATISTICS : CHARACTERIZATION AND ASYMPTOTICS

Sumitra Purkayastha

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements
for the award of the degree of
Doctor of Philosophy

CALCUTTA

1991

TO THE MEMORY OF MY FATHER

AND

TO MY MOTHER

ACKNOWLEDGEMENTS

This work has been done under the supervision of Prof. J.K. Ghosh. I find it difficult to express in a few words my deep sense of regard for him as my mentor, as he has played the most significant role in my academic development in the last few years or so. Working under his supervision was indeed an experience of pleasure and encouragement.

While working on this thesis, I had the privilege to work with Prof. R. Mukerjee and my friend Dr. S. K. Bhandari. I am grateful to them for the interest they showed in my work. I must also thank them for allowing me to include our joint works in this thesis.

I would be failing in my duties if I do not say a few words about my association with Prof. S. C. Bagchi. Over the years I carried out my research, I had numerous discussions with him from which I benefited immensely. In fact, much of the contents of Chapter 1 were developed under his direct guidance. To him I owe more than I can possibly express.

I would like also to mention that while working on the present version of Chapter 1, I received substantial help from Prof. J. C. Gupta and my friend Dr. D. Sengupta. Prof. S. N. Joshi also rendered his help when I began work on the 'Bahadur representation' problem. All of them deserve my sincere thanks.

After embarking upon my research career, I have been fortunate enough to get Shri Goutam Mukherjee and Shri Dhurjati Prasad Sengupta as my friends. Association with them made my hard days appear less formidable. It is with gratitude that I recall their love for me.

Talking on more human aspect, I must say that except for my mother's utmost patience and encouragement, writing this thesis would never come true.

Finally, I wish to express my sincere thanks for Shri Samir Kumar Chakraborty who put an enormous amount of effort in typing the thesis. Thanks are due also to Shri Muktalal Khanra for efficient cyclostyling of the thesis. I record my sincere appreciation for both of them.

Calcutta
June, 1991

Sumitra Purkayastha

CONTENTS

INTRODUCTION AND SUMMARY

In many natural and physical sciences the observations are in the form of directions in 2- or 3-dimensional Euclidean space or rotations of such a space. In the former case, it is customary to represent the observations by points on the unit ball in $\mathbb{R}^2$ or $\mathbb{R}^3$, or more generally, $\mathbb{R}^p$. In the latter case, the observations are represented by orthogonal matrices having determinant +1, or more generally, by $n \times p$ $(n \leq p)$ matrices A satisfying $AA^t = I_n$, i.e., by elements of Stiefel manifold. Examples of observations on directions in 2-dimensional space include those on wind direction or on flight direction of birds. Similarly, with directions in 3-dimensional space one comes across observations on arrival directions of showers of cosmic rays or on facing directions of conically folded planes. Other interesting examples of observations on directions in 2- or 3-dimensional spaces may be found in Mardia (1972), Batschelet (1981), and Fisher et al. (1987). On the other hand, examples of observations from Stiefel manifold include the specifications of elliptical cometary orbits by their perihelion and normal directions (Mardia (1975a), Jupp and Mardia (1979)) and similar 'orbits' in vectorcardiography (Downs (1972), Prentice (1986)).

However, it should be mentioned that the study of directional statistics gives rise to statistical problems which do not fit into the usual methods of statistical analysis one employs for observations from the real line or Euclidean space. In fact, all the issues which one explores with linear data demand separate attention when one works with directional statistics. Accordingly, a separate theory has developed to answer the various questions related to directional statistics. The relevant references are the books by Mardia (1972), Watson (1983), and Fisher et al. (1987). In addition to these books, the review papers by Mardia (1975a, 1975b, 1988), Rao (1984), and Jupp and Mardia (1989) provide surveys of what has been achieved in the study of directional statistics so far. See also Watson (1989).

This thesis explores some problems in directional statistics. The problems can broadly be classified into two categories ; characterization and asymptotics.

In what follows we mention briefly the contents of the thesis.

Chernoff (1981) proved that if $X \sim N(0,1)$, then for any absolutely continuous function g with $E[g^2(X)]$ finite,

$$V[g(X)] \leq E[\{g'(X)\}^2] , \qquad \ldots(*)$$

with equality if and only if $g(X)$ is linear in X. Borovkov and Utev (1983) characterized standard normal distribution via the inequality (*). Several other authors have established inequalities analogous to (*), which subsequently we refer to as Chernoff-type inequalities, in connection with distributions other than normal or in more general setting and studied the related characterization problems [vide Cacoullos (1982,1989), Cacoullos and Papathanasiou (1985,1986,1989), Chen (1982,1985), Chen and Lou (1987), Hu (1986), Hwang and Sheu (1987), Johnson (1988), Klaassen (1985), Papathanasiou (1989,1990), Prakasa Rao and Sreehari (1986,1987), Srivastava and Sreehari (1987,1990)].

Motivated by the works stated above, we have obtained in Chapter 1 a characterization of uniformity over circle by Chernoff-type inequality. Our work essentially comprises of obtaining a Chernoff-type inequality for $U[-\pi,\pi]$ distribution, along with the "Khintchine's Representation" of symmetric unimodal distribution which is subsequently made use of in order to study this kind of inequalities for distributions on $[-\pi,\pi]$ having symmetric unimodal densities. Finally, $U[-\pi,\pi]$ distribution is characterized among all distributions on $[-\pi,\pi]$ having symmetric unimodal densities through this kind of inequalities via a compactness argument. This indeed provides a characterization of uniformity over the circle.

In Chapter 2, we settle a maximum likelihood characterization problem. The first result of this kind was obtained by Teicher (1961) who, making rigorous a result attributed originally to Gauss (1809), obtained a characterization of the normal distribution. Similar results were obtained later by Ghosh and Rao (1971), who obtained a characterization of the Laplace distribution, and by Bingham and Mardia (1975), who obtained a characterization of the von Mises-Fisher distribution on the sphere. In the chapter under consideration, 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, this result is extended to the spherical case. It is also discussed why, in consideration of the results mentioned earlier in this context, the solution to the problem explored in the present chapter appears to be the natural one. It should be mentioned that the distribution that we have obtained gives a new directional distribution. Cabrera and Watson (1990) have made an application of the distribution, when considered on the sphere S^2 .

In Chapter 3, we settle yet another maximum likelihood characterization problem. It is proved that under mild conditions a location parameter (defined suitably) family of densities on the Stiefel manifold must be the von Mises-Fisher family, introduced by Downs (1972), if the polar component of the sum of observations is a maximum likelihood estimate of the location parameter. This extends the result of Bingham and Mardia (1975) for distributions on sphere, mentioned in the last paragraph, to distributions on Stiefel manifold.

In connection with the problem explored in Chapter 2, it should be mentioned that in order to settle it for distributions over circle we have proposed a definition of sample circular median which follows Fisher's (1985) notion of spherical median, and which is not the natural sample analogue of Mardia's (1972) definition of population circular median. In fact, it is illustrated through an example in Chapter 4 that the natural sample analogue of Mardia's definition of population circular median leads to problem of uniqueness of choice. In the chapter under consideration, we therefore study the definition of sample circular median proposed in Chapter 2 in detail. It is shown that apart from being computationally easy, sample circular median so defined can be regarded as a sample analogue of population circular median, as defined by Mardia. Moreover, the effect of translating all the samples by a fixed amount on the sample circular median is studied.

In Chapter 5, we undertake the study of asymptotics of sample circular median, developed in Chapter 4. One of the major tools for studying the large sample behaviour of a statistic consists in deriving an asymptotic linear representation of the same. For sample quantiles, such a linear representation was obtained by Bahadur (1966), which came to be known later as 'Bahadur representation of quantiles'. In the chapter under consideration, we obtain a Bahadur-type representation for sample circular median. In this context, we have established a result related to the population circular median which is very similar to the one connecting minimum mean absolute deviation and linear median. This result is used to establish strong consistency of sample circular median. As an application of the representation theorem, asymptotic normality of sample circular median — normalized suitably — is established.

In Chapter 6, it is demonstrated that for the circular distribution obtained in Chapter 2, sample circular median is asymptotically efficient (in the Hájek-LeCam framework) for estimating the location parameter. As regards the tool, using arguments involving contiguity, we establish a convolution theorem similar to the one due to Hájek (1970) and Inagaki (1970) that characterizes the limiting distribution of regular estimates. The final results follow from this convolution theorem and Anderson's inequality.

CHAPTER 1

A CHARACTERIZATION OF UNIFORMITY OVER THE CIRCLE BY CHERNOFF-TYPE INEQUALITY

1.1 Introduction

Chernoff (1981) proved that if $X \sim N(0,1)$, then for any absolutely continuous function g with $E[g^2(X)]$ finite,

$$V[g(X)] \leq E\left\{[g'(X)]^2\right\}, \qquad \ldots(1.1.1)$$

with equality if and only if $g(X)$ is linear in X. Motivated by this inequality, Barovkov and Utev (1983) obtained the following characterization of the normal distribution : for any random variable X with finite variance σ^2,

$$\sup_g \frac{V[g(X)]}{\sigma^2 E\left\{[g'(X)]^2\right\}} = 1,$$

where the supremum is taken over the class of absolutely continuous functions g such that $0 < V[g(X)] < \infty$, if and only if X has a normal distribution. Several other authors have established inequalities analogous to (1.1.1), which subsequently we refer to as Chernoff-type inequalities, in connection with distributions other than normal or in more general setting and studied the related characterization problems. Relevant references for these are Cacoullos (1982,1989), Cacoullos and Papathanasiou (1985,1986,1989), Chen (1982,1985), Chen and Lou (1987), Hu (1986), Hwang and Sheu (1987), Johnson (1988), Klaassen (1985), Papathanasiou (1989,1990), Prakasa Rao and Sreehari (1986, 1987), Srivastava and Sreehari (1987,1990). It should be mentioned that Chernoff-type inequalities have also been used to obtain alternative proofs of the central limit theorem (see Chen (1988), Cacoullos and Papathanasiou (1990)).

It is, therefore, worthwhile to study Chernoff-type inequalities and the related characterization problems for directional distributions. This chapter is devoted to this study. In fact, we have restricted our attention only to circular random variables, or to be more specific, to circular distributions admitting symmetric unimodal densities, since in the statistics of circular data this class of distributions is of paramount importance. Observe that this class of circular distributions can be identified with the class of distributions on $[-\pi,\pi]$ admitting symmetric unimodal densities, and hence this latter class will be used in all our results. In Section 1.2, we establish a Chernoff-type inequality for uniform distribution on $[-\pi,\pi]$, which is used later in Section 1.3 to study this kind of inequalities for distributions on $[-\pi,\pi]$ having symmetric unimodal densities. Finally in Section 1.4, uniform distribution on $[-\pi,\pi]$ is characterized within the class of distributions on $[-\pi,\pi]$ having symmetric unimodal densities through this kind of inequalities.

This chapter is a revised version of Purkayastha and Bhandari (1990).

1.2 A Chernoff-type inequality for uniform distribution on $[-\pi,\pi]$

Theorem 1.2.1 Suppose g is an absolutely continuous even function defined on $[-\pi,\pi]$ with $g(0) = 0$. Denote the derivative of g by g'. Then

$$\int_{-\pi}^{\pi} g^2(x)dx \leq 4 \int_{-\pi}^{\pi} [g'(x)]^2 \, dx, \qquad \ldots(1.2.1)$$

with equality if and only if

$$g(x) = c\left|\sin \frac{x}{2}\right| , \quad -\pi \leq x \leq \pi , \qquad \ldots(1.2.2)$$

for some constant c.

Proof. Observe first that g even implies

$$\int_{-\pi}^{\pi} g^2(x)dx = 2 \int_{0}^{\pi} g^2(x)dx, \qquad \ldots(1.2.3)$$

and also

$$g'(x) = -g'(-x) \quad \text{a.e.} . \qquad \ldots(1.2.4)$$

Further, from (1.2.4) we obtain

$$\int_{-\pi}^{\pi} [g'(x)]^2 \, dx = 2 \int_{0}^{\pi} [g'(x)]^2 \, dx , \qquad \ldots(1.2.5)$$

It, therefore, follows from (1.2.3) and (1.2.5) that in order to establish (1.2.1) it suffices to prove

$$\int_{0}^{\pi} g^2(x)dx \leq 4 \int_{0}^{\pi} [g'(x)]^2 \, dx , \qquad \ldots(1.2.6)$$

where g is an absolutely continuous function defined on $[0,\pi]$ with $g(0) = 0$. Observe that it follows immediately from absolute continuity of g,

$$\int_{0}^{\pi} g^2(x)dx < \infty .$$

Thus it is legitimate to assume moreover that

$$\int_{0}^{\pi} [g'(x)]^2 \, dx < \infty . \qquad \ldots(1.2.7)$$

Define $h : [-2\pi,2\pi] \longrightarrow \mathbb{R}$ by

$$\begin{aligned} h(x) &= -g(x+2\pi) && \text{if } -2\pi \leq x \leq -\pi , \\ &= -g(-x) && \text{if } -\pi \leq x \leq 0 , \\ &= g(x) && \text{if } 0 \leq x \leq \pi , \\ &= g(2\pi-x) && \text{if } \pi \leq x \leq 2\pi. \end{aligned} \qquad \ldots(1.2.8)$$

Define also $k : [-2\pi,2\pi] \longrightarrow \mathbb{R}$ by

$$
\begin{aligned}
k(x) &= -g'(x+2\pi) && \text{if } -2\pi \le x \le \pi , \\
&= g'(-x) && \text{if } -\pi \le x \le 0 , \\
&= g'(x) && \text{if } 0 \le x \le \pi , \\
&= -g'(2\pi-x) && \text{if } \pi \le \pi \le 2\pi .
\end{aligned}
$$

Observe first that h is an absolutely continuous odd function with derivative k and $h(-2\pi) = h(2\pi) = h(0) = g(0) = 0$. Moreover,

$$\int_{-2\pi}^{2\pi} h^2(x)dx = 4\int_0^{\pi} g^2(x)dx . \qquad \ldots(1.2.9)$$

Further,

$$\int_{-2\pi}^{2\pi} k^2(x)dx = 4\int_0^{\pi}[g'(x)]^2\, dx .$$

However, since k is the derivative of h, it follows that

$$\int_{-2\pi}^{2\pi} [h'(x)]^2\, dx = 4\int_0^{\pi}[g'(x)]^2\, dx . \qquad \ldots(1.2.10)$$

It, therefore, follows from (1.2.9) and (1.2.10) that in order to establish (1.2.6) it suffices to prove

$$\int_{-2\pi}^{2\pi} h^2(x)dx \le 4\int_{-2\pi}^{2\pi} [h'(x)]^2\, dx \qquad \ldots(1.2.11)$$

Observe first that since h is an odd function with $h(-2\pi) = h(2\pi) = 0$, the Fourier expansion of h gives

$$h(x) = \sum_{n=1}^{\infty} a_n \sin \frac{nx}{2}, \quad -2\pi \leq x \leq 2\pi, \qquad \ldots(1.2.12)$$

where a_n's are suitable constants and the equality in (1.2.12) means

$$\lim_{m \to \infty} \int_{-2\pi}^{2\pi} [h(x) - \sum_{n=1}^{m} a_n \sin \frac{nx}{2}]^2 \, dx = 0$$

Again, it follows from the definition of k that h' is an even function. Taken with (1.2.7), (1.2.10) and (1.2.12), this gives the following Fourier expansion of h' :

$$h'(x) = \sum_{n=1}^{\infty} \frac{na_n}{2} \cos \frac{nx}{2}, \quad -2\pi \leq x \leq 2\pi, \qquad \ldots(1.2.13)$$

where the notion of equality in (1.2.13) is same as that in (1.2.12) (vide Edwards (1979), Volume 1, p.32). Observe now that (1.2.12) implies

$$\int_{-2\pi}^{2\pi} h^2(x) dx = 2\pi \sum_{n=1}^{\infty} a_n^2 . \qquad \ldots(1.2.14)$$

Similarly, (1.2.13) implies

$$\int_{-2\pi}^{2\pi} [h'(x)]^2 \, dx = 2\pi \sum_{n=1}^{\infty} \frac{n^2 a_n^2}{4}, \qquad \ldots(1.2.15)$$

which in turn implies

$$\int_{-2\pi}^{2\pi} [h'(x)]^2 \, dx \geq \frac{1}{4} \cdot 2\pi \sum_{n=1}^{\infty} a_n^2 . \qquad \ldots(1.2.16)$$

The inequality in (1.2.11) is now an immediate consequence of (1.2.14) and (1.2.16), and in view of our early discussions (1.2.1) follows from (1.2.11).

Suppose now that equality occurs in (1.2.1) for some g, an absolutely continuous even function defined on $[-\pi,\pi]$ with $g(0) = 0$. Corresponding to this g define h as in (1.2.8). Observe now that

$$\int_{-2\pi}^{2\pi} h^2(x)dx = 4 \int_{-2\pi}^{2\pi} [h'(x)]^2 \, dx ,$$

for this h. In view of (1.2.14) and (1.2.15), this implies

$$\frac{1}{4} \sum_{n=1}^{\infty} a_n^2 = \sum_{n=1}^{\infty} \frac{n^2 a_n^2}{4} ,$$

which in turn implies

$$a_n = 0 \qquad \text{...(1.2.17)}$$

for every $n > 1$. From (1.2.12) and (1.2.17), we therefore obtain

$$h(x) = c \sin \frac{x}{2} , \quad -2\pi \leq x \leq 2\pi , \qquad \text{...(1.2.18)}$$

for some constant c. From (1.2.8), (1.2.18) and the fact that g is even, we now obtain

$$g(x) = c \left|\sin \frac{x}{2}\right| , \quad -\pi \leq x \leq \pi .$$

This completes the proof of our assertion.

Remark 1.2.1 If in Theorem 1.2.1, instead of considering the interval $[-\pi,\pi]$ we consider the interval $[-\pi a,\pi a]$ for some $a > 0$, essentially the same integral inequality in (1.2.1) holds with $4a^2$ appearing in place of 4. Moreover, the function g for which this inequality becomes an equality is

$$g(x) = c\left|\sin \frac{x}{2a}\right| , \quad -\pi a \leq x \leq \pi a ,$$

for some constant c.

1.3 Chernoff-type inequalities for symmetric unimodal densities on $[-\pi,\pi]$

In this section we study Chernoff-type inequalities for symmetric unimodal densities on $[-\pi,\pi]$, i.e., for densities f satisfying

(i) $f(x) \geq 0$, $\quad -\pi \leq x \leq \pi$,

(ii) $\int_{-\pi}^{\pi} f(x)dx = 1$,

(iii) $f(x) = f(-x)$, $-\pi \leq x \leq \pi$,

and

(iv) f is monotone increasing on $[-\pi,0]$ and monotone decreasing on $[0,\pi]$.

First we state a result without proof, which will be needed in the proofs of the subsequent results in this section. This result is indeed the 'Khintchine's Representation' for symmetric unimodal distributions. A proof of the result may be found in Dharmadhikari and Joag-dev (1988, Chapter 1, Section 2).

<u>Theorem 1.3.1</u> Given a random variable X with a distribution function F, the following statements are equivalent to one another :

(a) F is symmetric and unimodal

(b) There exist independent random variables V and Z such that V is uniform on $[-1,1]$, Z is nonnegative and the product VZ has distribution function F .

We now have the following result that extends inequality (1.2.1) in Theorem 1.2.1 to symmetric unimodal densities on $[-\pi,\pi]$.

<u>Theorem 1.3.2</u> Suppose f is a symmetric unimodal density on $[-\pi,\pi]$. Suppose, moreover, that g is an absolutely continuous even function defined on $[-\pi,\pi]$ with $g(0) = 0$. Denote the derivative of g by g' . Then

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx \leq 4 \int_{-\pi}^{\pi}[g'(x)]^2 f(x)dx. \qquad \ldots(1.3.1)$$

<u>Proof</u>. Observe first that since g is absolutely continuous on $[-\pi,\pi]$, it is bounded and hence the integral

appearing on the left hand side of (1.3.1) is finite. Therefore, we may assume, without loss of generality, that the integral appearing on the right hand side also is finite.

Denote now by X, a random variable having density f. It then follows from Theorem 1.3.1 that

$$X \overset{d}{=} VZ \,, \qquad \ldots(1.3.2)$$

where V is uniform on $[-\pi,\pi]$, Z is a nonnegative random variable independent of V, and '$\overset{d}{=}$' means 'has same distribution as'. Denote the distribution function of Z by G. Obviously

$$G(1) - G(0-) = 1 \,, \qquad \ldots(1.3.3)$$

since $P(|X| \leq \pi) = 1$.

It is now easy to see, using the representation in (1.3.1) and by conditioning on Z, that

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx = \int_0^1 E\{g^2(zV)\}\, dG(z), \qquad \ldots(1.3.4.1)$$

and

$$\int_{-\pi}^{\pi} [g'(x)]^2 f(x)dx = \int_0^1 E\{[g'(zV)]^2\}\, dG(z) \qquad \ldots(1.3.4.2)$$

Observe, moreover, that zV has uniform distribution on $[-\pi z, \pi z]$ for every $z \geq 0$, and hence it follows from

Remark 1.2.1 and $g(0) = 0$ that for every z satisfying $0 \leq z \leq 1$, we have

$$E\{g^2(zV)\} \leq 4z^2 E\{[g'(zV)]^2\} \leq 4E\{[g'(zV)]^2\}. \quad \ldots(1.3.5)$$

The inequality in (1.3.1) now follows immediately from (1.3.4.1), (1.3.4.2) and (1.3.5), thus completing the proof of our assertion.

In the last result of this section, we focus our attention on symmetric unimodal non-uniform densities on $[-\pi, \pi]$.

Theorem 1.3.3 Suppose f is a symmetric unimodal non-uniform density on $[-\pi, \pi]$. Suppose, moreover, that g is an absolutely continuous even function defined on $[-\pi, \pi]$ with $g(0) = 0$, and

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx > 0. \quad \ldots(1.3.6)$$

Denote the derivative of g by g'. Then

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx < 4\int_{-\pi}^{\pi}[g'(x)]^2 f(x)dx. \quad \ldots(1.3.7)$$

Proof. As in the proof of the preceding theorem, we assume, without loss of generality, that the integral appearing on the right hand side of (1.3.7) is finite. Moreover, we use the notations X, V, Z to denote the same random variables as in that proof.

Suppose, if possible, that (1.3.7) is false. In consideration of (1.3.1), this means

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx = 4 \int_{-\pi}^{\pi} [g'(x)]^2 f(x)dx . \qquad \ldots(1.3.8)$$

We now obtain, in view of (1.3.4.1), (1.3.4.2), (1.3.5) and (1.3.8) the following :

$$P(Z \varepsilon A) = 1 , \qquad \ldots(1.3.9)$$

where

$$A = \left\{ 0 \leq z \leq 1 \,\middle|\, E\{g^2(zV)]\} = 4z^2 E\{[g'(zV)]^2\} = 4E\{[g'(zV)]^2\} \right\}$$

We shall now argue by contradiction. Observe first that

$$P(0 < Z < 1) > 0 , \qquad \ldots(1.3.10)$$

since X is an absolutely continuous non-uniform random variable taking values in $[-\pi,\pi]$. Therefore, it follows from (1.3.9) and (1.3.10)

$$P(Z \varepsilon (0,1) \cap A) > 0 ,$$

and hence

$$(0,1) \cap A \neq \phi .$$

Choose and fix

$$z_0 \varepsilon (0,1) \cap A . \qquad \ldots(1.3.11)$$

Observe now that if $z \varepsilon (0,1) \cap A$, we have

$$E\{g^2(zV)\} = E\{[g'(zV)]^2\} = 0 ,$$

i.e.,

$$\int_{-\pi z}^{\pi z} g^2(x)dx = \int_{-\pi z}^{\pi z} [g'(x)]^2 dx = 0 .$$

This implies, in view of continuity of g,

$$g(x) = 0 , \quad -\pi z \leq x \leq \pi z . \qquad \ldots(1.3.12)$$

Again, if $1 \varepsilon A$, we have

$$E\{g^2(V)\} = 4 E\{[g'(V)]^2\} ,$$

i.e.,

$$\int_{-\pi}^{\pi} g^2(x)dx = 4 \int_{-\pi}^{\pi} [g'(x)]^2 dx .$$

In view of Theorem 1.2.1, this implies

$$g(x) = c\left|\sin \frac{x}{2}\right| , \quad -\pi \leq x \leq \pi , \qquad \ldots(1.3.13)$$

for some constant c. Define now

$$a = \sup A.$$

Obviously $0 < a \leq 1$. We shall consider the cases corresponding to $0 < a < 1$ and $a = 1$ separately.

<u>Case 1</u> $0 < a < 1$

In this case, we get hold of a sequence $\{a_n\}$ in A such that $0 < a_n < 1$ for every n, and $\lim_{n \to \infty} a_n = a$.

Observe now that (1.3.12) implies

$$g(x) = 0 , \quad -\pi a_n \leq x \leq \pi a_n ,$$

for every $n = 1,2,\ldots$. Since $a_n \longrightarrow a$, this implies in turn

$$g(x) = 0 , \quad -\pi a < x < \pi a .$$

However, it follows from the definition of a and (1.3.9) that

$$P(Z > a) = 0 ,$$

which implies

$$P(|X| > \pi a) = 0 .$$

Therefore,

$$\int_{-\pi}^{\pi} g^2(x)f(x)dx = \int_{-\pi a}^{\pi a} g^2(x)f(x)dx = 0 ,$$

contradicting (1.3.6).

<u>Case 2</u> $a = 1$

If $1 \varepsilon A$, we have in view of (1.3.13)

$$g(x) = c\left|\sin \frac{x}{2}\right| , \quad -\pi \leq x \leq \pi ,$$

for some constant c, and moreover in view of (1.3.11) and (1.3.12) we also have

$$g(x) = 0 , \quad -\pi z_o \leq x \leq \pi z_o .$$

Hence we must have $c = 0$, so that

$$g(x) = 0 , \quad -\pi \leq x \leq \pi ,$$

which implies

$$\int_{-\pi}^{\pi} g^2(x) f(x) dx = 0 ,$$

contradicting (1.3.6).

If, however, $1 \notin A$, we get hold of a sequence $\{a_n\}$ in A such that $0 < a_n < 1$ for every n, and $\lim_{n \to \infty} a_n = 1$. In this situation, employing arguments similar to those needed in Case 1 above, we obtain

$$\int_{-\pi}^{\pi} g^2(x) f(x) dx = 0 ,$$

contradicting (1.3.6).

Our assertion now follows in consideration of our early arguments.

The last result of this section will combine Theorem 1.2.1, Remark 1.2.1, Theorem 1.3.2 and Theorem 1.3.3. Before we state it, some relevant discussions are in order.

It is easy to see that given a symmetric unimodal density f on $[-\pi, \pi]$, the support of the probability distribution induced by f will be $[-\pi a, \pi a]$ for some a with $0 < a \leq 1$. However, in none of Theorem 1.2.1, Theorem 1.3.2

and Theorem 1.3.3 we put any restriction on the support of f. It will be seen in the next section that such a restriction is required in relation to the characterization problem studied there. The following result indeed puts restriction on the support of f. It should be mentioned beforehand that we shall provide only the statement of the result, and the proof will follow essentially by arguments leading to the proofs of Theorem 1.2.1, Theorem 1.3.2 and Theorem 1.3.3, and hence be omitted.

<u>Theorem 1.3.4</u> Suppose f is a symmetric unimodal density on $[-\pi a, \pi a]$ $(a > 0)$ with full support. Suppose, moreover, that g is an even function defined on $[-\pi a, \pi a]$ such that $g(0) = 0$, g is absolutely continuous on $(-\pi a, \pi a)$, and

$$0 < \int_{-\pi a}^{\pi a} g^2(x) f(x) dx < \infty . \qquad \ldots(1.3.14)$$

Denote the derivative of g by g'. Then

$$\int_{-\pi a}^{\pi a} g^2(x) f(x) dx \leq 4a^2 \int_{-\pi a}^{\pi a} [g'(x)]^2 f(x) dx, \qquad \ldots(1.3.15)$$

with equality if and only if f is the uniform density on $[-\pi a, \pi a]$ and

$$g(x) = c\left|\sin \frac{x}{2a}\right| , \quad -\pi a < x < \pi a ,$$

for some constant c.

Remark 1.3.1 It should be mentioned that the class of functions g considered in Theorem 1.2.1, Theorem 1.3.2 and Theorem 1.3.3 is smaller than the one considered in Theorem 1.3.4. That is why, we have to assume convergence of the integral in (1.3.14).

1.4 Characterization of uniformity on $[-\pi,\pi]$

In this section we obtain a characterization of uniformity on $[-\pi,\pi]$ among all distributions on $[-\pi,\pi]$ having symmetric unimodal densities.

Denote by $\mathcal{F}$, the class of symmetric unimodal densities on $[-\pi,\pi]$. Observe that given $f \in \mathcal{F}$, $\exists$ a unique number $s(f) \in (0,\pi]$ such that

$$
\begin{aligned}
f(x) &> 0 \quad \text{if } |x| < s(f) , \\
&= 0 \quad \text{if } |x| > s(f) .
\end{aligned}
$$

Define now for every $f \in \mathcal{F}$, the following class of functions :

$$
\mathcal{G}(f) = \left\{ g \;\middle|\; \begin{array}{l} g \text{ is an even function defined on } [-s(f),s(f)], \\ g \text{ is absolutely continuous on } (-s(f),s(f)), \\ \text{and } 0 < \int_{-s(f)}^{s(f)} g^2(x)f(x)dx < \infty \end{array} \right\} \qquad \ldots(1.4.1)
$$

Writing g' for the derivative of g, we define moreover

$$\mathcal{G}^*(f) = \left\{ g \in \mathcal{G}(f) \;\middle|\; \int_{-s(f)}^{s(f)} [g'(x)]^2 f(x)dx < \infty \right\}$$

...(1.4.2)

Define finally,

$$\Lambda(f) = \sup_{g \in \mathcal{G}(f)} \frac{\int_{-s(f)}^{s(f)} g^2(x)f(x)dx}{\int_{-s(f)}^{s(f)} [g'(x)]^2 f(x)dx}, \quad f \in \mathcal{F},$$

...(1.4.3)

where the ratio appearing on the right hand side of (1.4.3) is to be understood as zero, whenever the integral in the denominator diverges. With this understanding, it follows immediately that

$$\Lambda(f) = \sup_{g \in \mathcal{G}^*(f)} \frac{\int_{-s(f)}^{s(f)} g^2(x)f(x)dx}{\int_{-s(f)}^{s(f)} [g'(x)]^2 f(x)dx},$$

for every $f \in \mathcal{F}$. Observe finally that, since every $g \in \mathcal{G}(f)$ is an even function, we have

$$\Lambda(f) = \sup_{g \in \mathcal{G}^*(f)} \frac{\int_{0}^{s(f)} g^2(x)f(x)dx}{\int_{0}^{s(f)} [g'(x)]^2 f(x)dx},$$

for every $f \in \mathcal{F}$.

We now state two results without proof, which will be needed in our subsequent study of $\Lambda(\cdot)$. The proofs may be found in Halmos (1978, Chapter 15).

Proposition 1.4.1 Suppose that S is a measure space with a σ-finite measure μ . Suppose, moreover, that K is a real-valued measurable function on $S \times S$ such that K^2 is integrable with respect to the product measure $\mu \times \mu$. Then the equation

$$(A\phi)(s) = \int_S K(s,t)\phi(t)d\mu(t) , \quad s \in S , \quad \ldots(1.4.4)$$

defines a compact operator (with kernel K) on $L^2(\mu)$.

Proposition 1.4.2 Suppose H is a Hilbert space. Suppose, moreover, that

$$A : H \longrightarrow H$$

is a compact operator. Then there exists $v \in H$ with $||v|| = 1$ such that $||Av|| = ||A||$. In other words, compact operators on Hilbert spaces attain their norm.

Now choose and fix $f \in \mathcal{F}$. Take in Proposition 1.4.1, $S = [0,s(f)]$ with the (finite) measure μ defined as follows :

$$\mu(A) = \int_A f(x)dx ,$$

for Borel subsets A of $[0,s(f)]$. Define now $K : S \times S \longrightarrow \mathbb{R}$ as follows :

$$K(x,y) = \frac{1}{f(y)} \qquad \text{if } 0 \leq y \leq x < s(f),$$

...(1.4.5)

$$= 0 \qquad \text{otherwise.}$$

Observe that

$$\iint_{S\times S} K^2(x,y)d\mu(x)d\mu(y)$$

$$= \int_0^{s(f)} \int_0^x \frac{1}{f^2(y)} f(x)f(y)dx\ dy$$

$$= \int_0^{s(f)} \int_0^x \frac{f(x)}{f(y)} dx\ dy$$

$$\leq \int_0^{s(f)} \int_0^x dx\ dy$$

[$\because$ f is monotone decreasing on $[0,s(f)]$]

$$< \infty .$$

Thus, in view of the assertion of Proposition 1.4.1, it follows that the equation

$$(A\phi)(x) = \int_S K(x,y)\phi(y)d\mu(y) \ , \quad x \in S \ , \qquad \text{...(1.4.6)}$$

defines a compact operator (with kernel K in (1,4.5)) on $L^2(\mu)$. However, it is known that the integral appearing on the right-hand side of (1.4.6) is well-defined except for a set of x with μ-measure 0. We shall, therefore, compute this integral for every $x \in S$, since this will be needed later. Observe first in view of (1.4.5) and (1.4.6) that

$$(A\phi)(x) = \int_0^x \phi(y)dy \;, \qquad 0 \leq x < s(f),$$
$$= 0 \qquad , \qquad x = s(f). \qquad \ldots(1.4.7)$$

However, since f is monotone decreasing and strictly positive on $[0, s(f))$, we obtain for every x such that $0 \leq x < s(f)$,

$$\left| \int_0^x \phi(y)dy \right|$$
$$= \left| \int_0^x \phi(y) \frac{1}{f(y)} d\mu(y) \right|$$
$$\leq \int_0^x |\phi(y)| \frac{1}{f(y)} d\mu(y)$$
$$\leq \frac{1}{f(x)} \int_0^x |\phi(y)| \, d\mu(y) ,$$

which is finite since ϕ belongs to $L^2(\mu)$ and hence belongs also to $L^1(\mu)$. Thus the integral appearing on the right-hand side of (1.4.7) is well-defined for every x such that $0 \leq x < s(f)$. We, therefore obtain from (1.4.7) that

$$||A\phi||^2 = \int_0^{s(f)} \left[\int_0^x \phi(y)dy \right]^2 d\mu(x)$$
$$= \int_0^{s(f)} \left[\int_0^x \phi(y)dy \right]^2 f(x)dx .$$

Our next results connects $\Lambda(f)$ and $||A||$.

Lemma 1.4.1 $\Lambda(f) = ||A||^2$.

Proof. Observe first that $\Lambda(f)$ may also be written as follows :

$$\Lambda(f) = \sup_{g \varepsilon \mathcal{G}^*(f)} \frac{\int_0^{s(f)} [\int_0^x g'(y)dy]^2 f(x)dx}{\int_0^{s(f)} [g'(x)]^2 f(x)\, dx} \qquad \ldots(1.4.8)$$

The assertion now follows essentially from (1.4.8), and the discussion preceding the statement of the present lemma and following Proposition 1.4.2. We omit the details.

Lemma 1.4.2

$$\Lambda(f) = \frac{\int_{-s(f)}^{s(f)} g_o^2(x)f(x)dx}{\int_{-s(f)}^{s(f)} [g_o'(x)]^2 f(x)dx},$$

for some $g_o \varepsilon \mathcal{G}^*(f)$, depending on f.

Proof. Recall from (1.4.6), the definition of the operator A on $L^2(\mu)$. We have argued before that A is compact. Thus, it follows in view of Proposition 1.4.2 that there exists $\phi_o \varepsilon L^2(\mu)$ such that $||\phi_o|| = 1$ and $||A\phi_o||^2 = ||A||^2$. Obviously, ϕ_o depends on f. Now write $g_o = A\phi_o$ and extend the definition of g_o

from $[0,s(f)]$ to $[-s(f),s(f)]$ by stipulating $g_o(x) = g_o(-x)$, $|x| \leq s(f)$. It is now easy to verify that $g_o \in \mathcal{G}^*(f)$, and that

$$\frac{\int_{-s(f)}^{s(f)} g_o^2(x)f(x)\,dx}{\int_{-s(f)}^{s(f)} [g_o'(x)]^2 f(x)\,dx} = ||A\phi_o||^2 . \qquad \ldots(1.4.9)$$

The proof is now an immediate consequence of Lemma 1.4.1, the fact $||A\phi_o||^2 = ||A||^2$, and (1.4.9).

The next result characterizes the uniform density on $[-\pi,\pi]$ within the class $\mathcal{F}$.

Theorem 1.4.1 (a) For every $f \in \mathcal{F}$,

$$\Lambda(f) \leq 4 .$$

(b) $\Lambda(f) = 4$ for some $f \in \mathcal{F}$ if and only if f is the uniform density on $[-\pi,\pi]$.

Proof. (a) Observe first that it follows in view of the definition of $\Lambda(f)$ and Theorem 1.3.4,

$$\Lambda(f) \leq \frac{4[s(f)]^2}{\pi^2} . \qquad \ldots(1.4.10)$$

The proof now follows from (1.4.10) and noting that $0 < s(f) \leq \pi$.

(b) Suppose first that f is the uniform density on $[-\pi,\pi]$. Then, it follows from Theorem 1.3.4, by taking $a = 1$ there, that $\Lambda(f) = 4$.

Conversely, suppose that $\Lambda(f) = 4$ for some $f \in \mathcal{F}$. This implies, in view of (1.4.10), we must have $s(f) = \pi$. Moreover, we have from Lemma 1.4.2,

$$\Lambda(f) = \frac{\int_{-\pi}^{\pi} g_o^2(x)f(x)dx}{\int_{-\pi}^{\pi}[g_o'(x)]^2 f(x)dx} , \qquad \ldots(1.4.11)$$

for some $g_o \in \mathcal{G}^*(f)$, depending on f. Since $\Lambda(f) = 4$ and $s(f) = \pi$, (1.4.11) implies equality in (1.3.15) with this g_o, by taking $a = 1$. In view of Theorem 1.3.4, this in turn implies that f is the uniform density on $[-\pi, \pi]$.

This completes the proof of the theorem.

The next result also gives a characterization of uniformity on $[-\pi, \pi]$ within the class $\mathcal{F}$. This result is, however, slightly different from Theorem 1.4.1.

Define for every $f \in \mathcal{F}$, the following class of functions :

$$\mathcal{H}(f) = \Big\{ g \mid g \text{ is an absolutely continuous even function defined on } [-\pi,\pi] \text{ with } g(0) = 0, \text{ and } \int_{-\pi}^{\pi} g^2(x)f(x)dx > 0 \Big\} . \qquad \ldots(1.4.12)$$

Define, moreover,

$$\Gamma(f) = \sup_{g \varepsilon \mathcal{H}(f)} \frac{\int_{-\pi}^{\pi} g^2(x) f(x)\, dx}{\int_{-\pi}^{\pi} [g'(x)]^2 f(x)\, dx} \qquad \ldots(1.4.13)$$

Observe that for every $g \varepsilon \mathcal{H}(f)$.

$$\frac{\int_{-\pi}^{\pi} g^2(x) f(x)\, dx}{\int_{-\pi}^{\pi} [g'(x)]^2 f(x) dx} = \frac{\int_{-s(f)}^{s(f)} g^2(x) f(x)\, dx}{\int_{-s(f)}^{s(f)} [g'(x)]^2 f(x) dx},$$

and, moreover, the restriction of g to $[-s(f), s(f)]$ belongs to $\mathcal{G}(f)$. Hence ,

$$\Gamma(f) \leq \Lambda(f),$$

for every $f \varepsilon \mathcal{F}$. Observe, moreover, in view of Theorem 1.2.1, that when f is the uniform density on $[-\pi, \pi]$, we have

$$\Gamma(f) = 4 .$$

We, therefore, have the following theorem.

<u>Theorem 1.4.2</u> (a) For every $f \varepsilon \mathcal{F}$,

$$\Gamma(f) \leq 4 .$$

(b) $\Gamma(f) = 4$ for some $f \varepsilon \mathcal{F}$ if and only if f is the uniform density on $[-\pi, \pi]$.

<u>Remark 1.4.1</u> We mentioned in Section 1.1 that the present chapter is a revised version of Purkayastha and Bhandari (1990).

However, it should be mentioned that the characterization theorem obtained there is weaker than the one obtained here (Theorem 1.4.2), since there we restricted our attention only to those $g \in \mathcal{K}(f)$ which are concave on $[0,\pi]$. Moreover, our approach in that paper was different from the one in this chapter.

Remark 1.4.2 The results that we obtained in our early work on the problem explored in this chapter were reported in Purkayastha and Bhandari (1988). Later, while settling an open problem mentioned in that manuscript, Klaassen (1988) obtained results which are very close to those obtained in this chapter. His approach is, however, different from our approach in this chapter.

It should also be mentioned in this context that the inequality (1.2.1) in Theorem 1.2.1 follows from the fact that the norm of the Volterra integration operator on $L^2([0,1])$ is $2/\pi$ (vide Halmos (1978), Chapter 15). This was observed only recently when the final manuscript of the thesis was being typed. However, here also our approach is different.

CHAPTER 2

A ROTATIONALLY SYMMETRIC DIRECTIONAL DISTRIBUTION : OBTAINED THROUGH MAXIMUM LIKELIHOOD CHARACTERIZATION

2.1 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. This result is, however, attributed originally to Gauss (1809). Later, Ghosh and Rao (1971) solved the same problem with 'sample mean' replaced by 'sample median' and obtained a characterization of the Laplace distribution. [Vide Kagan et al. (1973), pp. 413-414 for a proof of this result.]

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. It should, however, be mentioned that when considered for circular data only, this result should be looked upon as a generalization and modern version of a result of von Mises (1918).

In view of the results stated above and also of the notion of spherical median, studied by Fisher (1985), it seems natural to ask whether it is possible to obtain a result similar to that of Bingham and Mardia (1975) with 'mean direction' replaced by 'median direction'. We explore this question in this chapter. In other words, our aim in this chapter is to characterize that rotationally symmetric directional distribution for which the median direction is a maximum likelihood estimate of the location parameter. In Section 2.2, we settle this problem for distributions on circle. This result is extended to higher dimensional spheres in Section 2.3. Finally, some general remarks in this context appear in Section 2.4.

As regards the proof of our result for distributions on circle, it should be mentioned that since there is no mathematically precise definition of median direction for sample observations from circle, following Fisher's (1985) notion of spherical median we have proposed such a definition (Definition 2.2.1).

It should also be mentioned that the distribution characterized in this chapter is a new directional distribution. An application of this distribution may be found in Cabrera and Watson (1990). It is discussed by them how this distribution might be used for analyzing reasonably concentrated directional data in three dimensions. In fact, they have established that data on the orientation of the orbits of the short-period comets fit this distribution well, but the Fisher distribution does not.

This chapter is a revised version of Purkayastha (1991a).

2.2 The circular case

We begin this section with the definition of circular median of a set of observations from S^1, the unit circle. However, before we state the definition formally, a few words motivating the definition are in order.

It is known that [vide Mardia (1972), p. 21 and p.219] the mean direction $\underline{y}_0$ of a set of observations $\underline{y}_1,\ldots,\underline{y}_n$ from S^p $(p \geq 1)$ minimizes

$$\sum_{i=1}^{n} ||\underline{y}_i - \underline{\xi}||_2^2 \qquad \ldots(2.2.1)$$

over $\underline{\xi} \in S^p$, where $||\underline{u} - \underline{v}||_2^2$ = square of the ℓ_2-norm of $\underline{u} - \underline{v}$. It is, therefore, obvious that if instead of taking the square of the ℓ_2-norm in (2.2.1) we take some other metric (or a power of it) and minimize over $\underline{\xi} \in S^p$, we might arrive at some other location of central tendency. One such metric is the geodesic distance : $d(\underline{u},\underline{v}) = \cos^{-1}(\underline{u}'\underline{v})$, where for $-1 \leq t \leq 1$, $\cos^{-1}(t)$ is the unique angle $\theta \in [0,\pi]$ such that $\cos\theta = t$.

<u>Definition 2.2.1</u> Let $\underline{x}_1,\ldots,\underline{x}_n \in S^1$. Then any point $\underline{x}_0 \in S^1$ is called a circular median of $\underline{x}_1,\ldots,\underline{x}_n$ if

$$\sum_{i=1}^{n} \cos^{-1}(\underline{x}_i' \, \underline{x}_o) = \min_{\underline{\xi} \in S^1} \sum_{i=1}^{n} \cos^{-1}(\underline{x}_i' \, \underline{\xi}). \quad \ldots(2.2.2)$$

Remark 2.2.1 The sum appearing on the right hand side of (2.2.2) is a continuous function in $\underline{\xi}$, defined over S^1 – a compact set. Hence it makes sense to talk of its minimum and define $\underline{x}_o$ accordingly. Observe, however, that $\underline{x}_o$ may not be unique.

Remark 2.2.2 The Definition 2.2.1 is actually the circular analogue of the spherical median given in Fisher (1985). It is also related to the one given in Mardia (1972, pp.28-33).

Remark 2.2.3 The Definition 2.2.1 will be studied in detail in Chapter 4.

Remark 2.2.4 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 location of central tendency.

Notation. For two points $\underline{a}, \underline{b} \in S^1$, $[\underline{a},\underline{b}]$ denotes the closed arc of S^1 with initial point $\underline{a}$, end point $\underline{b}$ and taken in clockwise sense. Thus, if we take, e.g., $\underline{a} = (0,1)$, $\underline{b} = (1,0)$, then $[\underline{a},\underline{b}] = \{(x,y) \in S^1 \mid x \geq 0, y \geq 0\}$.

Before writing down the conventions explicitly let us mention that in each of the following cases the points

$\underline{x}_1,\ldots,\underline{x}_n$ are supposed to be in clockwise order. It is also assumed that the points $\underline{x}_1,\ldots,\underline{x}_n$ are not all equal, since in that case it follows immediately from (2.2.2) that the sum appearing there on the right hand side attains its minimum only at $\underline{x}_1$; a situation boiling down to triviality. Denote now the angle, the arc $[\underline{x}_i,\underline{x}_{i+1}]$ subtends at the centre, by α_i , $i = 1,\ldots,n,(\underline{x}_{n+1} \equiv \underline{x}_1)$. It is indeed equal to the length of the arc $[\underline{x}_i, \underline{x}_{i+1}]$. Obviously, $\alpha_i \geq 0$ for every i, and $\alpha_1+\ldots+\alpha_n = 2\pi$. We also define $\underline{y}_i = -\underline{x}_i$, $i = 1,\ldots n$. Let us now state our conventions about choice of median direction.

A. Sample size $n = 2$: Assume, without loss of generality, that $\alpha_1 \leq \alpha_2$ (Fig. 2.2.1).

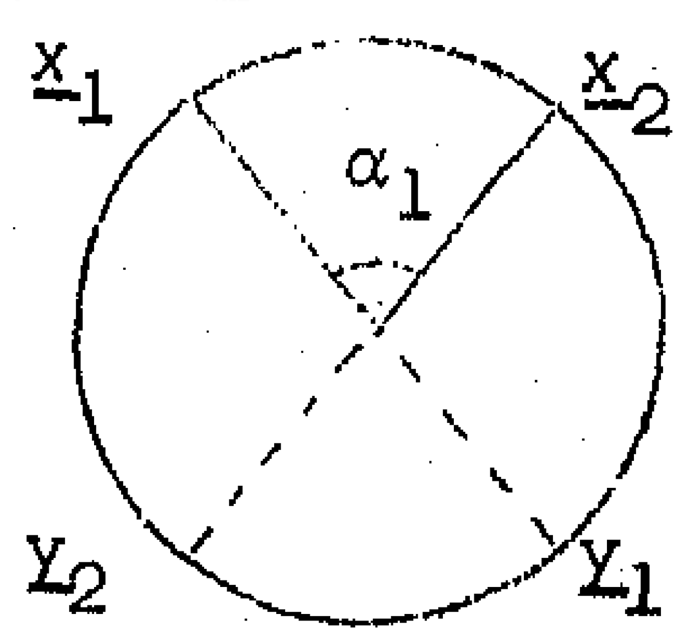


Fig. 2.2.1

Observe now that

$$\sum_{i=1}^{2} \cos^{-1}(\underline{x}_i'\underline{\xi}) = \alpha_1 \quad \text{if } \underline{\xi} \varepsilon [\underline{x}_1,\underline{x}_2],$$
$$= \alpha_1+2\cos^{-1}(\underline{x}_2'\underline{\xi}) \quad \text{if } \underline{\xi} \varepsilon [\underline{x}_2,\underline{y}_1],$$
$$= 2\pi - \alpha_1 \quad \text{if } \underline{\xi} \varepsilon [\underline{y}_1,\underline{y}_2], \quad \ldots(2.2.3)$$
$$= 2\pi-\alpha_1-2\cos^{-1}(\underline{y}_2'\underline{\xi}) \quad \text{if } \underline{\xi} \varepsilon [\underline{y}_2,\underline{x}_1].$$

To obtain now the subset of S^1 where the minimum of the sum appearing on the left hand side of (2.2.3) is attained we proceed as follows : first we minimize the sum separately on the sets $[\underline{x}_1, \underline{x}_2]$, $[\underline{x}_2, \underline{y}_1]$, $[\underline{y}_1, \underline{y}_2]$, $[\underline{y}_2, \underline{x}_1]$ and note the corresponding subsets where these minima are attained, and then we choose the minimum of these local minima and pick up the corresponding subset from the ones obtained before. The details, which are easy to verify, are as follows : the local minima are α_1, α_1, $2\pi - \alpha_1$, α_1 and the corresponding subsets where these minima are attained are $[\underline{x}_1, \underline{x}_2]$, $\{\underline{x}_2\}$, $[\underline{y}_1, \underline{y}_2]$, $\{\underline{x}_1\}$. Recall that $\alpha_1 \leq \alpha_2$. Thus, if $\alpha_1 < \alpha_2$, which happens if and only if $\alpha_1 < 2\pi - \alpha_1$, i.e., $\alpha_1 < \pi$, we have

$$\left\{ \underline{\mu} \in S^1 : \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi} \in S^1} \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\xi}) \right\} = [\underline{x}_1, \underline{x}_2] . \qquad \ldots(2.2.4)$$

On the other hand if $\alpha_1 = \alpha_2$, which happens if and only if $\alpha_1 = 2\pi - \alpha_1$, i.e., $\alpha_1 = \pi$, we have $\underline{x}_1 = -\underline{x}_2$ so that $[\underline{x}_1, \underline{x}_2] \cup [\underline{y}_1, \underline{y}_2] = S^1$, and hence

$$\left\{ \underline{\mu} \in S^1 : \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi} \in S^1} \cos^{-1}(\underline{x}_i' \underline{\xi}) \right\} = S^1 . \qquad \ldots(2.2.5)$$

In view of (2.2.4) and (2.2.5), our choice of median direction can now be stated as follows :

median direction = mid-point of $[\underline{x}_1, \underline{x}_2]$ if $\alpha_1 < \pi$,

= anyone of the mid-points of $[\underline{x}_1, \underline{x}_2]$ and $[\underline{x}_2, \underline{x}_1]$ if $\alpha_1 = \pi$.

...(2.2.6)

Observe that in the first situation the median direction and mean direction coincides.

B. <u>Sample size n = 3</u> : Assume, without loss of gene - rality, that max. $\{\alpha_1, \alpha_2, \alpha_3\} = \alpha_3$. We shall treat the two cases viz. (i) $\alpha_3 \geq \pi$ and (ii) $\alpha_3 < \pi$, separately. The two cases are illustrated in the following figures :

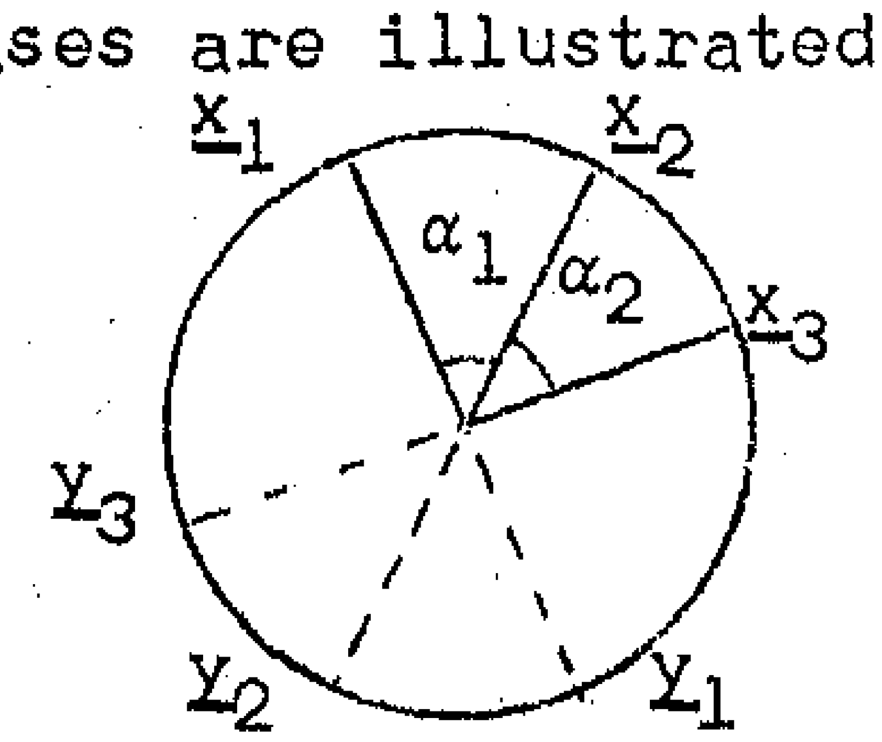


Fig. 2.2.2(a)

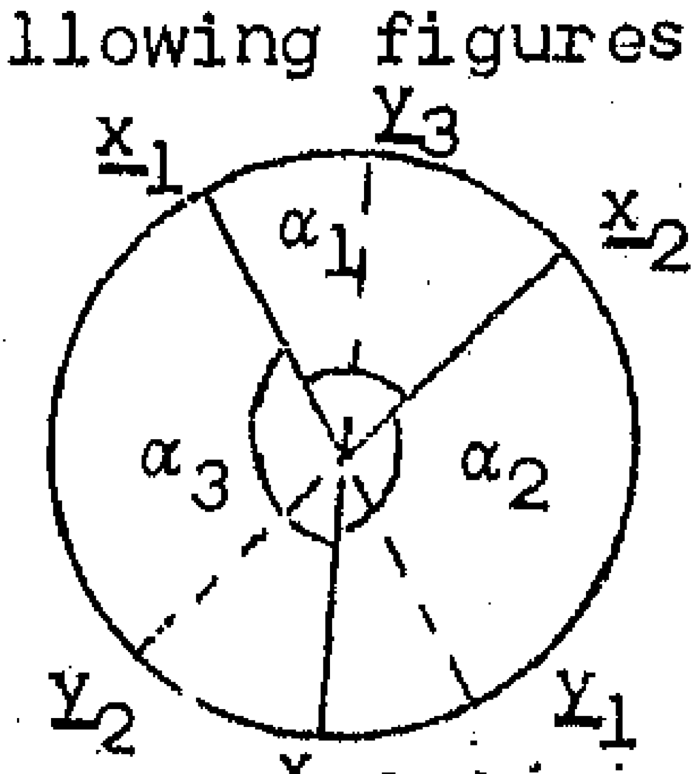


Fig. 2.2.2(b)

Observe now that

$$\sum_{i=1}^{3} \cos^{-1}(\underline{x}_i' \underline{\xi}) = \alpha_1 + \alpha_2 + \cos^{-1}(\underline{x}_2' \underline{\xi}) \quad \text{if } \underline{\xi} \in [\underline{x}_1, \underline{x}_3],$$
$$= \alpha_1 + 2\alpha_2 + 3\cos^{-1}(\underline{x}_3' \underline{\xi}) \quad \text{if } \underline{\xi} \in [\underline{x}_3, \underline{y}_1],$$
$$= 3\pi - \alpha_1 - \alpha_2 - \cos^{-1}(\underline{y}_2' \underline{\xi}) \quad \text{if } \underline{\xi} \in [\underline{y}_1, \underline{y}_3],$$
$$= 3\pi - \alpha_1 - 2\alpha_2 - 3\cos^{-1}(\underline{y}_3' \underline{\xi}) \quad \text{if } \underline{\xi} \in [\underline{y}_3, \underline{x}_1],$$

...(2.2.7)

when $\alpha_3 \geq \pi$, and

$$\begin{aligned}\sum_{i=1}^{3} \cos^{-1}(\underline{x}_i'\underline{\xi}) &= \alpha_1 + \cos^{-1}(\underline{x}_3'\underline{\xi}) \quad \text{if} \quad \underline{\xi} \in [\underline{x}_1, \underline{x}_2], \\ &= \alpha_2 + \cos^{-1}(\underline{x}_1'\underline{\xi}) \quad \text{if} \quad \underline{\xi} \in [\underline{x}_2, \underline{x}_3], \\ &= \alpha_3 + \cos^{-1}(\underline{x}_3'\underline{\xi}) \quad \text{if} \quad \underline{\xi} \in [\underline{x}_3, \underline{x}_1],\end{aligned} \qquad \ldots(2.2.8)$$

when $\alpha_3 < \pi$. From (2.2.7) and (2.2.8), using arguments similar to those employed in the case with sample size n=2, it follows that

$$\left\{ \underline{\mu} \in S^1 : \sum_{i=1}^{3} \cos^{-1}(\underline{x}_i'\underline{\mu}) = \min_{\underline{\xi} \in S^1} \sum_{i=1}^{3} \cos^{-1}(\underline{x}_i'\underline{\xi}) \right\} = \left\{ \underline{x}_2 \right\} \qquad \ldots(2.2.9)$$

when $\alpha_3 \geq \pi$, and

$$\left\{ \underline{\mu} \in S^1 : \sum_{i=1}^{3} \cos^{-1}(\underline{x}_i'\underline{\mu}) = \min_{\underline{\xi} \in S^1} \sum_{i=1}^{3} \cos^{-1}(\underline{x}_i'\underline{\xi}) \right\} = \begin{cases} \{\underline{x}_2\} & \text{if } \alpha_1 < \alpha_3,\ \alpha_2 < \alpha_3, \\ \{\underline{x}_1, \underline{x}_2\} & \text{if } \alpha_1 < \alpha_2 = \alpha_3, \\ \{\underline{x}_2, \underline{x}_3\} & \text{if } \alpha_2 < \alpha_1 = \alpha_3, \\ \{\underline{x}_1, \underline{x}_2, \underline{x}_3\} & \text{if } \alpha_1 = \alpha_2 = \alpha_3, \end{cases} \qquad \ldots(2.2.10)$$

when $\alpha_3 < \pi$.

In view of (2.2.9) and (2.2.10), our choice of median direction can now be stated as follows :

$$
\begin{aligned}
\text{median direction} &= \underline{x}_2 && \text{if } \alpha_1+\alpha_2 \le \pi, \\
&= \underline{x}_2 && \text{if } \alpha_1+\alpha_2 > \pi \text{ and } \alpha_1 < \alpha_3,\ \alpha_2 < \pi, \\
&= \text{anyone of } \underline{x}_1 \text{ and } \underline{x}_2 && \text{if } \alpha_1+\alpha_2 > \pi \text{ and } \alpha_1 < \alpha_2 = \alpha_3, \\
&= \text{anyone of } \underline{x}_2 \text{ and } \underline{x}_3 && \text{if } \alpha_1+\alpha_2 > \pi \text{ and } \alpha_2 < \alpha_1 = \alpha_3, \\
&= \text{anyone of } \underline{x}_1, \underline{x}_2 \text{ and } \underline{x}_3 && \text{if } \alpha_1 = \alpha_2 = \alpha_3 .
\end{aligned}
\qquad \ldots(2.2.11)
$$

C. Sample size $n = 4$: In this situation we shall treat the two cases viz. (i) $\max.\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} > \pi$ and (ii) $\max.\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \le \pi$, separately. The two cases are illustrated in the following figures :

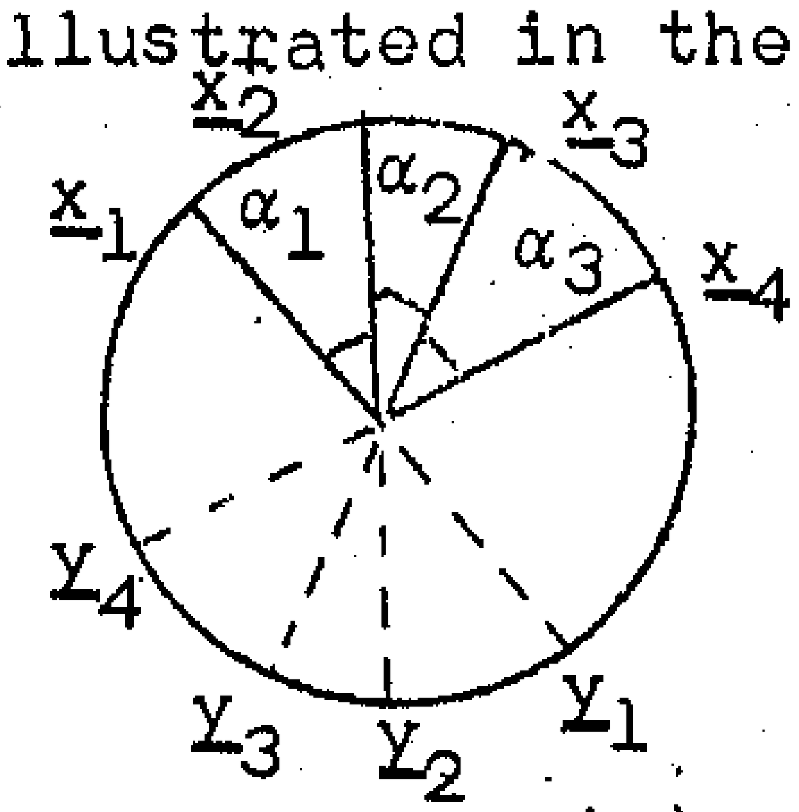


Fig. 2.2.3(a)

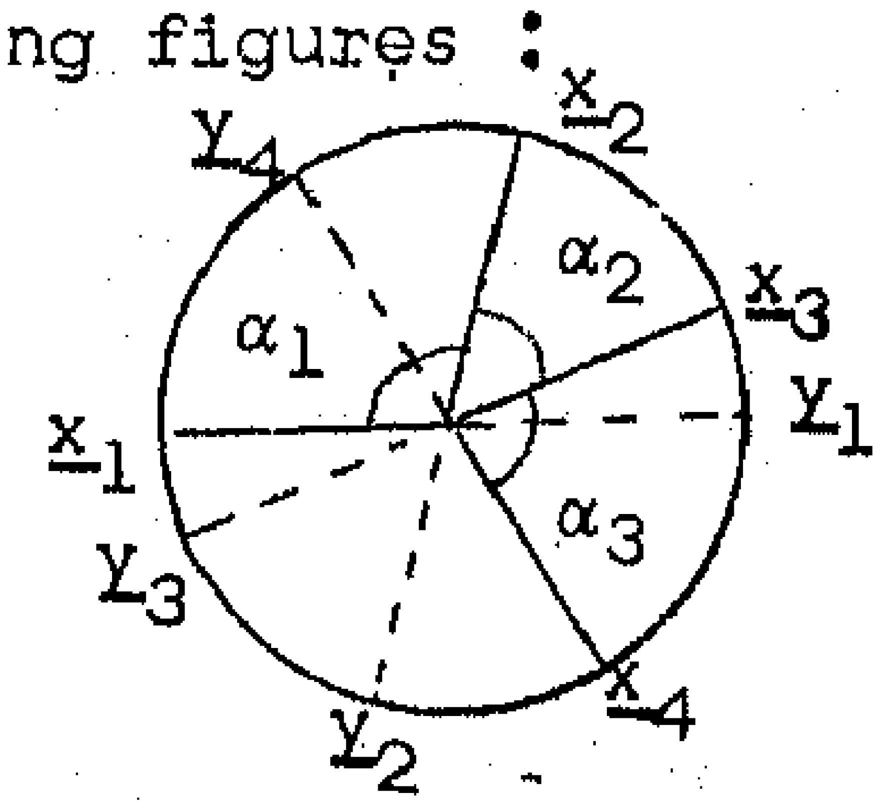


Fig. 2.2.3(b)

In the first case, observe first that we cannot have two of the α_i's greater than π since in that case the restriction $\alpha_1+\alpha_2+\alpha_3+\alpha_4 = 2\pi$ is violated. Thus there exists a unique i such that $\alpha_i > \pi$. Assume, without loss of generality, that $\alpha_4 > \pi$. Observe now that

$$\begin{aligned}
\sum_{i=1}^{4} \cos^{-1}(\underline{x}_i'\underline{\xi}) &= \alpha_1+2\alpha_2+\alpha_3+2\cos^{-1}(\underline{x}_2'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{x}_1,\underline{x}_2], \\
&= \alpha_1+2\alpha_2+\alpha_3 && \text{if } \underline{\xi}\varepsilon[\underline{x}_2,\underline{x}_3], \\
&= \alpha_1+2\alpha_2+\alpha_3+2\cos^{-1}(\underline{x}_3'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{x}_3,\underline{x}_4], \\
&= \alpha_1+2\alpha_2+3\alpha_3+4\cos^{-1}(\underline{x}_4'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{x}_4,\underline{y}_1], \\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3-2\cos^{-1}(\underline{y}_2'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{y}_1,\underline{y}_2], \\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3 && \text{if } \underline{\xi}\varepsilon[\underline{y}_2,\underline{y}_3], \\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3-2\cos^{-1}(\underline{y}_3'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{y}_3,\underline{y}_4], \\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3-4\cos^{-1}(\underline{y}_4'\underline{\xi}) && \text{if } \underline{\xi}\varepsilon[\underline{y}_4,\underline{x}_1].
\end{aligned} \qquad \ldots(2.2.12)$$

In the second case, we replace the condition $\max\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\} \leq \pi$ by a more tractable one before we obtain an explicit expression for the sum $\sum_{i=1}^{4} \cos^{-1}(\underline{x}_i'\underline{\xi})$. To achieve this, we make use of the following fact, which is easy to verify : if $t_1+t_2+t_3+t_4 = 2\pi$, then at least one of the following must be true -

(a) $t_1+t_2 \leq \pi$ and $t_2+t_3 \leq \pi$, ...(2.2.13.1)

(b) $t_2+t_3 \leq \pi$ and $t_3+t_4 \leq \pi$, ...(2.2.13.2)

(c) $t_3+t_4 \leq \pi$ and $t_4+t_1 \leq \pi$, ...(2.2.13.3)

(d) $t_4+t_1 \leq \pi$ and $t_1+t_2 \leq \pi$. ...(2.2.13.4)

In our case we have $t_i = \alpha_i$, $1 \le i \le 4$, and we assume, without loss of generality, that (2.2.13.1) holds. It is then easy to see that the condition $\max\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\} \le \pi$ is equivalent to the following : $\pi \le \alpha_1+\alpha_2+\alpha_3 \le 2\pi$, $\alpha_1+\alpha_2 \le \pi$, $\alpha_2+\alpha_3 \le \pi$. Observe now that

$$\begin{aligned}
\sum_{i=1}^{4} \cos^{-1}(\underline{x}_i'\underline{\xi}) &= 2\pi+\alpha_1-\alpha_3 && \text{if } \underline{\xi}\in[\underline{x}_1,\underline{y}_4],\\
&= \alpha_1+2\alpha_2+\alpha_3+2\cos^{-1}(\underline{x}_2'\underline{\xi}) && \text{if } \underline{\xi}\in[\underline{y}_4,\underline{x}_2],\\
&= \alpha_1+2\alpha_2+\alpha_3 && \text{if } \underline{\xi}\in[\underline{x}_2,\underline{x}_3],\\
&= \alpha_1+2\alpha_2+\alpha_3+2\cos^{-1}(\underline{x}_3'\underline{\xi}) && \text{if } \underline{\xi}\in[\underline{x}_3,\underline{y}_1],\\
&= 2\pi-\alpha_1+\alpha_3 && \text{if } \underline{\xi}\in[\underline{y}_1,\underline{x}_4],\\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3-2\cos^{-1}(\underline{y}_2'\underline{\xi}) && \text{if } \underline{\xi}\in[\underline{x}_4,\underline{y}_2],\\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3 && \text{if } \underline{\xi}\in[\underline{y}_2,\underline{y}_3],\\
&= 4\pi-\alpha_1-2\alpha_2-\alpha_3-2\cos^{-1}(\underline{y}_3'\underline{\xi}) && \text{if } \underline{\xi}\in[\underline{y}_3,\underline{x}_1].
\end{aligned} \quad \ldots(2.2.14)$$

From (2.2.12) and (2.2.14), using arguments similar to those employed in the case with sample size $n = 2$, it follows that

$$\left\{ \underline{\mu}\in S^1 : \sum_{i=1}^{4}\cos^{-1}(\underline{x}_i'\underline{\mu}) = \min_{\underline{\xi}\in S^1} \sum_{i=1}^{4}\cos^{-1}(\underline{x}_i'\underline{\xi}) \right\} = \begin{cases} [\underline{x}_2,\underline{x}_3] & \text{if } (\alpha_1,\alpha_2,\alpha_3)\ne(0,\pi,0),\\ S^1 & \text{if } \alpha_1=\alpha_3=0,\ \alpha_2=\pi, \end{cases} \quad \ldots(2.2.15)$$

when $\alpha_1+\alpha_2+\alpha_3 \leq \pi$, and

$$\left\{ \underline{\mu}\varepsilon S^1 : \sum_{i=1}^{4} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi}\varepsilon S^1} \sum_{i=1}^{4} \cos^{-1}(\underline{x}_i'\underline{\xi}) \right\}$$

$$= \begin{cases} [\underline{x}_2, \underline{x}_3] & \text{if } \alpha_1+\alpha_2<\pi,\ \alpha_2+\alpha_3<\pi, \\ [\underline{x}_2, \underline{x}_4] & \text{if } \alpha_1+\alpha_2=\pi,\ \alpha_2+\alpha_3<\pi, \\ [\underline{x}_1, \underline{x}_3] & \text{if } \alpha_1+\alpha_2<\pi,\ \alpha_2+\alpha_3=\pi, \\ S^1 & \text{if } \alpha_1+\alpha_2=\pi,\ \alpha_2+\alpha_3=\pi, \end{cases} \quad \ldots(2.2.16)$$

when $\pi < \alpha_1 + \alpha_2 + \alpha_3 < 2\pi$.

In view of (2.2.15) and (2.2.16), our choice of median direction can now be stated as follows :

median direction = mid-point of $[\underline{x}_2, \underline{x}_3]$ if $\alpha_1+\alpha_2+\alpha_3 \leq \pi$, $(\alpha_1, \alpha_2, \alpha_3) \neq (0, \pi, 0)$,

= anyone of the mid-points of $[\underline{x}_1, \underline{x}_3]$ and $[\underline{x}_3, \underline{x}_1]$ if $\alpha_1=\alpha_3=0$, $\alpha_2=\pi$,

= mid-point of $[\underline{x}_2, \underline{x}_3]$ if $\pi<\alpha_1+\alpha_2+\alpha_3 < 2\pi$, $\alpha_1+\alpha_2<\pi$, $\alpha_2+\alpha_3<\pi$,

= mid-point of $[\underline{x}_2, \underline{x}_4]$ if $\pi<\alpha_1+\alpha_2+\alpha_3<2\pi$, $\alpha_1+\alpha_2=\pi$, $\alpha_2+\alpha_3<\pi$,

= mid-point of $[\underline{x}_1, \underline{x}_3]$ if $\pi<\alpha_1+\alpha_2+\alpha_3<2\pi$, $\alpha_1+\alpha_2<\pi$, $\alpha_2+\alpha_3=\pi$,

= anyone of the mid-points of $[\underline{x}_2, \underline{x}_3]$ and $[\underline{x}_4, \underline{x}_1]$ if $\pi<\alpha_1+\alpha_2+\alpha_3<2\pi$, $\alpha_1+\alpha_2=\pi$, $\alpha_2+\alpha_3=\pi$, $\alpha_1>\alpha_2$,

= anyone of the mid-points of $[\underline{x}_1, \underline{x}_2]$ and $[\underline{x}_3, \underline{x}_4]$	if $\pi<\alpha_1+\alpha_2+\alpha_3<2\pi$, $\alpha_1+\alpha_2=\pi$, $\alpha_2+\alpha_3=\pi$, $\alpha_1<\alpha_2$.
= anyone of the mid-points of $[\underline{x}_1, \underline{x}_2]$, $[\underline{x}_2, \underline{x}_3]$, $[\underline{x}_3, \underline{x}_4]$ and $[\underline{x}_4, \underline{x}_1]$	if $\alpha_1=\alpha_2=\alpha_3=\alpha_4=\frac{\pi}{2}$.

...(2.2.17)

Before we conclude this remark we recall that in none of the cases above we assumed, to begin with, that $\cos^{-1}(\underline{x}_i'\underline{x}_{i+1}) = \alpha_i$, $i = 1,\ldots,n-1$. However, before we proceed to choose the median direction we have to choose a particular $\underline{x}_i$, and rename $\underline{x}_i, \underline{x}_{i+1},\ldots,\underline{x}_n, \underline{x}_1,\ldots,\underline{x}_{i-1}$ as $\underline{x}_1,\ldots,\underline{x}_n$ so that conditions like (i) $\alpha_1 \leq \alpha_2$ - for sample size $n = 2$, (ii) $\max\{\alpha_1,\alpha_2,\alpha_3\} = \alpha_3$ - for sample size $n = 3$, etc. are satisfied for the newly defined α_i's. As a consequence of this, finally it turns out that $\cos^{-1}(\underline{x}_i'\underline{x}_{i+1}) = \alpha_i$, $i=1,\ldots,n-1$.

We now state and prove the main result of this section. We recall that a density $p(\underline{x})$ on the sphere S^p which is rotationally symmetric about a fixed point $\underline{\theta} \varepsilon S^p$ has the form $f(\underline{x}'\underline{\theta})$ for some f [vide Watson (1983), p.92] : the justification of the term 'rotationally symmetric' lies in the fact that for every $(p+1)x(p+1)$ orthogonal matrix A with det (A) = 1 which satisfies $A\underline{\theta} = \underline{\theta}$, we have $p(\underline{x}) = p(\underline{y})$ whenever $\underline{y} = A\underline{x}$. Also, because of this geometric consideration we call $\underline{\theta}$ the location parameter for the class of distributions induced by the family of densities $\{f(\underline{x}'\underline{\theta}) \mid \underline{\theta} \varepsilon S^p\}$.

In passing we mention that for the special case, when $p = 1$, rotational symmetry is called circular symmetry.

Theorem 2.2.1 Let $\{p(\underline{x};\underline{\theta}) = f(\underline{x}'\underline{\theta}) \mid \underline{\theta} \in S^1\}$ 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 sample median direction be a maximum likelihood estimate of $\underline{\theta}$ for all samples of size $n = 4$, then

$$p(\underline{x};\underline{\theta}) = \frac{\lambda}{2(1-e^{-\lambda\pi})} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}, \quad \underline{x} \in S^1, \quad \ldots(2.2.18)$$

where λ is a positive constant.

Proof The assumption that the sample median direction is a maximum likelihood estimate of $\underline{\theta}$ for all samples of size $n = 4$ means the following :

$$\prod_{i=1}^{4} f(\underline{x}_i'\underline{x}_o) \geq \prod_{i=1}^{4} f(\underline{x}_i'\underline{\theta}) \quad \forall \ \underline{\theta} \in S^1, \quad \ldots(2.2.19)$$

and for all samples $(\underline{x}_1,\ldots,\underline{x}_4)$ of size $n = 4$; $\underline{x}_o$ being the sample median direction.

We shall now express (2.2.19) in terms of the function g, defined as $g(t) = f(\cos t)$, $0 \leq t \leq \pi$. To do it, choose $\underline{x}_i \in S^1$, $i = 1,\ldots,4$, such that $\underline{x}_i'\underline{x}_{i+1} = \cos \alpha_i$, $i=1,\ldots,3$, satisfy $\alpha_i \geq 0$, $0 \leq \alpha_1+\alpha_2+\alpha_3 \leq \pi$ (cf. Fig. 2.2.3(a)). Clearly, the median direction $\underline{x}_o$, which is the mid-point of $[\underline{x}_2,\underline{x}_3]$

in this case (vide Remark 2.2.4, part C), satisfies $\underline{x}_1'\underline{x}_o = \cos(\alpha_1+\alpha_2/2)$, $\underline{x}_2'\underline{x}_o = \underline{x}_3'\underline{x}_o = \cos\,\alpha_2/2$, $\underline{x}_4'\underline{x}_o = \cos(\alpha_3+\alpha_2/2)$. Choose now an arbitrary point $\underline{\theta} \in [\underline{x}_2, \underline{x}_o]$. Suppose $\underline{x}_2'\underline{\theta} = \cos x$. Obviously $0 \leq x \leq \alpha_2/2$, so that from (2.2.19) we obtain the following :

$$f[\cos(\alpha_1+\alpha_2/2)]f[\cos\,\alpha_2/2]f[\cos\,\alpha_2/2]f[\cos(\alpha_3+\alpha_2/2)]$$

$$\geq f[\cos(\alpha_1+\alpha_2/2+x)]f[\cos(\alpha_2/2+x)]f[\cos(\alpha_2/2-x)]f[\cos(\alpha_2/2+\alpha_3-x)]$$

for every x such that $0 \leq x \leq \alpha_2/2$, i.e.,

$$g(\alpha_1+\alpha_2/2)g^2(\alpha_2/2)g(\alpha_3+\alpha_2/2)$$

$$\geq g(\alpha_1+\alpha_2/2+x)g(\alpha_2/2+x)g(\alpha_2/2-x)g(\alpha_3+\alpha_2/2-x)$$

for every x such that $0 \leq x \leq \alpha_2/2$. Proceeding in this way we obtain the following :

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

$$g(\alpha_1+\alpha_2/2)g^2(\alpha_2/2)g(\alpha_3+\alpha_2/2)$$

$$\geq \begin{cases} g(\alpha_1+\alpha_2/2+x)g(\alpha_2/2+x)g(\alpha_2/2-x)g(\alpha_3+\alpha_2/2-x), & 0\leq x\leq \alpha_2/2 \quad \ldots(2.2.20.1) \\ g(\alpha_1+\alpha_2/2+x)g(\alpha_2/2+x)g(x-\alpha_2/2)g(\alpha_3+\alpha_2/2-x), & \alpha_2/2\leq x\leq \alpha_2/2+\alpha_3 \quad \ldots(2.2.20.2) \\ g(\alpha_1+\alpha_2/2+x)g(\alpha_2/2+x)g(x-\alpha_2/2)g(x-\alpha_2/2-\alpha_3), & \alpha_2/2+\alpha_3\leq x\leq \pi-(\alpha_1+\alpha_2/2) \quad \ldots(2.2.20.3) \end{cases}$$

and

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

$$g(\alpha_1+\alpha_2/2)g^2(\alpha_2/2)g(\alpha_3+\alpha_2/2)$$
$$\geq g(\alpha_1+\alpha_2/2+x)g(\alpha_2/2+x)g(\alpha_2/2-x)g(\alpha_2/2+\alpha_3-x), \quad 0 \leq x \leq \alpha_2/2. \qquad \ldots(2.2.21)$$

Observe now that if we choose $\alpha_1 = \alpha_3 = 0$ in (2.2.20.1), we obtain

$$g^4(\frac{\alpha_2}{2}) \geq g^2(\frac{\alpha_2}{2} + x)g^2(\frac{\alpha_2}{2} - x)$$

for every x such that $0 \leq x \leq \alpha_2/2$, where $0 \leq \alpha_2 \leq \pi$, i.e.,

$$g^2(\frac{\alpha_2}{2}) \geq g(y)g(\alpha_2-y) \qquad \ldots(2.2.22)$$

for every y such that $0 \leq y \leq \alpha_2$, where $0 \leq \alpha_2 \leq \pi$. Therefore, if $g(y) = \infty$ for some y such that $0 \leq y < \pi$, (2.2.22) implies that for every $\alpha_2 \in (y,\pi)$, either $g(\alpha_2-y) = 0$ or $g(\alpha_2/2) = \infty$. The former relation cannot hold because of the restriction on f put forth in the statement of the theorem and the latter relation cannot hold on a set of positive Lebesgue measure. Therefore,

$$g(t) < \infty, \quad 0 \leq t < \pi. \qquad \ldots(2.2.23)$$

Observe further that if we choose $\alpha_1 = \alpha_2 = \alpha_3 = 0$ in (2.2.20.3) (this is indeed the situation when all the 4 observations are equal), we obtain

$$g^4(0) \geq g^4(x), \quad 0 \leq x \leq \pi , \qquad \ldots(2.2.24)$$

which implies

$$g(0) > 0 ,$$

otherwise $g(x) = 0$ for all $x \in [0,\pi]$, a contradiction to the fact that f is a density on S^1 . We also have $g(t) > 0$ for every t such that $0 < t < \pi$. Hence, it is legitimate to define

$$h(t) = \log g(t), \quad 0 \leq t < \pi . \qquad \ldots(2.2.25)$$

From (2.2.20.1), (2.2.20.2), (2.2.21) and (2.2.25), we now obtain the following :

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

$$h(\alpha_1+\alpha_2/2)+2h(\alpha_2/2)+h(\alpha_3+\alpha_2/2)$$

$$\geq \begin{cases} h(\alpha_1+\alpha_2/2+x)+h(\alpha_2/2+x)+h(\alpha_2/2-x)+h(\alpha_2/2+\alpha_3-x), \\ \qquad 0 \leq x \leq \alpha_2/2, \qquad \ldots(2.2.26.1) \\ h(\alpha_1+\alpha_2/2+x)+h(\alpha_2/2+x)+h(x-\alpha_2/2)+h(\alpha_3+\alpha_2/2-x), \\ \qquad \alpha_2/2 \leq x \leq \alpha_2/2+\alpha_3, \qquad \ldots(2.2.26.2) \end{cases}$$

and

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

$\alpha_1+\alpha_2 < \pi$ and $\alpha_2+\alpha_3 < \pi$,

$$h(\alpha_1+\alpha_2/2)+2h(\alpha_2/2)+h(\alpha_3+\alpha_2/2)$$
$$\geq h(\alpha_1+\alpha_2/2+x)+h(\alpha_2/2+x)+h(\alpha_2/2-x)+h(\alpha_3+\alpha_2/2-x), \quad 0\leq x\leq \alpha_2/2 \qquad \ldots(2.2.27)$$

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

At first we prove that

$$h \text{ is concave on } [0, \tfrac{\pi}{2}] \qquad \ldots(2.2.28)$$

To see this, choose and fix $t_1, t_2 \in [0,\frac{\pi}{2}]$ with $t_1 < t_2$. Now use (2.2.26.1) with $\alpha_1 = \alpha_3 = 0$, $\alpha_2 = t_1 + t_2$ and $x = \frac{t_2-t_1}{2}$ to obtain

$$h(\frac{t_1+t_2}{2}) \geq \frac{h(t_1)+h(t_2)}{2},$$

in view of measurability of h which establishes (2.2.28).

Next we prove that

$$h(t) = -\lambda t+\mu , \quad 0 \leq t < \pi/2 , \qquad \ldots(2.2.29)$$

for some constants λ and μ. Observe that as a consequence of (2.2.28), h is differentiable on $(0,\pi/2)$, except (possibly) on a subset $\mathcal{D}$ of $(0,\pi/2)$ which is at most countable. Write $\mathcal{A} = (0,\pi/2) - \mathcal{D}$. Choose and fix $t_1, t_2 \in \mathcal{A}$ with $t_1 < t_2$. Choose, moreover, $t_3 \in \mathcal{A}$ with $t_3 < t_1$. In (2.2.26.1), take now $\alpha_1 = t_1-t_3$, $\alpha_2 = 2t_3$ and $\alpha_3 = t_2-t_3$ to obtain

$$h(t_1)+2h(t_3)+h(t_2)$$
$$\geq h(x+t_1)+h(t_3+x)+h(t_3-x)+h(t_2-x), \quad 0 \leq x < t_3.$$

Thus the function $h^* : [0,t_3) \longrightarrow \mathbb{R}$ 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,

$$\lim_{x \to 0+} \frac{h^*(x)-h^*(0)}{x} \leq 0 ,$$

which implies

$$h'(t_1) \leq h'(t_2).$$

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

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

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

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

This implies, in view of concavity of h and the fact that $\mathcal{A}$ is dense in $(0, \frac{\pi}{2})$, that h is differentiable everywhere on $(0, \frac{\pi}{2})$ with a constant derivative. This establishes assertion (2.2.29) for $0 < t < \frac{\pi}{2}$, i.e.,

$$h(t) = -\lambda t + \mu , \quad 0 < t < \frac{\pi}{2} . \qquad \ldots(2.2.30)$$

To complete the proof of (2.2.29), we now prove that $h(0) = \mu$. To see this, first choose three small positive numbers $\alpha_1, \alpha_2, \alpha_3$, and for this choice of α_i's, use (2.2.26.1) with $x = \frac{\alpha_2}{2}$ to obtain

$$h(\alpha_1+\frac{\alpha_2}{2})+2h(\frac{\alpha_2}{2})+h(\alpha_3+\frac{\alpha_2}{2}) \geq h(\alpha_1+\alpha_2)+h(\alpha_2)+h(0)+h(\alpha_3) \quad \ldots(2.2.31.1)$$

Employ (2.2.30) now in (2.2.31.1) to obtain

$$\mu \geq h(0) . \quad \ldots(2.2.31.2)$$

Again, choose two small positive numbers α_1, α_3 and $\alpha_2 = 0$, and for this choice of α_i's , use (2.2.26.2) with $x = \frac{\alpha_3}{2}$ to obtain

$$h(\alpha_1)+2h(0)+h(\alpha_3) \geq h(\alpha_1+\frac{\alpha_3}{2})+2h(\frac{\alpha_3}{2})+h(\frac{\alpha_3}{2}) \ldots(2.2.32.1)$$

Employ (2.2.30) now in (2.2.32.1) to obtain $h(0) \geq h(\frac{\alpha_3}{2}) = -\frac{\lambda\alpha_3}{2} + \mu$, from which we obtain

$$h(0) \geq \mu , \quad \ldots(2.2.32.2)$$

by taking limit as $\alpha_3 \longrightarrow 0+$. The fact that $h(0) = \mu$ is now an immediate consequence of (2.2.31.2) and (2.2.32.2) . Taken with (2.2.30), this establishes (2.2.29).

The next step consists in proving that

$$h \text{ is concave on } (\frac{\pi}{2}, \pi) . \quad \ldots(2.2.33)$$

To see this, choose and fix $t_1, t_2 \in (\frac{\pi}{2}, \pi)$ with $t_1 < t_2$. Now use (2.2.27) with $\alpha_1 = \alpha_3 = t_1, \alpha_2 = t_2 - t_1$ and $x = \frac{t_2 - t_1}{2}$ to obtain

$$2h(\frac{t_1+t_2}{2})+2h(\frac{t_2-t_1}{2}) \geq h(t_2)+h(t_2-t_1)+h(0)+h(t_1). \quad \ldots(2.2.34)$$

But our choice of t_1 and t_2 implies $0 < \frac{t_2-t_1}{2} < t_2-t_1 < \frac{\pi}{2}$, so that in view of (2.2.29), we obtain

$$2h(\frac{t_2-t_1}{2}) = h(t_2-t_1) + h(0) . \qquad \ldots(2.2.35)$$

From (2.2.34) and (2.2.35), we obtain

$$h(\frac{t_1+t_2}{2}) \geq \frac{h(t_1)+h(t_2)}{2} ,$$

in view of measurability of h which establishes (2.2.33).

The next step is analogous to (2.2.29). Employing arguments similar to those required to establish (2.2.29), we obtain from (2.2.33), (2.2.27) and (2.2.29)

$$h(t) = -\gamma t + \nu , \quad \frac{\pi}{2} < t < \pi , \qquad \ldots(2.2.36)$$

for some constants γ and ν.

Next we prove that

$$h \text{ is convex on } (\frac{\pi}{2} - \delta, \frac{\pi}{2} + \delta) , \qquad \ldots(2.2.37)$$

if we choose δ to be a sufficiently small positive number. To see this, choose and fix $t_1, t_2 \in (\frac{\pi}{2} - \delta, \frac{\pi}{2} + \delta)$ with $t_1 < t_2$. Choose now $\alpha_1, \alpha_2, \alpha_3 \in (0, \pi)$ such that $\alpha_1 + \frac{\alpha_2}{2} = t_1$, $\alpha_3 + \frac{\alpha_2}{2} = t_2$, $\alpha_3 - \alpha_2 < \alpha_2$, $\alpha_1 + \alpha_2 < \pi$ and $\alpha_2 + \alpha_3 < \pi$; such a choice is possible since δ is assumed to be a small positive number. For this choice of α_i's use now (2.2.26.1) with $x = \frac{\alpha_3-\alpha_1}{2}$ if $t_1+t_2 \leq \pi$, to obtain

$$h(t_1)+2h(\tfrac{\alpha_2}{2})+h(t_2) \geq 2h(\tfrac{t_1+t_2}{2})+h(\tfrac{\alpha_2}{2}+\tfrac{\alpha_3-\alpha_1}{2})+h(\tfrac{\alpha_2}{2}-\tfrac{\alpha_3-\alpha_1}{2}). \quad \ldots(2.2.38)$$

However, if $t_1+t_2 > \pi$, we use (2.2.27) with $x = \frac{\alpha_3-\alpha_1}{2}$ to obtain again (2.2.38). Observe now that subject to the restrictions stated earlier it is possible to choose the α_i's in a way so that 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 (2.2.29)

$$2h(\tfrac{\alpha_2}{2}) = h(\tfrac{\alpha_2}{2}+\tfrac{\alpha_3-\alpha_1}{2})+h(\tfrac{\alpha_2}{2}-\tfrac{\alpha_3-\alpha_1}{2}). \quad \ldots(2.2.39)$$

Thus, from (2.2.38) and (2.2.39), we obtain

$$\frac{h(t_1)+h(t_2)}{2} \geq h(\tfrac{t_1+t_2}{2}),$$

in view of measurability of h which implies (2.2.37).

Employing arguments similar to those required to establish (2.2.29), we now obtain from (2.2.37), (2.2.26.1), (2.2.27) and (2.2.29) that the graph of h on $(\frac{\pi}{2}-\delta, \frac{\pi}{2}+\delta)$ is a straight line. In view of (2.2.29) and (2.2.36), this implies

$$\lambda = \tau \quad \text{and} \quad \mu = \nu,$$

where λ, μ, τ, ν are the constants obtained in (2.2.29) and (2.2.36).

We, therefore, obtain

$$h(t) = -\lambda t+\mu, \quad 0 \leq t < \pi,$$

that is,

$$g(t) = \exp(-\lambda t + \mu), \quad 0 \leq t < \pi .$$

Moreover, right-continuity of $f(t)$ at $t = -1$ implies left-continuity of $g(t)$ at $t = \pi$. Hence,

$$g(t) = \exp(-\lambda t + \mu), \quad 0 \leq t \leq \pi .$$

Observe also that in view of (2.2.24) and stipulated non-uniformity of f, we obtain $\lambda > 0$.

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

$$p(\underline{x}; \underline{\theta}) = e^{\mu} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}, \quad \underline{x} \in S^1$$

where λ is a positive constant. Observe, however, that an easy integration yields

$$\int_{S^1} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})} \, d\underline{x} = \frac{2(1-e^{-\lambda\pi})}{\lambda},$$

so that

$$e^{\mu} = \frac{\lambda}{2(1-e^{-\lambda\pi})},$$

since $\int_{S^1} p(\underline{x};\underline{\theta}) \, d\underline{x} = 1$. This completes the proof of the theorem.

Remark 2.2.5 The theorem is false if we require the sample median direction to be a maximum likelihood estimate of $\underline{\theta}$ for all samples of size $n = 2$. To see this, consider the following class $\mathcal{F}_1$ of circularly symmetric non-uniform densities on S^1 :

$$\mathcal{F}_1 = \left\{ p(\underline{x};\underline{\theta}) = f(\underline{x}'\underline{\theta}) \,\middle|\, \underline{\theta} \in S^1 \right\} ,$$

where

$$f(t) = K \exp\left\{ h(\cos^{-1} t) \right\}, \; -1 \leq t \leq 1 ,$$

with $h : [0,\pi] \longrightarrow \mathbb{R}$ being defined as

$$\begin{aligned} h(t) &= -\lambda t + \mu , \quad 0 \leq t \leq \tfrac{\pi}{2} , \\ &= -\tau t + \nu, \quad \tfrac{\pi}{2} \leq t \leq \pi , \end{aligned}$$

where $\tau > \lambda > 0$, $\frac{\pi}{2} (\tau - \lambda) = \nu - \mu$ and

$$\frac{1}{K} = \frac{2e^{\mu}(1-e^{-\lambda\pi/2})}{\lambda} + \frac{2e^{\nu}(e^{-\tau\pi/2} - e^{-\tau\pi})}{\tau} .$$

Obviously f is not of the form described in (2.2.18). We are, therefore, required to prove that $\mathcal{F}_1$ indeed serves as a counter-example to the assertion of Theorem 2.2.1 with $n = 2$ samples, that is, we have to show that

$$\prod_{i=1}^{2} f(\underline{x}_i' \underline{x}_0) \geq \prod_{i=1}^{2} f(\underline{x}_i' \underline{\theta}) \quad \forall\, \underline{\theta} \in S^1 \qquad \ldots(2.2.40)$$

and for all samples $(\underline{x}_1, \underline{x}_2)$ of size $n = 2$; $\underline{x}_0$ being the sample median direction. Observe now that, in view of our convention about choice of median direction for samples of size $n = 2$ (described in part A of Remark 2.2.4) , (2.2.40) may be written alternatively as follows :

for every $0 \leq \alpha \leq \pi$,

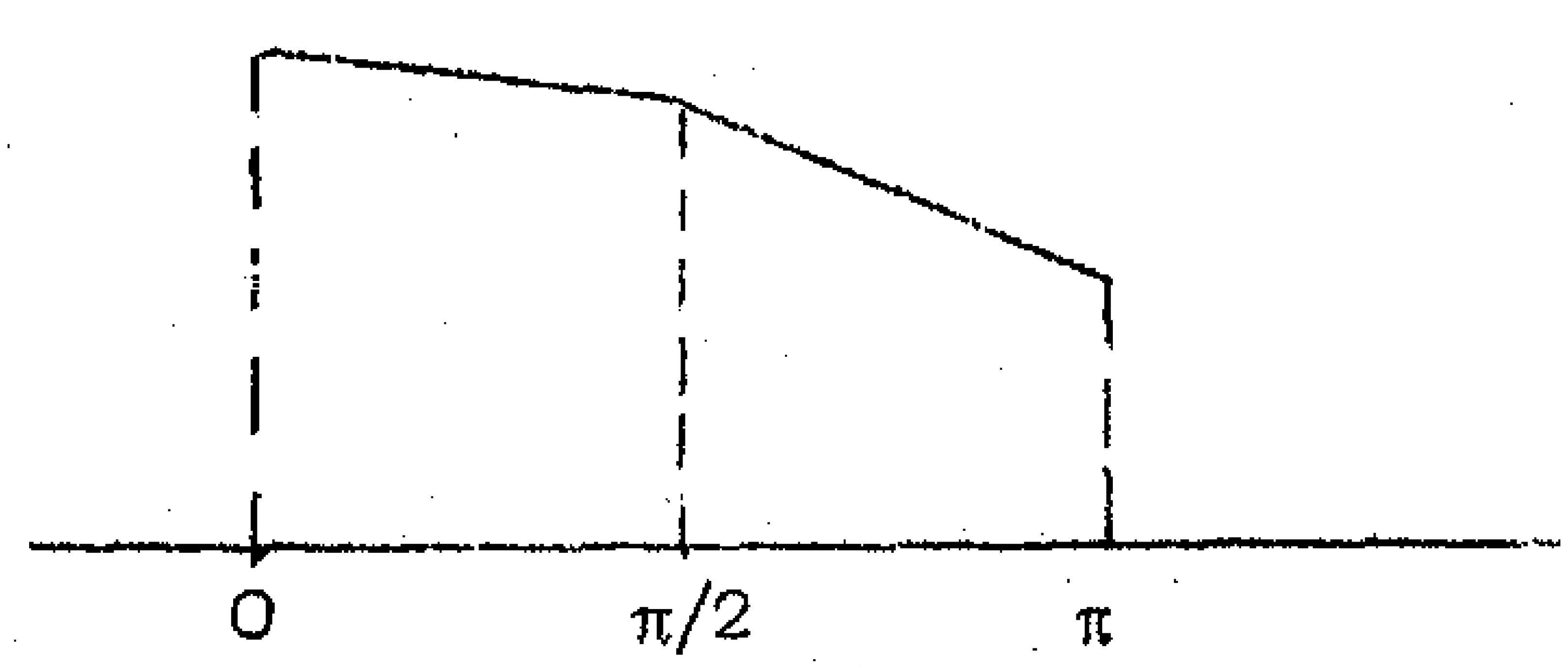
$$f^2(\cos \frac{\alpha}{2}) \geq \begin{cases} f(\cos x)f(\cos(\alpha-x)), & 0\leq x\leq\alpha, \\ f(\cos x)f(\cos(\alpha+x)), & 0\leq x\leq\pi-\alpha, \\ f(\cos(\pi-x))f(\cos(\pi-\alpha+x)), & 0\leq x\leq\alpha. \end{cases} \quad \ldots(2.2.41)$$

Again, in terms of h, (2.2.41) may be written as follows :

for every $0 \leq \alpha \leq \pi$,

$$2h(\frac{\alpha}{2}) \geq \begin{cases} h(x)+h(\alpha-x), & 0\leq x\leq\alpha, \\ h(x)+h(\alpha+x), & 0\leq x\leq\pi-\alpha, \\ h(\pi-x)+h(\pi-\alpha+x), & 0\leq x\leq\alpha. \end{cases} \quad \ldots(2.2.42)$$

We have to verify (2.2.42). Observe that the graph of h is as follows :



It is evident from the graph that h is a decreasing concave function and it is easy to verify that (2.2.42) follows immediately from this fact.

Before we conclude this remark we make the observation that in order to verify that $\mathcal{F}_1$ serves as a counterexample to the assertion of Theorem 2.2.1 with samples of size $n = 2$ we did **not** use the specific form of the function **h**, rather we used only the fact that h is a decreasing concave function. We, may, therefore, take any decreasing concave function in place of h which is not constant (this is to avoid uniformity) and is not of the form $h(t) = -\lambda t + \mu$, $0 \leq t \leq \pi$, for some constants λ and μ, λ positive.

<u>Remark 2.2.6</u> The theorem is false if we require the sample median direction to be a maximum likelihood estimate of $\underset{\sim}{\theta}$ for all samples of size $n = 3$. To see this, consider the following class $\mathcal{F}_2$ of circularly symmetric non-uniform densities on S^1 :

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

where

$$f(t) = K \exp\left\{ h(\cos^{-1} t) \right\}, \quad -1 \leq t \leq 1,$$

with $h : [0,\pi] \longrightarrow \mathbb{R}$ being defined as

$$h(t) = at^2 + bt + c,$$

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

$$\frac{1}{K} = 2 \int_0^{\pi} \exp(at^2 + bt + c)\, dt .$$

As we have argued in the previous remark, in view of our convention about choice of median direction for samples of size $n = 3$ (described in part B of Remark 2.2.4), we have to verify the following :

for every $0 \leq \alpha_1, \alpha_2 \leq \pi$ with $\alpha_1+\alpha_2 \leq \pi$,

$$h(0)+h(\alpha_1)+h(\alpha_2) \geq \begin{cases} h(x)+h(\alpha_1-x)+h(x+\alpha_2), & 0\leq x\leq\alpha_1, \\ h(x)+h(x-\alpha_1)+h(x+\alpha_2), & \alpha_1\leq x\leq\pi-\alpha_2, \\ h(x)+h(x-\alpha_1)+h(2\pi-x-\alpha_2), & \pi-\alpha_2\leq x\leq\pi ; \end{cases}$$

...(2.2.43.1)

for every $0 \leq \alpha_1, \alpha_2 \leq \pi$ with $\alpha_1 + \alpha_2 \geq \pi$, $\alpha_2 \geq \alpha_1$ and $\alpha_1 + 2\alpha_2 \leq \pi$,

$$h(0)+h(\alpha_1)+h(\alpha_2) \geq \begin{cases} h(x)+h(\alpha_1-x)+h(x+\alpha_2), & 0\leq x\leq\pi-\alpha_2, \\ h(x)+h(\alpha_1-x)+h(2\pi-\alpha_2-x), & \pi-\alpha_2\leq x\leq\alpha_1, \\ h(x)+h(x-\alpha_1)+h(2\pi-\alpha_2-x), & \alpha_1\leq x\leq\pi; \end{cases}$$

...(2.2.43.2)

and also

$$h(0)+h(\alpha_1)+h(\alpha_2) \geq \begin{cases} h(x)+h(x+\alpha_1)+h(\alpha_2-x), & 0\leq x\leq\pi-\alpha_1, \\ h(x)+h(2\pi-\alpha_1-x)+h(\alpha_2-x), & \pi-\alpha_1\leq x\leq\alpha_2 \\ h(x)+h(2\pi-\alpha_1-x)+h(x-\alpha_2), & \alpha_2\leq x\leq\pi . \end{cases}$$

...(2.2.43.3)

Denote the functions appearing on the right hand sides of (2.2.43.1)-(2.2.43.3) by $g^{(1)}_{\alpha_1,\alpha_2}(x)$, $g^{(2)}_{\alpha_1,\alpha_2}(x)$ and $g^{(3)}_{\alpha_1,\alpha_2}(x)$, respectively. We have to show that

$$g^{(i)}_{\alpha_1,\alpha_2}(0) \geq g^{(i)}_{\alpha_1,\alpha_2}(x) \quad \forall \ 0 \leq x \leq \pi, \quad \ldots(2.2.44)$$

for every $i = 1,2,3$. Observe that each $g^{(i)}_{\alpha_1,\alpha_2}$ is a continuous piecewise convex polynomial. (see the expression for $g^{(2)}_{\alpha_1,\alpha_2}(\cdot)$ below). Thus in order to establish (2.2.44) it suffices to show that

$$g^{(1)}_{\alpha_1,\alpha_2}(0) \geq g^{(1)}_{\alpha_1,\alpha_2}(\alpha_1),\ g^{(1)}_{\alpha_1,\alpha_2}(\pi-\alpha_2),\ g^{(1)}_{\alpha_1,\alpha_2}(\pi), \ldots(2.2.45.1)$$

$$g^{(2)}_{\alpha_1,\alpha_2}(0) \geq g^{(2)}_{\alpha_1,\alpha_2}(\pi-\alpha_2),\ g^{(2)}_{\alpha_1,\alpha_2}(\alpha_1),\ g^{(2)}_{\alpha_1,\alpha_2}(\pi), \ldots(2.2.45.2)$$

$$g^{(3)}_{\alpha_1,\alpha_2}(0) \geq g^{(3)}_{\alpha_1,\alpha_2}(\pi-\alpha_1),\ g^{(3)}_{\alpha_1,\alpha_2}(\alpha_2),\ g^{(3)}_{\alpha_1,\alpha_2}(\pi); \ldots(2.2.45.3)$$

the successive cases correspond to (2.2.43.1)-(2.2.43.3). We now proceed to establish (2.2.45.1)-(2.2.45.3). We begin with the observation that

$$h \text{ is a strictly decreasing function}, \qquad \ldots(2.2.46)$$

a fact, which is easy to verify. It is now easy to see that (2.2.45.1) is an immediate consequence of this fact. We omit the details. Let us now verify (2.2.45.2). Observe that

$$g^{(2)}_{\alpha_1,\alpha_2}(x) = 3ax^2+(b-2a\alpha_1+2a\alpha_2)x+(a\alpha_1^2+a\alpha_2^2+b\alpha_1+b\alpha_2+3c), \quad 0\leq x\leq \pi-\alpha_2,$$

$$= 3ax^2+(2a\alpha_2-2a\alpha_1-4a\pi-b)x +(a\alpha_1^2+a\alpha_2^2-4a\pi\alpha_2+4a\pi^2+b\alpha_1-b\alpha_2+2\pi b+3c), \quad \pi-\alpha_2\leq x\leq \alpha_1,$$

$$= 3ax^2+(2a\alpha_2-2a\alpha_1-4a\pi+b)x +(a\alpha_1^2+a\alpha_2^2-4a\pi\alpha_2+4a\pi^2-b\alpha_1-b\alpha_2+2\pi b+3c), \quad \alpha_1\leq x\leq \pi .$$

Observe also that because of the restrictions imposed on α_1 and α_2, under which $g^{(2)}_{\alpha_1,\alpha_2}(\cdot)$ is defined, it follows trivially from (2.2.46) that $g^{(2)}_{\alpha_1,\alpha_2}(0) \geq g^{(2)}_{\alpha_1,\alpha_2}(\alpha_1)$ and $g^{(2)}_{\alpha_1,\alpha_2}(\pi-\alpha_2) \geq g^{(2)}_{\alpha_1,\alpha_2}(\pi)$. Thus in order to verify (2.2.45.2) we need only to verify that $g^{(2)}_{\alpha_1,\alpha_2}(0) \geq g^{(2)}_{\alpha_1,\alpha_2}(\pi-\alpha_2)$, and since $g^{(2)}_{\alpha_1,\alpha_2}(x)$ is a quadratic polynomial on $[0,\pi-\alpha_2]$ with strictly positive leading coefficient ($\because a>0$), this happens if and only if

$$\pi - \alpha_2 \leq - \frac{b-2a\alpha_1+2a\alpha_2}{3a},$$

that is,

$$3a\pi + b \leq a(2\alpha_1+\alpha_2) . \qquad \ldots(2.2.47)$$

On the other hand, $\min\{2\alpha_1+\alpha_2: 0\leq\alpha_1, \alpha_2\leq\pi, \alpha_1+\alpha_2\geq \pi, \alpha_2\geq \alpha_1$ and $\alpha_1+2\alpha_2 \leq 2\pi\} = \pi$, which is easy to verify. Along with the facts $2a\pi+b < 0$ and $a > 0$, this implies (2.2.47). This completes the verification of (2.2.45.2). Verification of (2.2.45.3) is similar. So we omit it.

We now explore the question of validity of the assertion of Theorem 2.2.1 when sample size n is greater than 4. It is a trivial observation that when n is a multiple of 4, i.e., $n = 4k$ for some integer $k > 1$, the assertion of Theorem 2.2.1 holds true : to see this we take k copies of the sample $(\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4)$, observe that the median direction of the observations $\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4$ can be taken as the median direction of all the $4k$ observations, and finally proceed as in Theorem 2.2.1. We, therefore, consider the case when $n = 4k+2$ for some integer $k > 1$. First we address the issue of choice of median direction in this situation. Recall from part C of Remark 2.2.4 that while choosing median direction of 4 observations $\underline{x}_1, \ldots, \underline{x}_4$, we not only require them to be in clockwise order but demand moreover that α_i's (for the definition of α_i's, see the paragraph preceding part A of the same remark) satisfy certain restrictions, like (i) $\alpha_4 = \max.\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ and $\alpha_4 > \pi$ or (ii) $\max.\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \leq \pi$, $\alpha_1+\alpha_2 \leq \pi$ and $\alpha_2+\alpha_3 \leq \pi$. It is not difficult to see that with any of these restrictions median direction of the observations $\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4$ can be taken as the median direction of a set of observations consisting of k copies of $\underline{x}_1$ and $\underline{x}_4$, $k+1$

copies of $\underset{\sim}{x}_2$ and $\underset{\sim}{x}_3$. Now proceeding as in Theorem 2.2.1 it is easy to establish the assertion of this theorem for $n = 4k+2$ samples. We, therefore, have the following generalized version of Theorem 2.2.1.

Theorem 2.2.1' Let $\{p(\underset{\sim}{x};\ \underset{\sim}{\theta}) = f(\underset{\sim}{x}'\underset{\sim}{\theta}) \mid \underset{\sim}{\theta} \in S^1\}$ 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 sample median direction be a maximum likelihood estimate of $\underset{\sim}{\theta}$ for all samples of size n, where n is an even number greater than 2, then

$$p(\underset{\sim}{x};\ \underset{\sim}{\theta}) = \frac{\lambda}{2(1-e^{-\lambda\pi})}\, e^{-\lambda \cos^{-1}(\underset{\sim}{x}'\underset{\sim}{\theta})},\ x \in S^1, \quad \ldots(2.2.48)$$

where λ is a positive constant.

Remark 2.2.7 We have not explored the question of validity of the assertion of Theorem 2.2.1 when sample size n is an odd number greater than 3. In consideration of Remark 2.2.6, we conjecture that the answer to the question is in the negative.

Remark 2.2.8 We have seen in Remark 2.2.4 how to get rid of the non-uniqueness of sample median direction by choosing the median direction in a meaningful way. It should be pointed out that for the assertion of Theorem 2.2.1, or more generally Theorem 2.2.1', to hold this choice is crucial. In fact, if we require any median direction, i.e., any point on S^1 satisfying Definition 2.2.1, to be a maximum likelihood estimate of $\underline{\theta}$, then even for samples of size $n = 2$, the assertion of Theorem 2.2.1 holds true, so that Theorem 2.2.1 for samples of size $n = 4$ follows immediately and the counter-example described in Remark 2.2.5 ceases to be one.

In this context it is also worthwhile to mention that the result of Ghosh and Rao (vide Kagan et al. (1973), pp.413-414) depends crucially on their choice of median and if this special care is not taken then their counter-example for samples of size $n = 2$ (vide Ghosh and Rao (1971)) ceases to be one. This fact also serves as a motivation for our choice of median direction in Remark 2.2.4.

Remark 2.2.9 It is worth recording that essentially the same kind of arguments, as employed in obtaining Theorem 2.2.1' from Theorem 2.2.1, may be used to establish validity of the result of Ghosh and Rao (1971) for any even (> 2) sample size, which was proved originally for sample size $n = 4$.

2.3 The spherical case

We shall study the case for S^2 only. For S^p with $p > 2$, the derivations are essentially same.

We begin with the definition of spherical median of a set of observations from S^2 , the unit sphere in $\mathbb{R}^3$.

Definition 2.3.1 Let $\underline{x}_1,\ldots,\underline{x}_n \in S^2$. Then any point $\underline{x}_o \in S^2$ is called a spherical median of $\underline{x}_1,\ldots,\underline{x}_n$ if

$$\sum_{i=1}^{n} \cos^{-1}(\underline{x}_i'\underline{x}_o) = \min_{\underline{\xi}\in S^2} \sum_{i=1}^{n} \cos^{-1}(\underline{x}_i'\underline{\xi}) \; . \quad \ldots(2.3.1)$$

Remark 2.3.1 The Definition 2.3.1 is due to Fisher (1985).

The extension of Theorem 2.2.1, or more generally Theorem 2.2.1', to S^2 poses a special problem since the location of one possible median direction for every sample of size n=4 becomes difficult. However, it will be seen that in order to prove a result for S^2, which is analogous to Theorem 2.2.1, it is enough to consider all possible samples of size $n = 4$ lying on a great circle. As in Remark 2.2.4, we state in the remark below, our conventions about the choice of median direction of samples from S^2, with sample size $n = 2,4$, respectively.

Remark 2.3.2

A. Sample size $n = 2$: Suppose $\underline{x}_1, \underline{x}_2 \in S^2$. Suppose, moreover, that $\underline{x}_1 \neq \underline{x}_2$ and that $\underline{x}_1 \neq -\underline{x}_2$. Denote by $C[\underline{x}_1,\underline{x}_2]$, the great circle passing through $\underline{x}_1$ and $\underline{x}_2$. Observe that $\underline{x}_1 \neq -\underline{x}_2$ implies $\cos^{-1}(\underline{x}_1'\underline{x}_2) < \pi$. Therefore, it makes sense to talk of the smaller (in length) of the two closed arcs connecting $\underline{x}_1$ and $\underline{x}_2$, and taken along $C[\underline{x}_1,\underline{x}_2]$. We denote this arc by $[\underline{x}_1,\underline{x}_2]$. It is now easy to see in view of (2.2.3) and (2.2.4), that

$$\left\{ \underline{\mu} \in C[\underline{x}_1, \underline{x}_2] : \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi} \in C[\underline{x}_1, \underline{x}_2]} \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\xi}) \right\}$$

$$= [\underline{x}_1, \underline{x}_2] . \qquad \ldots(2.3.2)$$

Again, from triangle inequality it follows that for any $\underline{\xi} \in S^2 - C[\underline{x}_1, \underline{x}_2]$ we have

$$\cos^{-1}(\underline{x}_1' \underline{x}_2) < \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\xi}). \qquad \ldots(2.3.3)$$

Thus, from (2.3.2) and (2.3.3), it follows that

$$\left\{ \underline{\mu} \in S^2 : \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi} \in S^2} \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\xi}) \right\}$$

$$= [\underline{x}_1, \underline{x}_2] . \qquad \ldots(2.3.4)$$

On the other hand,

$$\left\{ \underline{\mu} \in S^2 : \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\mu}) = \min_{\underline{\xi} \in S^2} \sum_{i=1}^{2} \cos^{-1}(\underline{x}_i' \underline{\xi}) \right\}$$

$$= \begin{cases} \{\underline{x}_1\} & \text{if } \underline{x}_1 = \underline{x}_2 , \\ S^2 & \text{if } \underline{x}_1 = -\underline{x}_2 . \end{cases} \qquad \ldots(2.3.5)$$

In view of (2.3.4) and (2.3.5), our choice of median direction can now be stated as follows :

median direction = mid-point of $[\underline{x}_1, \underline{x}_2]$ if $\underline{x}_1 \neq \underline{x}_2$ and $\underline{x}_1 \neq -\underline{x}_2$,

= $\underline{x}_1$ if $\underline{x}_1 = \underline{x}_2$, ...(2.3.6)

= any point belonging to the great circle which is perpendicular to the line joining $\underline{x}_1$ and $\underline{x}_2$ if $\underline{x}_1 = -\underline{x}_2$.

B. Sample size n=4 : Suppose $\underline{x}_1,\ldots,\underline{x}_4 \varepsilon S^2$. Suppose, moreover, that $\underline{x}_1,\ldots,\underline{x}_4$ lie on a great circle. Denote the circular median of $\underline{x}_1,\ldots,\underline{x}_4$, described in part C of Remark 2.2.4, by $\underline{x}_o$. Then, employing an argument similar to that used in part A above, it makes sense to choose $\underline{x}_o$ as the spherical median of $\underline{x}_1,\ldots,\underline{x}_4$. On the other hand, if $\underline{x}_1,\ldots,\underline{x}_4$ do not lie on the same great circle, we choose any point $\underline{x}_o \varepsilon S^2$ satisfying (2.3.1) as the spherical median of $\underline{x}_1,\ldots,\underline{x}_4$.

We now state and prove the main result of this section. It should be mentioned that even though it is stated for an arbitrary sample size n, which is even and greater than 2, we shall prove it for n = 4 only. For other values of n, the assertion of the theorem can be established by using essentially the same kind of arguments as employed in obtaining Theorem 2.2.1' from Theorem 2.2.1. It must also be mentioned that in the statement of the theorem the expression 'the median direction', when considered for samples not lying on the same great circle, refers to an arbitrary $\underline{x}_o$ that satisfies (2.3.1). Such a fixation of choice is done only to rule out the possibility of the likelihood function being flat over some sets in which case the situation might lead to triviality (cf. Remark 2.2.8).

<u>Theorem 2.3.1</u> Let $\{ p(\underline{x};\underline{\theta}) = f(\underline{x}'\underline{\theta}) \mid \underline{\theta} \varepsilon S^2 \}$ be a class of rotationally symmetric non-uniform densities on S^2. Suppose $f(t) > 0$ for every $t \varepsilon (-1,1)$ and, moreover, $f(t)$ is right-continuous at $t = -1$. If the sample median direction be a maximum likelihood estimate of $\underline{\theta}$ for all samples of size n, where n is an even number greater than 2, then

$$p(\underline{x};\underline{\theta}) = \frac{\lambda^2 + 1}{2\pi(1+e^{-\lambda\pi})} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}, \quad \underline{x} \varepsilon S^2, \quad \ldots(2.3.7)$$

where λ is a positive constant.

<u>Proof</u>. We provide the proof for $n = 4$ only. It is essentially similar to that of Theorem 2.2.1.

Choose $\underline{x}_i = (x_{i1}, x_{i2}, x_{i3}) \varepsilon S^2$, $i = 1,\ldots,4$, such that $x_{i3} = 0$ for every i. It then follows from part B of Remark 2.3.2 that the median direction $\underline{x}_o = (x_{o1}, x_{o2}, x_{o3})$ of $\underline{x}_1,\ldots,\underline{x}_4$ satisfies $x_{o3} = 0$. Thus if we choose samples from the great circle $C_o = \{(x_1,x_2,x_3) \varepsilon S^2 \mid x_3 = 0\}$, then using this fact and the assumption that the sample median direction is a maximum likelihood estimate of $\underline{\theta}$ we obtain the following :

$$\prod_{i=1}^{4} f(\underline{x}_i'\underline{x}_o) \geq \prod_{i=1}^{4} f(\underline{x}_i'\underline{\theta}) \quad \forall \ \underline{\theta} \varepsilon C_o ,$$

and for all samples $(\underline{x}_1,\ldots,\underline{x}_n)$ of size $n = 4$ with $\underline{x}_i \varepsilon C_o$ for every i, $\underline{x}_o$ being the sample median direction.

As in Theorem 2.2.1, we now define $g(t) = f(\cos t)$, $0 \leq t \leq \pi$, and imitate the proof of that theorem. It is easy to see that the proof can be pushed through without any change in this case. We must, however, record one minor difference between the proof of Theorem 2.2.1 and the present proof. This is done below.

Recall from the proof of Theorem 2.2.1 that in order to establish (2.2.23) we needed the fact that the relation $g(t) = \infty$ cannot hold on a set of positive Lebesgue measure, and there this fact followed trivially. However, in this case we have

$$\int_{S^2} f(\underline{x}'\underline{\theta})\, d\underline{x} = 1 \qquad \forall\ \underline{\theta} \in S^2$$

$$\Rightarrow \quad \int_{S^2} f(\underline{x}'\underline{\theta}_0)\, d\underline{x} = 1, \qquad \text{where } \underline{\theta}_0 = (0,0,1)$$

$$\Rightarrow \quad \int_0^{2\pi}\int_0^{\pi} f(\cos t)\sin t \, ds\, dt = 1, \text{ introducing}$$

polar co-ordinates $\underline{x} = (\cos s \sin t, \sin s \sin t, \cos t)$, $0 \leq s < 2\pi$, $0 \leq t \leq \pi$

$$\Rightarrow \quad 0 < \int_0^{\pi} f(\cos t)\sin t \, dt < \infty$$

$$\Rightarrow \quad 0 < \int_0^{\pi} g(t)\sin t \, dt < \infty .$$

Hence, the relation $g(t) = \infty$ cannot hold on a set of positive Lebesgue measure. As regards the computation of the constant appearing on the right hand side of (2.3.7), it is done in the next remark. This completes the proof of the theorem.

Remark 2.2.3 For S^p with $p \geq 2$, the density obtained in (2.3.7) is as follows :

$$p(\underline{x};\underline{\theta}) = \frac{\Gamma(\frac{p}{2})K_{p-1}(\lambda)}{2\pi^{p/2}} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}, \ \underline{x} \in S^p, \quad \ldots(2.3.8)$$

where λ is a positive constant, and

$$K_m(\lambda) = \frac{\lambda(\lambda^2+2^2)(\lambda^2+4^2)\ldots(\lambda^2+m^2)}{m!(1-e^{-\lambda\pi})} \quad \text{if } m \text{ is even,}$$

$$\ldots(2.3.9)$$

$$= \frac{(\lambda^2+1^2)(\lambda^2+3^2)\ldots(\lambda^2+m^2)}{m!(1+e^{-\lambda\pi})} \quad \text{if } m \text{ is odd.}$$

We need to verify that the expression appearing on the right hand side of (2.3.8) is indeed a density function on S^p. In other words, we need to verify that

$$\int_{S^p} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})} d\underline{x} = \frac{2\pi^{p/2}}{\Gamma(\frac{p}{2})K_{p-1}(\lambda)} \quad \ldots(2.3.10)$$

To establish (2.3.10), fix $\underline{\theta} \in S^p$ and choose a $(p+1) \times (P+1)$ orthogonal matrix A such that $A\underline{\theta} = \underline{\theta}_0$, where $\underline{\theta}_0 = (0,0,\ldots,0,1)$. Write $\underline{y} = A\underline{x}$ and observe that

$$\int_{S^p} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}\, d\underline{x} = \int_{S^p} e^{-\lambda \cos^{-1}(\underline{y}'\underline{\theta}_o)}\, d\underline{y}. \quad \ldots(2.3.11)$$

Introduce now on the right hand side of (2.3.11) spherical co-ordinates $\phi_1,\ldots,\phi_p$ defined as follows :

$$y_1 = \sin\phi_1 \sin\phi_2 \ldots \sin\phi_{p-2} \sin\phi_{p-1} \cos\phi_p,$$

$$y_2 = \sin\phi_1 \sin\phi_2 \ldots \sin\phi_{p-2} \sin\phi_{p-1} \sin\phi_p,$$

$$y_3 = \sin\phi_1 \sin\phi_2 \ldots \sin\phi_{p-2} \cos\phi_{p-1},$$

$$\vdots$$

$$y_{p+1} = \cos\phi_1,$$

where $0 \le \phi_i < \pi$ for $i = 1,\ldots,p-1$ and $0 \le \phi_p < 2\pi$, and $\underline{y} = (y_1,\ldots,y_{p+1})$. This implies

$$\int_{S^p} e^{-\lambda \cos^{-1}(\underline{y}'\underline{\theta}_o)}\, d\underline{y}$$

$$= \int_0^\pi \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} e^{-\lambda\phi_1} \sin^{p-1}\phi_1 \sin^{p-2}\phi_2$$

$$\ldots \sin\phi_{p-1}\, d\phi_1 d\phi_2 \ldots d\phi_{p-1} d\phi_p$$

$$= 2\pi\left(\int_0^\pi e^{-\lambda\phi_1} \sin^{p-1}\phi_1 d\phi_1\right).\left(\int_0^\pi \sin^{p-2}\phi_2 d\phi_2\right)\ldots\left(\int_0^\pi \sin\phi_{p-1} d\phi_{p-1}\right)$$

if $p > 2$, and

$$\int_{S^p} e^{-\lambda \cos^{-1}(\underline{y}'\underline{\theta}_o)}\, d\underline{y} = \int_0^\pi \int_0^{2\pi} e^{-\lambda\phi_1} \sin\phi_1 d\phi_1 d\phi_2$$

if $p = 2$. It is now easy to see that (2.3.10) will follow if we can verify that

$$\int_0^{\pi} \sin^m x \; dx = \frac{\sqrt{\pi} \cdot \Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}, \quad m = 1,2,\ldots, \quad \ldots(2.3.12.1)$$

and

$$\int_0^{\pi} e^{-\lambda x} \sin^m x \; dx = K_m(\lambda), \quad m = 1,2,\ldots \; . \quad \ldots(2.3.12.2)$$

Observe that (2.3.12.1) is an immediate consequence of properties of Beta function and Gamma function, and in order to verify (2.3.12.2) we use the relation

$$\int e^{-\lambda x} \sin^m x \; dx = \frac{e^{-\lambda x} \sin^{m-1} x(\lambda \sin x - m \cos x)}{\lambda^2 + m^2}$$

$$+ \frac{m(m-1)}{\lambda^2 + m^2} \int e^{-\lambda x} \sin^{m-2} x \; dx,$$

which is easy to verify. We omit the details.

Remark 2.3.4 Theorem 2.3.1 is false if we require the sample median direction to be a maximum likelihood estimate of $\underline{\theta}$ for all samples of size $n = 2$. To see this, consider the following class $\mathcal{F}$ of rotationally symmetric non-uniform densities on S^2 :

$$\mathcal{F} = \left\{ p(\underline{x};\underline{\theta}) = f(\underline{x}'\underline{\theta}) \,\middle|\, \underline{\theta} \in S^2 \right\},$$

where

$$f(t) = K \exp\left\{h(\cos^{-1} t)\right\}, \quad -1 \leq t \leq 1 ,$$

with the same h as in Remark 2.2.5 and

$$\frac{1}{K} = 2\pi\left\{\frac{e^{\mu}(1-\lambda e^{-\lambda\pi/2})}{\lambda^2+1} + \frac{e^{\nu}(\tau e^{-\tau\pi/2} + e^{-\tau\pi})}{\tau^2+1}\right\}.$$

Obviously f is not of the form as described in (2.3.7). We are, therefore, required to verify that $\mathcal{F}$ indeed serves as a counter-example to the assertion of Theorem 2.3.1 with $n = 2$ samples, that is, we have to show that

$$\prod_{i=1}^{2} f(\underline{x}_i'\underline{x}_o) \geq \prod_{i=1}^{2} f(\underline{x}_i'\underline{\theta}) \quad \forall\ \underline{\theta} \in S^2 , \qquad \ldots(2.3.13)$$

and for all samples $(\underline{x}_1, \underline{x}_2)$ of size $n = 2$; $\underline{x}_o$ being the sample median direction.

Observe that if $\underline{x}_1 = \underline{x}_2$, then $\underline{x}_o = \underline{x}_1$, and if $\underline{x}_1 = -\underline{x}_2$ thus $\underline{x}_o$ satisfies $\underline{x}_1'\underline{x}_o = \underline{x}_2'\underline{x}_o = 0$ (vide part A of Remark 2.3.2). Thus (2.3.13) follows immediately from (2.2.41) if $\underline{x}_1 = \underline{x}_2$ or if $\underline{x}_1 = -\underline{x}_2$. We, therefore, establish (2.3.13) when neither $\underline{x}_1 = \underline{x}_2$ nor $\underline{x}_1 = -\underline{x}_2$. It is easy to see in this case that $\cos^{-1}(\underline{x}_1'\underline{x}_2) < \pi$. Then in view of our choice of median direction in this case, described in part A of Remark 2.3.2, and also in view of Remark 2.2.5, it follows that

$$\prod_{i=1}^{2} f(\underline{x}_i'\underline{x}_o) \geq \prod_{i=1}^{2} f(\underline{x}_i'\underline{\theta}) \quad \forall \ \underline{\theta} \varepsilon C[\underline{x}_1,\underline{x}_2] .$$

Thus, (2.3.13) will follow if we show that given $\underline{\theta} \varepsilon S^2 - C[\underline{x}_1,\underline{x}_2]$, $\exists \ \underline{\theta}^* \varepsilon [\underline{x}_1,\underline{x}_2]$ such that

$$\prod_{i=1}^{2} f(\underline{x}_i'\underline{\theta}^*) \geq \prod_{i=1}^{2} f(\underline{x}_i'\underline{\theta}) . \qquad \ldots(2.3.14)$$

Observe that from $\underline{x}_1 \neq \underline{x}_2$, $\underline{x}_1 \neq \underline{x}_2$ and $\underline{\theta} \varepsilon S^2-C[\underline{x}_1,\underline{x}_2]$, it is easy to see, using triangle inequality and little linear algebra, that

$$\cos^{-1}(\underline{x}_1'\underline{x}_2) < \cos^{-1}(\underline{x}_1'\underline{\theta}) + \cos^{-1}(\underline{x}_2'\underline{\theta}) .$$

On the arc $[\underline{x}_1,\underline{x}_2]$, choose now a point $\underline{\theta}^*$ such that

$$\cos^{-1}(\underline{x}_1'\underline{\theta}^*) = \frac{\cos^{-1}(\underline{x}_1'\underline{x}_2)\ \cos^{-1}(\underline{x}_1'\underline{\theta})}{\cos^{-1}(\underline{x}_1'\underline{\theta}) + \cos^{-1}(\underline{x}_2'\underline{\theta})}$$

and

$$\cos^{-1}(\underline{x}_2'\underline{\theta}^*) = \frac{\cos^{-1}(\underline{x}_1'\underline{x}_2)\ \cos^{-1}(\underline{x}_2'\underline{\theta})}{\cos^{-1}(\underline{x}_1'\underline{\theta})+\cos^{-1}(\underline{x}_2'\underline{\theta})} ;$$

this choice is possible since $\cos^{-1}(\underline{x}_1'\underline{\theta}^*) + \cos^{-1}(\underline{x}_2'\underline{\theta}^*) = \cos^{-1}(\underline{x}_1'\underline{x}_2)$. Observe also that $\cos^{-1}(\underline{x}_1'\underline{\theta}^*) < \cos^{-1}(\underline{x}_2'\underline{\theta})$ and $\cos^{-1}(\underline{x}_2'\underline{\theta}^*) < \cos^{-1}(\underline{x}_2'\underline{\theta})$. The inequality in (2.3.14) is now an immediate consequence of the fact that h is decreasing.

Remark 2.3.5 We have not discussed in Remark 2.3.2 about location of spherical median for samples of size $n = 3$. Therefore, the question of validity of the assertion of Theorem 2.3.1 for samples of size $n = 3$, or more generally $n = 2k+1$ with $k \geq 1$, remains open.

Remark 2.3.6 It is easy to see that the converse of Theorem 2.2.1' or Theorem 2.3.1 is also true, that is, if $\underline{X}$ has the density (2.2.18) or (2.3.7), then for independent and identically distributed observations $\underline{X}_1, \ldots, \underline{X}_n$ from $p(\underline{x};\underline{\theta})$, the sample median direction of $\underline{X}_1, \ldots, \underline{X}_n$ is a maximum likelihood estimate of $\underline{\theta}$.

Remark 2.3.7 We have assumed both in Theorem 2.2.1' and Theorem 2.3.1 that $f(t) > 0$ for every $t \in (-1,1)$. However, none of the results mentioned in Section 2.1 puts any such restriction on the density to be characterized. It is, therefore, of some interest to see if this assumption can be relaxed.

2.4 Some general remarks

Remark 2.4.1 The purpose of this remark is to provide a perspective to the problem considered in this chapter in which the solution obtained is the natural one.

Observe first that in each of the three results mentioned in Section 2.1, the sample space under consideration, which we denote by $\mathcal{X}$, is of the form $\mathcal{X} = G/H$, where G is a unimodular Lie group and H is a compact subgroup of G. This ensures the existence of a unique (upto multiplication by constant) G-invariant measure on $\mathcal{X}$, which is needed since we are dealing with a translation parameter problem. The translation parameter is also called location parameter. The parameter space is same as $\mathcal{X}$. Finally, the specific form of the maximum likelihood estimate of the location parameter can be thought of as that x_0 which minimizes

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

over $x \varepsilon \mathcal{X}$, where $x_i \varepsilon \mathcal{X}$, $i = 1,\ldots,n$, and d is (a power of) a G-invariant metric on $\mathcal{X}$. Thus, for example, in Teicher (1961), $G = \mathbb{R}$, $H = \{0\}$, $\mathcal{X} = \mathbb{R}$, $d(x,y) = (x-y)^2$, in Ghosh and Rao (1971), $G = \mathbb{R}$, $H = \{0\}$, $\mathcal{X} = \mathbb{R}$, $d(x,y) = |x-y|$, in Bingham and Mardia (1975), $G = O(p+1)$ = the set of all $(p+1) \times (p+1)$ orthogonal matrices, $H = O(p)$, $\mathcal{X} = S^p$, $d(\underset{\sim}{x},\underset{\sim}{y}) = ||\underset{\sim}{x}-\underset{\sim}{y}||_2^2$ = square of the ℓ_2-norm of $\underset{\sim}{x} - \underset{\sim}{y}$. The density characterizes then turns out to be of the form

$$A\, e^{-\lambda d(x,\theta)}, \quad x \varepsilon \mathcal{X},$$

where $\lambda > 0$ and A is a positive constant depending on λ [The von Mises-Fisher density $A\, e^{k(\underset{\sim}{x}'\underset{\sim}{\theta})}$ is easily seen to have the alternative representation $A\, e^{-\lambda ||\underset{\sim}{x} - \underset{\sim}{\theta}||^2}$, since $||\underset{\sim}{x}|| = ||\underset{\sim}{\theta}|| = 1$]. 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 chapter, we have $G = O(p+1)$, $H = O(p)$, i.e., $\mathcal{X} = S^p$ and $d(\underset{\sim}{x},\underset{\sim}{y}) = \cos^{-1}(\underset{\sim}{x}'\underset{\sim}{y})$ = the geodesic distance between $\underset{\sim}{x}$ and $\underset{\sim}{y}$. In view of the observations mentioned in the last paragraph it is, therefore, expected that the density characterized should have the form as in (2.2.18) and (2.3.7).

Remark 2.4.2 The previous remark may be used to generate densities on abstract spaces. In what follows, we give an example to illustrate this.

Suppose $G = O(p)$, $H = O(p-n)(n<p)$ or $\{I_p\}$, and $\mathcal{X} = G/H$. It then turns out that $\mathcal{X}$ is same as $O(n,p)(n\leq p)$, the set of all $n\times p$ matrices M satisfying $MM^+ = I_n$. Choose in place of d, the Riemannian metric on $O(n,p)$. It is now possible to think of a location of central tendency for observations $X_1,\ldots,X_N$ from $O(n,p)$, introduced as follows : minimize

$$\sum_{i=1}^{N} d(X_i,X)$$

over $X \in O(n,p)$, and call such an X_o where the minimum is attained a median orientation of $X_1,\ldots,X_N$.

Introduce now the following density on $O(n,p)$:

$$p(X;A) = K \exp\left\{ -\lambda\, d(X,A) \right\} , \quad X \varepsilon O(n,p), \; A \varepsilon O(n,p),$$

$\lambda > 0$, and K is a positive constant depending on λ . It is easy to see that for independent and identically distributed observations $X_1, \ldots, X_N$ from this density, their sample median orientation is a maximum likelihood estimate of A. We leave it as an open problem whether this fact characterizes the density p .

Remark 2.4.3 In connection with Remark 2.4.1 we would like also to add that it may be possible to prove a general theorem which encompasses the results mentioned in Section 2.1 and the one obtained in this chapter.

CHAPTER 3

MAXIMUM LIKELIHOOD CHARACTERIZATION OF THE VON MISES - FISHER MATRIX DISTRIBUTION

3.1 Introduction

In Chapter 2, we obtained a distribution over S^p through maximum likelihood characterization. In this chapter, we settle yet another problem which is similar in spirit.

Downs (1972), with a view to developing methods for summarizing and comparing the orientations of samples of orientable objects with each orientation being described by n distinguishable directions in p dimentions ($n \leq p$), extended the von Mises-Fisher distribution on S^p to Stiefel manifold . The resulting distribution came to be known as von Mises-Fisher matrix distribution. Khatri and Mardia (1977) investigated this distribution from distribution theoretic and inferential points of view. Issues related to inferential aspect of this distribution were explored also by Jupp and Mardia (1979).

In view of the result of Bingham and Mardia (1975) concerning maximum likelihood characterization of the von Mises - Fisher distribution on sphere, which was mentioned in Section 2.1, our aim in this chapter is to explore similar question in connection with the von Mises-Fisher matrix distribution. In Section 3.2, the problem is formulated. In Section 3.3, the main result is proved.

This chapter is a revised version of Purkayastha and Mukerjee (1990).

3.2 Formulation of the problem and some related issues

We begin this section with a brief discussion on polar decomposition of a matrix with full row rank. Discussion on polar decomposition of square matrices may be found in Halmos (1958, pp.169-170).

Recall from Remark 2.4.2 that the set of all $n \times p (n \leq p)$ matrices M satisfying $MM^t = I_n$ is denoted by $O(n,p)$. It is called the Stiefel manifold. For $n = p$, it coincides with the orthogonal group $O(n)$.

The following theorem records the facts about polar decomposition that we need for our purpose. A proof of this result may be found in Downs (1972).

Theorem 3.2.1 Let A be an $n \times p$ $(n \leq p)$ matrix with full row rank. Then we have the following :

(i) there is a unique decomposition

$$A = PU \qquad ...(3.2.1)$$

where P is an $n \times n$ positive definite matrix and $U \in O(n,p)$. In fact, $P = (AA^t)^{1/2}$, the square root of the positive definite matrix AA^t and $U = (AA^t)^{-1/2} A$. The representation of A in (3.2.1), in terms P and U, is called the polar decomposition of A. P is known as the elliptical component, and U the polar component, of A.

$$\text{(ii)} \quad tr((B-A)(B-A)^t) \geq tr((U-A)(U-A)^t) \quad \forall\ B \in D(n,p), \qquad \ldots(3.2.2)$$

with equality if and only if $B = U$.

Remark 3.2.1 In order to prove Theorem 3.2.1 (ii) above, Downs (1972) used the method of Lagrangian multiplier. It is, however, worthwhile to mention that an alternative proof may be given using the polar decomposition of A given in (3.2.1) and the spectral decomposition of P.

Remark 3.2.2 It is known that on $\mathcal{M}_{nxp}$ (without any restriction on n and p), the set of all nxp matrices, the following defines an inner product :

$$\langle C,D\rangle = tr(CD^t) , \quad C, D \in \mathcal{M}_{nxp} .$$

This induces the norm $||C|| = \{tr(CC^t)\}^{1/2}$ on $\mathcal{M}_{nxp}$. Thus, as observed in Downs (1972), Theorem 3.2.1 (ii) indeed brings out the relationship between polar decomposition and least squares. In what follows, we elucidate this issue.

Let $X_i \in O(n,p)$, $i = 1,\ldots,N$, and $X = \sum_{i=1}^{N} X_i$ be of full row rank. Consider now the problem of minimizing

$$\sum_{i=1}^{N} tr((B-X_i)(B-X_i)^t)$$

over $B \in O(n,p)$. Observe that this is equivalent to the problem of maximizing

$$\sum_{i=1}^{N} tr(BX_i^t) = tr\ BX^t$$

over $B \in O(n,p)$, since $BB^t = X_1X_1^t = \ldots = X_NX_N^t = I_n$. Again, this maximization problem is equivalent to one of minimizing

$$tr((B-X)(B-X)^t)$$

over $B \in O(n,p)$, since X is held fixed. In view of Theorem 3.2.1, this minimization problem admits a solution since X is of full row rank, the minimum being attained *only at* $B = U(X)$, *where*

$$X = P(X) \circ U(X)$$

is the polar decomposition of X. Observe, moreover, that $U(X) = (XX^t)^{-1/2} X$. Thus, if $X_1,\ldots,X_N$ describe orientations of a sample of orientable objects, it is reasonable to call $U(X)$ their mean orientation.

Consider now the special case when $n = 1$ and $p > 1$, i.e., when $O(n,p) = S^{p-1}$. It is easy to see in this case that $\sum_{i=1}^{N} X_i$ is of full row rank if and only if

$\sum_{i=1}^{N} X_i \neq \underset{\sim}{0}$, the zero vector, and in such a situation $U(X) = \frac{X}{||X||}$, the mean direction. Thus, mean orientation extends the notion of mean direction in a meaningful way.

Remark 3.2.3 We assumed in Theorem 3.2.1 that the matrix A is of full row rank. However, it should be mentioned that even if A is not of full row rank, it is possible to represent A as A = PU, where P is an nxn non-negative definite matrix, U ε O(n,p), and

$$tr((B-A)(B-A)^t) \geq tr((U-A)(U-A)^t) \quad \forall \; B \in O(n,p).$$

Let us now consider the following class of densities on O(n,p), the densities being considered with respect to the uniform distribution on O(n,p) (see Mardia and Khatri (1977) for the definition of this) :

$$\mathcal{F} = \left\{ p(X;A) = f\,[\,tr(AX^t)\,] \,\middle|\, A \in O(n,p) \right\}. \quad \ldots(3.2.3)$$

This class has the following property : p(X;A) = p(XB;A) for all pxp orthogonal matrix B with det (B) = 1 that satisfies AB = A. Because of this geometric consideration we call A the location parameter for the class of distributions induced by the family $\mathcal{F}$ (see the discussion preceding statement of Theorem 2.2.1). Moreover, $\mathcal{F}$ is the proper extension of a class of rotationally symmetric directional distribution to Stiefel manifold. In (3.2.3), we now take

$$f[tr(\Lambda X^t)] = K \exp\left\{ \lambda \, tr(\Lambda X^t) \right\}, \; X \in O(n,p), \qquad ...(3.2.4)$$

where λ is a positive constant and K is another positive constant depending only on λ. This a special case of the von Mises-Fisher matrix distribution given in Khatri and Mardia (1977) : there the parameter matrix is assumed to be an arbitrary $n \times p$ $(n \leq p)$ matrix, whereas here it is taken be a (positive) multiple of a matrix lying in $O(n,p)$. Then from what we have discussed in Remark 3.2.2, it follows trivially that for a random sample $X_1,\ldots,X_N$ from the von Mises-Fisher matrix distribution in (3.2.4), the maximum likelihood estimate of Λ is $U(X)$ provided X is of full row rank. In view of the maximum likelihood characterization of the von Mises-Fisher **distribution on** sphere, obtained by Bingham and Mardia (1975), we may now ask whether this fact characterizes the von Mises-Fisher matrix distribution.

We conclude this section by arguing in the light of Remark 2.4.1, why the answer to the question posed in the last paragraph is expected to be in the affirmative. Choose in Remark 2.4.1, $G = O(p)$, $H = O(p-n)(n < p)$ or $\{I_p\}$, $\mathcal{X} = G/H = O(n,p)$ (cf. Remark 2.4.2), and $d(X,Y) = tr[(X-Y)(X-Y)^t]$, $X, Y \in O(n,p)$. Then from what we have discussed earlier in this section and from Remark 2.4.1, it is expected that the density to be characterized should be of the form

$$K \exp\left\{ -\lambda \operatorname{tr}((\Lambda-X)(\Lambda-X)^t)\right\}, \quad X \in O(n,p),$$

for some constants λ and K, both positive. Observe that this density has the alternative expression as on the right hand side of (3.2.4), as $\Lambda\Lambda^t = XX^t = I_n$.

3.3 The main result

Theorem 3.3.1 Let $\left\{ p(X;\Lambda) = f[\operatorname{tr}(\Lambda X^t)] \,|\, \Lambda \in O(n,p)\right\}$ be a class of non-uniform densities on $O(n,p)$. Assume that f is lower semi-continuous at the point n. Furthermore, suppose that for every positive integral N and for all random samples $X_1,\ldots,X_N$, with $X = \sum_{i=1}^{N} X_i$ of full row rank, the polar component of X is a maximum likelihood estimate of Λ. Then

$$p(X;\Lambda) = K \exp\left\{ \lambda \operatorname{tr}(\Lambda X^t)\right\}, \quad X \in O(n,p),$$

where λ is a positive constant, and K is another positive constant depending only on λ.

Proof. For $n = 1$ and $p > 1$ (since with $n = p = 1$, the problem does not make any sense), our theorem follows from Theorem 1 and Theorem 2 of Bingham and Mardia (1975). Throughout the proof, we therefore consider the case for $n \geq 2$.

Observe that the condition regarding the maximum likelihood estimate is equivalent to the following : for every positive integral N and every choice of matrices $X_1, \ldots, X_N$, $A \varepsilon O(n,p)$ with $X = \sum_{i=1}^{N} X_i$ of full row rank, the relation

$$\prod_{i=1}^{N} f[tr(\hat{A}X_i^t) \geq \prod_{i=1}^{N} f[tr(AX_i^t)] \qquad \ldots(3.3.1)$$

holds, where $\hat{A} = (XX^t)^{-1/2} X$, the polar component of X.

Using (3.3.1), we now establish the following : for every positive integral N and every choice of matrices $C_1, \ldots, C_N$, $U \varepsilon O(n,n)$ with $C = \sum_{i=1}^{N} C_i$ positive definite, the relation

$$\prod_{i=1}^{N} f[tr(C_i)] \geq \prod_{i=1}^{N} f[tr(UC_i)] \qquad \ldots(3.3.2)$$

holds. To see how (3.3.2) follows from (3.3.1), choose $L = (I_n \vdots O) \varepsilon O(n,p)$, and define $X_i = C_i^t L$, $1 \leq i \leq N$. Then $X_i \varepsilon O(n,p)$, $1 \leq i \leq N$. Also, $X = \sum_{i=1}^{N} X_i = C^t L$ $= CL$, as C is positive definite and hence symmetric. Moreover, C is positive definite implies that CL is of full row rank, i.e., X is of full row rank, and $\hat{A} = (XX^t)^{-1/2} X$ $= (CLL^tC^t)^{-1/2} CL = (C^2)^{-1/2} CL = C^{-1}CL = L$. Then $\hat{A}X_i^t$ $= LL^tC_i = C_i$, $1 \leq i \leq N$. Observe now that for every $U \varepsilon O(n,n)$, $A = (U \vdots O) \varepsilon O(n,p)$ and that $AX_i^t = (U \vdots O)L^tC_i$ $= UC_i$, $1 \leq i \leq N$. The relation (3.3.2) is now an immediate consequence of (3.3.1).

The next step consists in proving the following lemma, which will be needed later.

Lemma 3.3.1 For every $x \in [-n,n]$, $\exists$ $U \in O(n,n)$ such that $tr(U) = x$.

Proof (of Lemma 3.3.1) For every real α, we define the the 2x2 matrix

$$H_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}.$$

First suppose n is even, say $n = 2m$ $(m \geq 1)$. Define

$$Q_{m\alpha} = I_m \otimes H_\alpha = \text{the Kronecker product of } I_m \text{ and } H_\alpha .$$

Then

$$\begin{aligned} Q_{m\alpha} Q_{m\alpha}^t &= (I_m \otimes H_\alpha)(I_m \otimes H_\alpha)^t \\ &= (I_m \otimes H_\alpha)(I_m \otimes H_\alpha^t) \\ &= I_m \otimes (H_\alpha H_\alpha^t) \\ &= I_m \otimes I_2 = I_{2m} = I_n , \end{aligned}$$

so that $Q_{m\alpha} \in O(n,n)$. Again,

$$tr(Q_{m\alpha}) = 2m \cos\alpha ,$$

and for any given $x \in [-n,n]$, $tr(Q_{m\alpha}) = x$ provided $\cos\alpha = \frac{x}{2m} = \frac{x}{n}$, a choice which is always possible .

Next suppose n is odd, say $n = 2m+1$ $(m \geq 1)$. Define

$$Q^*_{m\alpha}(u) = \begin{pmatrix} Q_{m\alpha} & \underline{0} \\ \underline{0}^t & u \end{pmatrix}$$

where u is a real number. It is easy to verify that $Q^*_{m\alpha}(u) \varepsilon O(n,n)$ for $u = -1, +1$, and that $tr(Q^*_{m\alpha}(u)) = 2m \cos \alpha + u$. Thus for any given $x \varepsilon [-(2m-1), 2m+1]$, $tr(Q^*_{m\alpha}(1)) = x$ provided $\cos \alpha = \frac{x-1}{2m}$, a choice which is always possible. Similarly, for any given $x \varepsilon [-(2m+1), 2m-1]$ $tr(Q^*_{m\alpha}(-1)) = x$ provided $\cos \alpha = \frac{x+1}{2m}$, a choice which is always possible.

This completes the proof of Lemma 3.3.1.

We shall now continue with the proof of our assertion. The proof being somewhat long, is broken into a number of steps.

First we prove that, for each $x \varepsilon [-n,n]$,

$$f(n) \geq f(x) . \qquad \ldots(3.3.3)$$

To see this, choose and fix $x \varepsilon [-n,n]$. Employ Lemma 3.3.1 to obtain $U_o \varepsilon O(n,n)$ such that $tr(U_o) = x$. Now in (3.3.2), take $N = 1$, $C_1 = I_n$ and $U = U_o$. This establishes (3.3.3).

Next we prove that, for each $x \varepsilon [-n,n]$

$$f(x) < \infty . \qquad \ldots(3.3.4)$$

In view of (3.3.3), it is enough to show that

$$f(n) < \infty . \qquad ...(3.3.5)$$

Choose $N = 2$, $U = C_1^t$ in (3.3.2). This gives us $f[tr(C_1)]f[tr(C_2)] \geq f(n)f[tr(C_1^tC_2)]$, for every $C_1, C_2 \varepsilon O(n,n)$ such that C_1+C_2 is positive definite. Hence if (3.3.5) does not hold then $f(n) = \infty$, and for every $C_1, C_2 \varepsilon O(n,n)$ such that C_1+C_2 is positive definite, one must have either (a) $f[tr(C_1^tC_2)] = 0$, or (b) $f[tr(C_1)]f[tr(C_2)] = \infty$. We now consider the cases corresponding to odd n and even n separately. First suppose that $n = 2m+1$ $(m \geq 1)$ and that (3.3.5) does not hold. Define $C_1 = Q^*_{m\alpha}(1)$, $C_2 = Q^*_{m(-\alpha)}(1)$, where $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$. It is easy to verify that $C_1, C_2 \varepsilon O(n,n)$ and $C_1+C_2 = \begin{pmatrix} 2\cos\alpha\, I_{2m} & \underline{0} \\ 0^t & 2 \end{pmatrix}$ is positive definite. Also, $C_1^tC_2 = Q^*_{m(-2\alpha)}(1)$ so that $tr(C_1^tC_2) = 2m\cos 2\alpha+1$. Therefore, it follows from what we obtained earlier that for $\alpha \varepsilon (-\frac{\pi}{2}, \frac{\pi}{2})$, 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 $\alpha \varepsilon (-\frac{\pi}{2}, \frac{\pi}{2})$, i.e., $f(x) = 0$ a.e. over $x \varepsilon (-(2m-1), 2m+1)$ and a contradiction is reached in consideration of lower semi-continuity of f at the point $n = 2m+1$ (cf. (3.3.8) below). Similarly, for even n $(= 2m, m \geq 1)$, we take $C_1 = Q_{m\alpha}$, $C_2 = Q_{m(-\alpha)}$, $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$, and proceed as in the case for odd n. This completes the proof of (3.3.5), and hence of (3.3.4).

In the next step we prove the following lemma . The statement of the lemma may be found in Bingham and Mardia (1975, p.389) also, where a proof does not appear. Even though it is easy, we provide it for the sake of completeness .

Lemma 3.3.2 For a given $\theta \varepsilon [0,\pi]$, there exists η satisfying

$$\text{(i) } -\tfrac{1}{2}\theta \leq \eta \leq 0 \text{ , (ii) } \cos\theta + 2\cos\eta > 0 \text{ , and (iii) } \sin\theta + 2\sin\eta = 0 \text{ .} \qquad \ldots(3.3.6)$$

Proof (of Lemma 3.3.2) Note that

$$\theta \varepsilon [0,\pi] \Rightarrow 0 \leq \frac{\sin\theta}{2} \leq \frac{1}{2} .$$

Hence there exists a unique α with $0 \leq \alpha \leq \frac{\pi}{6}$ such that $\sin\alpha = \frac{1}{2}\sin\theta$. Choose $\eta = -\alpha$. We prove that with this η, our purpose is served.

First observe that

$$\sin\theta + 2\sin\eta = \sin\theta - 2\sin\alpha = \sin\theta - \sin\theta = 0 ,$$

so that (iii) holds.

Next observe that

$$\cos\theta + 2\cos\eta = \cos\theta + 2\cos\alpha = \cos\theta + 2\sqrt{1-\sin^2\alpha}$$

(we take the positive square root as $0 \leq \alpha \leq \frac{\pi}{6}$)

$$= \cos\theta + 2\sqrt{1 - \frac{\sin^2\theta}{4}}$$

$$= \cos\theta + \sqrt{4 - \sin^2\theta}$$

$$\geq -1 + \sqrt{4 - \sin^2\theta}$$

$$= -1 + \sqrt{3 + \cos^2\theta}$$

$$\geq -1 + \sqrt{3} > 0 ,$$

so that (ii) holds.

Finally observe that

$$0 \leq \theta < \pi \Rightarrow 0 \leq \frac{\theta}{2} \leq \frac{\pi}{2}$$

$$\Rightarrow 0 \leq \sin\frac{\theta}{2} , \cos\frac{\theta}{2} \leq 1$$

$$\Rightarrow \sin\frac{\theta}{2}\cos\frac{\theta}{2} \leq \sin\frac{\theta}{2}$$

$$\Rightarrow \frac{\sin\theta}{2} \leq \sin\frac{\theta}{2}$$

$$\Rightarrow \sin\alpha \leq \sin\frac{\theta}{2}$$

$$\Rightarrow \alpha \leq \frac{\theta}{2}$$

$$\left(\text{since } 0 \leq \alpha \leq \frac{\pi}{6} \text{ and } 0 \leq \frac{\theta}{2} \leq \frac{\pi}{2}\right)$$

$$\Rightarrow \eta = -\alpha \geq -\frac{\theta}{2} .$$

On the other hand $\alpha \geq 0 \Rightarrow \eta = -\alpha \leq 0$. Thus (i) holds, and this completes the proof of Lemma 3.3.2.

Now we prove that, for each $x \in [-n,n]$,

$$f(x) > 0 . \qquad \ldots(3.3.7)$$

First note that

$$f(n) > 0 , \qquad \ldots (3.3.8)$$

for otherwise by (3.3.3), $f(x) = 0$ for each $x \varepsilon [-n,n]$, which is impossible as f is a density. In the rest of the proof (of (3.3.7)), we consider the cases corresponding to odd n and even n separately.

<u>Case 1</u> Suppose n is odd, say $n = 2m+1$ ($m \geq 1$). Define

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

If $\mathcal{B}$ is non-empty, then for each $\theta \varepsilon \mathcal{B}$, one can use Lemma 3.3.2 to obtain η satisfying

$$\text{(i)} -\frac{\theta}{2} \leq \eta \leq 0 , \quad \text{(ii)} \cos \theta + 2 \cos \eta > 0 ,$$

$$\text{and (iii)} \sin \theta + 2 \sin \eta = 0 .$$

Define now $C_1 = Q^*_{m\theta}(1)$, $C_2 = C_3 = Q^*_{m\eta}(1)$, $U = Q^*_{m\beta}(1)$, where $\beta = -\frac{1}{2}(\theta + \eta)$. It is easy to verify that C_1, C_2, C_3, $U \varepsilon O(n,n)$ and that $C_1+C_2+C_3$ is positive definite. Moreover, $UC_1 = Q^*_{m(\beta+\theta)}(1)$, $UC_2 = UC_3 = Q^*_{m(\beta+\eta)}(1)$. Thus it follows in view of (3.3.2) and (3.3.4) that

$$f[1 + 2m \cos (\frac{\theta - \eta}{2})] = 0 . \qquad \ldots (3.3.9)$$

Let $b = \inf \mathcal{B}$. Then $b \varepsilon [0,\pi]$. First suppose that $b > 0$ and choose a sequence $\{ \theta_i \}$ such that $\theta_i \varepsilon \mathcal{B}$ $\forall$ i, and $\lim_{i \to \infty} \theta_i = b$. For each i, $\exists \, \eta_i$ with

$-\frac{\theta_i}{2} \leq \eta_i \leq 0$, $\cos\theta_i + 2\cos\eta_i > 0$ and $\sin\theta_i + 2\sin\eta_i = 0$. Then, by (3.3.9) it follows that

$$f[1 + 2m\cos(\frac{\theta_i - \eta_i}{2})] = 0 ,$$

for every i. However,

$$0 \leq \frac{1}{2}(\theta_i - \eta_i) \leq \frac{3}{4}\theta_i \longrightarrow \frac{3}{4} b \quad \text{as} \quad i \longrightarrow \infty .$$

Hence $b = \inf \mathcal{B} \leq \frac{3}{4} b$, which is impossible if $b > 0$. Hence we must have $b = 0$. However, in this case we obtain a sequence $\{\theta_i'\}$ such that $\theta_i' \longrightarrow 0$, and

$$f(1 + 2m\cos\theta_i') = 0 ,$$

for every i. Writing $x_i = 1 + 2m\cos\theta_i'$, we obtain $x_i \longrightarrow 2m+1 = n$, and

$$f(x_i) = 0 ,$$

for every i. In view of (3.3.8) and lower semi-continuity of f at n, this leads us to a contradiction. Hence $\mathcal{B}$ is empty, which implies

$$f(x) > 0 \quad \forall\ x \in [-(2m-1), 2m+1]. \quad \ldots(3.3.10)$$

We shall now show that

$$f(x) > 0 ,$$

for $x \in [-(2m+1), -(2m-1)]$ also. If possible, let there exist $x_0 \in [-(2m+1), -(2m-1)]$ such that $f(x_0) = 0$. Let $\theta \in [0, \pi]$ be such that $\cos\theta = (x_0+1)/2m$, and corresponding to this θ, obtain η satisfying (3.3.6). Taking $N = 3$, $C_1 = Q^*_{m\theta}(-1)$, $C_2 = C_3 = Q^*_{m\eta}(1)$, $U = Q^*_{m(-\theta)}(-1)$ in (3.3.2), and using (3.3.4) one then gets

$$f(2m-1)\left\{f[1 + 2m\cos(\eta - \theta)]\right\}^2 = 0 .$$

However, since $1+2m\cos(\eta - \theta) \varepsilon [-(2m-1), 2m+1]$ and $2m-1 \varepsilon [-(2m-1), 2m+1]$ $(m \geq 1)$, in view of (3.3.10) this leads us to a contradiction. This establishes (3.3.7), when n is odd.

<u>Case 2</u> Suppose n is even. In this case, we proceed as we did to establish (3.3.10). The only difference lies in the fact that we have to take $C_1 = Q_{m\theta}(1)$, $C_2 = C_3 = Q_{m\eta}(1)$, and $U = Q_{m\beta}(1)$.

This completes the proof of (3.3.7).

The next step consists in proving the following stronger version of (3.3.2) : for every positive integral N' and every choice of matrices $C_1, \ldots, C_{N'}$, $U \varepsilon O(n,n)$ with $C = \sum_{i=1}^{N'} C_i$ non-neg.tive definite, the relation

$$\prod_{i=1}^{N'} f[tr(C_i)] \geq \prod_{i=1}^{N'} f[tr(UC_i)] \qquad \ldots(3.3.11)$$

holds. Observe first in view of (3.3.2) that in order to establish (3.3.11) it suffices to consider the case when $C = \sum_{i=1}^{N'} C_i$ is positive semidefinite. Obviously, then $I+sC$

is positive definite for every positive integral s. In (3.3.2), now take $N = 1+sN'$, and choose the C_i's such that one of them equals I_n and the rest are given by s copies of each of $C_1,\ldots,C_{N'}$. This gives

$$f(n)\left\{\prod_{i=1}^{N'} f[\mathrm{tr}(C_i)]\right\}^s \geq f[\mathrm{tr}(U)]\left\{\prod_{i=1}^{N'} f[\mathrm{tr}(UC_i)]\right\}^s .$$

Take the s-th root of this inequality and let $s \longrightarrow \infty$, noting that $f^{1/s}(x) \longrightarrow 1$ as $s \longrightarrow \infty$ for all x, because $0 < f(x) < \infty$. This establishes (3.3.11).

We now proceed to the final step of our proof. We consider the cases corresponding to odd n and even n separately.

Case 1 Suppose n is odd, say $n = 2m+1$ $(m \geq 1)$. Take in (3.3.11), $N' = N$, $C_i = Q^*_{m\theta_i}(1)$, $1 \leq i \leq N$, $U = Q^*_{m(-\alpha)}(1)$, where

$$\sum_{i=1}^{N} \cos\theta_i \geq 0 , \qquad \sum_{i=1}^{N} \sin\theta_i = 0 . \qquad \ldots(3.3.12)$$

This gives us the following : for every positive integral N and for every α,

$$\prod_{i=1}^{N} f(1+2m\cos\theta_i) \geq \prod_{i=1}^{N} f[1+2m\cos(\theta_i-\alpha)],$$

whenever the θ_i's satisfy (3.3.12). Writing $h(\theta) = \log f(1+2m\cos\theta)$, which is well-defined by (3.3.4) and (3.3.7), it follows that for each positive integral N and each α,

$$\sum_{i=1}^{N} h(\theta_i) \geq \sum_{i=1}^{N} h(\theta_i - \alpha) , \qquad \ldots(3.3.13)$$

whenever the θ_i's satisfy (3.3.12). The relation (3.3.13) is same as relation (4) in Bingham and Mardia (1975), and it is proved there that in this case we must have

$$h(\theta) = a \cos \theta + b, \quad \text{for every} \quad \theta ,$$

where $a(\geq 0)$ and b are some constants. By the definition of $h(\theta)$, we now obtain

$$f(x) = K \exp(\lambda x) , \quad \text{for} \quad x \in [-(2m-1),(2m+1)], \qquad \ldots(3.3.14)$$

where λ and K are some constants, with $\lambda \geq 0$ and $K > 0$. Again, by (3.3.11), for every $C, U \in O(n,n)$,

$$f[tr(C)]f[-tr(C)] \geq f[tr(UC)]f[-tr(UC)] ,$$

so that we obtain the following :

$$f(x)f(-x) \quad \text{remains constant over} \quad x \in [-n,n]. \qquad \ldots(3.3.15)$$

This, together with (3.3.14), implies that

$$f(x) = K \exp(\lambda x), \quad \text{for every} \quad x \in [-n,n],$$

where λ, K are constants, both positive, the positiveness of λ being a consequence of the stipulated non-uniformity of f. Finally, the fact that K depends only on λ is indeed a consequence of transitive group action of O(p) on O(n,p) and O(p)-invariance of Haar measure on O(n,p).

Case 2 Suppose n is even, say $n = 2m (m \geq 1)$. The proof is essentially same as that in Case 1. The only difference is the following : we have to take $C_i = Q_{m\theta_i}$, $1 \leq i \leq N$, and $U = Q_{m(-\alpha)}$. Moreover, we don't have to use (3.3.15).

This completes the proof of the theorem.

Remark 3.3.1 It is possible to give an alternative proof of Lemma 3.3.1 along the following line.

Observe first that $A \longrightarrow tr(U)$ is a real-valued continuous function on $\mathcal{M}_{nxn}$, the set of all nxn matrices. Also, $tr(I_n) = n$ and $tr(P_n) = n$, where $P_n = ((p_{ij}^{(n)}))$ is defined by

$$p_{ij}^{(n)} = 1 \quad \text{if } j = i+1,\ i > n,$$
$$= (-1)^{n-1} \quad \text{if } i = n,\ j = 1,$$
$$= 0 \quad \text{otherwise.}$$

However, since both I_n and P_n belong to SO(n), the set of all nxn orthogonal matrices with determinant + 1, in view of the fact that SO(n) is a connected subset of $\mathcal{M}_{nxn}$, we obtain by the intermediate-value property of continuous functions

$$[0,n] \subseteq \{ tr(A) : A \varepsilon SO(n) \} .$$

The proof can now be completed by using the facts that $A \varepsilon O(n,n) \Leftrightarrow -A \varepsilon O(n,n)$ and $tr(-A) = -tr(A)$.

It should be mentioned that the proof of Lemma 3.3.1 that we have provided earlier in this chapter introduces the matrix $Q_{m\theta}$ needed repeatedly later in the proof of the theorem and it is this need that prompted us to give that particular proof in spite of the fact that it is trivial.

Remark 3.3.2 It should be mentioned that the proof of the theorem generalizes the findings in Bingham and Mardia (1975). This generalization is not difficult when n is even. However, primarily because of the non-existence (other than I_n and $-I_n$) of odd-ordered orthogonal matrices with diagonal entries all equal, this generalization appears to be non-trivial when n is odd.

CHAPTER 4

A DEFINITION OF SAMPLE CIRCULAR MEDIAN

4.1 Introduction

We mentioned in Chapter 2 (vide Section 2.1) that there is no mathematically precise definition of median direction of sample observations from circle. The need for such a definition was, in fact, felt in connection with the problem explored in Section 2.2. Accordingly, following Fisher's (1985) notion of spherical median we proposed one such definition (vide Definition 2.2.1) that served our purpose there. We also proposed in Remark 2.2.3 to study this definition in detail. This chapter is devoted to this study. Some of these details will be needed in the next chapter where we shall study Bahadur representation and asymptotic normality of the sample circular median.

The problem of obtaining a suitable definition of sample circular median is, however, an issue of independent interest. We shall now discuss this. Suppose X is a random variable taking values in $[0,2\pi)$. In other words, X is a circular random variable (vide Mardia (1972), p.39). Denote the distribution function of X by F. Then $M \in [0,2\pi)$ is said to be a circular median (henceforth abbreviated to c-median) of X if

$$G(M) = \frac{1}{2} \qquad \ldots(4.1.1)$$

and

$$f(M) > f(M^*) , \qquad \ldots(4.1.2)$$

where

(a) $G : [0,2\pi) \longrightarrow [0,1]$ is defined as

$$\begin{aligned} G(x) &= F(x+\pi) - F(x) && \text{if } 0 \leq x < \pi, \\ &= 1 - \{F(x) - F(x-\pi)\} && \text{if } \pi \leq x < 2\pi, \end{aligned} \qquad \ldots(4.1.3)$$

(b) for any $x \in [0,2\pi)$, the antipodal point of x, denoted by x^*, is defined as

$$\begin{aligned} x^* &= x+\pi && \text{if } 0 \leq x < \pi , \\ &= x-\pi && \text{if } \pi \leq x < 2\pi , \end{aligned} \qquad \ldots(4.1.4)$$

and

(c) f is the density corresponding to F (Mardia (1972), p.28).

The idea underlying the definition is this : condition (4.1.1) helps us to get hold of a diameter that divides the circle into two arcs of equal probability and condition (4.1.2) tells us to choose that end point of the diameter as the c-median which has more mass. It should, however, be observed that in order to define c-median it is not necessary to have such a strong requirement as the existence of density, it suffices to have local smoothness of F at M and M^*, and then demand (4.1.2).

Suppose now that we are interested in defining sample c-median, denoted by m_n, based on n observations $x_1,\ldots,x_n$ from $[0,2\pi)$. This may be achieved by demanding, in accordance with (4.1.1) and (4.1.2), the following :

$$G_n(m_n) \simeq \frac{1}{2} \qquad \ldots(4.1.5)$$

and

$$\#\{i \mid d(x_i,m_n) \le \delta_n\} \ge \#\{i \mid d(x_i,m_n) \le \delta_n\}, \qquad \ldots(4.1.6)$$

where

(a) $G_n : [0,2\pi) \longrightarrow [0,1]$ is defined as

$$\begin{aligned} G_n(x) &= \frac{1}{n}\sum_{i=1}^{n} 1_{\{x<X_i\le x+\pi\}} && \text{if } 0\le x<\pi \\ &= 1-\frac{1}{n}\sum_{i=1}^{n} 1_{\{x-\pi<X_i\le x\}} && \text{if } \pi\le x<2\pi, \end{aligned} \qquad \ldots(4.1.7)$$

(b) $d(x,y) = \pi-\left|\pi-|x-y|\right|$, $x,y \in [0,2\pi)$ (cf. Section 4.2),

and

(c) $\{\delta_n : n \ge 1\}$ is a sequence of positive numbers which decreases monotonically to zero.

However, such a definition depends on the choice of a diameter that divides the observations into two sets of approximately equal size (cf. (4.1.5)) which might lead to problem of uniqueness of choice (vide Example 4.3.2), and also on the choice of $\{\delta_n\}$ which is purely subjective in nature.

Moreover, the latter choice can be shown to be valid <u>only</u> asymptotically -- in the sense of ensuring strong consistency -- thus offering not much practical value when the sample size is small. Hence this construction is less appealing to the practitioners.

It, therefore, appears from the discussion above that the problem of obtaining a suitable definition of sample c-median demands attention for its own sake. We explore this issue in this chapter. In Section 4.2, two results are proved which are used later in Section 4.3 where the definition is proposed. In Section 4.4, it is shown that sample c-median so defined enjoys some natural properties.

This chapter is a revised version of Purkayastha (1990a).

4.2 Two basic results

The relationship between the median of a real-valued random variable Z and minimization of $E|Z-\xi|$ ($E|Z|$ is assumed to be finite) over $\xi \in \mathbb{R}$ is well known. This provides, in particular, an alternative way of looking at sample median. The purpose of this section is to mimic this idea for obtaining a suitable definition of sample c-median.

We begin by stating the fact that for any integer $p \geq 1$, (S^p, d), where $d(\underset{\sim}{x},\underset{\sim}{y}) = \cos^{-1}(\underset{\sim}{x}'\underset{\sim}{y})$ [= the unique angle $\Theta \varepsilon [0,\pi]$ such that $\cos\Theta = \underset{\sim}{x}'\underset{\sim}{y}$], $\underset{\sim}{x}, \underset{\sim}{y} \varepsilon S^p$, is a compact metric space. This metric d is indeed the geodesic distance on S^p. For $p = 2$, the relevance of this metric in the study of spherical medians may be found in Fisher (1985).

We shall, however, focus our attention on the case corresponding to $p = 1$. Observe that if we identity $S^1 = \{(\cos x, \sin x) : 0 \leq x < 2\pi\}$ with $[0,2\pi)$ through the identification $(\cos x, \sin x) \equiv x$, the metric $\tilde{d}$ – induced by d on $[0,2\pi)$ –– has the form $\tilde{d}(x,y) = \cos^{-1}[\cos(x-y)] = \pi - \big|\pi - |x-y|\big|$, $x, y \varepsilon [0,2\pi)$. We agree to write d in place of $\tilde{d}$. It turns out that the same kind of relationship exists between the c-median of a circular random variable X and minimization of $E[d(X,y)]$ over $y \varepsilon [0,2\pi)$ as the one referred to in the first paragraph of this section. A result, which is very close in spirit to this fact, is stated in Mardia (1972, pp.30–31). In Section 5.3, we shall provide a proof of this fact. However, since there is no mathematically precise definition of sample c-median, this fact has remained unused in understanding sample c-median. We explore this issue in this section. To be more specific, we study the problem of minimizing the function $\tilde{d}(x)$, defined by

$$\tilde{d}(x) = \sum_{i=1}^{n} d(x,x_i), \quad 0 \leq x < 2\pi , \qquad \ldots(4.2.1)$$

over $x \in [0,2\pi)$, where $x_i \in [0,2\pi)$, $1 \leq i \leq n$.

Note that for every fixed $y \in [0,2\pi)$, the function $x \longmapsto d(x,y)$, $0 \leq x < 2\pi$, is continuous on $[0,2\pi)$. Moreover, $\lim_{x \to 2\pi -} d(x,y) = d(0,y)$. Therefore, it makes sense to talk of $\min_{0 \leq x < 2\pi} d(x,y)$, and hence of $\min_{0 \leq x < 2\pi} \tilde{d}(x)$.

<u>Lemma 4.2.1</u> There exists $t \in \{1,\ldots,n\}$ such that

$$\tilde{d}(x_t) = \min_{0 \leq x < 2\pi} \tilde{d}(x) .$$

<u>Proof.</u> The following are immediate consequences of the definition of the metric d:

$$\begin{aligned} d(x,y_0) &= y_0 - x & &\text{if } 0 \leq x \leq y_0 , \\ &= x - y_0 & &\text{if } y_0 \leq x \leq y_0^* , \qquad \ldots(4.2.2) \\ &= y_0 + 2\pi - x & &\text{if } y_0^* \leq x < 2\pi , \end{aligned}$$

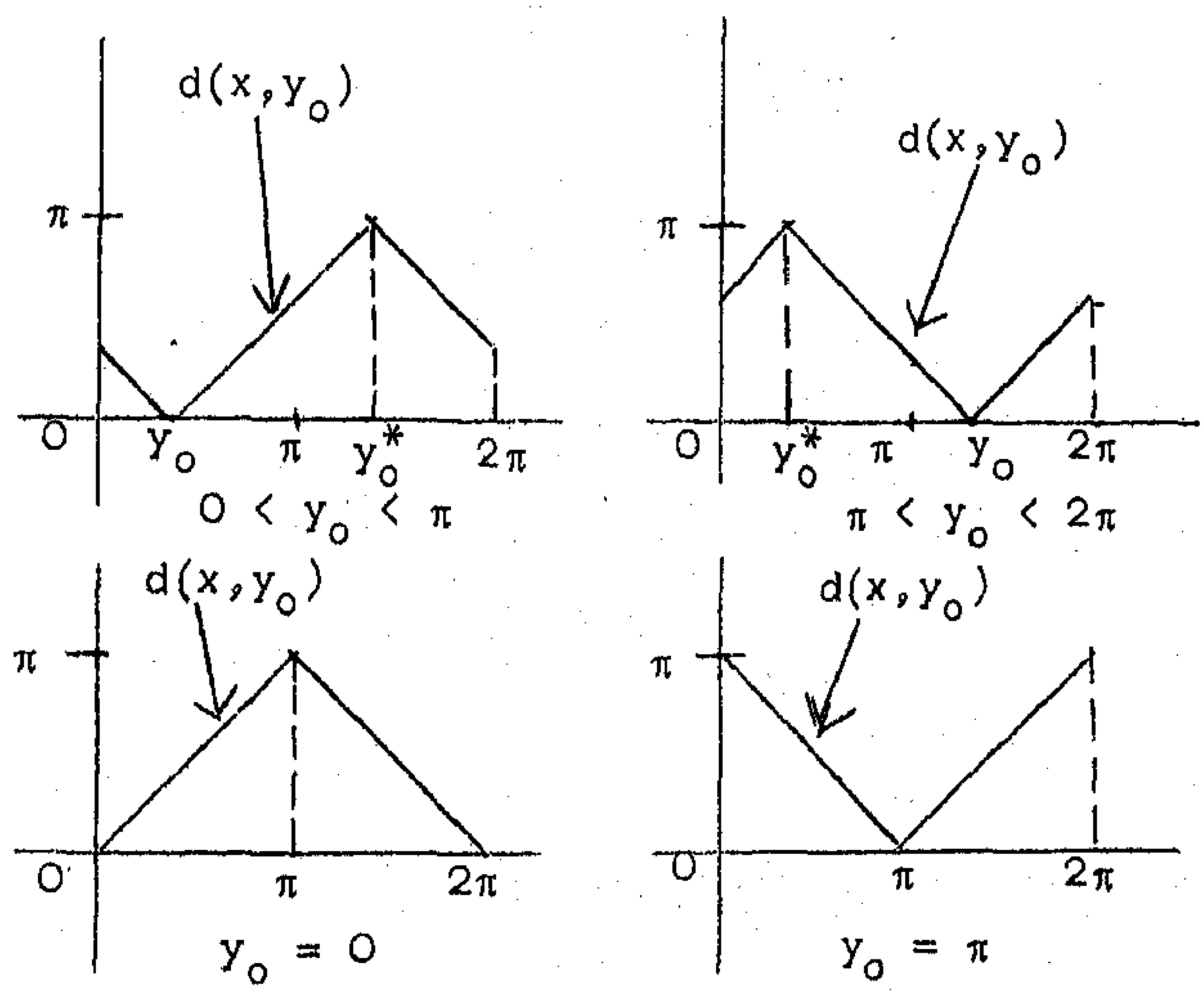
whenever $0 \leq y_0 < \pi$, and

$$\begin{aligned} d(x,y_0) &= 2\pi - y_0 + x & &\text{if } 0 \leq x \leq y_0^* , \\ &= y_0 - x & &\text{if } y_0^* \leq x \leq y_0 , \qquad \ldots(4.2.3) \\ &= x - y_0 & &\text{if } y_0 \leq x < 2\pi , \end{aligned}$$

whenever $\pi \leq y_0 < 2\pi$. Observe that (4.1.4) implies $y_0^* = y_0 + \pi$ in (4.2.2), and $y_0^* = y_0 - \pi$ in (4.2.3). The graph of the function

$$x \longrightarrow d(x,y_o) , \qquad 0 \le x < 2\pi ,$$

is as follows :



...(4.2.4)

Let us now write

$$\{x_1,\ldots,x_n\} \cup \{x_1^*,\ldots,x_n^*\} = \{z_1,\ldots,z_\ell\} ,$$

where $z_i \in [0,2\pi)$, $1 \le i \le \ell$, and $z_1 < z_2 < \cdots < z_\ell$.
Define the sets $A_1,\ldots,A_\ell$ by

$$A_j = [0,z_1] \cup [z_\ell,2\pi) , \qquad j = 1 ,$$

$$= [z_{j-1},z_j] \qquad , \qquad j = 2,\ldots,\ell .$$

It is easy to verify, using (4.2.2), (4.2.3) and (4.2.4), that the following are true :

(i) $\tilde{d}$ is linear on each of $[0,z_1],[z_1,z_2],$

$\ldots,[z_{\ell-1},z_\ell],[z_\ell,2\pi),$ ...(4.2.5)

and

(ii) $$\frac{\tilde{d}(0) - \tilde{d}(z_\ell)}{2\pi - z_\ell} = \frac{\tilde{d}(z_1) - \tilde{d}(0)}{z_1}, \text{ in case } z_1 > 0 . \qquad \ldots(4.2.6)$$

Therefore, we must have

$$\min_{0 \le x < 2\pi} \tilde{d}(x) = \tilde{d}(z_j) ,$$

for some $j \varepsilon \{1,\ldots,\ell\}$, say $j = j_o$.

The rest of the proof consists in establishing that

$$z_{j_o} = x_t , \qquad \ldots(4.2.7)$$

for some $t \varepsilon \{1,\ldots,n\}$. Suppose not. This implies $z_{j_o} \neq x_t$, for every $t = 1,\ldots,n$, so that by the definition of z_j's we must have $z_{j_o} = x_p^*$, for some p, say $p = p_o$. Observe now that we cannot have $x_i = x_{p_o}$ for every $i = 1,\ldots,n$, since if that was the case, we would have $\tilde{d}(x) = nd(x,x_{p_o})$ implying $\min_{0 \le x < 2\pi} \tilde{d}(x) = \tilde{d}(x_{p_o})$, a contradiction to the supposition that (4.2.7) is false. This implies the existence of an integer p' such that $x_{p'} \neq x_{p_o}$, so that it makes sense to define

$$\tilde{d}_1(x) = \sum_{i: x_i \neq x_{p_o}} d(x, x_i) , \quad 0 \leq x < 2\pi .$$

Now define the sets $B_1, \ldots, B_\ell$ by

$$\begin{aligned} B_i &= (z_\ell, 2\pi) \cup [0, z_2) && \text{if } i = 1 , \\ &= (z_{i-1}, z_{i+1}) && \text{if } i = 2, \ldots, \ell-1 , \\ &= (z_{\ell-1}, 2\pi) \cup [0, z_1) && \text{if } i = \ell . \end{aligned}$$

Observe that

$$B_i \cap \{ z_1, \ldots, z_\ell \} = \{ z_i \} ,$$

for every i . Moreover, $x_i \neq x_{p_o}$ implies $x_i^* \neq x_{p_o}^* = z_{j_o}$, and we already have $x_t \neq z_{j_o}$, for every t . Therefore, for every i such that $x_i \neq x_{p_o}$, the function

$$x \longrightarrow d(x, x_i), \quad 0 \leq x < 2\pi ,$$

is linear on B_{j_o} . Hence, $\tilde{d}_1(x)$ also is linear on B_{j_o} . This implies, in turn, that $\tilde{d}_1(x) \leq \tilde{d}_1(z_{j_o})$, for some $x \in B_{j_o} - \{z_{j_o}\}$, say $x = z_o$. Again, $d(x, x_{p_o}) \leq d(z_{j_o}, x_{p_o})$, for every $x \in [0, 2\pi)$, with equality if and only if $x = z_{j_o}$. This implies, in particular, $d(z_o, x_{p_o}) < d(z_{j_o}, x_{p_o})$. Thus

$$\tilde{c}\,(z_o) = \sum_{i: x_i \neq x_{p_o}} d(z_o, x_i) + \sum_{i: x_i = x_{p_o}} d(z_o, x_i)$$

$$= \tilde{d}_1(z_o) + n_{p_o}\, d(z_o, x_{p_o})$$

$$[n_{p_o} = \# \{ i : x_i = x_{p_o} \}, \text{ and is positive}]$$

$$< \tilde{d}_1(z_{j_o}) + n_{p_o}\, d(z_{j_o}, x_{p_o}) = \tilde{d}(z_{j_o}),$$

contradicting the definition of z_{j_o}. Hence (4.2.7) follows.

This completes the proof of our assertion.

Before we state and prove the next result, we introduce the following sets of integers :

(a) $N_i^{(n)} = \{1,\ldots,n\} - \{1,2,n\}$, $i = 1$,

$= \{1,\ldots,n\} - \{i-1,i,i+1\}$, $i = 2,\ldots,n-1$,

$= \{1,\ldots,n\} - \{1,n-1,n\}$, $i = n$;

n being an even number greater than 2 .

(b) $N_i^{(n)} = \{1,\ldots,n\} - \{i\}$, $i = 1,\ldots,n$;

n being an odd number greater than 2.

The sets will be needed to study the behaviour of the function $\tilde{d}$ outside neighbourhoods containing the sample observations.

<u>Lemma 4.2.2</u> Suppose $x_i \in [0,2\pi)$, $i = 1,\ldots,n$ $(n>2)$, satisfy

(i) $x_1,\ldots,x_n$ are all distinct,

and

(ii) $\{x_1,\ldots,x_n\} \cap \{x_1^*,\ldots,x_n^*\} = \phi$.

Suppose, moreover, that

(iii) for every $i = 1,\ldots,n$,

$$\tilde{d}(x_{(i)}) \neq \tilde{d}(x_{(j)})$$

for $j \in N_i^{(n)}$, where $x_{(1)} < \cdots < x_{(n)}$ are the x_i's arranged in increasing order.

Then

(a) for n even,

$$\left\{ y \in [0,2\pi) \mid \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \right\} = C_i, \text{ for some } i,$$

where

$$C_k = [0,x_{(1)}] \cup [x_{(n)},2\pi), \quad k = 1,$$
$$= [x_{(k-1)}, x_{(k)}], \quad k = 2,\ldots,n,$$

and

(b) for n odd,

$$\left\{ y \in [0,2\pi) \mid \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \right\} = \left\{ x_{(i)} \right\}, \text{ for some } i$$

<u>Note</u> The quantities or sets that appear in the proofs of Lemma 4.2.1 and Lemma 4.2.2 will be assumed to bear the same meaning.

<u>Proof of Lemma 4.2.2</u> (a) It is easy to see, in view of (4.2.5), (4.2.6) and $\lim_{x \to 2\pi-} \tilde{d}(x) = \tilde{d}(0)$, that there exist

constants $g_1, g_2, \ldots, g_\ell$ and $c_1, c_2, \ldots, c_\ell$ such that

$$\tilde{d}(x) = g_j x + c_j, \quad x \in A_j, \quad j = 1, \ldots, \ell \qquad \ldots(4.2.8)$$

[see proof of Lemma 4.2.1 for the definition of the sets A_j]. Observe now that for every $i = 1, \ldots, n$, $x \longmapsto d(x, x_{(i)})$, $0 \le x < 2\pi$, is a piecewise linear continuous function with every line segment having gradient +1 or -1. Therefore, every g_j, being a sum of even (=n) number of +1 or -1, is an even number -- positive, negative or zero.

We have also seen in the proof of Lemma 4.2.1 that $\tilde{d}_1$ is linear on B_{j_o}, where $j_o = \min. \{ j : \tilde{d}(z_j) = \min_{0 \le x < 2\pi} \tilde{d}(x) \}$. Therefore, there exist real constants g and c such that

$$\tilde{d}_1(x) = gx + c, \quad x \in B_{j_o}. \qquad \ldots(4.2.9)$$

Observe now that

$$\tilde{d}(x) = \tilde{d}_1(x) + d(x, x_{t_o}),$$

for every $0 \le x < 2\pi$, and hence in particular for $x \in B_{j_o}$, where $x_{t_o} = z_{j_o}$. On the other hand, since $B_{j_o} \subseteq A_{j_o} \cup A_{j_o+1}$ ($A_{\ell+1} \equiv A_1$), it follows from (4.2.2), (4.2.3), (4.2.8), (4.2.9), and the definition of A_j's and B_i's that

$$g_{j_o} = g - 1 \quad \text{and} \quad g_{j_o+1} = g+1$$

($g_{\ell+1} \equiv g_1$). This implies, in turn,

$$g_{j_o+1} - g_{j_o} = 2 \,. \qquad \ldots(4.2.10)$$

The next step consists in proving that either

$$g_{j_o} = 0 \text{ and } g_{j_o+1} = 2 \qquad \ldots(4.2.11.1)$$

or

$$g_{j_o} = -2 \text{ and } g_{j_o+1} = 0 \,. \qquad \ldots(4.2.11.2)$$

Observe in view of (4.2.10) that in order to establish (4.2.11.1) or (4.2.11.2), it suffices to establish

$$g_{j_o} \leq 0 \text{ and } g_{j_o+1} \geq 0 \,. \qquad \ldots(4.2.12)$$

Suppose not. Then either

$$g_{j_o} > 0 \qquad \ldots(4.2.13.1)$$

or

$$g_{j_o+1} < 0 \,. \qquad \ldots(4.2.13.2)$$

Note that when (4.2.13.1) is true, we obtain from (4.2.8) and the fact $z_{j_o} \in \underset{\sim}{A}_{j_o}$, that there exists $y \in A_{j_o}$ such that $y \neq z_{j_o}$ and $\tilde{d}(y) < \tilde{d}(z_{j_o})$, contradicting the definition of z_{j_o}. Again, when (4.2.13.2) is true, we are lead to the same contradiction. This contradiction, therefore, establishes (4.2.12), and hence (4.2.11.1) and (4.2.11.2).

Let us now observe in consideration of (4.2.8) and (4.2.11.1) that $\tilde{d}$ is constant on A_{j_o} . Since both z_{j_o-1} and z_{j_o} $(z_o \equiv z_{\ell+1})$ belong to A_{j_o} , this implies , in particular, that $\tilde{d}(z_{j_o-1}) = \tilde{d}(z_{j_o})$. In turn, this implies $\tilde{d}(z_{j_o-1}) = \min\limits_{0 \leq x < 2\pi} \tilde{d}(x)$. Employing now arguments similar to those needed in Lemma 4.2.1 to prove $z_{j_o} = x_{(t_o)}$, we obtain $z_{j_o-1} = x_{(t)}$ for some $1 \leq t \leq n$ with $t \neq t_o$, say $t = t_o'$. In view of the definition of z_j's and the conditions (i) and (ii) in the statement of the lemma, this implies $x_{(t_o')} = x_{(t_o-1)} (x_{(o)} \equiv x_{(n)})$. We, therefore, obtain

$$C_{t_o} \subseteq \{ y \in [0,2\pi) \mid \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \} .$$

Similarly, it follows from (4.2.11.2) that

$$C_{t_o+1} \subseteq \{ y \in [0,2\pi) \mid \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \}$$

$(C_{n+1} \equiv C_1)$. From what we have done so far it is now evident that

$$\{ y \in [0,2\pi) \mid \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \} = \bigcup_{r=1}^{s} C_{i_r} ,$$

where $\{i_1,\ldots,i_s\} \subseteq \{1,\ldots,n\}$. However, in consideration of condition (iii) in the statement of the lemma it is immediate that $|\{i_1,\ldots,i_s\}| = 1$. This proves part (a) of our assertion.

(b) The proof of this part proceeds essentially along the same line as in the proof of part (a). We, therefore, provide the outline only.

We begin with the observation that every g_j, being a sum of odd (= n) number of +1 or -1 , is an odd number - positive or negative. Therefore, we must have

$$\left\{ y \in [0,2\pi) \,\middle|\, \tilde{d}(y) = \min_{0 \le x < 2\pi} \tilde{d}(x) \right\}$$

$$= \text{a non-empty subset of } \{ z_1, \ldots, z_\ell \} .$$

Observe also from Lemma 4.2.1 that every z_j satisfying $\tilde{d}(z_j) = \min_{0 \le x < 2\pi} \tilde{d}(x)$ must also satisfy $z_j = x_t$ for some $t \in \{1, \ldots, n\}$. Hence

$$\left\{ y \in [0,2\pi) \,\middle|\, \tilde{d}(y) = \min_{0 \le x < 2\pi} \tilde{d}(x) \right\}$$

$$= \text{a non-empty subset of } \{ x_{(1)}, \ldots, x_{(n)} \} .$$

Suppose

$$\left\{ y \in [0,2\pi) \,\middle|\, \tilde{d}(y) = \min_{0 \le x < 2\pi} \tilde{d}(x) \right\} = \{ x_{(i_1)}, \ldots, x_{(i_s)} \} ,$$

where $1 \le i_1 < \ldots < i_s \le n$. However, in consideration of condition (iii) in the statement of the lemma we must have $s = 1$. This proves part (b) of our assertion.

<u>Remark 4.2.1</u> Suppose in part (b) of Lemma 4.2.2, z_{j_o} satisfies $\tilde{d}(z_{j_o}) = \min_{0 \le x < 2\pi} \tilde{d}(x)$. Then the relation analogous to (4.2.11.1) and (4.2.11.2) that arises in this connection is

$$g_{j_0} = -1 \quad \text{and} \quad g_{j_0+1} = 1 . \qquad \ldots(4.2.14)$$

A proof of this fact is essentially similar to that of (4.2.11.1) and (4.2.11.2).

Remark 4.2.2 In Lemma 4.2.2, it is assumed that n is greater than 2. Therefore, it is of some interest to ask whether the assertion holds with $n = 1$ and $n = 2$.

If $n = 1$, we have only x_1 and so

$$\left\{ y \in [0,2\pi) \,\middle|\, \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \right\} = \{ x_1 \} . \qquad \ldots(4.2.15)$$

If $n = 2$, it does not make sense to talk of $N_1^{(2)}$ and $N_2^{(2)}$. The assertion, however, holds good. In fact, we can establish little more :

$$\left\{ y \in [0,2\pi) \,\middle|\, \tilde{d}(y) = \min_{0 \leq x < 2\pi} \tilde{d}(x) \right\} = C_1 \quad \text{if} \quad |x_1 - x_2| > \pi$$

$$\ldots(4.2.16)$$

$$= C_2 \quad \text{if} \quad |x_1 - x_2| < \pi .$$

4.3 The definition

We begin this section by introducing the following sets :

$$U_n = \left\{ (x_{(1)}, \ldots, x_{(n)}) \in \mathbb{R}^n \mid 0 \leq x_{(1)} \leq \ldots \leq x_{(n)} < 2\pi \right\} .$$

$$A_n(\underset{\sim}{x}^{(n)}) = \left\{ y \in [0,2\pi) \,\middle|\, \sum_1^n d(y, x_{(i)}) = \min_{0 \leq x < 2\pi} \sum_1^n d(x, x_{(i)}) \right\} ;$$

$$\underset{\sim}{x}^{(n)} \equiv (x_{(1)}, \ldots, x_{(n)}) \in U_n .$$

$$C_{i,n}(\underset{\sim}{x}^{(n)}) = \begin{cases} [0, x_{(1)}] \cup [x_{(n)}, 2\pi), & i = 1, \\ [x_{(i-1)}, x_{(i)}] & , \ i = 2, \ldots, n, \end{cases}$$

when n is even, and

$$= \{ x_{(i)} \} , \quad i = 1, \ldots, n ,$$

when n is odd .

$$E_{i,n} = \{ \underset{\sim}{x}^{(n)} \in U_n \mid s \neq t \Rightarrow |x_{(s)} - x_{(t)}| > 0 \text{ and } \neq \pi, \quad A_n(\underset{\sim}{x}^{(n)}) = C_{i,n}(\underset{\sim}{x}^{(n)}) \} , \quad 1 \leq i \leq n, \ n \geq 1 .$$

$$E'_{i,n} = \{ \underset{\sim}{x}^{(n)} \in U_n \mid s \neq t \Rightarrow |x_{(s)} - x_{(t)}| > 0 \text{ and } \neq \pi, \quad \sum_{j=1}^{n} d(x_{(i)}, x_{(j)}) \neq \sum_{j=1}^{n} d(x_{(p)}, x_{(j)}) \ \forall \, p \in N_i^{(n)} \} , \quad 1 \leq i \leq n, \ n \geq 2.$$

$$E_n = \bigcup_{i=1}^{n} E_{i,n} , \quad E'_n = \bigcap_{i=1}^{n} E'_{i,n} .$$

Observe now that for every $i = 1, \ldots, n$, with $n > 2$,

(i) $\quad E_{i,n} \subseteq E'_{i,n} \cap E'_{i+1,n}$ ...(4.3.1)

$(E'_{n+1,n} \equiv E'_{1,n})$, when n is even ; and

(ii) $E_{i,n} \subseteq E'_{i,n}$, ...(4.3.2)

when n is odd. Moreover, it is easy to see in view of Lemma 4.2.2 that for every $n > 2$,

$$E'_n \subsetneq E_n . \qquad \text{...(4.3.3)}$$

With Lemma 4.2.1 and Lemma 4.2.2 in mind, we are now prepared to define circular median for n (>2) observations from $[0,2\pi)$.

Definition 4.3.1 Suppose $x_1,\ldots,x_n \in [0,2\pi)$ $(n > 2)$. Denote the x_i's , arranged in increasing order, by $x_{(1)},\ldots,x_{(n)}$. The c-median of $x_1,\ldots,x_n$, denoted by m_n , is defined as follows :

(a) if n is even ,

$$m_n = \frac{x_{(1)}+2\pi+x_{(n)}}{2} \pmod{2\pi} \quad \text{if } \underset{\sim}{x}^{(n)} \in E_{1,n} ,$$

$$= \frac{x_{(i-1)}+x_{(i)}}{2} \quad \text{if } \underset{\sim}{x}^{(n)} \in E_{i,n},\ i=2,\ldots,n ,$$

$$= x_{(j)} \quad \text{if } \underset{\sim}{x}^{(n)} \notin E_n , \text{ and } j=\min.\left\{ t \mid x_{(t)} \in A_n(\underset{\sim}{x}^{(n)}) \right\},$$

where $\underset{\sim}{x}^{(n)} = (x_{(1)},\ldots,x_{(n)})$,

(b) if n is odd ,

$$m_n = x_{(i)} \quad \text{if } \underset{\sim}{x}^{(n)} \in E_{i,n},\ i = 1,\ldots,n ,$$

$$= x_{(j)} \quad \text{if} \quad \underset{\sim}{x}^{(n)} \notin E_n \text{ , and}$$

$$j = \min. \left\{ t \,|\, x_{(t)} \in A_n(\underset{\sim}{x}^{(n)}) \right\} .$$

Remark 4.3.1 For $n = 1$, obviously the definition runs as follows :

$$m_n = x_1 .$$

For $n = 2$, making use of (4.2.16) we define

$$m_n = \frac{x_{(1)}+2\pi+x_{(2)}}{2} \qquad \text{if} \quad x_{(2)} - x_{(1)} > \pi ,$$

$$= \frac{x_{(1)}+x_{(2)}}{2} \qquad \text{if} \quad 0 \leq x_{(2)}-x_{(1)} < \pi ,$$

$$= x_{(1)} \qquad \text{otherwise.}$$

Remark 4.3.2 The idea behind the definition of m_n , for n even, is the following : recall that $[0,2\pi)$ is a representation of S^1 , the unit circle. Thus, for $\underset{\sim}{x}^{(n)} \in E_n$, $C_{i,n}(\underset{\sim}{x}^{(n)})$ is indeed an 'arc' for every $i = 1,\ldots,n$. The c-median m_n is then defined as the mid-point of $C_{i,n}(\underset{\sim}{x}^{(n)})$ for $\underset{\sim}{x}^{(n)} \in E_n$.

Remark 4.3.3 The proposed definition involves only computing

$$\sum_{i=1}^{n} \left\{ \pi - \left| \pi - |x_{(t)} - x_{(i)}| \right| \right\}$$

for $t = 1,\ldots,n$, and then proceed as described there.

It should be observed that while defining the c - median we encounter samples for which $\underset{\sim}{x}^{(n)} \notin E_n$, and in this case we have to choose the c-median suitably. The following result asserts that for independent and identically distributed observations from a continuous circular distribution, such samples have got probability zero.

Proposition 4.3.1 Suppose $X_1, \ldots, X_n$ $(n > 2)$ are independent and identically distributed continuous circular random variables defined over $(\Omega, \mathcal{A}, P)$. Denote the X_i's, arranged in increasing order, by $X_{(1)} \leq \ldots \leq X_{(n)}$, and write $\underset{\sim}{X}^{(n)} = (X_{(1)}, \ldots, X_{(n)})$. Then

(a) $P(\underset{\sim}{X}^{(n)} \varepsilon E'_{i,n}) = 1$, for every $i = 1, \ldots, n$,

(b) $P(\underset{\sim}{X}^{(n)} \varepsilon E'_n) = 1$,

and

(c) $P(\underset{\sim}{X}^{(n)} \varepsilon E_n) = 1$.

Remark 4.3.4 The statement of the theorem presuppose that for every $i = 1, \ldots, n$,

$$E'_{i,n} \varepsilon \mathcal{B}(U_n),$$

and for every n,

$$E_n \varepsilon \mathcal{B}(U_n),$$

where $\mathcal{B}(U_n) = \{A \cap U_n \mid A \varepsilon \mathcal{B}(\mathbb{R}^n)\}$. This is easy to verify.

Proof of Proposition 4.3.1 Note in view of (4.3.1)-(4.3.3) that it suffices to establish part (a) only.

Observe that $\underset{\sim}{X}^{(n)} \notin E'_{i,n}$ if and only if $\underset{\sim}{X}^{(n)}$ belongs to a union of finitely many sets, each such set consisting of $\underset{\sim}{x}^{(n)}$ with $x_{(i)}$'s satisfying some linear restriction and hence having probability zero. Hence the proof.

Remark 4.3.5 In Proposition 4.3.1, it is assumed that n is greater than 2 . Therefore, it is of some interest to ask whether the assertion hold for $n = 1$ and $n = 2$.

In these cases, it does not make any sense to talk of (a) and (b), since we cannot happen define $E'_{i,n}$. It should, however, be observed in view of (4.2.15) and (4.2.16) that (c) holds good.

In the following result, M_n denotes the sample c-median of n independent and identically distributed circular random variables $X_1, \ldots, X_n$ defined over $(\Omega, \mathcal{A}, P)$.

Lemma 4.3.1 M_n is measurable.

Proof. The proof is easy. Hence we omit it.

We conclude this section by two examples, where we apply our definition to obtain the sample c-median. The examples are, however, preceded by the following notations : for independent and identically distributed observations $X_1, \ldots, X_n$

from a circular distribution F , define

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}}, \qquad 0 \le x < 2\pi ,$$

$$G_n(x) = \begin{cases} F_n(x+\pi) - F_n(x), & 0 \le x < \pi , \\ 1-\{F_n(x)-F_n(x-\pi)\}, & \pi \le x < 2\pi, \end{cases}$$

and

$$D_n(x) = \frac{1}{n}\sum_{i=1}^{n} d(x,X_i) , \quad 0 \le x < 2\pi .$$

Example 4.3.1 We obtain the sample c-median for the observations given in Example 1.1 of Mardia (1972, p.2). The observations are

$$X_{(1)} = c.43 , \quad X_{(2)} = c.45 , \quad X_{(3)} = c.52 ,$$

$$X_{(4)} = c.61 , \quad X_{(5)} = c.75 , \quad X_{(6)} = c.88 ,$$

$$X_{(7)} = c.88 , \quad X_{(8)} = c.279, \quad X_{(9)} = c.357.$$

[The constant c is introduced to represent the angles in radian.]

Observe now that

$$D_n(X_{(1)}) = c'.326 , \quad D_n(X_{(2)}) = c'.320 , \quad D_n(X_{(3)}) = c'.313,$$

$$D_n(X_{(4)}) = c'.322 , \quad D_n(X_{(5)}) = c'.364 , \quad D_n(X_{(6)}) = c'.429,$$

$$D_n(X_{(7)}) = c'.429 , \quad D_n(X_{(8)}) = c'.992 , \quad D_n(X_{(9)}) = c'.581 ,$$

for some constant c' . Therefore, by our proposed definition

$$M_n = c.52$$

with $G_n(M_n) = \frac{4}{9}$.

It should be mentioned that the c-median that we have obtained is same as that obtained in Mardia (1972, p.29) by representing the data by points on the circumference of a circle (vide Mardia (1972), p.2) and this inspecting visually. However, such a strategy may not work always. The next example illustrates this issue .

Example 4.3.2 Suppose the observations are

$X_{(1)} = c.8$, $X_{(2)} = c.27$, $X_{(3)} = c.29$, $X_{(4)} = c.41$,
$X_{(5)} = c.72$, $X_{(6)} = c.138$, $X_{(7)} = c.202$, $X_{(8)} = c.206$,
$X_{(9)} = c.277$, $X_{(10)} = c.340$.

Observe now that $G_n(x) = \frac{1}{2}$ if either $x \varepsilon [X_{(1)}, X^*_{(7)})$ or if $x \varepsilon [X_{(3)}, X_{(4)}]$. Thus the choice of a diameter that divides the observations into two sets of approximately equal size poses problem of uniqueness.

However, observing that

$D_n(X_{(1)}) = c^+.714$, $D_n(X_{(2)}) = c^+.702$, $D_n(X_{(3)}) = c^+.698$,
$D_n(X_{(4)}) = c^+.693$, $D_n(X_{(5)}) = c^+.760$, $D_n(X_{(6)}) = c^+.942$,

$D_n(X_{(7)}) = c'.1086$, $D_n(X_{(8)}) = c'.1094$, $D_n(X_{(9)}) = c'.940$,
$D_n(X_{(10)}) = c'.770$,

it follows immediately

$$\left\{ y \in [0,2\pi) \mid D_n(y) = \min_{0 \le x < 2\pi} D_n(x) \right\} = [X_{(3)}, X_{(4)}] \text{ .}$$

In fact

$$\underset{\sim}{X}^{(n)} \in E_{4,10} \text{ .}$$

Therefore, by our proposed definition

$$M_n = c.\ \frac{29+41}{2} = c.35$$

with

$$G_n(M_n) = \frac{1}{2} \text{ .}$$

Remark 4.3.6 Suppose we have a sample of size 8, the data being 2 copies of 1.2,2.3,2.6,2.9 - all in radian. Then, by our proposed definition $M_8 = 2.3$. If, however, we had a sample of size 4, the data being 1.2,2.3,2.6,2.9; the c-median M_n would be $(2.3+2.6)/2 = 2.45$. It is, therefore, natural to demand that M_8 should also be 2.45. The particular choice of m_n in Definition 4.3.1, when $\underset{\sim}{x}^{(n)} \notin E_n$, is made only to ensure measurability of M_n. For $\underset{\sim}{x}^{(n)} \notin E_n$, we may replace it by any other choice from $A_n(\underset{\sim}{x}^{(n)})$ that ensures measurability of M_n.

4.4 Some properties of sample circular median

The purpose of this section is to demonstrate that sample c-median defined in the preceding section is a proper sample analogue of population c-median defined in Section 4.1. The effect of translating all the samples by a fixed amount on the sample c-median is also studied.

Throughout this section, $X_1, \ldots, X_n$ denote a set of n independent and identically distributed continuous circular random variables defined over $(\Omega, \mathcal{A}, P)$. Denote by F their common distribution function, and by M_n their sample c-median.

Theorem 4.4.1 (a) If n is even,

$$G_n(M_n) = \frac{1}{2} \text{ a.s.}$$

(b) If n is odd,

$$G_n(M_n) = \frac{1}{2} - \frac{1}{2n} \text{ a.s.}$$

Remark 4.4.1 The statement of the theorem presupposes that

(a) $\left\{ G_n(M_n) = \frac{1}{2} \right\} \varepsilon \mathcal{A}$, for n even,

and

(b) $\left\{ G_n(M_n) = \frac{1}{2} - \frac{1}{2n} \right\} \varepsilon \mathcal{A}$, for n odd.

This is easy to verify.

Proof of Theorem 4.4.1 (a) We shall first prove our assertion for $n > 2$. Note in view of Proposition 4.3.1, part (b) that it suffices to show $\underset{\sim}{X}^{(n)} \in E_n'$ implies $G_n(M_n) = \frac{1}{2}$.

Choose $\omega \in \Omega$ such that $\underset{\sim}{X}^{(n)}(\omega) = (X_{(1)}(\omega), \ldots, X_{(n)}(\omega)) \in E_n'$. However, since $E_n' \subseteq E_n$ and $E_n = \bigcup_{i=1}^{n} E_{i,n}$, $\omega \in E_{i,n}$ for some $1 \leq i \leq n$, say $i = i_o$. This implies, in particular, the following :

$$D_n(x) \text{ restricted to } C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) \text{ is a constant.} \qquad \ldots(4.4.1)$$

Observe now that for $x \in [0,2\pi) - \{X_{(1)}(\omega), \ldots, X_{(n)}(\omega), X^*_{(1)}(\omega), \ldots, X^*_{(n)}(\omega)\}$,

$$D_n(x) = \sum_{i:0\leq X_{(i)}(\omega)<x} (x - X_{(i)}(\omega)) + \sum_{i:x<X_{(i)}(\omega)<x+\pi} (X_{(i)}(\omega) - x)$$

$$+ \sum_{i:x+\pi<X_{(i)}(\omega)<2\pi} \{2\pi - (X_{(i)}(\omega) - x)\} \text{ if } 0\leq x<\pi,$$

$$= \sum_{i:0\leq X_{(i)}(\omega)<x-\pi} \{2\pi - (x - X_{(i)}(\omega))\} + \sum_{i:x-\pi<X_{(i)}(\omega)<x} (x - X_{(i)}(\omega)) + \sum_{i:x<X_{(i)}(\omega)<2\pi} (X_{(i)}(\omega) - x)$$

$$\text{if } \pi\leq x<2\pi,$$

or, in other words,

$$D_n(x) = x.[\#\{i:0\leq X_{(i)}(\omega)<x\} - \#\{i:x<X_{(i)}(\omega)<x+\pi\}$$
$$+ \#\{i: x+\pi<X_{(i)}(\omega)<2\pi\}]$$

$$- \sum_{i:0 \leq X_{(i)}(\omega) < x} X_{(i)}(\omega) + \sum_{i:x < X_{(i)}(\omega) < x+\pi} X_{(i)}(\omega) + \sum_{i:x+\pi < X_{(i)}(\omega) < x} (2\pi - X_{(i)}(\omega)) ,$$

if $0 \leq x < \pi$,

...(4.4.2.1)

$$= x.[- \# \{ i: 0 \leq X_{(i)}(\omega) < x-\pi \} + \# \{ i: x-\pi < X_{(i)}(\omega) < x \} - \# \{ i: x < X_{(i)}(\omega) < 2\pi \}]$$

$$+ \sum_{i:0 \leq X_{(i)}(\omega) < x-\pi} (2\pi + X_{(i)}(\omega)) - \sum_{i:x-\pi < X_{(i)}(\omega) < x} X_{(i)}(\omega) \quad \sum_{i:x < X_{(i)}(\omega) < 2\pi} X_{(i)}(\omega) ,$$

if $\pi \leq x < 2\pi$.

...(4.4.2.2)

Observe, moreover, that for $x \in [0,2\pi) - \{ X_{(1)}(\omega), \ldots, X_{(n)}(\omega), X^*_{(i)}(\omega), \ldots, X^*_{(n)}(\omega) \}$,

$$\# \{ i: 0 \leq X_{(i)}(\omega) < x \} - \# \{ i: x < X_{(i)}(\omega) < x+\pi \} + \# \{ i: x+\pi < X_{(i)}(\omega) < 2\pi \}$$

$$= nF_n(x) - n \{ F_n(x+\pi) - F_n(x) \} + n \{ 1 - F_n(x+\pi) \}$$

$$= n - 2n \{ F_n(x+\pi) - F_n(x) \}$$

$$= n \{ 1 - 2G_n(x) \} , \qquad \ldots(4.4.3.1)$$

if $0 \leq x < \pi$, and similarly

$$- \# \{ i: 0 \leq X_{(i)}(\omega) < x-\pi \} + \# \{ i: x-\pi < X_{(i)}(\omega) < x \} - \# \{ i: x < X_{(i)}(\omega) < 2\pi \}$$

$$= n \{ 1 - 2G_n(x) \} , \qquad \ldots(4.4.3.2)$$

if $\pi \leq x < 2\pi$.

Let us recall now from the proof of Lemma 4.2.2, part (a) and Definition 4.3.1 that $M_n \varepsilon C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) - \{X_{(1)}(\omega),\ldots,X_{(n)}(w),X^*_{(1)}(\omega),\ldots,X^*_{(n)}(\omega)\}$. In consideration of (4.4.1), (4.4.2.1), (4.4.2.2), (4.4.3.1) and (4.4.3.2), this implies $G_n(M_n) = \frac{1}{2}$; thus establishing our assertion for $n > 2$.

In order to complete the proof of the assertion we now consider the case for $n = 2$. Note in view of Remark 4.3.5 that it suffices to show $\underset{\sim}{X}^{(n)} \varepsilon E_n$ implies $G_n(M_n) = \frac{1}{2}$.

In this case we choose $\omega \varepsilon \Omega$ such that $\underset{\sim}{X}^{(n)}(\omega) \varepsilon E_n$, which implies $\underset{\sim}{X}^{(n)}(\omega) \varepsilon E_{i,n}$ for some $1 \leq i \leq n$, say $i = i_o$. We now imitate the proof of our assertion for $n>2$. However, in this case the fact $M_n \varepsilon C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) - \{X_{(1)}(\omega),\ldots,X_{(n)}(\omega),X^*_{(1)}(\omega),\ldots,X^*_{(n)}(\omega)\}$ follows immediately from (4.2.16) and Remark 4.3.1.

This completes the proof of part (a).

(b) The proof of this part proceeds essentially along the same line as in the proof of part (a). We, therefore, provide the outline only.

Let us first observe that for $n= 1$, $M_n = X_1$ and $F_n(x) = 1_{\{X_1 \leq x\}}$, so that $G_n(M_n) = 0$. Hence the assertion follows trivially.

We now consider the case when $n > 1$. As in the proof of part (a), we choose $\omega \in \Omega$ such that $\underset{\sim}{X}^{(n)}(\omega) \in E_n'$. Hence $\underset{\sim}{X}^{(n)} \in E_{i,n}$ for some $1 \leq i \leq n$, say $i = i_o$. In view of the proof of Lemma 4.2.2 - part (b), (4.4.2.1), (4.4.2.2), (4.4.3.1), (4.4.3.2) and (4.2.14), it now follows that

$$n \{ 1 - 2G_n(x) \} = 1, \qquad \ldots(4.4.4)$$

for every x such that $0 < x - X_{(i_o)} < \delta$, where δ is a sufficiently small positive number. However, $X_{(i_o)} = M_n$, so that employing right-continuity of G we obtain from (4.4.4) that $G_n(M_n) = \frac{1}{2} - \frac{1}{2n}$.

This completes the proof of part (b).

The next result of this section establishes a sample analogue of (4.1.2), when the sample size n is even. Suppose first that $n > 2$. Choose $\omega \in \Omega$ such that $\underset{\sim}{X}^{(n)}(\omega) \in E_n'$. Hence, $\underset{\sim}{X}^{(n)}(\omega) \in E_{i,n}$ for some i such that $1 \leq i \leq n$, say $i = i_1$. Obviously i_1 depends on ω through $\underset{\sim}{X}^{(n)}(\omega)$. Let us now recall from Lemma 4.2.2, part (a) and Definition 4.3.1 that

$$M_n \in C_{i_1,n}(\underset{\sim}{X}^{(n)}(\omega)) . \qquad \ldots(4.4.5)$$

Observe in view of the identity

$$d(x,y) + d(x,y^*) = \pi \quad \forall \ x, y \in [0,2\pi),$$

that M_n^*, the antipodal point of M_n (see 4.1.4 for definition), satisfies

$$D_n(M_n^*) = \max_{0 \leq x < 2\pi} D_n(x) .$$

Moreover, D_n restricted to $C_{i_1,n}(\underset{\sim}{X}^{(n)}(\omega))$ behaves like a constant function. Therefore, $M_n^* \notin C_{i_1,n}(\underset{\sim}{X}^{(n)}(\omega))$, since otherwise it follows that D_n is a constant function, which in turn contradicts the fact $C_{i_1,n}(\underset{\sim}{X}^{(n)}(\omega))$ is a proper subset of $[0,2\pi)$. Thus, we obtain

$$M_n^* \varepsilon C_{i_2,n}(\underset{\sim}{X}^{(n)}(\omega)) , \qquad \ldots(4.4.6)$$

where $1 \leq i_2 \leq n$ and $i_2 \neq i_1$. Employing similar arguments, we can obtain relations analogous to (4.4.5) and (4.4.6), also when $n = 2$. Because each $C_{i,n}(\underset{\sim}{X}^{(n)}(\omega))$ is an 'arc', in view of (4.1.2) we should expect that

$$(\text{length of } C_{i_1,n}(\underset{\sim}{X}^{(n)}(\omega)) < (\text{length of } C_{i_2,n}(\underset{\sim}{X}^{(n)}(\omega)), \qquad \ldots(4.4.7)$$

for almost all $\underset{\sim}{X}^{(n)}$.

Before we establish (4.4.7) in our next result, we introduce the following quantities :

$$L(C_{i,n}(\underset{\sim}{x}^{(n)})) = 2\pi - x_{(n)} + x_{(1)}, \quad i = 1,$$
$$= x_{(i)} - x_{(i-1)} \quad , \quad i = 2,\ldots,n,$$

for $\underset{\sim}{x}^{(n)} \varepsilon E_n$.

$$LI(M_n) = L(C_{i,n}(\underset{\sim}{X}^{(n)})) \quad \text{if } \underset{\sim}{X}^{(n)} \varepsilon E_{i,n}, \ i=1,\ldots,n$$
$$= 0 \quad \text{otherwise.}$$

$$LI(M_n^*) = L(C_{j,n}(\underset{\sim}{X}^{(n)})) \quad \text{if } \underset{\sim}{X}^{(n)} \varepsilon E_n \text{ and } M_n^* \varepsilon C_{j,n}(\underset{\sim}{X}^{(n)})$$
$$= 0 \quad \text{otherwise.}$$

Observe that $L(C_{i,n}(\underset{\sim}{x}^{(n)}))$ stands for the length of $C_{i,n}(\underset{\sim}{x}^{(n)})$ and $LI(M_n)$ (respectively, $LI(M_n^*)$) stands for the length of the 'arc' containing M_n (respectively, M_n^*).

<u>Theorem 4.4.2</u> Let n be even. Then

$$LI(M_n) < LI(M_n^*) \quad \text{a.s.} \qquad \ldots(4.4.8)$$

<u>Remark 4.4.2</u> The statement of the theorem presupposes that both $LI(M_n)$ and $LI(M_n^*)$ are measurable. This is easy to verify.

<u>Proof of Theorem 4.4.2</u> We shall first prove our assertion for $n > 2$. Note in view of Proposition 4.3.1, part (b) that it suffices to show $\underset{\sim}{X}^{(n)} \varepsilon E_n'$ implies $LI(M_n) < LI(M_n^*)$.

As in the proof of Theorem 4.4.1 - part (a) , choose $\omega \varepsilon \Omega$ such that $M_n \varepsilon C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega))$, where $1 \leq i_o \leq n$. Suppose, moreover, that $M_n^* \varepsilon C_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega))$ where $1 \leq j_o \leq n$ and $j_o \neq i_o$. Recall now from the proofs of Lemma 4.2.1 and Lemma 4.2.2 - part (a) that $C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) = A_{t_o}$ for some t_o such that $1 \leq t_o \leq \ell$ (here $\ell = 2n$), and

$$A_{t_o} \cap \{X_{(1)}(\omega),\ldots,X_{(n)}(\omega),X^*_{(1)}(\omega),\ldots,X^*_{(n)}(\omega)\}$$
$$= \{X_{(i_o-1)}(\omega),X_{(i_o)}(\omega)\} \qquad \ldots(4.4.9)$$

$(X_{(o)} \equiv X_{(n)})$. Moreover, from the definition of A_j's , it follows that $A^*_{t_o} = A_{s_o}$ for some s_o such that $1 \leq s_o \leq \ell$, so that in view of (4.4.9) we obtain

$$A_{s_o} \cap \{X_{(1)}(\omega),\ldots,X_{(n)}(\omega),X^*_{(1)}(\omega),\ldots,X^*_{(n)}(\omega)\}$$
$$= \{X^*_{(i_o-1)}(\omega),X^*_{(i_o)}(\omega)\} . \qquad \ldots(4.4.10)$$

Observe now that $M_n \varepsilon C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega))$ implies

$$M^*_n \varepsilon A_{s_o} . \qquad \ldots(4.4.11)$$

Moreover,

$$M^*_n \varepsilon C_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) . \qquad \ldots(4.4.12)$$

In consideration of the facts that each $C_{i,n}(\underset{\sim}{X}^{(n)}(\omega))$ is an 'arc' for every $1 \leq i \leq n$, and A_j is an 'arc' for every $1 \leq j \leq \ell$, we now obtain from (4.4.10), (4.4.11) and (4.4.12),

$$A_{s_o} \subsetneq C_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) ,$$

which implies in turn,

$$A^*_{s_o} \subsetneq C^*_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) ,$$

that is,

$$C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) \subsetneq C^*_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) \text{ .} \qquad \ldots(4.4.13)$$

However, since both $C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega))$ and $C^*_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega))$ are 'closed arcs', it follows from (4.4.13) that

$$L(C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) < L(C^*_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) \text{ .} \qquad \ldots(4.4.14)$$

But it is easy to see that $L(C^*_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) = L(C_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega))$, so that we obtain from (4.4.14) above

$$L(C_{i_o,n}(\underset{\sim}{X}^{(n)}(\omega)) < L(C_{j_o,n}(\underset{\sim}{X}^{(n)}(\omega)) \text{ .}$$

We have thus proved that

$$\underset{\sim}{X}^{(n)} \varepsilon E'_n \Rightarrow LI(M_n) < LI(M^*_n) \text{ ,}$$

which in turn implies (4.4.8). This establishes our assertion for $n > 2$.

In order to complete the proof of the assertion we now consider the case for $n = 2$. Note in view of Remark 4.3.5 that it suffices to show $\underset{\sim}{X}^{(n)} \varepsilon E_n$ implies $LI(M_n) < LI(M^*_n)$.

Observe in view of (4.2.16) that

$$E_{1,2} = \left\{ \underset{\sim}{x}^{(2)} \varepsilon U_2 \;\middle|\; x_{(2)} - x_{(1)} > \pi \right\} \text{,}$$

so that $\underset{\sim}{X}^{(2)} \varepsilon E_{1,n}$ implies $LI(M_n) = 2\pi - (X_{(2)} - X_{(1)})$

and $LI(M_n^*) = X_{(2)} - X_{(1)}$. Since $X_{(2)} - X_{(1)} > \pi$, we therefore obtain $LI(M_n) < LI(M_n^*)$. Similarly, $\underset{\sim}{X}^{(2)} \varepsilon E_{2,2}$ implies $LI(M_n) < LI(M_n^*)$.

This completes the proof of the theorem.

The final result of this section studies the effect of translating all the sample observations by a fixed amount on sample c-median. However, before we state and prove the result we must make precise the notion of translation in this situation. Since we are working with observations from the unit circle, it is only natural that we should choose the set of orthogonal transformations as the class of translations. But, we are representing the unit circle by $[0,2\pi)$, hence it is necessary that we get hold of those bijections on $[0,2\pi)$ which correspond to the orthogonal transformations. This is done below.

It is known that the set $O(2)$ of all 2x2 orthogonal matrices can be written as

$$O(2) = \{ A_\alpha : 0 \leq \alpha < 2\pi \} \cup \{ B_\alpha : 0 \leq \alpha < 2\pi \} ,$$

where

$$A_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} ,$$

and

$$B_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ -\sin\alpha & -\cos\alpha \end{pmatrix}.$$

Hence if $\underset{\sim}{u}' \in S^1$ be such that $u = (\cos\theta\ ,\ \sin\theta)'$ for some $0 \leq \theta < 2\pi$, then $A_\alpha \underset{\sim}{u} = (\cos(\theta+\alpha), \sin(\theta+\alpha))'$, and $B_\alpha \underset{\sim}{u} = (\cos(\theta+\alpha), -\sin(\theta+\alpha))'$. It is now obvious that if we denote the bijection on $[0,2\pi)$ that corresponds to A_α by T_α , and the one that corresponds to B_α by T^*_α , we must have

$$T_\alpha(x) = x+\alpha(\text{mod } 2\pi)\ ,\quad 0 \leq x < 2\pi\ , \qquad \ldots(4.4.15.1)$$

and

$$T^*_\alpha(x) = 2\pi - T_\alpha(x)\ ,\quad 0 \leq x < 2\pi\ , \qquad \ldots(4.4.15.2)$$

It is now easy to see in view of the distance preserving property of orthogonal transformations that for any $x, y \in [0,2\pi)$, we have

$$d(x,y) = d(T_\alpha(x), T_\alpha(y)) = d(T^*_\alpha(x), T^*_\alpha(y)), \quad \ldots(4.4.16)$$

for all $0 \leq \alpha < 2\pi$.

<u>Theorem 4.4.3</u> Denoting the sample c-median based on n independent and identically distributed observations $X_1,\ldots,X_n$ by $M_n(X_1,\ldots,X_n)$, we have

(a) $P(M_n(T_\alpha(X_1),\ldots,T_\alpha(X_n)) = T_\alpha(M_n(X_1,\ldots,X_n))) = 1$

and

(b) $P(M_n(T_\alpha^*(X_1),\ldots,T_\alpha^*(X_n)) = T_\alpha^*(M_n(X_1,\ldots,X_n))) = 1$,

for every $0 \leq \alpha < 2\pi$.

Remark 4.4.3 The statement of the theorem presupposes that for every $0 \leq \alpha < 2\pi$, both the sets

$$\{M_n(T_\alpha(X_1),\ldots,T_\alpha(X_n)) = T_\alpha(M_n(X_1,\ldots,X_n))\}$$

and

$$\{M_n(T_\alpha^*(X_1),\ldots,T_\alpha^*(X_n)) = T_\alpha^*(M_n(X_1,\ldots,X_n))\}$$

are in $\mathcal{A}$. This is easy to verify.

Proof of Theorem 4.4.3 (a) Note in view of Proposition 4.3.1, part (a) and Remark 4.3.5 that it suffices to prove

$$M_n(T_\alpha(X_1),\ldots,T_\alpha(X_n)) = T_\alpha(M_n(X_1,\ldots,X_n))$$

for all $(X_1,\ldots,X_n)$ such that $\underset{\sim}{X}^{(n)} \in E_n$. However, verification of this consists in routine algebraic manipulation. Hence we omit it.

(b) The proof of this part is essentially similar to that of the preceding part.

Remark 4.4.4 In consideration of a similar result about sample (linear) median, it is of some interest to see whether for independent and identically distributed observations from a circular distribution having circularly symmetric unimodal density, distribution of sample c-median, defined in this chapter, also has a circularly symmetric unimodal density.

CHAPTER 5

A BAHADUR-TYPE REPRESENTATION OF SAMPLE CIRCULAR MEDIAN

5.1 Introduction

In Chapter 4, we developed a definition of sample circular median which is computationally easy and enjoys some natural properties. It is now natural that we undertake the study of some of its asymptotics. Such a study has gained additional importance in view of the fact that in Chapter 2 we obtained a directional distribution for which the median direction is a maximum likelihood estimate of the location parameter. This chapter is devoted to this study.

One of the major tools for studying the large sample behaviour of a statistic consists in deriving an asymptotic linear representation of the same. For sample quantiles, such a linear representation was obtained by Bahadur (1966), which came to be known later as 'Bahadur representation of quantiles'. Kiefer (1967) calculated the exact order of the remainder term in the representation obtained by Bahadur, Sen (1968) extended Bahadur's result to random variables which are neither independent nor identically distributed. Later, Ghosh (1971) obtained a much simpler proof of a weaker version of Bahadur's result, which involves fewer assumptions than Bahadur's.

Less restrictive conditions for Bahadur representation of sample quantiles were obtained also by de Haan and Taconis - Haantjes (1979).

It is, therefore, worthwhile to investigate whether sample c-median admits any such representation. This investigation is the subject matter of this chapter. In Section 5.2, we state the basic set of assumptions under which the representation theorem will be established. In Section 5.3, a property of population c-median is established which is used later in Section 5.4 where strong consistency of sample C-median is proved. In Section 5.5, the main representation theorem is obtained. As an application of this result, we establish in Section 5.6 asymptotic normality of sample c-median —— normalized suitably.

This chapter is a revised version of Purkayastha (1990b).

5.2 The basic set-up

Let $\{X_i : i \geq 1\}$ be a sequence of independent and identically distributed circular random variabled defined over some probability space $(\Omega, \mathcal{A}, P)$. This means, $0 \leq X_i(\omega) < 2\pi$ for all $\omega \in \Omega$, for every $i \geq 1$. Denote the distribution function of X_1 by F. The following set of assumptions constitute the set-up in which we shall be working :

(A.1) F is continuous.

(A.2) There exists a unique $\theta \varepsilon [0,\pi)$ such that $G(\theta) = \frac{1}{2}$ (see (4.1.3) for the definition of $G(\cdot)$. We denote this unique θ by θ_0.

(A.3) There exists $\varepsilon > 0$ such that $F(x)$ is differentiable at every $x \varepsilon (\theta_0-\varepsilon,\theta_0+\varepsilon)U(\theta_0+\pi-\varepsilon, \theta_0+\pi+\varepsilon)$ if $0 < \theta_0 < \pi$. If, however, $\theta_0 = 0$, $F(x)$ is differentiable at every $x \varepsilon [0,\varepsilon) U (\pi-\varepsilon,\pi+\varepsilon)U(2\pi-\varepsilon,2\pi]$. In this case, $F'(x)$ at $x = 0$ means indeed the right-derivative and at $x = 2\pi$ it means the left-derivative. Further, we stipulate that $F'(0) = F'(2\pi)$. We denote $F'(x)$ whenever it exists, by $f(x)$.

(A.4) There exists a positive constant K such that

<u>either</u>

$$|f(x)-f(\theta_0)| < K|x - \theta_0|^{1/2} \quad \text{if} \quad |x-\theta_0| < \varepsilon,$$

and

$$|f(x)-f(\theta_0+\pi)|<K|x-\theta_0-\pi|^{1/2} \quad \text{if} \quad |x-\theta_0-\pi|<\varepsilon,$$

whenever $0 < \theta_0 < \pi$,

<u>or</u>

$$|f(x)-f(0)| < K|x|^{1/2} \quad \text{if} \quad 0 \leq x < \varepsilon,$$

and

$$|f(x)-f(\pi)| < K|x-\pi|^{1/2} \quad \text{if} \quad |x-\pi| < \varepsilon,$$

and

$$|f(x)-f(2\pi)| < K\ |x-2\pi|^{1/2} \quad \text{if} \quad 2\pi-\varepsilon < x \leq 2\pi,$$

whenever $\Theta_0 = 0$.

(A.5) $f(\Theta_0) \neq f(\Theta_0+\pi)$.

Let us now define

$$M = \Theta_0 \qquad \text{if} \quad f(\Theta_0) > f(\Theta_0+\pi) \ ,$$
$$\qquad = \Theta_0+\pi \quad \text{if} \quad f(\Theta_0) < f(\Theta_0+\pi) \ . \qquad \ldots(5.2.1)$$

Thus,

$$f(M) > f(M^*) \qquad \ldots(5.2.2)$$

where M^* is the antipodal point of M (see (4.1.3) for the definition of antipodal point). Observe now in view of the definition of G that G satisfies

$$G(x) + G(x^*) = 1 \ , \qquad \ldots(5.2.3)$$

for every $x \in [0,2\pi)$. Hence we obtain from (A.2) and (5.2.1),

$$G(M) = \frac{1}{2} \ . \qquad \ldots(5.2.4)$$

It, therefore, follows from (5.2.4) and (5.2.2) that M is a c-median of X_1 . We should, however, mention that our definition of population c-median is slightly different from that given in Mardia (1972, p.30). The difference lies in the fact that Mardia demands F to have a density, whereas we demand F to be differentiable only at Θ_o and $\Theta_o+\pi$.

Let us now argue that M is the unique population c-median of X_1 . Observe in view of (A.2) and (5.2.3) that

$$\{ x \in [0,2\pi) | G(x) = \tfrac{1}{2} \} = \{\Theta_o, \Theta_o+\pi\} , \qquad \ldots(5.2.5.1)$$

i.e.,

$$\{ x \in [0,2\pi) | G(x) = \tfrac{1}{2} \} = \{M, M^*\} . \qquad \ldots(5.2.5.2)$$

Thus if M_1 be a population c-median of X_1 , we must have either $M_1 = M$ or $M_1 = M^*$. However, (5.2.2) implies that we cannot have $M_1 = M^*$. Hence we must have $M_1 = M$, which means M is the unique population c-median of X_1 .

We conclude this section with the following result that studies the behaviour of the function $G(x)$.

Lemma 5.2.1 Assume (A.1), (A.2), (A.3) and (A.5).

(a) Let $0 < M < \pi$. Then

$$\begin{aligned} G(x) &> \tfrac{1}{2} \quad \text{if } x \in [0,M) \cup (M+\pi, 2\pi), \\ &= \tfrac{1}{2} \quad \text{if } x \in \{M, M+\pi\} , \qquad \ldots(5.2.6) \\ &< \tfrac{1}{2} \quad \text{if } x \in (M, M+\pi). \end{aligned}$$

If, in addition, (A.4) holds, there exists $\delta > 0$ such that

$$\begin{aligned} &G \text{ is strictly decreasing on } [M-\delta, M+\delta], \\ \text{and} \qquad &\qquad\qquad\qquad\qquad\qquad\qquad \ldots(5.2.7) \\ &\text{strictly increasing on } [M+\pi-\delta, M+\pi+\delta]. \end{aligned}$$

(b) Let $M = 0$. Then

$$\begin{aligned} G(x) &< \tfrac{1}{2} \quad \text{if } x \in (0,\pi) , \\ &= \tfrac{1}{2} \quad \text{if } x \in \{0,\pi\}, \qquad \ldots(5.2.8) \\ &> \tfrac{1}{2} \quad \text{if } x \in [\pi,2\pi) . \end{aligned}$$

If, in addition, (A.4) holds, there exists $\delta > 0$ such that

$$\begin{aligned} &G \text{ is strictly decreasing on } [0, \delta], \\ &\text{strictly increasing on } [\pi-\delta,\pi+\delta], \qquad \ldots(5.2.9) \end{aligned}$$

and

$$\text{strictly decreasing on } [2\pi-\delta, 2\pi).$$

(c) Let $\pi < M < 2\pi$. Then

$$\begin{aligned} G(x) &< \tfrac{1}{2} \quad \text{if } x \in [0,M-\pi)\cup(M,2\pi), \\ &= \tfrac{1}{2} \quad \text{if } x \in \{M-\pi,M\}, \qquad \ldots(5.2.10) \\ &> \tfrac{1}{2} \quad \text{if } x \in (M-\pi,M) . \end{aligned}$$

If, in addition, (A.4) holds, there exists $\delta > 0$ such that

$$\begin{aligned} &G \text{ is strictly increasing on } [M-\pi-\delta, M-\pi+\delta], \\ \text{and} \qquad &\qquad\qquad\qquad\qquad\qquad\qquad \ldots(5.2.11) \\ &\text{strictly decreasing on } [M-\delta, M+\delta] . \end{aligned}$$

(d) Let $M = \pi$. Then

$$G(x) < \frac{1}{2} \quad \text{if } x \in (0,\pi),$$
$$= \frac{1}{2} \quad \text{if } x \in \{0,\pi\}, \qquad \ldots(5.2.12)$$
$$> \frac{1}{2} \quad \text{if } x \in (\pi,2\pi).$$

If, in addition, (A.4) holds, there exists $\delta > 0$ such that

G is strictly increasing on $[0,\delta]$,

and strictly decreasing on $[\pi-\delta,\pi+\delta]$, ...(5.2.13)

strictly increasing on $[2\pi-\delta,2\pi)$.

Proof. We shall provide the proof for part (a) only, the proofs for the other parts being essentially similar.

We begin with the observations that (A.1) implies G is continuous on $[0,2\pi)$ and (5.2.5.2) implies

$$\{x \in [0,2\pi) \mid G(x) = \tfrac{1}{2}\} = \{M, M+\pi\}.$$

Moreover, (A.3) implies G is differentiable at M with $G'(M) = f(M+\pi) - f(M)$, which, in view of (5.2.2), is negative.

From the preceding paragraph it is therefore clear, in view of intermediate-value property of continuous functions, that

$$x \in (M,M+\pi) \Rightarrow G(x) < \frac{1}{2}. \qquad \ldots(5.2.14)$$

On the other hand, $x \in [0,M) \cup (M+\pi,2\pi)$ if and only if $x^* \in (M,M+\pi)$. In view of (5.2.3) and (5.2.14), we therefore obtain

$$x \in [0,M) \cup (M+\pi, 2\pi) \Rightarrow G(x) > \frac{1}{2} .$$

This completes the proof of (5.2.6).

Let us now assume, in addition, that (A.4) holds. Observe that (A.3) implies G is differentiable at every $x \in (M-\varepsilon, M+\varepsilon)$ with derivative $G'(x) = f(x+\pi)-f(x)$. In turn, this implies in view of (A.4), G' is continuous at M. However, $G'(M) = f(M+\pi)-f(M)$ is negative. Hence, there exists $\delta > 0$ such that

$$x \in (M-\delta, M+\delta) \Rightarrow G'(x) < 0 . \qquad \ldots(5.2.15)$$

The fact that G is strictly decreasing on $[M-\delta, M+\delta]$ is now an immediate consequence of (5.2.15) and mean-value theorem. In turn, this implies $1-G(x)$ is strictly increasing on $[M-\delta, M+\delta]$. However, we have from (5.2.3), $1-G(x) = G(x^*)$, and moreover $x \in [M-\delta, M+\delta]$ if and only if $x^* \in [M+\pi-\delta, M+\pi+\delta]$. Hence, G is strictly increasing on $[M+\pi-\delta, M+\pi+\delta]$.

This completes the proof of our assertion.

Remark 5.2.1 It is worth mentioning that in order to establish (5.2.6) we don't need (A.3) in its full force but only differentiability of F at θ_0 and $\theta_0+\pi$.

5.3 A property of population circular median

We mentioned in Section 4.2 that the same kind of relationship exists between the c-median of a circular random variable X and $\min_{0\le y<2\pi} E\{d(X,y)\}$, where $d(x,y) = \pi-|\pi-|x-y||$ for $x,y \in [0,2\pi)$, as the one between linear median and minimum mean absolute deviation. As proposed there, we explore this issue in this section.

Consider the function $D : [0,2\pi) \longrightarrow \mathbb{R}$ defined as follows :

$$D(y) = E\{d(X_1,y)\}.$$

Lemma 5.3.1 (a) $D(y)+D(y^*) = \pi$, for every $y \in [0,2\pi)$.

(b) $\lim_{y \to 2\pi-} D(y) = D(0)$ and D is uniformly continuous on $[0,2\pi)$.

(c) D is differentiable on $[0,2\pi)$ with

$$D'(y) = 1-2G(y), \quad 0 \le y < 2\pi .$$

[For $y = 0$, $D'(y)$ is indeed the right-derivative].

(d) Suppose there exists $y_0 \in [0,2\pi)$ such that

$$D(y_0) = \min.\{D(y) : d(y,y_0) \le \varepsilon\}$$

or

$$= \max.\{D(y) : d(y,y_0) \le \varepsilon\}$$

for some $\varepsilon > 0$. Then $G(y_0) = \frac{1}{2}$.

Proof. (a) It is easy to see that for every $x \in [0,2\pi)$,

$$d(x,y) + d(x,y^*) = \pi$$

for all $y \in [0,2\pi)$. Hence

$$D(y)+D(y^*) = E\left\{d(X_1,y)+d(X_1,y^*)\right\}$$
$$= \pi ,$$

for every $y \in [0,2\pi)$.

(b) We have

$$D(y) = \int_0^{2\pi} d(x,y)dF(x)$$

for $0 \leq y < 2\pi$. Observe now that for every fixed $x \in [0,2\pi)$, the function

$$y \longrightarrow d(x,y), \quad 0 \leq y < 2\pi ,$$

is continuous on $[0,2\pi)$ and satisfies $\lim_{y \to 2\pi-} d(x,y) = d(x,0)$. Moreover, $0 \leq d(x,y) \leq \pi$ for all $x,y \in [0,2\pi)$. Therefore, a straightforward application of dominated convergence theorem implies D is continuous on $[0,2\pi)$ and $\lim_{y \to 2\pi-} D(y) = D(0)$. The fact that D is uniformly continuous on $[0,2\pi)$ now follows immediately.

(c) We shall only prove that

$$\lim_{\delta \to 0+} \frac{D(y+\delta)-D(y)}{\delta} = 1-2G(y) \qquad \ldots(5.3.1)$$

for every $y \in [0,\pi)$. The other parts of the assertion will follow along essentially the same line.

Let us first establish that

$$D(y) = y\left\{1-2(F(y+\pi)-F(y))\right\} -E(X_1)+2\int_y^{y+\pi} x\, dF(x)+2\pi\left\{1-F(y+\pi)\right\} \qquad \ldots(5.3.2)$$

for $0 \leq y \leq \pi$. To see this, choose and fix $y \in [0,\pi]$.
Then,

$$D(y) = \int_0^{2\pi} \left\{ \pi - \left|\pi - |x-y|\right| \right\} dF(x)$$

$$= \int_0^{y+\pi} |x-y|\, dF(x) + \int_{y+\pi}^{2\pi} \left\{ 2\pi - |x-y| \right\} dF(x)$$

$$= \int_0^{y} (y-x)dF(x) + \int_y^{y+\pi} (x-y)dF(x) + 2\pi \int_{y+\pi}^{2\pi} dF(x) - \int_{y+\pi}^{2\pi} (x-y)dF(x)$$

$$= yF(y) - \int_0^y x\, dF(x) + \int_y^{y+\pi} x\, dF(x) - y\left\{F(y+\pi)-F(y)\right\}$$

$$+ 2\pi\left\{1-F(y+\pi)\right\} - \int_{y+\pi}^{2\pi} x\, dF(x) + y\int_{y+\pi}^{2\pi} dF(x)$$

$$= yF(y) + [-\int_0^y x\, dF(x) + \int_y^{y+\pi} x\, dF(x) - \int_{y+\pi}^{2\pi} x\, dF(x)] - y\left\{F(y+\pi)\right.$$

$$\left. -F(y)\right\} + 2\pi\left\{1-F(y+\pi)\right\} + y\left\{1-F(y+\pi)\right\}$$

$$= y[1-2F(y+\pi)+2F(y)] + [-E(X_1) + 2\int_y^{y+\pi} x\, dF(x)] + 2\pi\left\{1-F(y+\pi)\right\}$$

$$= y\left\{1-2(F(y+\pi)-F(y))\right\} - E(X_1) + 2\int_y^{y+\pi} x\, dF(x) + 2\pi\left\{1-F(y+\pi)\right\} .$$

Hence (5.3.2) follows.

The next step consists in proving the following :

$$D(y+\delta) = D(y) + 2\int_{y+\pi}^{y+\pi+\delta} \left\{x-(y+\pi)\right\} dF(x) - 2\int_y^{y+\delta} (x-y)dF(x)$$

$$+ \delta\left\{1-2G(y+\delta)\right\} \qquad \ldots(5.3.3)$$

for $y \varepsilon [0,\pi)$ and $\delta > 0$ such that $y+\delta \varepsilon [0,\pi]$. To see this, choose and fix $y \varepsilon [0,\pi)$ and $\delta > 0$ such that $y+\delta \varepsilon [0,\pi]$. Then, from (5.3.2) we get

$$D(y+\delta) = (y+\delta)\{1-2(F(y+\delta+\pi)-F(y+\delta))\} - E(X_1)+2\int_{y+\delta}^{y+\delta+\pi} x\, dF(x) + 2\pi\{1-F(y+\delta+\pi)\}$$

$$= y\{1-2(F(y+\delta+\pi)-F(y+\delta))\} + \delta\{1-2(F(y+\delta+\pi)-F(y+\delta))\} - E(X_1)+2[\int_{y}^{y+\pi} x\, dF(x) + \int_{y+\pi}^{y+\delta+\pi} x\, dF(x) - \int_{y}^{y+\delta} x\, dF(x)] + 2\pi\{1-F(y+\delta+\pi)\}$$

$$= y\{1-2(F(y+\pi)-F(y))\} - 2y\{(F(y+\delta+\pi)-F(y+\delta)) - (F(y+\pi)-F(y))\} + \delta\{1-2G(y+\delta)\} - E(X_1)+2\int_{y}^{y+\pi} x\, dF(x) + 2\int_{y+\pi}^{y+\delta+\pi} x\, dF(x) - 2\int_{y}^{y+\delta} x\, dF(x) + 2\pi\{1-F(y+\pi)\} - 2\pi\{F(y+\delta+\pi)-F(y+\pi)\}$$

$$= y\{1-2(F(y+\pi) - F(y))\} - E(X_1)+2\int_{y}^{y+\pi} x\, dF(x) + 2\pi\{1-F(y+\pi)\} - 2y\{F(y+\delta+\pi)-F(y+\pi)\} + 2y\{F(y+\delta)-F(y)\} + 2\int_{y+\pi}^{y+\delta+\pi} x\, dF(x) - 2\int_{y}^{y+\delta} x\, dF(x) - 2\pi\{F(y+\delta+\pi) - F(y+\pi)\} + \delta\{1-2G(y+\delta)\}$$

$$= \; D(y)+2\int_{y+\pi}^{y+\pi+\delta}\{x-(y+\pi)\}\,dF(x)-2\int_{y}^{y+\delta}(x-y)\,dF(x)$$

$$+\,\delta\{1-2G(y+\delta)\}\;.$$

Hence (5.3.3) follows.

We now establish (5.3.1). Choose and fix $y\,\varepsilon\,[0,\pi)$. Then, from (5.3.3) we obtain

$$\frac{D(y+\delta)-D(y)}{\delta}=\frac{2}{\delta}\int_{y+\pi}^{y+\pi+\delta}\{x-(y+\pi)\}\,dF(x)-\frac{2}{\delta}\int_{y}^{y+\delta}(x-y)\,dF(x)$$

$$+\{1-2G(y+\delta)\}, \qquad \ldots(5.3.4)$$

for all δ, positive and sufficiently small. Observe now that

$$0\leq\frac{2}{\delta}\int_{y+\pi}^{y+\pi+\delta}\{x-(y+\pi)\}\,dF(x)\leq\frac{2}{\delta}\,.\,\delta\,.\{F(y+\pi+\delta)-P(y+\pi)\},$$

so that in view of continuity of F, we get

$$\lim_{\delta\to 0+}\;\frac{2}{\delta}\int_{y+\pi}^{y+\pi+\delta}\{x-(y+\pi)\}\,dF(x)=0\;. \qquad \ldots(5.3.5.1)$$

Similarly,

$$\lim_{\delta\to 0+}\;\frac{2}{\delta}\int_{y}^{y+\delta}(x-y)\,dF(x)=0\;. \qquad \ldots(5.3.5.2)$$

The assertion (5.3.1) is now an immediate consequence of (5.3.4), (5.3.5.1), (5.3.5.2), and continuity of F.

(d) We shall provide the proof for $y_o \in (0,2\pi)$. The proof for $y_o = 0$ is essentially similar.

Observe that if $y_o \in (0,2\pi)$, there exists $\eta > 0$ such that

$$d(x,y_o) \leq \varepsilon \iff |x-y_o| \leq \eta \ .$$

(Here, we are assuming of course that $\varepsilon > 0$ is sufficiently small).

Therefore, by elementary analysis, the derivative of D at y_o -- which is known to exist by part (c) of the present lemma -- must vanish. This means

$$D'(y_o) = 1 - 2G(y_o) = 0 \ ,$$

implying

$$G(y_o) = \frac{1}{2} \ .$$

This completes the proof of the lemma.

Remark 5.3.1 It is worth mentioning that parts (a) and (b) of the above lemma are true for any circular random variable, i.e., without any restriction on F. Parts (c) and (d), however, require the additional assumption of continuity of F, i.e., (A.1).

Remark 5.3.2 In Mardia (1972, p.31), a result similar to part (c) of Lemma 5.3.1 can be found.

Lemma 5.3.2 Assume (A.1), (A.2), (A.3), and (A.5).

(a) Let $0 < M < \pi$. Then D is strictly decreasing on $[0,M]$, strictly increasing on $[M,M+\pi]$, and strictly decreasing on $[M+\pi,2\pi)$.

(b) Let $M = 0$. Then D is strictly increasing on $[0,\pi]$ and strictly decreasing on $[\pi,2\pi)$.

(c) Let $\pi < M < 2\pi$. Then D is strictly increasing on $[0,M-\pi]$, strictly decreasing on $[M-\pi,M]$, and strictly increasing on $[M,2\pi)$.

(d) Let $M = \pi$. Then D is strictly decreasing on $[0,\pi]$ and strictly increasing on $[\pi,2\pi)$.

Proof. The proof is an immediate consequence of Lemma 5.3.1, part (c) and Lemma 5.2.1. We omit the details.

We shall conclude this section with the following theorem, which is indeed the goal we have set out with. Before we state it, we note that in view of part (b) of Lemma 5.3.1 it makes sense to talk of $\min_{0\le y<2\pi} D(y)$. The statement of the theorem assumes (A.1), (A.2), (A.5), and (A.3) not in its full force but only differentiability of F at θ_0 and $\theta_0+\pi$.

Theorem 5.3.1 $\{z \in [0,2\pi) \mid D(z) = \min_{0\le y<2\pi} D(y)\} = \{M\}$.

Proof. The proof is an immediate consequence of Lemma 5.3.2 and Lemma 5.3.1, part (b). We omit the details.

Remark 5.3.3 In Mardia (1972, p.31), a result similar to the above theorem can be found.

5.4 Strong consistency of sample circular median

In this section we establish strong consistency of sample C-median, defined in Section 4.3.

The proof of our result will be based on the following theorem.

Theorem 5.4.1 (Sverdrup-Thygeson (1981)) Given a probability space $(\Omega, \mathcal{A}, P)$ and a compact metric space (S, d), let $\mathcal{B}$ be a σ-algebra containing the open subsets of S, and let $Z_1, Z_2, \ldots$ be independent, identically distributed random variables from $(\Omega, \mathcal{A}, P)$ into $(S, \mathcal{B})$, with their common probability distribution on $(S, \mathcal{B})$ denoted by p. Suppose

(i) there exists $m \in S$ such that

$$\left\{ y \in S \mid \int d(y,z)dp(z) = \inf_{x \in S} \int d(x,z)dp(z) \right\} = \{ m \}$$

and

(ii) for every integer $n \geq 1$, there exists a measurable transformation M_n from $(\Omega, \mathcal{A}, P)$ into $(S, \mathcal{B})$ such that

$$\frac{1}{n} \sum_{1}^{n} d(M_n(\omega), Z_i(\omega)) = \inf_{x \in S} \frac{1}{n} \sum_{1}^{n} d(x, Z_i(\omega))$$

for every $\omega \in \Omega$.

Then

$$P(\{\omega \in \Omega \mid \lim_{n\to\infty} d(M_n(\omega),m) = 0\}) = 1.$$

<u>Remark 5.4.1</u> The original result of Sverdrup-Thygeson (1981) is somewhat more general than the one we have stated here.

The next result establishes strong consistency of sample c-median. For the result to hold, we need the same assumptions as the ones we did in Theorem 5.3.1.

<u>Theorem 5.4.2</u> $d(M_n,M) \longrightarrow 0$ a.s.

<u>Proof.</u> We stated in Section 4.2 that $([0,2\pi),d)$ with $d(x,y) = \pi-\left|\pi-|x-y|\right|$, $x, y \in [0,2\pi)$, is a compact metric space. Hence, we choose $S = [0,2\pi)$ and $d(x,y) = \pi-\left|\pi-|x-y|\right|$, $x, y \in [0,2\pi)$ in the statement of Theorem 5.4.1. Choose, moreover, $Z_i = X_i$ for every $i \geq 1$, there. Denote the probability distribution on $([0,2\pi),\mathcal{B})$ corresponding to the distribution function F by p. Observe that in view of Theorem 5.3.1, condition (i) in the statement of Theorem 5.4.1 is satisfied if we choose $m = M$, the population c-median of X_1. Moreover, our definition of sample c-median, developed in Chapter 4, states that condition (ii) also in the statement of Theorem 5.4.1 is satisfied with M_n = sample c-median of $X_1,\ldots,X_n$. It is now obvious that our assertion is an immediate consequence of that of Theorem 5.4.1.

5.5 The main representation theorem

The purpose of this section is to prove the main result of this chapter.

The first result of this section (Lemma 5.5.1) is true for any circular distribution F^*. A proof of this result, however, depends on the following celebrated exponential-type probability inequality for Kolmogorov-Smirnov distance.

Theorem 5.5.1 (Dvoretzky, Kiefer, Wolfowitz (1956)) Let H^* be a distribution function defined on $\mathbb{R}$. There exists a finite positive constant C (not depending on H^*) such that

$$P(\sup_{x \in \mathbb{R}} |H_n^*(x) - H(x)| > d) \leq C\, e^{-2nd^2}, \quad d > 0, \quad n=1,2,\ldots,$$

where H_n^* denotes the empirical distribution function corresponding to n independent and identically distributed observations $X_1,\ldots,X_n$ from H^*.

Lemma 5.5.1 Let $X_1,\ldots,X_n$ be independent and identically distributed observations from a circular distribution F^*. Denote by G^* and G_n^*, the functions analogous to G in (4.1.3) and G_n in (4.3.5), in this case. There exists a finite positive constant C^* (not depending on F^*) such that

$$P(\sup_{0\le x<2\pi} |G_n^*(x)-G^*(x)| > d) \le C^* e^{-\frac{nd^2}{2}} , \quad d > 0, \ n=1,2,\ldots ;$$

Proof. Observe that

$$G_n^*(x)-G^*(x) = \{F_n^*(x+\pi)-F^*(x+\pi)\} - \{F_n^*(x)-F^*(x)\}, \qquad 0\le x<\pi,$$
$$= -\{F_n^*(x)-F^*(x)\} + \{F_n^*(x-\pi)-F^*(x-\pi)\}, \quad \pi\le x<2\pi.$$

(Cf. (4.3.4) for the definition of the function F_n^*).

This implies

$$\sup_{0\le x<2\pi} |G_n^*(x)-G^*(x)| = \sup_{0\le x<\pi} | \{F_n^*(x+\pi)-F^*(x+\pi)\} - \{F_n^*(x)-F^*(x)\} |.$$

Therefore,

$$P(\sup_{0\le x<2\pi} |G_n^*(x)-G^*(x)| > d)$$

$$\le P(\sup_{0\le x<\pi} |F_n^*(x+\pi)-F^*(x+\pi)| > \tfrac{d}{2}) + P(\sup_{0\le x<\pi} |F_n^*(x)-F^*(x)| > \tfrac{d}{2})$$

$$\le 2P(\sup_{0\le x<2\pi} |F_n^*(x)-F^*(x)| > \tfrac{d}{2}) ,$$

which is

$$\le 2C e^{-\frac{2nd^2}{n}} ,$$

by Theorem 5.5.1 . Writing $C^* = 2C$, the proof follows.

Before we prove our next result (Lemma 5.5.2), we need some more definitions. Assume (A.1), (A.2), (A.3), and (A.5). Observe that the function G, defined in (4.1.3), is differentiable at every $x \in \{z \in [0,2\pi) | d(M,z) < \varepsilon\} \cup \{z \in [0,2\pi) | d(M^*,z) < \varepsilon\}$ (this $\varepsilon > 0$ is same as that in (A.3)) with derivative g', given by

$$g(x) = f(x^*) - f(x) .$$

Define a sequence of positive numbers $\{\varepsilon_n : n \geq 2\}$ by

$$\varepsilon_n = \frac{8(\log n)^{1/2}}{|g(M)|n^{1/2}} .$$

Define the following subsets $\{I_n : n \geq 2\}$ and $\{J_n : n \geq 2\}$ of $[0,2\pi)$ by

$$I_n = \{x \in [0,2\pi) \,|\, d(x,M) \leq \varepsilon_n\} , \quad J_n = \{x \in [0,2\pi) \,|\, d(x,M^*) \leq \varepsilon_n\} .$$

We now state a result of elementary analysis which will be used to prove Lemma 5.5.2. The proof is easy and hence omitted.

Proposition 5.5.1 Let $\phi : [a,b] \longrightarrow \mathbb{R}$ be a continuous function with $\phi(a) = \phi(b) = 0$ and $\phi(x) > 0 \;\forall\; x \in (a,b)$. Suppose $\exists\; \delta_1 > 0$ and $\delta_2 > 0$ such that

(i) ϕ is strictly increasing on $[a,a+\delta_1]$

and

(ii) ϕ is strictly decreasing on $[b-\delta_2,b]$.

Then $\exists\; \delta_0 > 0$ such that

$$\inf_{a+\delta \leq x \leq b-\delta} \phi(x) = \phi(a+\delta) \quad \text{or} \quad \phi(b-\delta)$$

for all $\delta < \delta_0$.

Lemma 5.5.2 Assume (A.1) - (A.5). Denote the sample c-median of $X_1,\ldots,X_n$ by M_n . Then with probability 1,

$M_n \varepsilon I_n$ for all n, sufficiently large. ...(5.5.1)

Proof. The proof is provided for the case $0 < M < \pi$ only. The proofs for the other cases are essentially similar except for a few minor modifications.

Note in view of the definitions of I_n and J_n that for every $n \geq 3$,

$$I_{n+1} \subsetneq I_n \quad \text{and} \quad J_{n+1} \subsetneq J_n ,$$

and, moreover,

$$\bigcap_{n=2}^{\infty} I_n = \{M\} , \quad \bigcap_{n=2}^{\infty} J_n = \{M^*\} .$$

It is, therefore, clear that if we prove

$$P(M_n \varepsilon (I_n \cup J_n)^c \text{ infinitely often}) = 0, \quad ...(5.5.2)$$

(where $(I_n \cup J_n)^c \equiv [0,2\pi)-(I_n \cup J_n)$), in view of Theorem 5.4.2 this will imply (5.5.1). However, observe in consideration of Borel-Cantelli lemma that in order to establish (5.5.2) it suffices to prove

$$\sum_{n=2}^{\infty} P(M_n \varepsilon (I_n \cup J_n)^c) < \infty . \quad ...(5.5.3)$$

We shall, therefore, prove (5.5.3).

Note first that $\exists$ an integer N_0 such that

$$I_n = [M-\varepsilon_n, M+\varepsilon_n], \quad J_n = [M^*-\varepsilon_n, M^*+\varepsilon_n]$$

for every $n \geq N_o$. Observe now that since G is continuous on $[0,2\pi)$, in view of Lemma 5.2.1, part (a) and Proposition 5.5.1, we obtain

$$\inf_{M+\varepsilon_n \leq x \leq M^*-\varepsilon_n} |G(x) - \tfrac{1}{2}| = |G(M+\varepsilon_n) - \tfrac{1}{2}| \text{ or } |G(M^*-\varepsilon_n) - \tfrac{1}{2}| \qquad \ldots(5.5.4)$$

for every $n \geq N_1$, where $N_1 > N_o$. Again, since $G(x)+G(x^*) = 1$ for every $x \in [0,2\pi)$, and because $x \in [0,M-\varepsilon_n]U[M^*+\varepsilon_n,2\pi)$ if and only if $x^* \varepsilon[M+\varepsilon_n,M^*-\varepsilon_n]$, from (5.5.4) we obtain immediately

$$\inf_{\substack{0 \leq x \leq M-\varepsilon_n \\ \text{and} \\ M^*+\varepsilon_n \leq x < 2\pi}} |G(x) - \tfrac{1}{2}| = |G(M+\varepsilon_n) - \tfrac{1}{2}| \text{ or } |G(M^*-\varepsilon_n) - \tfrac{1}{2}| \qquad \ldots(5.5.5)$$

for every $n \geq N_1$. However, it is easy to see that

$$(I_n \cup J_n)^c = [0,M-\varepsilon_n]U[M+\varepsilon_n,M^*-\varepsilon_n]U[M^*+\varepsilon_n,2\pi),$$

for every $n \geq N_1$. It, therefore, follows from (5.5.4) and (5.5.5),

$$\inf_{x \,\varepsilon (I_n \cup J_n)^c} |G(x) - \tfrac{1}{2}| = |G(M+\varepsilon_n) - \tfrac{1}{2}| \text{ or } |G(M^*-\varepsilon_n) - \tfrac{1}{2}| \qquad \ldots(5.5.6)$$

for every $n \geq N_1$. Observe now that differentiability of G at M implies

$$G(M+\varepsilon_n) = G(M) + \varepsilon_n\, g(M) + o(\varepsilon_n), \quad n \longrightarrow \infty$$

$$\Rightarrow \quad G(M+\varepsilon_n) = \frac{1}{2} + \varepsilon_n[g(M) + \frac{o(\varepsilon_n)}{\varepsilon_n}], \qquad n \longrightarrow \infty$$

$$\Rightarrow \quad G(M+\varepsilon_n) - \frac{1}{2} < \varepsilon_n \cdot \frac{g(M)}{2},$$

for every $n \geq N_2$, where $N_2 > N_1$. However, since $g(M) = f(M^*) - f(M)$ is negative (cf. (5.2.2)), this implies

$$|G(M+\varepsilon_n) - \tfrac{1}{2}| > \varepsilon_n \cdot \frac{|g(M)|}{2} = \frac{4(\log n)^{1/2}}{n^{1/2}} \qquad \ldots(5.5.7.1)$$

for every $n \geq N_2$. Similarly, from differentiability of G at M^* we obtain

$$|G(M^*-\varepsilon_n) - \tfrac{1}{2}| > \frac{4(\log n)^{1/2}}{n^{1/2}} \qquad \ldots(5.5.7.2)$$

for every $n \geq N_3$, where $N_3 > N_2$. From (5.5.6), (5.5.7.1) and (5.5.7.2), we thus get

$$\inf_{x \in (I_n \cup J_n)^c} |G(x) - \tfrac{1}{2}| > \frac{4(\log n)^{1/2}}{n^{1/2}} \qquad \ldots(5.5.8)$$

for every $n \geq N_3$.

The next step consists in proving that

$$P(M_n \in (I_n \cup J_n)^c) \leq P(|G_n(M_n) - G(M_n)| > \frac{2(\log n)^{1/2}}{n^{1/2}}) \qquad \ldots(5.5.9)$$

for every $n \geq N_3$. To see this, choose and fix an odd number $n \geq N_3$. Then

$$P(M_n \in (I_n \cup J_n)^c)$$

$$= P(M_n \in (I_n \cup J_n)^c, G_n(M_n) = \tfrac{1}{2} - \tfrac{1}{2n})$$

(By Theorem 4.4.1, part (b))

$$\leq P(|G(M_n) - \tfrac{1}{2}| > \frac{4(\log n)^{1/2}}{n^{1/2}}, G_n(M_n) = \tfrac{1}{2} - \tfrac{1}{2n})$$

(By (5.5.8) above)

$$= P(|G(M_n) - G_n(M_n) - \tfrac{1}{2n}| > \frac{4(\log n)^{1/2}}{n^{1/2}})$$

(By Theorem 4.4.1, part (b))

$$\leq P(|G_n(M_n) - G(M_n)| > \frac{2(\log n)^{1/2}}{n^{1/2}}),$$

the last inequality being a consequence of the fact that for two real numbers x and y $|x-y| \leq \frac{2(\log n)^{1/2}}{n^{1/2}}$ for an integer $n \geq 2$ implies $|x-y-\frac{1}{2n}| < \frac{4(\log n)^{1/2}}{n^{1/2}}$. Similarly, for an even number $n \geq N_3$, (5.5.9) can be established by using (5.5.8) and Theorem 4.4.1, part (a). This establishes (5.5.9).

Observe in the final step of the proof that for every $n \geq N_3$, it follows from (5.5.9)

$$P(M_n \in (I_n \cup J_n)^c)$$

$$\leq P(\sup_{0 \leq x < 2\pi} |G_n(x) - G(x)| > \frac{2(\log n)^{1/2}}{n^{1/2}})$$

$$\leq C^* \exp(-n \cdot \frac{4 \log n}{n} \cdot \tfrac{1}{2}) \qquad \text{(By Lemma 5.5.1)}$$

$$= \frac{C^*}{n^2},$$

so that (5.5.3) follows immediately, completing the proof of our assertion.

Lemma 5.5.3 Assume (A.1) - (A.5). Then the following expansion for the function $G(x)$ holds ;

(a) if $0 < M < 2\pi$,

$$G(x) = G(M) + (x-M)g(M) + O(|x-M|^{3/2}), \quad |x-M| < \varepsilon ,$$

(b) if $M = 0$,

$$\begin{aligned} G(x) &= G(M) + (x-M)g(M) + O(|x-M|^{3/2}), \quad 0 \leq x < \varepsilon, \\ &= G(M)+(x-2\pi)g(M)+O(|x-2\pi|^{3/2}) , \quad 2\pi-\varepsilon < x < 2\pi. \end{aligned}$$

Proof. For $0 < M < 2\pi$, the proof follows immediately from mean-value theorem and (A.4). For $M = 0$, essentially the same proof can be pushed through, the only difference lies in the fact that we have to use $\lim_{x \to 2\pi-} G(x) = G(0)$. We omit the details.

Let us now recall that the Bahadur representation for linear median given indeed a representation for

$$|\hat{\xi}_{1/2,n} - \xi_{1/2}| \operatorname{sgn}(\hat{\xi}_{1/2,n} - \xi_{1/2}) ,$$

where $\xi_{1/2}$ is the population median and $\hat{\xi}_{1/2,n}$ is the sample median based on a random sample of size n. In our case, we shall replace $|\hat{\xi}_{1/2,n} - \xi_{1/2}|$ by $d(M_n, M)$. As regards the 'sgn' part, we observe that there is no natural order on the circle. Therefore, we have to force one such notion into the picture that serves our purpose. The following definition is made towards this end.

Definition 5.5.1 We define a sequence of random variables $\{Z_n : n \geq 1\}$ as follows :

$$Z_n = d(M_n, M) \quad \text{if} \quad M < M_n \leq M+\pi ,$$
$$= -d(M_n, M) \quad \text{otherwise},$$

whenever $0 \leq M < \pi$, and

$$Z_n = -d(M_n, M) \quad \text{if} \quad M-\pi < M_n \leq M ,$$
$$= d(M_n, M) \quad \text{otherwise},$$

whenever $\pi \leq M < 2\pi$.

Observe that Z_n is positive (respectively, negative) if in order to move from M to M_n along the geodesic connecting M and M_n one has to traverse in an anti-clockwise (respectively, clockwise) direction.

Lemma 5.5.4 Assume (A.1) - (A.5). Then with probability 1,

$$G(M_n)-G(M) = Z_n g(M)+O(\{d(M_n, M)\}^{3/2}), \quad n \longrightarrow \infty .$$

Proof. We provide the proof for $0 < M < 2\pi$ only. For $M = 0$, the proof is essentially similar.

Note first that we have from Theorem 5.4.2,

$$\lim_{n \to \infty} d(M_n, M) = 0 \quad \text{a.s.}$$

Since $0 < M < 2\pi$, this implies the following : with probability 1,

$$d(M_n,M) = |M_n-M| \quad \text{and} \quad Z_n = M_n-M \qquad \ldots(5.5.10)$$

for all n, sufficiently large.

The proof is now an immediate consequence of Lemma 5.5.3 and (5.5.10) above.

The following result is a slightly modified version of Lemma 1 of Bahadur (1966). The proof of Bahadur can be adapted here without any substantial modification. We , therefore, state the result only.

Lemma 5.5.5 Suppose that $\tilde{F}$ is a distribution function defined over IR. Suppose $x_0 \in$ IR is such that $\tilde{F}$ is differentiable on $(x_0,x_0+\delta)$ for some $\delta > 0$ and $\tilde{F}$ is differentiable from the right at x_0. Write $\tilde{F}'(x) = \tilde{f}(x)$, $x \in I = (x_0,x_0+\delta)$. Suppose $\tilde{f}$ is continuous on I, and moreover, $\lim\limits_{x \to x_0+} \tilde{f}(x) = \tilde{F}'_{+(x_0)}$. Let $\{a_n\}$ be a sequence of positive constants such that

$$a_n \sim c_0 n^{-1/2} (\log n)^{1/2} , \quad n \longrightarrow \infty .$$

Put

$$\tilde{H}_n = \sup_{0 \le x \le a_n} |[\tilde{F}_n(x_0+x) - \tilde{F}_n(x_0)] - [\tilde{F}(x_0+x) - \tilde{F}(x_0)]|$$

$[\tilde{F}_n(\cdot)$ = empirical distribution function]

Then, with probability 1,

$$\tilde{H}_n = O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty .$$

<u>Remark 5.5.1</u> If in the above lemma, instead of demanding $\tilde{F}$ to be differentiable continuously on $[x_o, x_o+\delta)$, we assume $\tilde{F}$ to be differentiable continuously on $(x_o-\delta, x_o]$, the resulting assertion (with supremum taken over $-a_n \leq x \leq 0$) holds good.

<u>Lemma 5.5.6</u> Assume (A.1) - (A.5). Define

$$H_n = \sup_{x \in I_n} |[G_n(x) - G_n(M)] - [G(x) - G(M)]|$$

Then, with probability 1,

$$H_n = O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty .$$

<u>Proof</u>. We shall provide the proof corresponding to $0 \leq M < \pi$ only. For $\pi \leq M < 2\pi$, the proof is essentially similar. However, it will be seen that the proof corresponding to $0 < M < \pi$ is slightly different from the proof corresponding to $M = 0$.

<u>Case (a)</u> $0 < M < \pi$.

Observe first that there exists an integer $N > 0$ such that

$$I_n = [M - \varepsilon_n , M + \varepsilon_n] \subsetneq (0, \pi)$$

for every $n \geq N$. Hence

$$G_n(x) = F_n(x + \pi) - F_n(x) \ , \quad x \in I_n \ ,$$

and

$$G_n(M) = F_n(M + \pi) - F_n(M)$$

for every $n \geq N$. Also,

$$G(x) = F(x+\pi) - F(x), \quad x \in I_n, \quad n \geq N \ ,$$

and

$$G(M) = F(M+\pi) - F(M) \ .$$

Therefore, for every $x \in I_n$, $n \geq N$,

$$[G_n(x) - G_n(M)] - [G(x) - G(M)]$$

$$= \{[F_n(x+\pi)-F(M+\pi)]-[F(x+\pi)-F(M+\pi)]\}$$

$$- \{[F_n(x)-F_n(M)]-[F(x)-F(M)]\}$$

so that for every $n \geq N$,

$$H_n = \sup_{x \in I_n} |[G_n(x)-G_n(M)]-[G(x)-G(M)]|$$

$$\leq \sup_{x \in I_n} |[F_n(x+\pi)-F_n(M+\pi)]-[F(x+\pi)-F(M+\pi)]|$$

$$+ \sup_{x \in I_n} |[F_n(x)-F_n(M)]-[F(x)-F(x)]|$$

$$= \sup_{|x| \leq \varepsilon_n} |[F_n(M+\pi+x)-F_n(M+\pi)]-[F(M+\pi+x)-F(M+\pi)]|$$

$$+ \sup_{|x| \leq \varepsilon_n} |[F_n(M+\pi)-F_n(M)]-[F(M+\pi)-F(M)]|$$

Observe now that in view of (A.3) and (A.4), F is differentiable continuously on both $(M-\varepsilon, M+\varepsilon)$ and $(M+\pi-\varepsilon, M+\pi+\varepsilon)$. It, therefore, follows from Lemma 5.5.5 and Remark 5.5.1, with probability 1,

$$H_n = O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty .$$

Case (b) $M = 0$

Observe first that there exists an integer $N > 0$ such that

$$I_n = [2\pi-\varepsilon_n, 2\pi) \cup [0, \varepsilon_n]$$

for every $n \geq N$. Hence

$$\begin{aligned} G_n(x) &= F_n(x+\pi)-F(x), & 0 \leq x \leq \varepsilon_n , \\ &= 1 - \{F_n(x)-F_n(x-\pi)\}, & 2\pi-\varepsilon_n \leq x < 2\pi , \end{aligned}$$

and

$$G_n(M) = F_n(\pi)-F_n(0) ,$$

for every $n \geq N$. Also,

$$\begin{aligned} G(x) &= F(x+\pi)-F(x) , & 0 \leq x \leq \varepsilon_n \\ &= 1 - \{F(x)-F(x-\pi)\} , & 2\pi-\varepsilon_n \leq x < 2\pi , \end{aligned}$$

and

$$G(M) = F(\pi)-F(0) .$$

Observe, moreover, that in view of continuity of F,

$$1-G_n(M) = F_n(2\pi)-F_n(\pi) \quad a.s.$$

for every $n \geq 1$, and

$$1-G(M) = F(2\pi)-F(\pi) .$$

It, therefore, follows that for every $n \geq N$,

$$[G_n(x)-G_n(M)]-[G(x)-G(M)]$$
$$= \{[F_n(x+\pi)-F_n(\pi)]-[F(x+\pi)-F(\pi)]\} - \{[F_n(x)-F_n(0)]-[F(x)-F(0)]\}$$

whenever $0 \leq x \leq \varepsilon_n$, and

$$[G_n(x)-G_n(M)]-[G(x)-G(M)]$$
$$= -\{[F_n(x)-F_n(2\pi)]-[F(x)-F(2\pi)]\} + \{[F_n(x-\pi)-F_n(\pi)]-[F(x-\pi)-F(x)] \quad a.s.,$$

whenever $2\pi-\varepsilon_n \leq x < 2\pi$ (the almost sure set not being dependent on x).

Observe now that in view of (A.3) and (A.4), F is differentiable continuously on each of $[0,\varepsilon)$, $(2\pi-\varepsilon,2\pi]$ and $(\pi-\varepsilon,\pi+\varepsilon)$. Hence, employing Lemma 5.5.5 and Remark 5.5.1, the proof follows.

We conclude this section with the main representation theorem, the goal that we have set out with.

<u>Theorem 5.5.2</u> Assume (A.1) - (A.5). Then

$$Z_n = \frac{\frac{1}{2} - G_n(M)}{g(M)} + R_n ,$$

where with probability 1 ,

$$R_n = O(n^{-3/4}(\log n)^{3/4}) , \quad n \longrightarrow \infty .$$

Proof. In view of Lemma 5.5.2, note first that we have the following : with probability 1 ,

$$d(M_n,M) = O(\frac{(\log n)^{1/2}}{n^{1/2}}) , \quad n \longrightarrow \infty . \qquad \ldots(5.5.11)$$

Hence, from (5.5.11) and Lemma 5.5.4 we obtain the following: with probability 1 ,

$$G(M_n)-G(M) = Z_n g(M)+O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty . \qquad \ldots(5.5.12)$$

Further, from (5.5.11), (5.5.12) and Lemma 5.5.6 we obtain the following : with probability 1 ,

$$G_n(M_n)-G_n(M) = Z_n g(M)+O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty . \qquad \ldots(5.5.13)$$

However, in view of Theorem 4.4.1, we have

$$G_n(M_n) = \frac{1}{2} + \frac{c_n}{n} \quad \text{a.s.} ,$$

where

$$c_n = 0 \quad \text{for } n \text{ even} ,$$

$$= -\frac{1}{2} \quad \text{for } n \text{ odd} .$$

Therefore, from (5.5.13) and (5.5.14), we obtain

$$Z_n g(M) = \frac{1}{2} - G_n(M) + R_n,$$

where with probability 1,

$$R_n = O(n^{-3/4}(\log n)^{3/4}), \quad n \longrightarrow \infty .$$

This completes the proof of our assertion.

Remark 5.5.2 *In view of Ghosh's (1971) simpler proof of* a weaker version of Bahadur's result, it is of some interest to see whether the following weaker version of Theorem 5.5.2 can be established :

$$Z_n = \frac{\frac{1}{2} - G_n(M)}{g(M)} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

under (A.1), (A.2), (A.5) and (A.3) not in its full force but only differentiability of F at θ_o and $\theta_o + \pi$.

5.6 An application

As an application of Theorem 5.5.2, we have the following result which establishes asymptotic normality of $\{Z_n\}$.

Theorem 5.6.1 Assume (A.1) - (A.5). Then

$$\sqrt{n}\, Z_n \xrightarrow{d} N\left(0, \frac{1}{4g^2(M)}\right)$$

Proof. We have

$$G_n(M) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{M < X_i \leq M+\pi\}} \quad \text{if} \quad 0 \leq M < \pi ,$$

$$= 1 - \frac{1}{n} \sum_{j=1}^{n} 1_{\{M-\pi \leq X_i \leq M\}} \quad \text{if} \quad \pi \leq M < 2\pi .$$

However, since

$$P(M < X_i \leq M+\pi) = G(M) = \frac{1}{2}$$

for $0 \leq M < \pi$, and

$$P(M-\pi < X_i \leq M) = 1 - G(M) = \frac{1}{2}$$

for $\pi \leq M < 2\pi$, we obtain by central limit theorem

$$\sqrt{n}\, G_n(M) \xrightarrow{d} N(0, \tfrac{1}{4}) . \qquad \ldots(5.6.1)$$

The proof now follows by employing Slutsky's Theorem in Theorem 5.5.2 and (5.6.1) .

Remark 5.6.1 In each of Theorem 5.4.2, Theorem 5.5.2 and Theorem 5.6.1, the underlying metric on $[0,2\pi)$ is the one induced by the geodesic distance on S^1. It, therefore, seems natural to ask whether the assertions of these theorems hold if instead we make use of the metric on $[0,2\pi)$ induced by the Euclidean metric on $\mathbb{R}^2$ restricted to S^1 . We argue below that in each of the cases the answer is in the affirmative.

Observe that for $\underset{\sim}{u}, \underset{\sim}{v} \in S^1$, the following relation exists between the Euclidean distance $||\underset{\sim}{u} - \underset{\sim}{v}||_2$ and their geodesic distance $\cos^{-1}(\underset{\sim}{u}' \underset{\sim}{v})$:

$$||\underset{\sim}{u} - \underset{\sim}{v}||_2 = 2 \sin \left[\frac{\cos^{-1}(\underset{\sim}{u}' \underset{\sim}{v})}{2}\right] , \qquad \ldots(5.6.2)$$

since

$$\begin{aligned} ||\underset{\sim}{u} - \underset{\sim}{v}||_2^2 &= 2 - 2(\underset{\sim}{u}' \underset{\sim}{v}) \\ &= 2 \left\{1 - \cos[\cos^{-1}(\underset{\sim}{u}' \underset{\sim}{v})]\right\} \\ &= 4 \sin^2 \left[\frac{\cos^{-1}(\underset{\sim}{u}' \underset{\sim}{v})}{2}\right] . \end{aligned}$$

Denote now by d_1, the metric on $[0,2\pi)$, induced by the Euclidean metric on $\mathbb{R}^2$ restricted to S^1. It, then, follows from (5.6.2) that

$$d_1(x,y) = 2 \sin \frac{d(x,y)}{2} \quad \forall\, x,y \in [0,2\pi]. \qquad \ldots(5.6.3)$$

Let us now redefine Z_n in Definition 5.5.1 by making use of d_1, and not d. Denote this new Z_n by Z_n'. It is easy to see, in view of (5.6.3), that for every $n \geq 1$,

$$d_1(M_n,M) = 2 \sin \frac{d(M_n,M)}{2} \qquad \ldots(5.6.4.1)$$

and

$$Z_n' = 2 \sin \frac{Z_n}{2} . \qquad \ldots(5.6.4.2)$$

Observe first in view of (5.6.4.1) and Theorem 5.4.2 that

$$d_1(M_n,M) \longrightarrow 0 \quad a.s.$$

Observe next from Definition 5.5.1 that $|Z_n| = d(M_n,M)$ for every $n \geq 1$. Therefore, from Theorem 5.4.2 we obtain

$$Z_n \longrightarrow 0 \quad a.s. \qquad \ldots(5.6.5)$$

However, since

$$|\sin x - x| = O(|x|^3), \quad x \longrightarrow 0,$$

we obtain from (5.6.5) above the following : with probability 1,

$$|2 \sin \frac{Z_n}{2} - Z_n| = O(|Z_n|^3), \quad n \longrightarrow \infty,$$

i.e., with probability 1,

$$|Z_n' - Z_n| = O(|Z_n|^3), \quad n \longrightarrow \infty. \qquad \ldots(5.6.6)$$

On the other hand, we have from Lemma 5.5.2 the following : with probability 1,

$$|Z_n| = O(\frac{(\log n)^{1/2}}{n^{1/2}}), \quad n \longrightarrow \infty,$$

so that from (5.6.6) above we obtain,

$$|Z_n' - Z_n| = O(\frac{(\log n)^{3/2}}{n^{3/2}}), \quad n \longrightarrow \infty, \qquad \ldots(5.6.7)$$

with probability 1. Taken with Theorem 5.5.2, this implies

$$Z_n' = \frac{\frac{1}{2} - G_n(M)}{g(M)} + R_n \qquad \ldots(5.6.8)$$

where, with probability 1 ,

$$R_n = O\left(\frac{(\log n)^{3/4}}{n^{3/4}}\right), \quad n \longrightarrow \infty .$$

Finally, it follows from (5.6.8) along exactly the same line as in the proof of Theorem 5.6.1 that

$$\sqrt{n}\, Z_n' \xrightarrow{d} N\left(0, \frac{1}{4g^2(M)}\right) .$$

Remark 5.6.2 This remark is made to discuss applicability of Theorem 5.5.2 and Theorem 5.6.1, given a particular distribution.

It turns out that most of the standard circular distributions are symmetric and unimodal about a fixed point of the circle. This guarantees existence of unique c-median (vide Mardia (1972), pp.46-47). Thus assumptions (A.1) - (A.3) and (A.5) are satisfied for most of the circular distributions admitting density. As regards assumption (A.4), it really requires f to be Lipschitz continuous of order $\frac{1}{2}$ at θ_0 and $\theta_0+\pi$ (with Lipschitz continuity being defined suitably when $\theta_0 = 0$) . A sufficient condition that implies (A.4) is the following :

(A.4)' f is differentiable both from left and right at both θ_0 and $\theta_0+\pi$, if $0 < \theta_0 < \pi$. If, however, $\theta_0 = 0$, f is differentiable from the right at 0, from the left at 2π and both from left and right at π.

Observe that it is easier to check assumption (A.4)' mathematically than to check assumption (A.4).

It should be mentioned that the circular distribution obtained in Chapter 2 satisfies (A.1), (A.2), (A.3), (A.4)' and (A.5). This fact will be used in the next chapter.

<u>Remark 5.6.3</u> In Ducharme and Milasevic (1987), henceforth abbreviated to DM, a result similar to Theorem 5.6.1 may be found. However, it appears that there are some mistakes and gaps in their approach. In what follows, we discuss this briefly.

DM restrict their attention to symmetric and unimodal circular distributions, and assume without loss of generality that the population c-median $M(\mu$, in the notation of DM) lies in $[0,\pi)$. However, it is a mistake to take the sample c-median $M_n(\mu_n$, in the notation of DM) also in $[0,\pi)$ since such a strategy does not take care of our relation (4.1.2) and uses prior knowledge about the population c-median. The difficulty becomes more clear when $M = 0$. In this situation, with a high probability one would get samples with M_n close to 2π, as in Figure 5.6.1, whereas it is suggested in DM that for

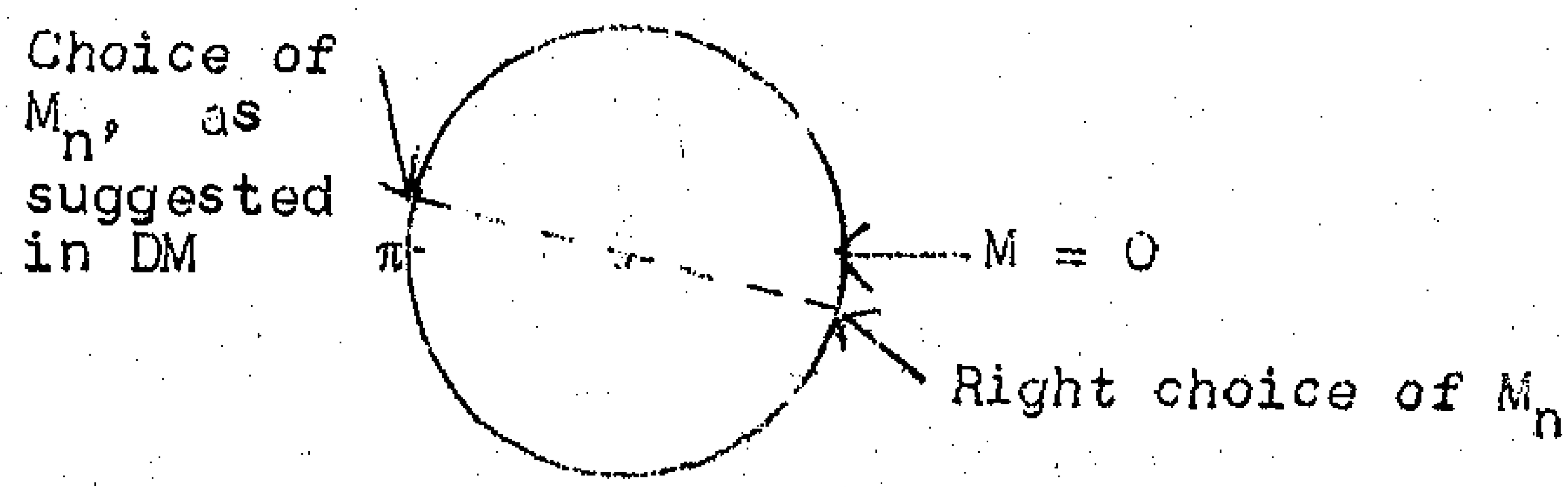


Fig. 5.6.1

such samples the sample c-median should be so taken as to lie very near to π . As a consequence, it is not true of this definition of sample c-median that strong consistency can be established along the lines as in Pollard (1984, p.7). It should also be pointed out that the first approximation in Section 2 of DM lacks rigorous justification and hence so also do the subsequent steps leading to asymptotic normality of M_n .

<u>Remark 5.6.4</u> It may be possible to prove Theorem 5.5.2 (and hence Theorem 5.6.1) by using techniques involving empirical process and quantile process.

CHAPTER 6

ON THE ASYMPTOTIC EFFICIENCY OF SAMPLE CIRCULAR MEDIAN

6.1 Introduction

In Chapter 2, we obtained a circularly symmetric directional distribution by showing that in the class of circularly symmetric directional distributions on circle it is the only distribution for which the sample circular median is a maximum likelihood estimate of the location parameter. In this chapter, we demonstrate asymptotic efficiency of sample circular median as an estimate of the location parameter of this distribution.

The modern study of efficiency in estimation theory or what would perhaps be called today first order asymptotic efficiency, began with the seminal paper of LeCam (1953). For a beautiful proof of LeCam's result on the usual lower bound holding almost everywhere see Bahadur (1964). Later, two routes have been followed in the study : one due to Rao(1961,1962, 1963) confined attention to estimates which are asymptotically normal uniformly on compacts, the other due to LeCam (1960) and Hájek (1972) introduced the notion of local asymptotic minimaxity. In Hájek (1970) and Inagaki (1970) one has a representation of the asymptotic distribution of an estimate as a convolution, which also may be used to prove asymptotic efficiency of an estimate. Other important papers in this area are LeCam (1972), Wolfowitz (1965), and Millar (1983). A lucid account of this development of Hájek-LeCam theory of efficient estimation may be found in Ghosh (1985).

It should be mentioned that much of the work on efficiency was prompted by an attempt to understand how well the maximum likelihood estimate performs as the sample size n goes to infinity. It is, therefore, worthwhile to study sample circular median as an estimate of the location parameter of the circular distribution obtained in Chapter 2 from the point of view of asymptotic efficiency. In Section 6.2, we establish a few facts about our model and distribution of the corresponding sample circular median. In Section 6.3, a convolution theorem similar to the one due to Hajek (1970) and Inagaki (1970) that characterizes the limiting distributions of regular estimates is obtained. In Section 6.3, the main results are stated.

This chapter is a revised version of Purkayastha (1991b).

6.2 Some facts about the model and the distribution of the corresponding sample circular median

We obtained in Chapter 2, the following distribution on S^1 :

$$p(\underline{x};\underline{\theta}) = \frac{\lambda}{2(1-e^{-\lambda\pi})} e^{-\lambda \cos^{-1}(\underline{x}'\underline{\theta})}, \quad \underline{x} \in S^1, \quad \underline{\theta} \in S^1, \qquad \ldots(6.2.1)$$

and λ is a positive constant. However, if we identify S^1 and $[0,2\pi)$ through the identification mentioned in Section 4.2, this distribution induces the following distribution over $[0,2\pi)$:

$$p(x;\theta) = \frac{\lambda}{2(1-e^{-\lambda\pi})} e^{-\lambda d(x,\theta)}, \quad x \in [0,2\pi), \quad \theta \in [0,2\pi), \qquad \ldots(6.2.2)$$

λ is a positive constant, and $d(x,\theta) = \pi-\big|\pi-|x-\theta|\big|$. In consonance with the terminology used for circular distributions, θ is called the location parameter of the distribution in (6.2.2). Throughout the rest of the chapter, $p(x;\theta)$ will refer to the density in (6.2.2) above.

Consider now a sequence of independent and identically distributed circular random variables $\{X_i : i \geq 1\}$ of which n are observed. The common density is $p(x;\theta)$ (λ known). Denote the sample circular median of $X_1,\ldots,X_n$ by $M_n(X_1,\ldots,X_n)$, or M_n, in short.

Observe now that the density $p(x;\theta)$ is circularly symmetric about θ, and moreover it is unimodal with mode at θ. It, therefore, follows that θ is the unique c-median of X_1 (vide Mardia (1972), pp.46-47). It should also be observed that $p(x;\theta)$ is differentiable both from left and right at both θ and θ^*, where θ^* is the antipodal point of θ (see (4.1.4) for definition). Thus in view of Remark 5.6.2, assumptions (A.1) - (A.5) of Section 5.2 are satisfied by the distribution function of X_1.

The first result of this section studies the effect of translation on the density $p(x;\theta)$. The translations under consideration are the bijections on $[0,2\pi)$ corresponding to the 2x2 orthogonal matrices. Recall from the discussion preceding Theorem 4.4.3 that we denoted the bijections corresponding to the 2x2 orthogonal matrices with determinant +1 (respectively, -1) by T_α (respectively, T_α^*). See (4.4.15.1) and (4.4.15.2) for the rigorous definitions of T_α and T_α^*.

<u>Lemma 6.2.1</u> (a) $T_\alpha(X_1)$ has density $p(x;T_\alpha(\theta))$ for every $0 \leq \alpha < 2\pi$.

(b) $T_\alpha^*(X_1)$ has density $p(x;T_\alpha^*(\theta))$ for every $0 \leq \alpha < 2\pi$.

<u>Proof</u>. We provide the proof for part (a) only, the proof for part (b) is essentially similar.

Observe that

$$T_\alpha(x) = x+\alpha \qquad \text{if } 0 \leq x < 2\pi-\alpha ,$$
$$= x+\alpha-2\pi \qquad \text{if } 2\pi-\alpha \leq x < 2\pi .$$

Moreover, from (4.4.16) we have, $d(x,\theta) = d(T_\alpha(x), T_\alpha(\theta))$ for every x. In view of these facts, our assertion follows immediately.

Denote now by P_θ, the probability distribution on $[0,2\pi)$, induced by the density $p(x;\theta)$. The n-fold product of P_θ is denoted by P_θ^n. Fix $\theta_0 \in [0,2\pi)$. The next result of this section establishes local asymptotic normality of the family $\{P_\theta^n : \theta \in \textcircled{H}\}$ at θ_0, where $\textcircled{H} = [0,2\pi)$. However, since there is no standard notion of local asymptotic normality for distributions over circle, our notion essentially adapts the one for distributions over $\mathbb{R}$ to the circle. The following notation is introduced towards this end : for $\theta \in [0,2\pi)$ and $-\pi < a < \pi$, we define

$$(\theta + a) = \theta + a \pmod{2\pi} .$$

We now have the following result.

<u>Lemma 6.2.2</u> Fix $\theta_0 \in [0,2\pi)$. Then

$$\log \frac{dP^n_{(\theta_0 + \frac{h}{\sqrt{n}})}}{dP^n_{\theta_0}} = h\,\Delta_n(\theta_0) - \frac{h^2\lambda^2}{2} + \Psi_n(\theta_0;h) ,$$

where $\dfrac{dP^n_{(\theta_0+\frac{h}{\sqrt{n}})}}{dP^n_{\theta_0}}$ is the density of the absolutely

continuous part of $P^n_{(\theta_0+\frac{h}{\sqrt{n}})}$ with respect to $P^n_{\theta_0}$,

$$\mathcal{L}(\Delta_n(\theta_o) | P^n_{\theta_o}) \Rightarrow N(0,\lambda^2),$$

and

$$\Psi_n(\theta_o;h) \xrightarrow{P^n_{\theta_o}} 0 \text{ for all } h.$$

Proof. We provide the proof for $0 < \theta_o < \pi$ only. The proof for other values of θ_o is essentially same. Assume, without loss of generality, further than $h > 0$.

Write $\theta_n = (\theta_o + \frac{h}{\sqrt{n}})$, and observe that $\theta_n = \theta_o + \frac{h}{\sqrt{n}}$ for all n, sufficiently large. Also, $\theta_o^* = \theta_o + \pi$ and $\theta_n^* = \theta_n + \pi$, where for any $x \in [0,2\pi)$, x^* denotes its antipodal point (see (4.1.4) for definition). Observe now that

$$\log \frac{dP^n_{\theta_n}}{dP^n_{\theta_o}} = -\lambda[\sum_1^n d(x_i,\theta_n) - \sum_1^n d(x_i,\theta_o)], \qquad \ldots(6.2.3)$$

Observe further that

$$-\lambda \sum_1^n [d(X_i,\theta_n) - d(X_i,\theta_o)]$$

$$= 2\lambda \sum_1^n (X_i - \theta_n)\, 1_{\{\theta_o < X_i \leq \theta_n\}} + 2\lambda \sum_1^n (\theta_n^* - X_i)\, 1_{\{\theta_o^* < X_i \leq \theta_n^*\}}$$

$$+ \frac{\lambda h}{\sqrt{n}} \sum_1^n 1_{\{\theta_o < X_i \leq \theta_o^*\}} - \frac{\lambda h}{\sqrt{n}} \sum_1^n 1_{\{0 \leq X_i \leq \theta_o\} \cup \{\theta_o^* < X_i < 2\pi\}},$$

$$\ldots(6.2.4)$$

for all n, sufficiently large. Define now

$$\phi(X_i;\theta_o) = \frac{\lambda}{2} \quad \text{if} \quad \theta_o < X_i \leq \theta_o^*,$$

$$= -\frac{\lambda}{2} \quad \text{otherwise},$$

$$U_{in} = (X_i - \theta_n)\, 1_{\{\theta_o < X_i \leq \theta_n\}}, \quad 1 \leq i \leq n,$$

and

$$V_{in} = (\theta_n^* - X_i)\, 1_{\{\theta_o^* < X_i \leq \theta_n^*\}}, \quad 1 \leq i \leq n,$$

where n is sufficiently large. Therefore, we obtain from (6.2.3)

$$-\lambda \sum_1^n [d(X_i,\theta_n) - d(X_i,\theta_o)]$$

$$= 2\lambda \sum_1^n U_{in} + 2\lambda \sum_1^n V_{in} + \frac{2h}{\sqrt{n}} \sum_1^n \phi(X_i;\theta_o) \,. \qquad \ldots(6.2.5)$$

It is now easy to verify that

$$E_{\theta_o}(U_{in}) = -\frac{1}{2\lambda(1-e^{-\lambda\pi})}\left[e^{-\frac{\lambda h}{\sqrt{n}}} - 1 + \frac{\lambda h}{\sqrt{n}}\right], \qquad \ldots(6.2.6.1)$$

$$E_{\theta_o}(U_{in}^2) = \frac{1}{2\lambda^2(1-e^{-\lambda\pi})}\left[e^{-\frac{\lambda h}{\sqrt{n}}}\left[-2e^{-\frac{\lambda h}{\sqrt{n}}} + \frac{h^2\lambda^2}{n} - \frac{2\lambda h}{\sqrt{n}} + 2\right]\right.,$$

$$\ldots(6.2.6.2)$$

$$E_{\theta_o}(V_{in}) = \frac{1}{2\lambda(1-e^{-\lambda\pi})}\left[e^{\frac{\lambda h}{\sqrt{n}}} - 1 - \frac{\lambda h}{\sqrt{n}}\right], \qquad \ldots(6.2.6.3)$$

$$E_{\theta_o}(V_{in}^2) = \frac{1}{2\lambda^2(1-e^{-\lambda\pi})}\left[2e^{\frac{\lambda h}{\sqrt{n}}} - 2 - \frac{2\lambda h}{\sqrt{n}} - \frac{\lambda^2 h^2}{n}\right], \qquad \ldots(6.2.6.4)$$

$$E_{\theta_o}(\phi(X_i;\theta_o)) = 0, \qquad \ldots(6.2.7.1)$$

and

$$E_{\theta_o}(\phi^2(X_i;\theta_o)) = \frac{\lambda^2}{4}. \qquad \ldots(6.2.7.2)$$

Consequently, it follows that

$$\lim_{n\to\infty}[nE_{\theta_o}(U_{in})] = \frac{-\lambda h^2}{2(1-e^{-\lambda\pi})}, \quad \lim_{n\to\infty}[nE_{\theta_o}(V_{in})] = \frac{\lambda h^2 e^{-\lambda\pi}}{2(1-e^{-\lambda\pi})},$$

and $\ldots(6.2.8)$

$$\lim_{n\to\infty}[nE_{\theta_o}(U_{in}^2)] = 0, \qquad \lim_{n\to\infty}[nE_{\theta_o}(V_{in}^2)] = 0.$$

Define now

$$U_n = \sum_1^n U_{in} \quad \text{and} \quad V_n = \sum_1^n V_{in}, \qquad \ldots(6.2.9.1)$$

and

$$\Delta_n(\theta_o) = \frac{2}{\sqrt{n}}\sum_1^n \phi(X_i;\theta_o). \qquad \ldots(6.2.9.2)$$

It is easy to see from (6.2.6.1) - (6.2.6.4) and (6.2.8) that

$$2\lambda[U_n + V_n] \xrightarrow{P_{\theta_o}^n} -\frac{h^2\lambda^2}{2}. \qquad \ldots(6.2.10.1)$$

Moreover, it follows from (6.2.7.1) and (6.2.7.2) that

$$\mathcal{L}(\Delta_n(\theta_o)|P^n_{\theta_o}) \Rightarrow N(0,\lambda^2) . \qquad \ldots(6.2.10.2)$$

We now define

$$\psi_n(\theta_o;h) = 2\lambda[U_n + V_n] + \frac{h^2\lambda^2}{2} , \qquad \ldots(6.2.11)$$

and observe that the proof follows from (6.2.5), (6.2.10.1), (6.2.10.2) and (6.2.11).

<u>Lemma 6.2.3</u> Fix $\theta_o \varepsilon [0,2\pi)$. Then for any $h \varepsilon IR$, $\{P^n_{(\theta_o+\frac{h}{\sqrt{n}})} : n \geq 1\}$ and $\{P^n_{\theta_o} : n \geq 1\}$ are contiguous.

<u>Proof.</u> It follows from the preceding lemma that

$$\mathcal{L}\left(\log \frac{dP^n_{(\theta_o+\frac{h}{\sqrt{n}})}}{dP^n_{\theta_o}} \middle| P^n_{\theta_o}\right) \Rightarrow N\left(-\frac{h^2\lambda^2}{2} , h^2\lambda^2\right) = \mathcal{L}(\Lambda), \text{ say.}$$

However, since

$$E(e^{\Lambda}) = 1 ,$$

our assertion follows immediately (vide Roussas (1972), p.11).

The next result gives the effect of translation on the distribution of $M_n(X_1,\ldots,X_n)$.

<u>Lemma 6.2.4</u> For every $0 \leq \alpha < 2\pi$,

(a) $T_\alpha(M_n(X_1,\ldots,X_n)) \stackrel{d}{=} M_n(Y_1,\ldots,Y_n)$, where $Y_1,\ldots,Y_n$ are independent and identically distributed circular random variables with Y_1 having density $p(\cdot,T_\alpha(\Theta))$,

$$\text{(b)}\quad T^*_\alpha(M_n(X_1,\ldots,X_n)) \overset{d}{=} M_n(Y_1,\ldots,Y_n) ,$$

where $Y_1,\ldots,Y_n$ are independent and identically distributed circular random variables with Y_1 having density $p(\cdot,T^*_\alpha(\Theta))$.

Proof. The proof is an immediate consequence of Theorem 4.4.3 and Lemma 6.2.1 . We omit the details.

Before we state the next result, we make the following definition. In this definition, T_n, depending on $X_1,\ldots,X_n$, is an estimate of Θ .

Definition 6.2.1 We define a sequence of random variables $\{Z_n(T_n,\Theta) : n \geq 1\}$ as follows :

$$\begin{aligned} Z_n(T_n,\Theta) &= d(T_n,\Theta) && \text{if } \Theta < T_n \leq \Theta + \pi , \\ &= -d(T_n,\Theta) && \text{otherwise,} \end{aligned}$$

whenever $0 \leq \Theta < \pi$, and

$$\begin{aligned} Z_n(T_n,\Theta) &= -d(T_n,\Theta) && \text{if } \Theta - \pi < T_n \leq \Theta , \\ &= d(T_n,\Theta) && \text{otherwise ,} \end{aligned}$$

whenever $\pi \leq \Theta < 2\pi$ (cf. Definition 5.5.1).

Observe now that a natural choice for T_n is M_n, the sample c-median, since it is a maximum likelihood estimate of θ. We now have the following fact about $Z_n(M_n,\theta)$.

Lemma 6.2.5 The distribution of $Z_n(M_n,\theta)$ under P_θ^n does not depend on θ.

Proof. Choose $\theta, \theta' \varepsilon [0,2\pi)$ with $\theta \neq \theta'$. Choose $\alpha \varepsilon [0,2\pi)$ such that $T_\alpha(\theta) = \theta'$; the fact that such a choice is possible is easy to see. Let $X_1,\ldots,X_n$ be independent and identically distributed circular random variables with X_1 having density $p(x;\theta)$.

Observe now in view of the definition of $Z_n(M_n,\theta)$ and (4.4.16) that

$$Z_n(M_n(X_1,\ldots,X_n),\theta) = Z_n(T_\alpha(M_n(X_1,\ldots,X_n)),\ T_\alpha(\theta)). \qquad \ldots(6.2.12)$$

Our proof is now an immediate consequence of (6.2.12), Theorem 4.4.3 and Lemma 6.2.4. We omit the details.

Lemma 6.2.6 $\mathcal{L}(\sqrt{n}\ Z_n(M_n,\theta) \mid P_\theta^n) \Rightarrow N(0, \frac{1}{\lambda^2})$.

Proof. We have stated at the beginning of this chapter that assumptions (A.1) - (A.5) of Section 5.2 are satisfied by the distribution function of X_1. Observe, moreover, that

$$p(\theta,\theta) - p(\theta^*,\theta) = \frac{\lambda}{2} \qquad \ldots(6.2.13)$$

The proof is now an immediate consequence of Theorem 5.6.1 and (6.2.13).

6.3 A convolution theorem

In this section, we prove a result which is similar to the Hájek-Inagaki convolution theorem (Hájek (1970), Inagaki (1970)).

Suppose $T_n(X_1,\ldots,X_n)$, or T_n in short, is an estimate of θ based on $X_1,\ldots,X_n$. Suppose, moreover, that T_n satisfies

$$\mathcal{L}(\sqrt{n}\, Z_n(T_n,(\theta + \frac{h}{\sqrt{n}})) \,|\, P^n_{(\theta + \frac{h}{\sqrt{n}})}) \Rightarrow G_\theta, \ldots(6.3.1)$$

where G_θ is some probability distribution over $\mathbb{R}$, not depending on h. We now have the following result that characterizes G_θ.

Theorem 6.3.1 $G_\theta = N(0,\frac{1}{\lambda^2}) * H_\theta$,

where H_θ is some probability distribution over $\mathbb{R}$, and '*' denotes convolution.

Proof. Write $T_n^* = \sqrt{n}\, Z_n(T_n,\theta)$. Then (6.3.1) implies, with $h = 0$,

$$\mathcal{L}(T_n^* \,|\, P^n_\theta) \Rightarrow G_\theta .$$

Choose a subsequence $\{n_k\}$ of $\{n\}$ such that

$$\mathcal{L}((T^*_{n_k}, \Delta_{n_k}(\theta)) | P^n_\theta) \Rightarrow G_{T,\Delta} , \qquad \ldots(6.3.2)$$

where $\Delta_n(\theta)$ is given by Lemma 6.2.2 (write θ in place of θ_o) . Since the marginals of $G_{T,\Delta}$ are probability distributions, $G_{T,\Delta}$ also is a probability distribution. Write now in Lemma 6.2.2, $\Lambda_n(h)$ in place of $\log \frac{dP^n_{(\theta + \frac{h}{\sqrt{n}})}}{dP^n_\theta}$. It then follows from Lemma 6.2.2 and (6.3.2) above that

$$\mathcal{L}((T^*_{n_k}, \Lambda_{n_k}(h)) \mid P^{n_k}_\theta) \Rightarrow G_{T,\Lambda} , \qquad \ldots(6.3.3)$$

where $\Lambda = h\Delta - \frac{h^2\lambda^2}{2}$. However, Lemma 6.2.3 implies that $\{P^n_{(\theta + \frac{h}{\sqrt{n}})} : n \geq 1\}$ and $\{P^n_\theta : n \geq 1\}$ are contiguous.

Combined with (6.3.3) above, this implies

$$\mathcal{L}((T^*_{n_k}, \Lambda_{n_k}(h)) \mid P^{n_k}_{(\theta + \frac{h}{\sqrt{n_k}})}) \Rightarrow G'_{T,\Lambda} , \qquad \ldots(6.3.4)$$

where $\frac{dG'(t,\lambda)}{dG(t,\lambda)} = e^\lambda$ (vide Roussas (1972), pp.33-34)).

Observe now that (6.3.1) implies, with $h = 0$,

$$Z_n(T_n, \theta) \xrightarrow{P_\theta^n} 0 ,$$

which implies, in turn,

$$\lim_{n\to\infty} P_\theta^n \left(\left\{ Z_n(T_n, (\theta + \frac{h}{\sqrt{n}})) - Z_n(T_n, \theta) + \frac{h}{\sqrt{n}} = 0 \right\} \right) = 0 .$$

From this, we obtain in view of Lemma 6.2.3,

$$\lim_{n\to\infty} P^n_{(\theta + \frac{h}{\sqrt{n}})} \left(\left\{ \sqrt{n}\, Z_n(T_n, (\theta + \frac{h}{\sqrt{n}})) - \sqrt{n}\, Z_n(T_n, \theta) + h = 0 \right\} \right) = 0. \qquad \ldots(6.3.5)$$

Taken with (6.3.1), this implies

$$\mathcal{L}\left(\sqrt{n}\, Z_n(T_n, \theta) - h \;\middle|\; P^n_{(\theta + \frac{h}{\sqrt{n}})}\right) \Rightarrow G_\theta ,$$

that is,

$$\int e^{is(\sqrt{n}\, Z_n(T_n, \theta) - h)}\, dP^n_{(\theta + \frac{h}{\sqrt{n}})} \longrightarrow f(s), \qquad \ldots(6.3.6)$$

where $f(s)$ is the characteristic function of G_θ . On the other hand,

$$\int e^{is(\sqrt{n_k}\, Z_{n_k}(T_{n_k},\Theta)-h)}\, dP^{n_k}_{(\Theta + \frac{h}{\sqrt{n_k}})}$$

$$= e^{-ish} \int e^{isT^*_{n_k}}\, dP^{n}_{(\Theta + \frac{h}{\sqrt{n_k}})}$$

$$\longrightarrow E_{G'}(e^{isT-ish}) , \quad \text{as} \quad k \longrightarrow \infty , \qquad \ldots(6.3.7)$$

the convergence being a consequence of (6.3.4). From (6.3.4) it also follows that

$$E_{G'}(e^{isT-ish}) = E_G(e^{isT-ish+\Lambda}) . \qquad \ldots(6.3.8)$$

Moreover, it follows from (6.3.3) that

$$E_G(e^{isT-ish+\Lambda}) = E_G(e^{isT-ish+h\Delta-\frac{h^2\lambda^2}{2}}). \qquad \ldots(6.3.9)$$

We now obtain from (6.3.6)-(6.3.9) above the following :

$$E_G(e^{isT-ish+h\Delta-\frac{h^2\lambda^2}{2}}) = f(s) \quad \forall \quad h \in \mathbb{R} . \qquad \ldots(6.3.10)$$

However, it is known that if two analytic functions agree on $\mathbb{R}$, they agree on $\mathbb{C}$ also. So, replacing h by ih on both sides of (6.3.10) above, we obtain

$$E_G(e^{isT+sh+ih\Delta+\frac{h^2\lambda^2}{2}}) = f(s) \quad \forall \quad h \in \mathbb{R}. \qquad \ldots(6.3.11)$$

Choose now $h = -\frac{s}{\lambda^2}$ on the left hand side of (6.3.11) to obtain

$$e^{-\frac{s^2}{2\lambda^2}} E_G(e^{isT-\frac{is\Delta}{\lambda^2}}) = f(s) \quad \forall \; s \in \mathbb{R}.$$

Therefore, $f(s)$ is the product of characteristic functions of $T - \frac{\Delta}{\lambda^2}$ and $N(0,\frac{1}{\lambda^2})$. Hence,

$$G_\Theta = N(0, \frac{1}{\lambda^2}) * H_\Theta ,$$

for some probability distribution H_Θ on $\mathbb{R}$, as asserted.

Remark 6.3.1 Our proof of Theorem 6.3.1 essentially follows Bickel's proof of Hájek - Inagaki convolution theorem (vide Roussas (1972), pp.136-138). It is easy to see that if $0 < \Theta < 2\pi$,

$$(\Theta + \frac{h}{\sqrt{n}}) = \Theta + \frac{h}{\sqrt{n}}$$

for all n, sufficiently large, and

$$\lim_{n\to\infty} P^n_\Theta (\{ Z_n(T_n,\Theta) - (T_n - \Theta) = 0 \}) = 0 .$$

Therefore, our assertion now follows immediately from the Hájek - Inagaki convolution theorem if $0 < \Theta < 2\pi$. If, however, $\Theta = 0$, some more care is needed to adapt Bickel's proof. The convergence in (6.3.5) serves this purpose.

Remark 6.3.2 It should be mentioned that (6.3.1) holds true if we choose $T_n = M_n$, the sample c-median. This follows immediately from Lemma 6.2.5 and Lemma 6.2.6.

6.4 The main results

In this section, we establish asymptotic efficiency of sample c-median M_n within the class of estimators T_n satisfying (6.3.1). We shall only state the results, the proofs follow by using standard arguments involving Anderson's (1955) inequality and Theorem 6.3.1.

Theorem 6.4.1 Suppose A is a symmetric (i.e., $A = -A$) convex subset of $\mathbb{R}$. Then

$$\lim_{n\to\infty} P_\theta(\{\sqrt{n}\, Z_n(M_n,\theta) \in A\}) \geq \overline{\lim_{n\to\infty}} P_\theta(\{\sqrt{n}\, Z_n(T_n,\theta) \in A\}),$$

where T_n is an estimate satisfying (6.3.1).

Theorem 6.4.2 Suppose $L : \mathbb{R} \to [0,\infty)$ is a function satisfying

(i) $L(0) = 0$,

(ii) $L(x) = L(-x)$ for every $x \in \mathbb{R}$,

(iii) $\{x : L(x) < c\}$ is a convex subset of $\mathbb{R}$ for every $c > 0$,

and

(iv) L is bounded.

Then

$$\underline{\lim_{n\to\infty}}\, E_\theta[L(\sqrt{n}\, Z_n(T_n,\theta))] \geq \lim_{n\to\infty} E_\theta\, [L(\sqrt{n}\, Z_n(M_n,\theta))],$$

where T_n is an estimate satisfying (6.3.1).

Remark 6.4.1. If we redefine $Z_n(T_n,\theta)$ by using the metric d_1 in (5.6.3) and not d, the assertions of Lemma 6.2.5, Theorem 6.3.1, Theorem 6.4.1 and Theorem 6.4.2 hold true (cf. Remark 5.6.1).

REFERENCES

Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Am. Math. Soc., 6, 170-176.

Bahadur, R.R. (1964). On Fisher's bound for asymptotic variances. Ann. Math. Statist., 35, 1545-1552.

Bahadur, R.R. (1966). A note on quantiles in large samples. Ann. Math. Statist., 37, 577-580.

Batschelet, E. (1981). Circular Statistics in Biology. Academic Press, London.

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

Borovkov, A.A. and Utev, S.A. (1983). On an inequality and a related characterization of the normal distribution. Theor. Probab. Appl., 28, 219-228.

Cabrera, J. and Watson, G.S. (1990). On a spherical median related distribution. Commun. Statist. — Theory Meth., 19, 1973-1986.

Cacoullos, T. (1982). On upper and lower bounds for the variance of a function of a random variable. Ann. Probab., 10, 799-809.

Cacoullos, T. (1989). Dual Poincaré-type inequalities via the Cramér-Rao and the Cauchy-Schwarz inequalities and related characterizations. In Statistical Data Analysis and Inference, (Y. Dodge, ed.), 239-250. North Holland, Amsterdam.

Cacoullos, T. and Papathanasiou, V. (1985). On upper bounds for the variance of functions of random variables. Statist. Probab. Lett., 3, 175-184.

Cacoullos, T. and Papathanasiou, V. (1986). Bounds for the variance of functions of random variables by orthogonal polynomials and Bhattacharya bounds. Statist. Probab. Lett., 4, 21-23.

Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. Statist. Probab. Lett., 7, 351-356.

Cacoullos, T. and Papathanasiou, V. (1990). Another proof of the central limit theorem by variance bounds. Preprint, University of Athens.

Chen, L.H.Y. (1982). An inequality for the multivariate normal distribution. Jour. Mult. Anal., 12, 306-315.

Chen, L.H.Y. (1985). Poincaré-type inequalities via stochastic integrals. Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 69, 251-277.

Chen, L.H.Y. (1988). The central limit theorem and Poincaré-type inequalities. Ann. Probab., 16, 300-304.

Chen, L.H.Y. and Lou, J.H. (1987). Characterization of probability distributions by Poincaré-type inequalities. Ann. Inst. Henri Poincaré, 23, 91-110.

Chernoff, H. (1981). A note on an inequality involving the normal distribution. Ann. Probab., 9, 533-535.

deHaan, L. and Taconis-Haantjes, E. (1979). On Bahadur's representations of sample quantiles. Ann. Inst. Statist. Math., 31, Pt.A, 299-308.

Dharmadhikari, S. and Joag-dev, K. (1988). Unimodality, Convexity, and Applications. Academic Press, Boston.

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

Ducharme, G.R. and Milasevic, P. (1987). Some asymptotic properties of the circular median. Commun. Statist. — Theory Meth., 16, 659-664.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist., 27, 642-669.

Edwards, R.E. (1979). Fourier Series. A Modern Introduction. Volume 1. Second Edition. Springer-Verlag, New York.

Fisher, N.I. (1985). Spherical medians. Jour. Royal Statist. Soc., B 47, 342-348.

Fisher, N.I., Lewis, T. and Embleton, B.J.J. (1987). Statistical Analysis of Spherical Data. Cambridge University Press.

Gauss, C.F. (1809). Theoria Motus Corporum Coelestium. Perthes, Hamburg. Translation reprinted as Theory of Motions of the Heavenly Bodies Moving about the Sun in Conic Sections. Dover, New York, 1963.

Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist., 42, 1957-1961.

Ghosh, J.K. (1985). Efficiency of estimates - Part 1. Sankhyā, A 47, 310-325.

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, 91-92.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 14, 323-330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. Proc. 6th Berkeley Symp., 1, (L.M. LeCam et al. eds.), 175-194. University of California Press, Berkeley and Los Angeles.

Halmos, P.R. (1958). Finite Dimensional Vector Spaces. Van Norstrand, New York.

Halmos, P.R. (1978). A Hilbert Space Problem Book. Narosa Publishing House, New Delhi.

Hu, C.Y. (1986). An inequality involving the rectangular distribution. Bull. Inst. Math., Acad. Sinica, 4, 21-23.

Hwang, C.R. and Sheu S. J. (1987). A generalization of Chernoff inequality via stochastic analysis. Probab. Th. Rel. Fields, 75, 149-157.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. Ann. Inst. Statist. Math., 22, 1-13.

Johnson, R.W. (1988). Smoothed regression estimates with Pearson noise. Tech. Report 1231, Naval Ocean Systems Center.

Jupp, P.E. and Mardia, K.V. (1979). Maximum likelihood estimation for the matrix von Mises-Fisher and Bingham distributions. Ann. Statist., 7, 599-606.

Jupp, P.E. and Mardia, K.V. (1989). A unified view of the theory of directional statistics, 1975-1988. Int. Statist. Rev., 57, 261-294.

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

Khatri, C.G. and Mardia, K.V. (1977). The von Mises-Fisher distribution in orientation statistics. Jour. Royal Statist. Soc., B 39, 95-106.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. Ann. Math. Statist., 38, 1323-1342.

Klaassen, C.A.J. (1985). On an inequality of Chernoff. Ann. Probab., 13, 966-974.

Klaassen, C.A.J. (1988). Personal communication.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood and Bayes estimates. Univ. Calif. Publ. Statist., 1, 277-330.

LeCam, L. (1960). Locally asymptotically normal families of distributions. Univ. Calif. Publ. Statist., 3, 37-98.

LeCam, L. (1972). Limits of experiments. Proc. 6th Berkeley Symp., 1, (L.M. LeCam et al. eds.), 245-261. University of California Press, Berkeley and Los Angeles.

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

Mardia, K.V. (1975a). Statistics of directional data (with discussion). Jour. Royal Statist. Soc., B 37, 349-393.

Mardia, K.V. (1975b). Characterization of directional distribution. In Statistical Distributions in Scientific Work, 3, (G.P. Patil et al. eds.), 365-385. Reidel, Dordrecht.

Mardia, K.V. (1988). Directional data analysis : an overview. Jour. Appl. Statist., 15, 115-122.

Mardia, K.V. and Khatri, C.G. (1977). Uniform distribution on a Stiefel manifold. Jour. Mult. Anal., 7, 468-473.

Millar, P.W. (1983). The minimax principle in asymptotic theory. Ecole d'Ele de Probabilities de Saint-Flour XI (P.L. Hennequin, ed.), Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Papathanasiou, V. (1989). Variance bounds by a generalization of the Cauchy-Schwarz inequality. Statist. Probab. Lett., 7, 29-33.

Papathanasiou, V. (1990). Characterizations of power series and the factorial series distributions. To appear in Sankhyā, Ser. A.

Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlag, New York.

Prakasa Rao, B.L.S. and Sreehari, M. (1986). Another characterization of multivariate normal distribution. Statist. Probab. Lett., 4, 209-210.

Prakasa Rao, B.L.S. and Sreehari, M. (1987). On a characterization of Poisson distribution through inequalities of Chernoff-type. Aust. Jour. Statist., 29, 38-41.

Prentics, M.J. (1982). Antipodally symmetric distributions for orientation statistics. Jour. Statist. Plan. Inf., 6, 205-214.

Prentics, M.J. (1986). Orientation statistics without parametric assumptions. Jour. Royal Statist. Soc., B 48, 214-222.

Purkayastha, S. (1990a). A definition of sample circular median. Tech. Report 14/90, Stat. Math. Div., Indian Statistical Institute, Calcutta (Submitted for publication).

Purkayastha, S. (1990b). A Bahadur-type representation of circular median and an application. Tech. Report 15/90, Stat. Math. Div., Indian Statistical Institute, Calcutta (Submitted for publication).

Purkayastha, S. (1991a). A rotationally symmetric directional distribution : obtained through maximum likelihood characterization. Sankhyā, A 53, 70-83.

Purkayastha, S. (1991b). On the asymptotic efficiency of sample circular median (To be submitted).

Purkayastha, S. and Bhandari, S.K. (1988). Characterization of uniform distributions by inequalities of Chernoff-type. Tech. Report 3/88, Stat. Math. Div., Indian Statistical Institute, Calcutta.

Purkayastha, S. and Bhandari, S.K. (1990). Characterization of uniform distributions by inequalities of Chernoff-type. Sankhyā, A 52, 376-382.

Purkayastha, S. and Mukerjee, R. (1989). Maximum likelihood characterization of the von Mises-Fisher matrix distribution. To appear in Sankhyā, Ser. A.

Rao, C.R. (1961). Asymptotic efficiency and limiting information. Proc. 4th Berkeley Symp., 1, (J. Neyman, ed.), 531-545. University of California Press, Berkeley.

Rao, C.R. (1962). Efficient estimates and optimum inference procedures in large samples. Jour. Royal Statist. Soc., B 24, 46-72.

Rao, C.R. (1963). Criteria of estimation in large samples. Sankhyā, A 25, 189-206.

Rao, J.S. (1984). Nonparametric methods in directional data analysis. In Handbook of Statistics, 4 : Nonparametric Methods, (P.R. Krishnaiah and P.K. Sen eds.), 755-770. Elsevier, Amsterdam.

Roussas, G.G. (1972). Contiguity of Probability Measures : Applications in Statistics. Cambridge University Press, Cambridge.

Sen, P.K. (1968). Asymptotic normality of sample quantiles for m-dependent processes. Ann. Math. Statist., 39, 1724-1730.

Srivastava, D.K. and Sreehari, M. (1987). Characterization of a family of discrete distributions via a Chernoff-type inequality. Statist. Probab. Lett., 5, 293-294.

Srivastava, D.K. and Sreehari, M. (1990). Characterization results based on variance bounds. Sankhyā, A 52, 297-302.

Sverdrúp-Thygeson, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. Ann. Statist., 6, 141-145.

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

von Mises, R. (1918). Uber die 'Ganzzahligkeit' der Atom-gewicht und verwandte Fragen. Physikal. Z., 19, 490-500.

Watson, G.S. (1983). Statistics on Spheres. University of Arkansas Lecture notes in the Mathematical Sciences, 6, Wiley, New York.

Watson, G.S. (1989). Past, current and future problems with directional data. Draft of Leuven talk.

Wolfowitz, J. (1965). Asymptotic efficiency of the maximum likelihood estimator. Theor. Probab. Appl., 10, 247-260.