

## Spatial median and directional data

BY GILLES R. DUCHARME

*Département de mathématiques et de statistique, Université de Montréal, Montréal, P.Q. H3C 3J7,  
Canada*

AND PHILIP MILASEVIC

*Institut de mathématiques, Université de Fribourg, 1700 Fribourg, Switzerland*

### SUMMARY

We introduce the normalized spatial median as an estimator of location for rotationally symmetric distributions on the hypersphere. We investigate some of its asymptotic properties and use them to obtain confidence regions for the modal direction of a distribution on the hypersphere. These results are then applied to the von Mises-Fisher distribution and to a contamination model. It is seen that the normalized spatial median can perform more efficiently than the normalized mean in presence of outliers.

*Some key words:* Asymptotic relative efficiency; Directional data; Outlier; Rotational symmetry; Spatial median; von Mises-Fisher distribution.

### 1. INTRODUCTION

Statistical analysis of directional data usually requires the identification of a modal direction for the random process generating the data. This modal direction is particularly meaningful when the underlying distribution is rotationally symmetric about a central axis, for then it represents the natural location parameter of the distribution.

Denote by  $\|\cdot\|$  the euclidean norm on  $\mathbb{R}^p$  and let  $\mathcal{R}_p(\theta)$  be the class of all distributions on the unit sphere in  $\mathbb{R}^p$  having a nonzero expectation,  $\mu$ , and a density that is rotationally symmetric about the unit vector  $\theta$ . Since then  $\theta = \mu / \|\mu\|$ , the normalized sample mean  $\hat{\theta} = \bar{X} / \|\bar{X}\|$  is a consistent estimator of  $\theta$ . This estimator is the maximum likelihood estimator of  $\theta$  if the distribution is von Mises-Fisher.

An alternative estimator, denoted here by  $\theta^*$ , is the normalized eigenvector associated with the largest eigenvalue of the sample's second-moment matrix. Although originally intended for axial data,  $\theta^*$  can be used for directional data provided the indeterminacy about the direction can be resolved. When  $p = 3$  and the distribution is von Mises-Fisher the asymptotic relative efficiency of  $\theta^*$  with respect to  $\theta$  is close to one if the distribution is sufficiently concentrated about  $\theta$  (Watson, 1983, p. 186).

In this paper, we propose an estimator of  $\theta$  based on the spatial median. In § 2 we introduce the estimator and in § 3 we prove its consistency, derive its asymptotic distribution and obtain confidence cones for  $\theta$ . In § 4 asymptotic relative efficiency is computed for two types of distributions and finally, in § 5, a real data set is investigated.

### 2. THE SPATIAL MEDIAN

Let  $F$  be a probability distribution on  $\mathbb{R}^p$  such that  $\int \|x\| dF(x) < \infty$ . A point  $\eta$  is a spatial median of  $F$  if it minimizes  $\Phi(\alpha) = \int \|x - \alpha\| dF(x)$ .

It follows from the tightness of  $F$  that such a spatial median always exists. When  $p = 1$ , it is equal to the usual median and thus may not be unique. For  $p > 1$  however, it can be shown that if  $F$  is not concentrated on a line then the spatial median is unique; see also Haldane (1948).

If  $F$  is an atomless distribution on the sphere then  $\|\eta\| < 1$ . Indeed, since  $\|\eta\| \leq 1$  there exists a unit vector  $\eta_0$  and  $\lambda \in [0, 1]$  such that  $\eta = \lambda\eta_0$ . The hypothesis now implies that

$$\int x' \eta_0 / \|x - \lambda\eta_0\| dF(x) = \lambda \int 1 / \|x - \lambda\eta_0\| dF(x)$$

and the result follows because  $|x' \eta_0| < 1$  almost surely with respect to  $F$ . If  $F \in \mathcal{R}_p(\theta)$ , which we shall henceforth suppose, then  $\eta = \pm \|\eta\| \theta$ . We may assume without loss of generality that  $\theta = \eta / \|\eta\|$ .

We now propose as an estimator of  $\theta$  the normalized spatial median  $\tilde{\theta} = \tilde{\eta} / \|\tilde{\eta}\|$ , where  $\tilde{\eta}$  is the spatial median of the empirical distribution function based on a random sample from  $F$ . A FORTRAN program for the computation of  $\tilde{\eta}$  is given by Gower (1974).

Now  $\tilde{\eta}$  is equivariant under the group of rotations. It is not scale invariant but this is of no concern for directional data. Note also that since  $\tilde{\eta}$  minimizes the sum of the norms whereas  $\bar{X}$  minimizes the sum of the squared norms one may expect  $\tilde{\theta}$  to be less sensitive to outliers than  $\hat{\theta}$ ; see §§ 4, 5.

### 3. ASYMPTOTIC PROPERTIES

#### 3.1. Consistency and asymptotic normality

First note that, from the results of § 2, the spatial median  $\eta$  is the unique solution of

$$\int \psi(x - \alpha) dF(x) = 0,$$

where  $\psi(x) = x / \|x\|$ . From Huber (1967) and the results of § 2, it can be shown that  $\tilde{\eta}$  is a strongly consistent estimator of  $\eta$  and that in distribution

$$n^{1/2}(\tilde{\eta} - \eta) \rightarrow N(0, C^{-1} \Lambda C^{-1}), \tag{3.1}$$

with  $\Lambda = E\{\psi(X - \eta)\psi'(X - \eta)\}$  and  $C = E\{[I - \psi(X - \eta)\psi'(X - \eta)] / \|X - \eta\|\}$ , where  $I$  is the  $p \times p$  identity matrix. A little algebra yields

$$C = E\{(1 - s^2) / \|X - \eta\|\} \theta \theta' + E\{(p - 2 + s^2) / \{(p - 1)\|X - \eta\|\}\} (I - \theta \theta'),$$

$$\Lambda = \text{var}(s) \theta \theta' + [1 - \text{var}(s)] / (p - 1) (I - \theta \theta'),$$

where  $s = \theta' \psi(X - \eta)$ . It now follows from (3.1) that  $n^{1/2}(\tilde{\eta} - \eta) \rightarrow N(0, V)$  in distribution with

$$V = \rho \theta \theta' + \beta (I - \theta \theta'), \quad \rho = \text{var}(s) / [E\{(1 - s^2) / \|X - \eta\|\}]^2,$$

$$\beta = \{1 - \text{var}(s)\} (p - 1) / [E\{(p - 2 + s^2) / \|X - \eta\|\}]^2.$$

Finally, expanding  $\psi$  in a neighbourhood of  $\eta$  and using the  $\delta$ -theorem, we conclude that in distribution

$$n^{1/2}(\tilde{\theta} - \theta) \rightarrow N\{0, (\beta / \|\eta\|^2)(I - \theta \theta')\}. \tag{3.2}$$

#### 3.2. Confidence cones

We now exploit the results of the last section to establish confidence regions about the parameter  $\theta$ . Since this parameter is a unit vector and  $\psi(X - \eta)$  has a rotationally symmetric distribution about it, these regions should naturally be cones with vertex at the origin and semi-axis  $\theta$ . We present two approaches for constructing such cones.

A first approach is to use (3.2) to build up a Wald statistic. For this, notice that  $(\|\eta\|^2 / \beta)(I - \theta \theta')$  is a generalized inverse of the covariance matrix in (3.2), so that

$$n(\|\eta\|^2 / \beta)(1 + \tilde{\theta}' \theta)(1 - \tilde{\theta}' \theta) \rightarrow \chi_{p-1}^2$$

in distribution. If we denote by  $x_\alpha$  the upper  $\alpha$ -quantile of the  $\chi^2_{p-1}$  distribution, this leads to

$$\{\theta: n(\|\tilde{\eta}\|^2/\tilde{\beta})(1+\tilde{\theta}'\theta)(1-\theta'\theta) < x_\alpha\},$$

where  $\tilde{\beta}$  is a consistent estimator of  $\beta$ , as a confidence cone with asymptotic level  $1-\alpha$ .

A second approach, along the lines of Watson (1983, p. 138), is based on a Taylor expansion of  $\tilde{\theta}'\theta$  about  $\eta$  up to the quadratic term. Using standard results on the distribution of quadratic forms, it is then found from (3.2) that  $2n(\|\tilde{\eta}\|^2/\tilde{\beta})(I-\tilde{\theta}'\theta)$  is asymptotically distributed as a  $\chi^2_{p-1}$ . Thus the confidence cone  $\{\theta: 2n(\|\tilde{\eta}\|^2/\tilde{\beta})(1-\tilde{\theta}'\theta) < x_\alpha\}$  will have asymptotically a coverage probability of  $1-\alpha$ .

Since  $1+\tilde{\theta}'\theta \rightarrow 2$  in probability, both approaches lead to similar confidence regions. However, a Monte Carlo simulation performed by the authors seems to indicate that, for small samples, the first approach is preferable, because less upwardly biased.

To use these confidence cones in practice, we need a consistent estimator of  $\beta$  or, equivalently, of  $\text{var}(s)$ ,  $E\{\|X-\eta\|^{-1}\}$  and  $E\{s^2\|X-\eta\|^{-1}\}$ . The following theorem, whose proof follows the argument of Watson (1983, p. 138), provides such estimators.

**THEOREM 3.1.** *Let  $X_1, \dots, X_n$  be a random sample from  $F \in \mathcal{R}_p(\theta)$ . Then, in probability,*

$$n^{-1} \sum \|X_i - \tilde{\eta}\|^{-1} \rightarrow E(\|X_i - \eta\|^{-1}), \quad n^{-1} \sum \{(X_i - \tilde{\eta})' \tilde{\theta}\}^2 \|X_i - \tilde{\eta}\|^{-2} \rightarrow \text{var}(s),$$

$$n^{-1} \sum \{(X_i - \tilde{\eta})' \tilde{\theta}\}^2 \|X_i - \tilde{\eta}\|^{-3} \rightarrow E(s^2 \|X_i - \eta\|^{-1}).$$

4. ASYMPTOTIC RELATIVE EFFICIENCY

Let  $F = F(\theta, k)$  be the von Mises-Fisher distribution with mode  $\theta$  and concentration parameter  $k > 0$ . The density of  $F$  is then proportional to  $\exp(k\theta'x)$  and the maximum likelihood estimator of  $\theta$  is  $\hat{\theta} = \bar{X}/\|\bar{X}\|$ . Watson (1983, p. 153) shows that  $n^{1/2}(\hat{\theta} - \theta)$  is asymptotically normal with mean 0 and covariance matrix  $\{kA_p(k)\}^{-1}(I - \theta\theta')$ , where  $A_p(k) = I_{\frac{1}{2}p}(k)/I_{\frac{1}{2}p-1}(k)$  and  $I_q(k)$  is the modified Bessel function of the first kind and of order  $q$ . A comparison with (3.2) suggests defining the asymptotic relative efficiency of  $\tilde{\theta}$  with respect to  $\hat{\theta}$  as  $\|\eta\|^2/\{kA_p(k)\beta\}$ . The values of this efficiency for  $p=2, 3$  and  $4$  and  $k=0.1, 0.5, 1, 3, 5, 10$  and  $100$  are listed in Table 1. It can be observed that the loss of efficiency with respect to  $\hat{\theta}$  is moderately small, particularly for largely dispersed distributions, i.e. with small  $k$ , and that the efficiency increases with the dimension of the observations space. Moreover, for small values of  $k$ ,  $\tilde{\theta}$  is more efficient than the estimator  $\theta^*$  mentioned in § 1 (Watson, 1983, p. 186).

Now suppose that  $F$  has the form  $(1-\varepsilon)F(\theta, k) + \varepsilon F(\theta, 10^{-3}k)$  for some  $k > 0$  and  $0 \leq \varepsilon \leq 1$ . Table 2 contains some values of the asymptotic relative efficiency of  $\tilde{\theta}$  with respect to  $\hat{\theta}$  for this

Table 1. *Asymptotic relative efficiency of the normalized spatial median with respect to the normalized mean for the distribution  $F(\theta, k)$*

$p$	$k=0.1$	$k=0.5$	$k=1$	$k=3$	$k=5$	$k=10$	$k=100$
2	1.0	0.95	0.86	0.71	0.69	0.67	0.64
3	1.0	0.99	0.96	0.87	0.84	0.81	0.79
4	1.0	1.0	0.98	0.93	0.90	0.88	0.85

Table 2. *Asymptotic relative efficiency of the normalized spatial median with respect to the normalized mean for the distribution  $(1-\varepsilon)F(\theta, k) + \varepsilon F(\theta, 10^{-3}k)$  when  $p=3$*

$\varepsilon$	$k=0.5$	$k=1$	$k=3$	$k=4$	$k=7$	$k=10$	$k=20$	$k=50$
0.05	0.99	0.97	0.91	0.90	0.92	0.95	1.07	1.47
0.1	0.99	0.98	0.95	0.96	1.02	1.09	1.36	2.19
0.2	0.99	1.00	1.04	1.09	1.24	1.41	2.00	3.76

contamination model when  $p = 3$ . It can be seen that in this situation  $\tilde{\theta}$  can do much better than  $\hat{\theta}$ . Notice also that for  $\varepsilon = 0.05$  and  $0.1$  the efficiency first decreases and then increases with  $k$  and that the variation is more pronounced for  $\varepsilon = 0.05$ . This can be explained as follows: the smaller the contamination, the better the normalized mean can cope with it for moderate values of  $k$ . But as  $k$  grows large the outliers are further out of the rest of the data and have a more disturbing influence. When the contamination is larger, for example,  $\varepsilon = 0.2$ , the efficiency steadily rises.

#### 5. AN EXAMPLE

Ferguson, Landreth & McKeown (1967) have investigated the homing ability of the northern cricket frog. Fourteen frogs were collected from the mud flats of an abandoned stream meander near Indianola, Mississippi and were enclosed for 30 hours within a dark chamber. They were then released and the direction taken by them recorded. The data set (in degrees) is: 104, 110, 117, 121, 127, 130, 136, 145, 152, 178, 184, 192, 200 and 316. The last value is suspected to be an outlier.

For these data we obtain  $\hat{\theta} = 146.0$ ,  $\theta^* = 140.7$  and  $\tilde{\theta} = 135.6$  with 95% confidence interval (125.5, 166.5), (117.2, 165.6) and (114.3, 157.2) respectively. If the last point is omitted then  $\hat{\theta} = 145.0$ ,  $\theta^* = 141.4$  and  $\tilde{\theta} = 135.4$ . The outlier has thus little effect on the magnitude of all three estimators.

Now suppose that the last data point is moved around the circle. Then  $\hat{\theta}$  varies from 139.9 to 150.2,  $\theta^*$  from 137.0 to 145.8 while  $\tilde{\theta}$  is less affected and varies from 132.1 to 139.3. For example, if the last data point is changed to 10 then  $\hat{\theta} = 141.2$ ,  $\theta^* = 145.8$ , while  $\tilde{\theta}$  stands at 134.1. Thus, for this example, the normalized spatial median appears to be less affected by outlier points than the other two estimators.

#### REFERENCES

- FERGUSON, D. E., LANDRETH, H. F. & McKEOWN, J. P. (1967). Sun compass orientation of the northern cricket frog, *Acris crepitans*. *Anim. Behav.* **15**, 45-53.
- GOWER, J. C. (1974). The median centre. *Appl. Statist.* **23**, 466-70.
- HALDANE, J. B. S. (1948). A note on the median of a multivariate distribution. *Biometrika* **35**, 414-5.
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. 5th Berkeley Symp.* **1**, 221-33.
- WATSON, G. N. (1983). *Statistics on Spheres*. New York: Wiley.

[Received October 1985. Revised June 1986]