The Annals of Statistics
1992, Vol. 20, No. 2, 917–928

ROBUST ESTIMATION OF THE CONCENTRATION PARAMETER OF THE VON MISES–FISHER DISTRIBUTION[1]

BY DAIJIN KO

Virginia Commonwealth University

We introduce a simple procedure for obtaining robust estimates of the concentration parameter of the von Mises–Fisher distribution on the q-dimensional unit sphere. The procedure is based on the median deviation from a location parameter. Its influence function is derived and standardized bias robustness is proved. The asymptotic efficiency is calculated and an example is given.

1. Introduction. The von Mises–Fisher distribution is most frequently used as a model for samples of circular or spherical data. The maximum likelihood estimators (m.l.e.) of the (directional) location and concentration parameters of the distribution are known to be nonrobust [Fisher (1982); Kimber (1985); Watson (1986); Ko and Guttorp (1988)]; therefore, inference based on them is bound to be nonrobust. Several attempts have been made to robustify the estimator, but they are dependent upon a robust initial estimator [Lenth (1981)] or limited to spherical data with a high concentration parameter [Fisher (1982)]. In this paper we develop a simple estimator for the concentration parameter which is comparable to the median absolute deviation in linear data. It may be used as an exploratory analysis as well as a starting value for more efficient robust estimators for concentration and (directional) location parameters.

The paper is organized as follows. In Section 2 an estimator based on the median deviation is introduced. The asymptotic efficiency of the estimator for the von Mises–Fisher distribution is derived in Section 3. In Section 4 we derive the influence curve and prove standardized bias robustness. In Section 5 we discusses the case of unknown location. An example is given in Section 6.

2. Estimator based on median deviation. The most commonly used distribution in directional data analysis is the von Mises–Fisher distribution. The density of the q-dimensional von Mises–Fisher distribution $\mathrm{VF}_q(\cdot;\mu,\kappa)$, with location parameter μ and concentration parameter κ on the unit sphere Ω_q in the q-dimensional Euclidean space R^q is given by

$$f_q(x;\mu,\kappa) = a_q^{-1}(\kappa)\exp(\kappa\mu^T x),$$

where $\mu, x \in \Omega_q$, $\kappa \in [0,\infty)$, $a_q^{-1}(\kappa) = \kappa^{q/2-1}/[(2\pi)^{q/2} I_{(q/2)-1}(\kappa)]$ and $I_p(\cdot)$ is

Received April 1990; revised September 1991.

[1]Work supported in part by NCI Grant P30CA16059.

AMS 1980 *subject classifications*. Primary 62F35, 62F10, 62H12.

Key words and phrases. Concentration parameter, directional data, robustness, SB robustness, von Mises–Fisher distribution.

the modified Bessel function of the first kind and order p. Let $A_q(\cdot) = I_{q/2}(\cdot)/I_{q/2-1}(\cdot)$. Let $x_1, \dots, x_n$ be n independent realizations of a random unit vector X whose distribution is $\mathrm{VF}_q(\cdot; \mu, \kappa)$. The m.l.e. $\hat{\kappa}$ of κ is given by Mardia (1972) as

$$\hat{\kappa} = \frac{1}{A_q}\left(\frac{1}{n}\sum_{i=1}^{n} \mu^T x_i\right). \tag{2.1}$$

Replacing μ in (2.1) with its m.l.e. $\hat{\mu}$ given by $\Sigma x_i / |\Sigma x_i|$ results in $\hat{\kappa} = A_q^{-1}(|(1/n)\Sigma x_i|)$, where $|\cdot|$ is the Euclidean norm in R^q

For $q = 2$, Lenth (1981) proposed an estimator $\hat{\kappa}_L = d^2/\mathrm{med}_i(2(1 - \hat{\mu}_c^T x_i))$, where d is the upper quartile of the standardized normal distribution, which is approximately 0.6724, and $\hat{\mu}_c$ is the circular median. Since $\sqrt{2(1 - \mu^T x_i)}$ is equal to $|\mu - x_i|$, it is distributed approximately as the absolute value of a normal random variable with variance $1/\kappa$ when κ is large; hence $\hat{\kappa}_L$ is a reasonable estimator for large κ. For small κ, however, this approximation is not as good. For example, when X is a two-dimensional random unit vector following the von Mises distribution with parameter $\kappa = 1$, $d^2/[\mathrm{med}\, 2(1 - \mu^T X)]$ is only 0.74. A weighted version of (2.1) was also proposed by Lenth (1981) as

$$\hat{\kappa}_w = \frac{1}{A_q} \frac{\Sigma w_i \mu^T x_i}{\Sigma w_i}.$$

Here the w_i's are weights used in computing a robust circular location estimate $\tilde{\mu}$. They are dependent on the unknown parameter κ and, hence, to calculate $\hat{\kappa}_w$ we need a robust initial estimate of κ.

For $q = 3$, using the fact that the distribution of $1 - \mu^T X$ is approximately exponential with mean κ for $\kappa \geq 2.5$, Fisher (1982) proposed the following L estimator $\hat{\kappa}_{w,r}$:

$$\hat{\kappa}_{w,r} = \frac{n - r + 1}{\Sigma_1^{n-r+1} c_{(i)} + (r+1)c_{(n-r)}},$$

where $c_i = 1 - \mu^T x_i$ and $c_{(1)} \leq \cdots \leq c_{(n)}$ are ordered c_i's. This estimator is useful and the approximation is very good for moderately large κ (say $\kappa \geq 2.5$), but it cannot be used for $q \neq 3$.

The mean operator $(1/n)\Sigma$ in (2.1) is very sensitive to outliers, causing the m.l.e. $\hat{\kappa}$ to be nonrobust. An alternative is the *median*, which is more insensitive to the outliers. Let $C_q(\kappa)$ be the median of $\mu^T X$ where X is a random unit q-dimensional vector with von Mises–Fisher distribution $\mathrm{VF}_q(\cdot; \mu, \kappa)$. Replacing A_q and $(1/n)\Sigma$ with C_q and med_i, respectively, in (2.1), we obtain an estimator of κ,

$$\hat{\kappa}_m = C_q^{-1}\left[\operatorname*{med}_i\left(\mu^T x_i\right)\right], \tag{2.2}$$

which is Fisher consistent and robust. One may use the median of $1 - \mu^T X$ instead of $\mu^T X$. If $M_q(\kappa)$ is the median of $2(1 - \mu^T X)$, then $M_q(\kappa) =$

$2(1 - C_q(\kappa))$ and the estimator becomes $M_q^{-1}[\text{med}_i\, 2(1 - \mu^T x_i)]$. The median of $\sqrt{2(1 - \mu^T X)}$ is, in fact, the median deviation from the location parameter μ with respect to the Euclidean metric $|\cdot|$ on R^q.

We can evaluate the function C_q and C_q^{-1} using the distribution of $T = \mu^T X$. When the distribution of X is $\text{VF}_q(\cdot\,; \mu, \kappa)$, the density g_q of T is given as

$$g_q(t) = a_q^{*-1}(\kappa) e^{\kappa t}(1 - t^2)^{(q-3)/2}, \qquad -1 < t < 1,$$

where $a_q^{*-1}(\kappa) = w_{q-1} a_q^{-1}(\kappa)$ with $w_q = 2\pi^{q/2}/\Gamma(q/2)$, the area of Ω_q [Watson (1983), page 136]. Therefore, $C_q(\kappa)$ is the solution of the equation

$$\int_{-1}^{C_q(\kappa)} g_q(t)\, dt = \frac{1}{2}. \tag{2.3}$$

When $q = 3$, $C_q(\kappa) = (1/\kappa)\log(\cosh \kappa)$ and $C_q'(\kappa) = -(1/\kappa^2)\log(\cosh \kappa) + (1/\kappa)\tanh \kappa$. For $q \neq 3$, the function C_q^{-1} as well as C_q' and C_q can be calculated by numerical integration. The values of C_q^{-1} are evaluated in Tables 1 and 2 for $q = 2$ and 3.

PROPOSITION 1. (i) $2\kappa(1 - C_q(\kappa)) \to m_{(q-1)}$ *as* $\kappa \to \infty$ *where* $m_{(q-1)}$ *is the median of the chi-square distribution with* $q - 1$ *degrees of freedom; hence we can approximate* $C_q^{-1}(x)$ *by* $m_{(q-1)}/2(1 - x)$ *for* x *close to* 1.

(ii) $C_q^{-1}(x)/x \to (q - 1)$ *as* $x \to 0$; *hence* $C_q^{-1}(x) \approx (q - 1)x$ *for small* x.

PROOF. (i) follows from the fact that $2\kappa(1 - \mu^T X)$ converges to χ_{q-1}^2 in distribution as $\kappa \to \infty$ [Watson (1984)].

By differentiating (2.3) with respect to κ, we have

$$C_q'(\kappa) = \frac{A_q(\kappa)/2 - \int_{-1}^{C_q(\kappa)} a_q^{*-1}(\kappa) t e^{\kappa t}(1 - t^2)^{(q-3)/2}\, dt}{a_q^{*-1}(\kappa) e^{\kappa C_q(\kappa)}\left(1 - C_q(\kappa)^2\right)^{(q-3)/2}}. \tag{2.4}$$

Since $C_q(\kappa) \to 0$ and $A_q(\kappa) \to 0$ as $\kappa \to 0$, by evaluating (2.4) at $\kappa = 0$, we have $C_q'(\kappa) \to (q - 1)^{-1}$ as $\kappa \to 0$.

(ii) follows from noting that $(C_q^{-1})'(x) = 1/C_q'(C_q^{-1}(x))$ and the first order Taylor expansion of C_q^{-1}. □

In calculating $m_{(q-1)}$ in Proposition 1(i), we can use the fact that $m_{(1)} \approx 0.45453$ and $m_{(2)} = 2\log 2$. For $q > 3$, $m_{(q-1)}$ is approximated by $(q - 1)[1 - (2/9)/(q - 1)]^3$ [Wilson and Hilferty (1931)].

3. Asymptotic properties. Let $X_1, \ldots, X_n$ be i.i.d. unit random vectors with distribution $\text{VF}_q(\cdot\,; \mu, \kappa)$. Let $\hat{\kappa}_m = C_q^{-1}(\text{med}_i(\mu^T X_i))$. Then we have the following proposition.

PROPOSITION 2.

$$\sqrt{n}\,(\hat{\kappa}_m - \kappa) \to_d N\left(0, \left[2C_q'(\kappa) g_q(C_q(\kappa))\right]^{-2}\right). \tag{3.1}$$

TABLE 1
C_2^{-1}

x	$C_2^{-1}(x)$	x	$C_2^{-1}(x)$	x	$C_2^{-1}(x)$	x	$C_2^{-1}(x)$
0.01	0.0100	0.30	0.3162	0.59	0.7521	0.88	2.2311
0.02	0.0200	0.31	0.3280	0.60	0.7731	0.89	2.3990
0.03	0.0300	0.32	0.3399	0.61	0.7948	0.90	2.5998
0.04	0.0400	0.33	0.3519	0.62	0.8173	0.905	2.7160
0.05	0.0501	0.34	0.3641	0.63	0.8405	0.910	2.8452
0.06	0.0601	0.35	0.3765	0.64	0.8647	0.915	2.9897
0.07	0.0702	0.36	0.3890	0.65	0.8897	0.920	3.1525
0.08	0.0803	0.37	0.4017	0.66	0.9158	0.925	3.3374
0.09	0.0904	0.38	0.4146	0.67	0.9430	0.930	3.5492
0.10	0.1006	0.39	0.4277	0.68	0.9713	0.935	3.7944
0.11	0.1107	0.40	0.4410	0.69	1.0010	0.940	4.0812
0.12	0.1210	0.41	0.4545	0.70	1.0321	0.945	4.4212
0.13	0.1312	0.42	0.4683	0.71	1.0648	0.950	4.8304
0.14	0.1415	0.43	0.4822	0.72	1.0992	0.955	5.3317
0.15	0.1519	0.44	0.4964	0.73	1.1356	0.960	5.9597
0.16	0.1623	0.45	0.5109	0.74	1.1741	0.965	6.7686
0.17	0.1728	0.46	0.5257	0.75	1.2149	0.970	7.8485
0.18	0.1833	0.47	0.5408	0.76	1.2585	0.975	9.3619
0.19	0.1939	0.48	0.5562	0.77	1.3050	0.980	11.6337
0.20	0.2046	0.49	0.5719	0.78	1.3550	0.985	15.4221
0.21	0.2153	0.50	0.5879	0.79	1.4088	0.990	23.0017
0.22	0.2262	0.51	0.6043	0.80	1.4671	0.991	25.5287
0.23	0.2371	0.52	0.6212	0.81	1.5306	0.992	28.6874
0.24	0.2481	0.53	0.6384	0.82	1.6001	0.993	32.7489
0.25	0.2592	0.54	0.6561	0.83	1.6768	0.994	38.1643
0.26	0.2703	0.55	0.6742	0.84	1.7619	0.995	45.7461
0.27	0.2816	0.56	0.6929	0.85	1.8573	0.996	57.1190
0.28	0.2930	0.57	0.7120	0.86	1.9651	0.997	76.0742
0.29	0.3046	0.58	0.7318	0.87	2.0884		

PROOF. Since $\sqrt{n}(\mathrm{med}_i(\mu^T X_i) - C_q(\kappa)) \to N(0, \sigma^2)$ in distribution with $\sigma^2 = [2g_q(C_q(\kappa))]^{-2}$, the result follows from the δ method. □

The asymptotic variance of $\sqrt{n}\,\hat{\kappa}$, where $\hat{\kappa}$ is the m.l.e., is $1/A'_q(\kappa)$. Hence the asymptotic efficiency of the estimator $\hat{\kappa}_m$ is given by

$$4\left[C'_q(\kappa)g_q\left(C_q(\kappa)\right)\right]^2 A'_q(\kappa).$$

Table 3 gives the values of efficiencies for $\kappa = 0.1, 0.5, 1, 3, 5, 10, 50$ and for $q = 2$ and 3.

Let us consider the case of $q = 2$. Let $X = (\cos\Theta, \sin\Theta)^T$ and $\mu = (\cos\theta_0, \sin\theta_0)^T$ where $\theta_0 - \pi < \Theta \le \theta_0 + \pi$. For large κ, $\Theta - \theta_0$ is distributed approximately $N(0, \kappa^{-1})$ [Mardia (1972), page 60]. One may use an estimate $\tilde{\kappa} = [\Phi^{-1}(0.75)/\mathrm{med}|\Theta - \theta_0|]^2$ using the fact that MAD = med$|\Theta -$

TABLE 2
C_3^{-1}

x	$C_3^{-1}(x)$	x	$C_3^{-1}(x)$	x	$C_3^{-1}(x)$	x	$C_3^{-1}(x)$
0.01	0.0200	0.26	0.5450	0.51	1.2552	0.76	2.8749
0.02	0.0400	0.27	0.5682	0.52	1.2927	0.77	3.0030
0.03	0.0600	0.28	0.5916	0.53	1.3313	0.78	3.1422
0.04	0.0801	0.29	0.6153	0.54	1.3711	0.79	3.2942
0.05	0.1002	0.30	0.6393	0.55	1.4122	0.80	3.4608
0.06	0.1203	0.31	0.6637	0.56	1.4547	0.81	3.6445
0.07	0.1405	0.32	0.6883	0.57	1.4987	0.82	3.8483
0.08	0.1607	0.33	0.7134	0.58	1.5442	0.83	4.0756
0.09	0.1810	0.34	0.7388	0.59	1.5915	0.84	4.3311
0.10	0.2013	0.35	0.7646	0.60	1.6406	0.85	4.6203
0.11	0.2218	0.36	0.7909	0.61	1.6917	0.86	4.9507
0.12	0.2423	0.37	0.8176	0.62	1.7450	0.87	5.3317
0.13	0.2630	0.38	0.8447	0.63	1.8006	0.88	5.7761
0.14	0.2837	0.39	0.8724	0.64	1.8587	0.89	6.3013
0.15	0.3046	0.40	0.9005	0.65	1.9196	0.90	6.9315
0.16	0.3256	0.41	0.9293	0.66	1.9835	0.91	7.7016
0.17	0.3467	0.42	0.9586	0.67	2.0507	0.92	8.6643
0.18	0.3680	0.43	0.9885	0.68	2.1215	0.93	9.9021
0.19	0.3895	0.44	1.0191	0.69	2.1963	0.94	11.5525
0.20	0.4111	0.45	1.0504	0.70	2.2755	0.95	13.8629
0.21	0.4329	0.46	1.0824	0.71	2.3595	0.96	17.3287
0.22	0.4549	0.47	1.1152	0.72	2.4490	0.97	23.1049
0.23	0.4771	0.48	1.1488	0.73	2.5444	0.98	34.6574
0.24	0.4995	0.49	1.1833	0.74	2.6467	0.99	69.3147
0.25	0.5221	0.50	1.2188	0.75	2.7565		

$\theta_0|/\Phi^{-1}(0.75)$ is a consistent estimate of $\kappa^{-1/2}$. Here, MAD stands for median absolute deviation. Now

$$\hat{\kappa}_m = C_2^{-1}[\operatorname{med}\cos(\Theta - \theta_0)]$$

is approximated by $m_{(1)}/[2(1 - \operatorname{med}\cos(\Theta - \theta_0)]$ by Proposition 1. Using a Taylor expansion of the cosine function, we can approximate $\hat{\kappa}_m$ by $\Phi^{-1}(0.75)^2/\operatorname{med}|\Theta - \theta_0|^2$, which is MAD^{-2}. Therefore the asymptotic efficiency of $\hat{\kappa}_m$, when κ is large, is the same as MAD at the normal distribution.

TABLE 3
Asymptotic efficiencies of $\hat{\kappa}_m$

q	$\kappa = 0.1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 3$	$\kappa = 5$	$\kappa = 10$	$\kappa = 50$
2	0.81	0.73	0.59	0.36	0.36	0.37	0.37
3	0.75	0.73	0.67	0.51	0.48	0.48	0.48

As in the location-scale problem in linear data, the estimator $\hat{\kappa}_m$ is favored for its robustness not for its efficiency. We can improve the efficiency by introducing M estimators with an initial value $\hat{\kappa}_m$ [Ko and Chang (1991)].

4. Standardized bias (SB) robustness of $\hat{\kappa}_m$. Ko and Guttorp (1988) proposed SB robustness in assessing robustness of estimators for directional data. They argue that even for the same amount of bias the problem is more serious if the main mass of data is more concentrated, where more accuracy is needed. They measure this relative seriousness by standardized gross error sensitivity (SGES), defined by

$$\gamma^*(T, \mathbf{F}, S) = \sup_{\mathbf{F}} \gamma(T, F)/S(F),$$

where $\gamma(T, F)/S(F)$ is the gross error sensitivity (GES) of the functional T at F, $S(F)$ is the measure of concentration of the main mass of the data and the supremum is taken over a family $\mathbf{F}$ of distributions. The same definition may be stated in terms of the breakdown function using the fact that the GES is the slope of the breakdown function at 0 [He and Simpson (1989)].

A reasonable choice of $S(F)$ is the inverse of Fisher information [or equivalently the Cramér–Rao (CR) bound for the standard error of T]. Then $\gamma^*(T, \mathbf{F}, S)$ is the supremum of the information standardized sensitivities [Hampel, Ronchetti, Rousseeuw and Stahel (1986)] over the family $\mathbf{F}$ of distributions. An estimator T is called *SB robust* at $\mathbf{F}$ if the SGES is bounded. If an estimator is SB robust at $\mathbf{F}$, then the ratio between the GES and the CR error bound is bounded over the family $\mathbf{F}$ of distributions.

For the concentration parameter κ from the von Mises–Fisher distribution, the inverse of Fisher information is $[A'_q(\kappa)]^{-1/2}$ and we have the following proposition.

PROPOSITION 3. *The estimator $\hat{\kappa}_m$ is SB robust at the family of the von Mises distributions.*

PROOF. Using the influence function of the median and the chain rule, we can derive the influence function of $\hat{\kappa}_m$ at z and $\mathrm{VF}_q(\cdot; \mu, \kappa)$ as

$$\mathrm{IF}(z; \hat{\kappa}_m, \mathrm{VF}_q) = \frac{1}{C'_q(\kappa)} \frac{\operatorname{sign}\left[\mu^T z - C_q(\kappa)\right]}{2g_q(C_q(\kappa))}.$$

Therefore the GES is $[2C'_q(\kappa)g_q(C_q(\kappa))]^{-1}$ and SGES γ^* at the von Mises–Fisher family $\mathbf{F}$ is

$$\gamma^*(\hat{\kappa}_m, \mathbf{F}, S) = \sup_{\kappa} \frac{\left(A'_q(\kappa)\right)^{1/2}}{2C'_q(\kappa)g_q(C_q(\kappa))}.$$

$C_q(\kappa) \to 0$, $A'_q(\kappa) \to q^{-1}$ and $C'_q(\kappa) \to (q-1)^{-1}$ as $\kappa \to 0$. Hence γ^* stays bounded as $\kappa \to 0$. When $\kappa \to \infty$, $C_q(\kappa) = 1 - O(\kappa^{-1})$, $C'_q(\kappa) = O(\kappa^{-2})$ and $A'_q(\kappa) = O(\kappa^{-2})$. Since $I_{q/2-1}(\kappa) = (2\pi)^{-1/2}\kappa^{-1/2}e^{\kappa}(1 - O(\kappa^{-1}))$ for large κ, we have that $a_q^*(\kappa) = w_{q-1}^{-1}(2\pi)^{(q-1)/2}\kappa^{(-q+1)/2}e^{\kappa}(1 + O(\kappa^{-1}))$ and $g_q(C_q(\kappa)) = O(\kappa)$. Therefore $\gamma^*(\hat{\kappa}_m, \mathbf{F}, S) = O(1)$. □

Ko and Guttorp (1988) show that the m.l.e. $\hat{\kappa}$ is not SB robust. In fact, $\gamma(\hat{\kappa}, F)/S(F)$ when F is $\mathrm{VF}_q(\cdot; \mu; \kappa)$ is $O(\kappa)$ and hence $\gamma^*(\hat{\kappa}, \mathbf{F}) = \infty$.

5. When location is unknown. The location vector μ is usually unknown. In this case, we may substitute an estimate of μ. In this section, we show that if the location estimator is SB robust, the estimator $\hat{\kappa}_m$ based on the estimated location is also SB robust. We also prove that the asymptotic property obtained in Section 3 holds when the estimated location is used.

For any location estimator M with $M(F) = \mu \in \Omega_q$, $M((1-\varepsilon)F + \varepsilon\delta_z)^T M((1-\varepsilon)F + \varepsilon\delta_z) = 1$. By differentiating with respect to ε at $\varepsilon = 0$, we have $\mathrm{IF}(z; M, F)^T\mu = 0$. That is, as vectors in R^q, $\mathrm{IF}(z; M, F)$ is tangent to Ω_q at μ. Since Fisher information can be formulated as the metric of the inner product on the tangent space at μ [see Kass (1989), page 198], the information standardized sensitivity is $\sup_z[\langle \mathrm{IF}(z; M, F), \mathrm{IF}(z; M, F)\rangle_{M(F)}]^{1/2}$ and the SGES at $\mathbf{F}$ is $\gamma^*(M, \mathbf{F}, S) = \sup_{\mathbf{F}} \sup_z[\langle \mathrm{IF}(z; M, F), \mathrm{IF}(z; M, F)\rangle_{M(F)}]^{1/2}$, where $\langle \nu, w\rangle_\mu$ denotes the Fisher information metric at μ. For the rotationally symmetric distribution with density $f(\mu^T x)$ with respect to the surface measure on Ω_q, $\gamma^*(M, F, S) = \sup_{\mathbf{F}} \sup_z \sqrt{\kappa A_q(\kappa)}\, |\mathrm{IF}(z; M, F)|$; therefore the estimator M is SB robust at $\mathbf{F}$ if and only if $\sup_z |\mathrm{IF}(z; M, F)| = O(\kappa^{-1/2})$. See Ko and Chang (1991) for details. In the following, $M(X)$ and $K_m(X)$ denote $M(F)$ and $K_m(F)$, respectively, where X is a random unit vector with distribution F.

Let K_m be a functional defined by

$$K_m(F) = C_q^{-1}\left[\mathrm{med}\left(M(X)^T X\right)\right].$$

Let $\hat{K}_m = K_m(F_n)$ and $\hat{\mu} = M(F_n)$ where F_n is the empirical distribution. $\hat{K}_m = K_m(F_n)$ is obtained by replacing μ by an estimate $M(F_n)$ in (2.2) in Section 3.

PROPOSITION 4. *K_m is SB robust at* $\mathbf{F}$ *if M is Fisher consistent and SB robust at* $\mathbf{F}$.

PROOF. See the Appendix.

Suppose further that $\hat{\mu}$ is a $\sqrt{n}$ consistent estimate of μ. The limiting distribution of $\hat{K}_m$ is the same as that of $\hat{\kappa}_m$ as follows.

PROPOSITION 5. $\sqrt{n}(\hat{K}_m - \kappa) \to_d N(0, [2C_q'(\kappa)g_q(C_q(\kappa))]^{-2})$.

PROOF. See the Appendix.

If M is B robust, that is, it has a bounded influence function, but not SB robust (e.g., the directional mean), then the term $\mathrm{IF}_M^T(X - \mu)$ in (7.1) in the

Appendix is $O(\kappa^{-a})$, not $O(\kappa^{-1})$ for some $0 < a < 1$ and becomes the dominating term in calculating the influence function of MD, the functional corresponding to $\text{med}_i\, \mu^T X_i$. In fact, in this case, the influence function of MD is

$$\text{IF}(z; \text{MD}, \text{VF}_q) = \frac{1}{2g_q\big(C_q(\kappa) + O(\kappa^{-1})\big)} + O(\kappa^{-a})$$

and the SGES is $\sup_\kappa O(\kappa^{1-a})$, which is ∞. Therefore we do not recommend the directional mean for the location estimate in calculating $\hat{K}_m$.

The circular/spherical median [Mardia (1972); Fisher (1985); Ducharme and Milasevic (1987)] is SB robust [Ko and Guttorp (1988); He and Simpson (1989); Ko and Chang (1991)] and can be used as an estimator of location for K_m, which is a monotone function of the median distance from the estimated spherical median. For the computation of the spherical median, especially for $q \geq 3$, see Fisher (1985) and Gower (1974).

Another interesting example for the location estimator is the least median square (LMS) estimator $\hat{\mu}_{\text{LMS}}$ [Rousseeuw (1984)], which is defined as the center of the smallest circular cap on Ω_q which covers more than half of the data. Then the $\text{med}_i\, \hat{\mu}_{\text{LMS}}^T x_i = 1 - r^2/2$, where r is the radius (in R^q) of the smallest circle. On Ω_2, $\text{med}_i\, \hat{\mu}_{\text{LMS}}^T x_i$ coincides with the cosine of half of the circular interquartile range. In linear data, the LMS location estimate converges slowly with rate $n^{-1/3}$; therefore, its influence function and gross error sensitivity are not well defined and the argument of Proposition 3 cannot be used to prove or disprove SB robustness. However, the scale estimator $S(F)$ based on the LMS estimate defined as the median of $|X - \hat{\mu}_{\text{LMS}}|$, where the distribution of X is F, is an S estimator, which has better continuity properties than the LMS location estimator. In fact, $S(F_n)$ has the usual rates of normal convergence [Grübel (1988)] and is approximately most bias robust [Martin and Zamar (1989)]. Since $\text{med}|X - \hat{\mu}_{\text{LMS}}|^2 = \inf_{\mu \in \Omega_q} \text{med}|X - \mu|^2$, the estimate $C_q^{-1}(\text{med}\, \hat{\mu}_{\text{LMS}}^T X)$ can be viewed as an S estimator and is expected to have similar robustness properties. However, SB robustness of the estimator K_m based on $\hat{\mu}_{\text{LMS}}$ is yet to be proved.

6. Examples. Ferguson, Landreth and McKeown (1967) have investigated the homing ability of the northern cricket frog. These data have been reanalyzed by Collett (1980), Ducharme and Milasevic (1987) and Rousseeuw and Leroy (1987). The data are (in degrees) 104, 110, 117, 121, 127, 130, 136, 145, 152, 178, 184, 192, 200, 316. The directional mean angle is 146°, where the circular median is 133° and the circular LMS estimate is 120°. The true home direction is 122°. $\mu^T x_i = \cos(\phi_i - \phi)$, where μ and x_i are points on Ω_2 corresponding to angles ϕ_i and ϕ.

With the true home direction 122°, the m.l.e. $\hat{\kappa}$ is 1.8, while the estimate $\hat{\kappa}_m$ based on the median deviation is 3.83. Ninety-five percent (resp. 99%) of the data for the von Mises distribution fall in the range of 122° $\pm$ 111.6° (resp.

159.2°) when $\hat{\kappa}$ is used and 122° ± 63.4° (resp. 88.5°) when $\hat{\kappa}_m$ is used. The angle 316° is an extreme point beyond 1% by both estimates. The points 192° and 200° are identified as extreme points beyond 5% only when $\hat{\kappa}_m$ is used.

Without the extreme point 316°, the m.l.e. is 2.73 and $\hat{\kappa}_m$ is 4.92. None was found to be an extreme point beyond 5% when the m.l.e. is used; 178°, 184°, 192° were identified as extreme points beyond 5% and 200° was observed as an extreme point beyond 1% when $\hat{\kappa}_m$ is used. The m.l.e. picked up the most outlying point 316° as a 1% extreme point but failed to detect other extreme points.

$\hat{K}_m$ is estimated as 3.69 when the circular median is used; it is 3.73 when the LMS estimate is used; the m.l.e. of κ is 2.18. Ninety-five percent (resp. 99%) of the data for the von Mises distribution fall in the ranges of 133° ± 65° (resp. 91.2°), 120° ± 64.5° (resp. 90.5°) and 146° ± 95.8° (resp. 144.3°) when these estimates of κ and μ are used. Using the LMS estimate and the corresponding $\hat{K}_m$, we identify 192° and 200° as 5% extreme points and 316° as a 1% extreme point. If the circular median and the corresponding $\hat{K}_m$ are used, we identify 200° as a 5% extreme point and 316° as a 1% extreme point, where the inference based on m.l.e. identifies only 316° as a 1% extreme point.

APPENDIX

PROOF OF PROPOSITION 4. Let X be a random unit vector with the distribution $\mathrm{VF}_q(\cdot\,; \mu, \kappa)$ and $X_{\varepsilon,z} = (1 - W)X + W_z$, where X and W are independent, W is a Bernoulli random variable with $\Pr\{W = 1\} = \varepsilon$ and z is a unit vector. Then $X_{\varepsilon,z}$ is a random unit vector with the contaminated distribution $(1 - \varepsilon)\mathrm{VF}_q + \varepsilon\delta_z$ where δ_z denotes the point mass of 1 at z. Let $D_{\varepsilon,z} = M(X_{\varepsilon,z})^T X_{\varepsilon,z}$, $D = M(X)^T X = \mu^T X$ and let g_q be the density of D. Let IF_M denote $\mathrm{IF}(z; M, \mathrm{VF}_q)$. For small $\varepsilon > 0$,

$$
\begin{aligned}
(7.1)\quad D_{\varepsilon,z} - D &= (1 - W)\left[\varepsilon \mathrm{IF}_M^T X + o(\varepsilon)\right] + W\left(M(X_{\varepsilon,z})^T z - \mu^T X\right)\\
&= (1 - W)\left[\varepsilon \mathrm{IF}_M^T (X - \mu) + o(\varepsilon)\right] + W\left(M(X_{\varepsilon,z})^T z - \mu^T X\right)
\end{aligned}
$$

because $\mathrm{IF}_M^T \mu = 0$. Since M is SB robust, $\sup_z |\mathrm{IF}(z, M, \mathrm{VF}_q)| = O(\kappa^{-1/2})$ and also $|X - \mu| = O_p(\kappa^{-1/2})$. Therefore,

$$D_{\varepsilon,z} = (1 - W)\left[D + \varepsilon O_p(\kappa^{-1}) + o(\varepsilon)\right] + W\left(M(X_{\varepsilon,z})^T z\right).$$

Using the influence function of the *median* we have

$$\mathrm{med}[(1 - W)Y + Wz^*] - \mathrm{med}(Y) = \varepsilon\left[\frac{1}{2g(\mathrm{med}(Y))} + o(\varepsilon)\right]$$

for any random variable Y with density g. By letting $Y = D + \varepsilon O_p(\kappa^{-1}) + o(\varepsilon)$

and $z^* = M(X_{\varepsilon,z})^T z$ we have

$$\begin{aligned}
&\varepsilon \mathrm{IF}(z; \mathrm{MD}, \mathrm{VF}_q)\\
&\quad = \mathrm{med}(D_{\varepsilon,z}) - \mathrm{med}(D)\\
&\quad = [\mathrm{med}((1 - W)Y + Wz^*) - \mathrm{med}(Y)] + [\mathrm{med}(Y) - \mathrm{med}(D)]\\
&\quad = \varepsilon\left[\frac{1}{2g(\mathrm{med}(Y))} + o(\varepsilon)\right] + \varepsilon O(\kappa^{-1}) + o(\varepsilon),
\end{aligned}$$

where g is a density function of Y and MD is the functional corresponding to med D. As $\varepsilon \to 0$, the distribution of Y converges to the distribution of D and the median of Y can be written as $C_q(\kappa) + O(\kappa^{-1})$. Therefore, the influence function is given by

$$\mathrm{IF}(z; \mathrm{MD}, \mathrm{VF}_q) = \frac{1}{2g_q\big(C_q(\kappa) + O(\kappa^{-1})\big)} + O(\kappa^{-1}).$$

Since $K_m = C_q^{-1} \circ \mathrm{MD}$, $2g_q(C_q(\kappa) + O(\kappa^{-1})) = O(\kappa)$ and $C_q'(\kappa) = O(\kappa^{-2})$, by the chain rule, we have

$$\begin{aligned}
\mathrm{IF}(z; K_m, \mathrm{VF}_q) &= C_q'(\kappa)^{-1}\left[\frac{1}{2g_q\big(C_q(\kappa) + O(\kappa^{-1})\big)} + O(\kappa^{-1})\right]\\
&= O(\kappa^2)\big[O(\kappa^{-1}) + O(\kappa^{-1})\big] = O(\kappa).
\end{aligned}$$

Since $A_q'(\kappa) = O(\kappa^{-2})$,

$$\begin{aligned}
\gamma^*(K_m, \mathbf{F}, S) &= \sup_{\kappa} \sup_{z} \big(A_q'(\kappa)\big)^{1/2} C_q'(\kappa)^{-1} \big|\mathrm{IF}(z; \mathrm{MD}, \mathrm{VF}_q)\big|\\
&= O(\kappa^{-1}) O(\kappa) = O(1).
\end{aligned}$$

Therefore K_m is SB robust. □

Proof of Proposition 5. Since $|X_i - \mu|^2 = 2(1 - \mu^T X_i)$, by Proposition 2, it is enough to show that $\sqrt{n}\ \mathrm{med}_i |X_i - \hat{\mu}|^2 = \sqrt{n}\ \mathrm{med}_i |X_i - \mu|^2 + o_p(1)$. Define

$$\underset{i}{\mathrm{med}}\, Z_i = F_n^{-1}(\tfrac{1}{2}) = \inf\{t | F_n(t) \geq \tfrac{1}{2}\},$$

where F_n is the same distribution function based on $Z_1, \dots, Z_n$. By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&|X_i - \mu|^2 - 2|X_i - \mu|\,|\mu - \hat{\mu}| + |\mu - \hat{\mu}|^2\\
&\quad \leq |X_i - \hat{\mu}|^2 \leq |X_i - \mu|^2 + 2|X_i - \mu|\,|\mu - \hat{\mu}| + |\mu - \hat{\mu}|^2
\end{aligned} \tag{7.2}$$

for all i. Since $\tilde{h}(x) = x^2 + 2|\mu - \hat{\mu}|x$ is a monotone function of x on $(0, \infty)$,

$$\operatorname*{med}_i \left(|X_i - \mu|^2 + 2|X_i - \mu|\,|\mu - \hat{\mu}|\right)$$
$$= \operatorname*{med}_i |X_i - \mu|^2 + 2\left(\operatorname*{med}_i |X_i - \mu|\right)|\mu - \hat{\mu}|.$$

Let $|X_{(1)} - \mu| \le \cdots \le |X_{(n)} - \mu|$. For any fixed $\alpha \in (0, 1/2)$, say $\alpha = 1/4$,

$$|X_{([n\alpha])} - \mu|^2 \xrightarrow{\text{a.s.}} H^{-1}(\alpha) > 0,$$

where H is the distribution function of $|X_i - \mu|^2$ and $[n\alpha]$ is the integer part of $n\alpha$. Therefore, as $n \to \infty$, $2|\mu - \hat{\mu}| < |X_{([n\alpha])} - \mu|$ with probability tending to 1. (Most of the identities and inequalities in the rest of the proof hold with probability tending to 1.) Note that the function $h(x) = x^2 - |\mu - \hat{\mu}|x$ is increasing on $(|\mu - \hat{\mu}|, \infty)$ and $h(x) \le h(s)$ for all $0 \le x \le s$ where $s \in (2|\mu - \hat{\mu}|, \infty)$. Hence, for large n, the order of $|X_{([n\alpha])} - \mu| \le \cdots \le |X_{(n)} - \mu|$ is maintained under the transformation $h(\cdot)$ and also

$$|X_{(j)} - \mu|^2 - 2|X_{(j)} - \mu|\,|\mu - \hat{\mu}| \le |X_{([n\alpha])} - \mu|^2 - 2|X_{([n\alpha])} - \mu|\,|\mu - \hat{\mu}|$$

for all $j \le [n\alpha]$; that is, all the lower $100\alpha\%$ order statistics of $h(|X_i - \mu|)$ are not greater than $h(|X_{([n\alpha])} - \mu|)$. Therefore, $100(1 - \alpha)\%$ upper order statistics are

$$|X_{([n\alpha])} - \mu|^2 - 2|X_{([n\alpha])} - \mu|\,|\mu - \hat{\mu}|$$
$$\le \cdots \le |X_{(n)} - \mu|^2 - 2|X_{(n)} - \mu|\,|\mu - \hat{\mu}|$$

and hence

$$\operatorname*{med}_i \left(|X_i - \mu|^2 - 2|X_i - \mu|\,|\mu - \hat{\mu}|\right)$$
$$= \operatorname*{med}_i |X_i - \mu|^2 - 2\left(\operatorname*{med}_i |X_i - \mu|\right)|\mu - \hat{\mu}|.$$

Since $\sqrt{n}\ \mathrm{med}_i\, |X_i - \mu|$ and $\sqrt{n}\, |\mu - \hat{\mu}|$ are $O_P(1)$ and $|\mu - \hat{\mu}|$ is $o_p(1)$,

$$\sqrt{n} \operatorname*{med}_i \left(|X_i - \mu|^2 - 2|X_i - \mu|\,|\mu - \hat{\mu}| + |\mu - \hat{\mu}|^2\right)$$
$$= \sqrt{n} \operatorname*{med}_i |X_i - \mu|^2 + o_p(1)$$

and

$$\sqrt{n} \operatorname*{med}_i \left(|X_i - \mu|^2 + 2|X_i - \mu|\,|\mu - \hat{\mu}| + |\mu - \hat{\mu}|^2\right)$$
$$= \sqrt{n} \operatorname*{med}_i |X_i - \mu|^2 + o_p(1).$$

By (7.2), $\sqrt{n}\ \mathrm{med}_i\, |X_i - \hat{\mu}|^2 = \sqrt{n}\ \mathrm{med}_i\, |X_i - \mu|^2 + o_p(1)$ and the result follows from (3.1). The arguments hold for the other definition of sample median such as $\mathrm{med}_i\, Z_i = \sup\{t | F_n(t) \le 1/2\}$ and therefore for the usual definition of sample median $\mathrm{med}_i\, Z_i = [\inf\{t | F_n(t) \ge 1/2\} + \sup\{t | F_n(t) \le 1/2\}]/2$. □

Acknowledgments. I am grateful to Professor Sung Choi and Professor Chris Gennings for their reading of and valuable comments on an earlier draft of the article and to the referees for their helpful suggestions and corrections.

REFERENCES

COLLETT, D. (1980). Outliers in circular data. *J. Roy. Statist. Soc. Ser. C* **29** 50–57.

DUCHARME, G. R. and MILASEVIC, P. (1987). Spatial median and directional data. *Biometrika* **74** 212–215.

FERGUSON, D. E., LANDRETH, H. F. and MCKEOWN, J. P. (1967). Sun compass orientation of the northern cricket frog, *Acris crepitans*. *Animal Behavior* **15** 45–53.

FISHER, N. I. (1982). Robust estimation of the concentration parameter of Fisher's distribution on the sphere. *J. Roy. Statist. Soc. Ser. C* **31** 152–154.

FISHER, N. I. (1985). Spherical median. *J. Roy. Statist. Soc. Ser. B* **47** 342–348.

GOWER, J. D. (1974). The median center. *J. Roy. Statist. Soc. Ser. C* **23** 446–470.

GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*. Wiley, New York.

HE, X. and SIMPSON, D. (1989). Robust direction estimation. Univ. Illinois, unpublished manuscript.

KASS, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4** 188–234.

KIMBER, A. C. (1985). A note on the detection and accommodation of outliers relative to Fisher's distribution on the sphere. *J. Roy. Statist. Soc. Ser. C* **34** 169–172.

KO, D. and CHANG, T. (1991). Robust *M*-estimators on sphere. Technical Report 91-01, Dept. Biostatistics, Virginia Commonwealth Univ.

KO, D. and GUTTORP, P. (1988). Robustness of estimators for directional data. *Ann. Statist.* **16** 609–618.

LENTH, R. V. (1981). Robust measure of location for directional data. *Technometrics* **23** 77–81.

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

MARTIN, R. D. and ZAMAR, R. H. (1989). Bias robust estimation of scale when location is unknown. Univ. Washington, unpublished manuscript.

ROUSSEEUW, P. J. (1984). Least median square regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

WATSON, G. S. (1984). The theory of concentrated Langevin distributions. *J. Multivariate Anal.* **14** 74–82.

WATSON, G. S. (1986). Some estimation theory on the sphere. *Ann. Inst. Statist. Math.* **38** 263–275.

WILSON, E. B. and HILFERTY, M. M. (1931). Distribution of chi-square. *Proc. Nat. Acad. Sci. U.S.A.* **17** 684–688.

DEPARTMENT OF BIOSTATISTICS
MEDICAL COLLEGE OF VIRGINIA
VIRGINIA COMMONWEALTH UNIVERSITY
RICHMOND, VIRGINIA 23298-0032