OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Biometrika Trust

A Practical Affine Equivariant Multivariate Median
Author(s): Thomas P. Hettmansperger and Ronald H. Randles
Source: *Biometrika*, Vol. 89, No. 4 (Dec., 2002), pp. 851-860
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/4140542
Accessed: 05-02-2016 05:54 UTC

REFERENCES

Linked references are available on JSTOR for this article:
http://www.jstor.org/stable/4140542?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Biometrika Trust and Oxford University Press are collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

http://www.jstor.org

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

Biometrika (2002), **89**, 4, *pp.* 851–860
© 2002 Biometrika Trust
Printed in Great Britain

A practical affine equivariant multivariate median

BY THOMAS P. HETTMANSPERGER

Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

tph@stat.psu.edu

AND RONALD H. RANDLES

Department of Statistics, University of Florida, Gainesville, Florida 32611, U.S.A.

rrandles@stat.ufl.edu

SUMMARY

A robust affine equivariant estimator of location for multivariate data is proposed which becomes the univariate median for data of dimension one. The estimator is robust in the sense that it has a bounded influence function, a positive breakdown value and has high efficiency compared to the sample mean for heavy-tailed distributions. Perhaps its greatest strength is that, unlike other affine equivariant multivariate medians, it is easily computed for data in any practical dimension.

Some key words: Bounded influence; Breakdown; L_1 median; Multivariate location estimate; Multivariate median; Oja median; Robust.

1. INTRODUCTION

Our objective is to find a robust estimator of location for multivariate data with the desirable properties of affine equivariance, a bounded influence function, a positive breakdown value, high efficiency, computability in high dimensions, the availability of estimators of standard error and an interpretable location functional. Let $x_1, \ldots, x_n$ denote a random sample of $k \times 1$ vectors, $x_i = (x_{i1}, \ldots, x_{ik})'$, from some continuous population. Our approach to finding an estimator of a $(k \times 1)$-dimensional location vector θ combines an L_1, or spatial, median with an M-estimator of scatter. The method is similar to the transformation-retransformation approach of Chakraborty et al. (1998), and uses Tyler's (1987) M-estimator of scatter. The L_1 median is equivariant under orthogonal transformations of the data but not affine equivariant. It is the minimiser of $\Sigma \| x_i - \theta \|$, where $\|.\|$ denotes Euclidean distance.

Small (1990) traces the history of the multivariate L_1 median back to Hayford (1902). Gini & Galvani (1929) introduced the median into the statistical literature and then it was rediscovered by Haldane (1948). Gower (1974) termed it the mediancentre and discussed some of its properties while providing an algorithm for the bivariate case; see also Bedall & Zimmermann (1979) for an algorithm in higher dimensions. Brown (1983) refers to it as the spatial median and showed that for spherical normal models the efficiency increases as the dimension increases, beginning with 0·785 in the bivariate case. It has 50% breakdown and bounded influence; see Lopuhaa & Rousseeuw (1991) and

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

Kemperman (1987). It is unique in dimensions greater than one; see also Hettmansperger & McKean (1998, p. 344).

Tyler's (1987) M-estimator of scatter can be written as $(\hat{A}'\hat{A})^{-1}$, and Tyler provides an algorithm for the rapid computation of $\hat{A}$ based on data centred at a location estimate. Tyler argues that his scatter estimator is the 'most robust' estimator of scatter in an elliptical model. Adrover (1998) shows that Tyler's scatter matrix is minimax bias-robust. Further evidence of the stability of Tyler's scatter matrix can be found in Duembgen (1998) and Maronna et al. (1992), in which the condition number of the matrix is studied. No scaling or tuning constant has to be specified for the determination of $\hat{A}$. In an unpublished technical report from Rutgers University, L. Duembgen and D. Tyler show that the breakdown point for Tyler's scatter estimator, roughly speaking the fraction of contamination that must be mixed into the model to move the estimator beyond any bound, is generally between $1/(k+1)$ and $1/k$, where k is the number of dimensions. In addition, they show that breakdown only occurs when the contamination is concentrated in a low-dimensional subspace. For further discussion of breakdown of M-estimators, see Maronna (1976), Huber (1981, Ch. 8) and Hampel et al. (1986, Ch. 5). We will use Tyler's $\hat{A}$ as a transformation matrix to define $z_i = \hat{A}x_i$ for $i = 1, 2, \ldots, n$, where $x_1, \ldots, x_n$ are the original k-dimensional data points. Then our estimator of θ is $\hat{A}^{-1}$ applied to the L_1 median of $z_1, \ldots, z_n$.

Other affine equivariant multivariate location estimators have appeared in the literature. For example, Small (1990) provides a discussion of the Tukey half-space depth median and the Liu simplicial depth median. Maronna (1976), Huber (1981, Ch. 8) and Hampel et al. (1986, Ch. 5) examine affine equivariant M-estimators. These are subject to decreasing breakdown value as the dimension increases. In § 2 we show that our proposed estimator is an M-estimator. Another example is the Oja (1983) median. This estimator is computationally very intensive and has zero breakdown for some special configurations of contamination. However, faster algorithms for higher dimensions are currently under development; see Ronkainen et al. (2002), and see Niinimaa et al. (1990) and Niinimaa & Oja (1995) for details about breakdown. An additional example is given by the minimum volume ellipsoid described by Rousseeuw & Leroy (1987, p. 258). This estimator has 50% breakdown but is computationally very intensive and has poor efficiency properties. It is best used in diagnostics. Rocke & Woodruff (1993) provide an excellent overview of the computation of robust estimators including the S-estimators of Rousseeuw & Yohai (1984) as well as M-estimators and minimum volume ellipsoid or minimum covariance determinant estimators.

2. THE ESTIMATOR

The k-dimensional location estimator $\hat{\theta}$ is chosen to be the solution of

$$S(\theta, \hat{A}_\theta) \equiv n^{-1} \sum_{i=1}^{n} \frac{\hat{A}_\theta(x_i - \theta)}{\|\hat{A}_\theta(x_i - \theta)\|} = 0, \tag{1}$$

in which $\hat{A}_\theta$ is a $k \times k$ upper triangular positive definite matrix, with a one in the upper left-hand element, chosen to satisfy

$$n^{-1} \sum_{i=1}^{n} \frac{A_\theta(x_i - \theta)(x_i - \theta)' A_\theta'}{\|A_\theta(x_i - \theta)\|^2} = k^{-1} I, \tag{2}$$

where I denotes the $k \times k$ identity matrix. Equation (1) shows that $\hat{\theta}$ is the point at which

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

the mean unit vector of the transformed data, centred at $\hat{\theta}$, is the zero vector. The transformation $\hat{A}$ is chosen in (2) so that the sample variance-covariance matrix of the unit vectors of the transformed data is k^{-1} times the identity matrix, i.e. the variance-covariance matrix of a random variable that is uniformly distributed on the unit k-sphere. Note that both criteria (1) and (2) use the directions of the data points from θ and not their distances from θ. Thus it is clearly appropriate to consider this point $\hat{\theta}$ to be a directional centre of the data. The transformation $\hat{A}_\theta$ was described and developed by Tyler (1987), who also mentioned, but did not develop, the associated location vector $\hat{\theta}$. Maronna et al. (1992) include the Tyler scatter matrix centred with the above proposed location estimator in a comparison of several scatter matrix estimators. The estimator $\hat{\theta}$ is also the location estimator that corresponds to a directional multivariate sign test described by Randles (2000). The transformation $\hat{A}_\theta$ chosen to satisfy (2) is unique up to multiplication by some positive constant, but the constant plays no role in the solution of equation (1). Without loss of generality we have taken the upper left-hand element of $\hat{A}_\theta$ to be one, scaling the matrix appropriately and thus making its solution unique. When the data are univariate, that is $k=1$, then $\hat{A}_\theta \equiv 1$ and $\hat{\theta}$ denotes the usual univariate sample median.

From Maronna (1976) simultaneous M-estimators of location and scatter can be expressed as solutions of

$$\frac{1}{n}\sum_{i=1}^{n} u_1(d_i)A(x_i-\theta)=0, \quad \frac{1}{n}\sum_{i=1}^{n} u_2(d_i)A(x_i-\theta)(x_i-\theta)'A'=I,$$

where $d_i = \| A(x_i-\theta)\|^2$. Then the estimators studied here correspond to $u_1(d)=1/\sqrt{d}$ and $u_2(d)=k/d$.

Proposition. *The location estimator described by* (1) *and* (2) *is affine equivariant; that is, if $\hat{\theta}_x$ is the estimator based on $x_1, \ldots, x_n$, and if $\hat{\theta}_y$ is the estimator based on $y_1, \ldots, y_n$, where $y_i = Dx_i + d$ for D an arbitrary $k \times k$ nonsingular matrix and d an arbitrary $k \times 1$ vector, then*

$$\hat{\theta}_y = D\hat{\theta}_x + d.$$

The proof of this proposition follows because the estimators are M-estimators.

The process of computing $(\hat{\theta}, \hat{A}_{\hat{\theta}})$ involves two routines. The first routine finds the θ-value that solves equation (1) using a fixed value for $\hat{A}$. This is accomplished by forming $y_i = \hat{A}x_i$ and finding $\hat{\theta}_y$ as the θ-value that minimises $\sum_{i=1}^{n} \| y_i - \theta \|$. The desired solution to (1) is then $\hat{\theta}_x = \hat{A}^{-1}\hat{\theta}_y$. This is the transformation-retransformation process described by Chakraborty et al. (1998).

The second routine, which is a bit more complicated, finds the $\hat{A}_\theta$ value that solves equation (2) using a fixed value for θ. It is an iterative process that begins with

$$S_0 = n^{-1}\sum_{i=1}^{n} (x_i-\theta)(x_i-\theta)'/\| x_i - \theta \|^2$$

and then forms $\hat{A}_0 = \mathrm{Chol}(S_0^{-1})$. Here $\mathrm{Chol}(M)$ denotes the upper triangular Choleski factorisation of the matrix M, divided by the upper left-hand element of that upper triangular matrix. At the tth iteration, form

$$\begin{aligned} \hat{A}_{dt} &= \hat{A}_{t-1}\hat{A}_{t-2}\ldots\hat{A}_0, \\ V_{it} &= \hat{A}_{dt}(x_i-\theta)/\| \hat{A}_{dt}(x_i-\theta)\|, \\ S_t &= n^{-1}\sum_{i=1}^{n} V_{it}V_{it}'. \end{aligned} \tag{3}$$

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

If $\|S_t - k^{-1}I\|$ is sufficiently small, where $\|M\|$ here is the square root of the sum of the squared elements in the matrix M, then stop and set $\hat{A}_\theta = \hat{A}_{dt}$. If not, compute $\hat{A}_t = \text{Chol}(S_t^{-1})$ and go back to (3). This process was described by Tyler (1987). Use of modern computing languages with built-in minimisation routines and matrix operations enables both of these routines to provide solutions very rapidly.

The process for determining $(\hat{\theta}, \hat{A}_{\hat{\theta}})$ alternates between these two routines to obtain the simultaneous solution in an iterative manner. To obtain the starting values, let $\theta_{0i} = x_i$, the ith data point. Use the second routine to obtain $\hat{A}_{0i}$ for this fixed value of θ. The starting $(\hat{\theta}_0, \hat{A}_0)$ is the $(\theta_{0i}, \hat{A}_{0i})$ pair that minimises

$$S(\theta_{0i}, \hat{A}_{0i})' S(\theta_{0i}, \hat{A}_{0i}).$$

This inner product times nk is the value of the multivariate sign test statistic proposed by Randles (2000) using θ_{0i} as the null hypothesis location value.

In general, the rth stage of the solution process uses a fixed $\hat{A}_{r-1}$ and the first routine to determine $\hat{\theta}_r$, and then uses the fixed $\hat{\theta}_r$ and the second routine to determine $\hat{A}_r$. The process is repeated until $\hat{\theta}_r$ converges.

The estimator $\hat{\theta}$ comes with an associated transformation $\hat{A}_{\hat{\theta}}$. While it does not directly estimate the variance-covariance matrix, $(\hat{A}_{\hat{\theta}}' \hat{A}_{\hat{\theta}})^{-1}$ is an estimator of a scatter matrix, and when the underlying distribution is elliptically symmetric this scatter matrix is proportional to the variance-covariance matrix. In either case we consider $d(x) = (x - \hat{\theta})' \hat{A}_{\hat{\theta}}' \hat{A}_{\hat{\theta}} (x - \hat{\theta})$ as an appropriate measure of distance, particularly if the population is elliptically symmetric.

Conditions that guarantee the existence and uniqueness of simultaneous solutions to M-estimation equations have been proposed by Maronna (1976), Huber (1981, p. 215), Tyler (1988) and Kent & Tyler (1991). The equations defining the proposed $(\hat{\theta}, \hat{A})$ do not fully satisfy any given set of conditions. Tyler (1987) pointed this out and it remains an open question. For fixed $\hat{A}$ there exists a unique solution for $\hat{\theta}$, and, for fixed $\hat{\theta}$, $\hat{A}$ exists and is unique up to multiplication by a scalar. In our examples and simulations, the above algorithms always converged, but there is no general proof of convergence for simultaneous solutions. Huber (1981, p. 238) indicates that convergence holds for $\hat{A}$ when $\hat{\theta}$ is fixed and vice versa. Rather than use the double-loop algorithm, it is possible to use a single loop and define

$$\theta_{t+1} = \left\{ \sum_{i=1}^{n} u_1(d_{it}) \right\}^{-1} \sum_{i=1}^{n} u_1(d_{it}) x_i$$

$$\Sigma_{t+1} = n^{-1} \sum_{i=1}^{n} u_2(d_{it})(x_i - \theta_{t+1})(x - \theta_{t+1})',$$

where $d_{it} = \|A_t(x_i - \theta_t)\|^2$ and $\Sigma_t = (A_t' A_t)^{-1}$. Hampel et al. (1986, p. 289) say that this approach may be slow and numerically unstable. While it may be possible to pursue this approach to computation, we have not done so here since the double-loop algorithm described above seems to be quite efficient; see also Rocke & Woodruff (1993), in which the single-loop algorithm is used to compute simultaneous M-estimators of location and scatter.

3. LARGE-SAMPLE PROPERTIES

In this section we develop the large-sample properties of $\hat{\theta}$ under the assumption that $X_1, \ldots, X_n$ is a random sample from a population that is continuous and directionally

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

symmetric around θ_0, that is that $(X-\theta_0)/\|X-\theta_0\|$ has the same distribution as $-(X-\theta_0)/\|-(X-\theta_0)\|$. This is a weaker assumption about the underlying population than assuming it is symmetric around θ_0, that is that $(X-\theta_0)$ and $-(X-\theta_0)$ have the same distribution, or assuming it is elliptically symmetric, for example. It is easy to show, under the assumption of directional symmetry around θ_0, that $A(X-\theta_0)/\|A(X-\theta_0)\|$ has the same distribution as $-A(X-\theta_0)/\|-A(X-\theta_0)\|$ for every nonsingular $k\times k$ matrix A, and thus

$$E\left(\frac{A(X-\theta_0)}{\|A(X-\theta_0)\|}\right)=0. \tag{4}$$

Tyler (1987) argues that, corresponding to a fixed θ_0, there is a unique, positive definite, upper triangular A_0 with a one in its upper left-hand entry satisfying

$$E\left(\frac{A_0(X-\theta_0)(X-\theta_0)'A_0'}{\|A_0(X-\theta_0)\|^2}\right)=k^{-1}I. \tag{5}$$

Thus, (θ_0, A_0) satisfies both (4) and (5). Moreover, they are the unique values which simultaneously satisfy both. Suppose there were some (θ_1, A_1) which also satisfied (4) and (5) with $\theta_1 \neq \theta_0$. Minimising $E\{\|A(X-\theta)\|\}$ with respect to θ, for a fixed A, is equivalent to solving (4). The minimisation of this norm produces a unique θ-value. However, θ_0 satisfies (4) for every A, including A_1. This contradicts $\theta_1 \neq \theta_0$. Therefore, under the assumption of directional symmetry, the population parameters (θ_0, A_0) satisfying (4) and (5) are unique.

The asymptotic distribution theory of $\hat\theta$ and $\hat A_{\hat\theta}$ is sketched in the Appendix. It shows that

$$\surd n(\hat\theta-\theta_0)\rightarrow N(0, H_{11}^{-1}\Sigma_{11}^{*}H_{11}^{-1'}),$$

in distribution, where $\Sigma_{11}^{*}=\operatorname{cov}(V)$ and

$$H_{11}=E\left\{\frac{1}{\|A_0(X-\theta_0)\|}(VV'-I)A_0\right\} \tag{6}$$

for $V=A_0(X-\theta)/\|A_0(X-\theta)\|$. The derivation also shows that the influence function of $\hat\theta$ is

$$-H_{11}^{-1}\frac{A_0(x-\theta_0)}{\|A_0(x-\theta_0)\|},$$

which is clearly bounded, an important aspect of the robustness of $\hat\theta$.

If we assume that the underlying distribution is elliptically symmetric with a density proportional to $\{\det(A_1)\}f_0\{\|A_1(x-\theta_0)\|^2\}$, where A_1 is a nonsingular matrix, then

$$\surd n(\hat\theta-\theta_0)\rightarrow N[0, (A_1'A_1)^{-1}k\{(k-1)E(R^{-1})\}^{-2}],$$

where $R=\|A(X-\theta)\|^{-1}$ has a density proportional to $r^{k-1}f_0(r^2)$. In this situation,

$$\frac{k}{n(k-1)^2}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|\hat A_{\hat\theta}(X_i-\hat\theta)\|}\right\}^{-2}(\hat A_{\hat\theta}'\hat A_{\hat\theta})^{-1},$$

is an easily computed estimator of the asymptotic variance-covariance matrix of $\hat\theta$.

Since both estimators are affine equivariant, the asymptotic relative efficiency of $\hat\theta$

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

compared to the sample mean $\bar{X}$ is also easily obtained for elliptically symmetric models and it does not depend on A_1. The efficiency expression is

$$\text{ARE}(\hat{\theta}, \bar{X}) = (k-1)^2 k^{-2} E(R^2)\{E(R^{-1})\}^2.$$

This has been evaluated for multivariate t distributions by Hettmansperger et al. (1994, Table 2) and for power family models by Randles (1989, Table 1). These verify that $\hat{\theta}$ is superior to $\bar{X}$ when the underlying distribution is heavy-tailed, and that it becomes more competitive with $\bar{X}$ for lighter-tailed distributions as the dimension increases.

4. Breakdown properties

In theoretical work about breakdown, the model is contaminated via a contaminating distribution and the functional defining the estimator is driven beyond all bounds. The theoretical notion was first discussed by Hampel (1971). The finite sample versions were described by Donoho & Huber (1983), who consider two types of breakdown. The first is the fraction of data that must be contaminated for breakdown and the second is the amount of data that must be added to the original data in order that the estimator breaks down. In the unpublished report cited earlier, Duembgen and Tyler consider all types of breakdown, model- and sample-based, for Tyler's M-estimator of scatter with known location. In all cases, for n sufficient large, such as $n > k^2$, the breakdown values are between $1/(k+1)$ and $1/k$, where k is the number of dimensions. This corresponds to results for general affine equivariant M-estimators of multivariate location and scatter first discussed by Maronna (1976); see also Huber (1981, p. 227).

Hence, for the unweighted approach described in this paper the breakdown is taken to be in the interval $[1/(k+1), 1/k]$ irrespective of the type of breakdown considered. The main point is that the estimator has positive breakdown. Note that the estimator proposed here is easily computed, unlike the Oja median, and has the same efficiency for elliptical models. The discouraging point is that the breakdown can be quite low for high dimensions.

In order to offset the decreasing breakdown we investigated weighted versions of both the location and scatter estimators. We experimented with various weightings and settled on the one presented here. Note that as the dimension increases the weights become more severe.

We solve the following system cyclically:

$$\left(\sum w_i\right)^{-1} \sum w_i \| A(X_i - \theta) \|^{-1} A(X_i - \theta) = 0,$$

$$\left(\sum w_i\right)^{-1} \sum w_i \{ \| A(X_i - \theta) \|^{-1} A(X_i - \theta)\}\{ \| A(X_i - \theta) \|^{-1} A(X_i - \theta)\}' = k^{-1} I,$$

$$w_i = \min(1, \exp[-(k-1)^2\{(d_i - M)/M\}]), \quad d_i = (X_i - \theta)' A' A (X_i - \theta),$$

$$M = \text{med}\{d_i, i = 1, \ldots, n\}.$$

These equations converge quite fast and pose no additional computational problem beyond those of the original unweighted equations. In fact, this system is equivalent to redescending M-estimation equations with

$$u_1(d) = d^{-\frac{1}{2}} \min(1, \exp[-(k-1)^2\{d - M)/M\}]), \quad u_2(d) = k d^{-\frac{1}{2}} u_1(d),$$

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

where M is data-dependent in this case. Hence, there may be multiple solutions and we no longer interpret a solution as a multivariate median even though, when $k=1$, the scheme produces a univariate median. When $k>1$, the point being estimated, while not a multivariate median, will at least be the point of symmetry, θ, if the population is symmetric. Furthermore, as Tyler (1991) points out, W. A. Stahel, in an unpublished research report from the Fachgruppe für Statistik at the Eidgenössische Technische Hochschule in Zürich, shows that at least one solution will have breakdown bounded above by $1/k$. It is not clear how to determine a 'good' solution. Nevertheless, the proposed weighted estimators do seem to be more resistant to location shift contamination. Note also that Kent & Tyler (1991) advocate re-descending M-estimators with weight functions that only approach but do not equal zero. This is the case with the estimators proposed here. Their class of M-estimators is not as general, however, as that of Huber (1981).

To check the breakdown values we generated samples of various sizes from different k-dimensional spherical normal distributions. We contaminated the data by adding 200 000 to the first coordinate of points chosen for contamination. This location shift contamination is one of several types studied by Rocke & Woodruff (1993). They found this type to be both challenging and realistic. In Table 1 we have included the case $k=10$ and $n=50$ to correspond to Rocke & Woodruff's Table 2. The weighted estimator has better location shift breakdown than the unweighted version, but its breakdown performance for the other contamination models studied by Rocke & Woodruff is the same as the unweighted, in being consistent with $1/k$ to $1/(k+1)$, except in the radial outlier model for which it has breakdown $\frac{1}{2}$, as with the others they studied.

Table 1. *Empirical breakdown points*

k	$[1/(k+1), 1/k]$	n	Breakdown [1]	Breakdown [2]
2	[0·33, 0·50]	15	0·33	—
		50	0·30	—
		100	0·31	0·48
5	[0·17, 0·20]	30	0·17	—
		100	0·16	0·40
10	[0·099, 0·100]	50	0·14	0·24

Breakdown [1], unweighted estimate location shift breakdown values: the smallest fraction needed to break down the solution. Breakdown [2], weighted estimate location shift breakdown values.

As in Rocke & Woodruff (1993), we tried good and bad starting points. The starting point did not affect the breakdown performance of the unweighted estimator, but it did affect that of the weighted estimator. For example, when $k=10$ and $n=50$, the breakdown was $\frac{10}{50}$ for a bad starting point and $\frac{13}{50}$ for a good one. The performance described in Table 1 uses the starting value recommended earlier in the paper.

From this small study we see that the weighted version of the estimator achieves higher location shift breakdown. Its computation remains easy in high dimensions and it retains the other desirable properties. We would recommend a bootstrap estimate of the covariance matrix of the weighted estimator since the weights depend on the data and the usual asymptotic formulae will no longer hold.

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

5. An example

We consider data from Andrews & Herzberg (1985, pp. 310–4) on skull measurements of species of grey kangaroos. Seven of the measurements made on each skull were basilar length, occipitonasal length, nasal length, nasal width, crest width, mandible width and mandible depth. For the 50 sample vectors of dimension seven from the *Macropus giganteus* species, we computed four location estimates, the sample mean, the componentwise median, the affine invariant estimate $\hat{\theta}$ described in this paper and the weighted version of that statistic $\hat{\theta}_w$. The results are shown in Table 2(a); those from 45 observations from the species *Macropus fuliginosus melanops* are shown in Table 2(b). While other affine equivariant extensions of the sample median to multivariate data would struggle with this number of seven-dimensional data points, these estimates were obtained instantly with a program written in OX. They are also easily computed using S-Plus.

Table 2. *Sample location vectors of kangaroo skull data for Macropus giganteus and fuliginosus melanops species*

(*a*) *Macropus giganteus*

Sample mean	(1491·4	1585·1	702·90	245·44	110·12	135·04	193·76)
Component median	(1490·5	1570·0	700·50	243·50	113·00	136·00	194·50)
$\hat{\theta}$	(1477·4	1572·3	694·92	243·77	111·67	134·49	192·31)
$\hat{\theta}_w$	(1443·9	1542·8	679·22	240·32	115·90	133·42	188·45)

(*b*) *Macropus fuliginosus melanops*

Sample mean	(1483·1	1564·8	673·40	230·13	113·82	134·78	190·00)
Component median	(1487·0	1588·0	681·00	235·00	120·00	135·00	196·00)
$\hat{\theta}$	(1471·6	1556·8	669·90	228·78	115·73	133·50	188·93)
$\hat{\theta}_w$	(1454·4	1549·3	667·37	227·80	116·24	131·14	188·27)

Acknowledgement

The authors would like to express their appreciation to the referees for their careful reading and suggested improvements.

Appendix

Asymptotic theory

Let $A = (a_{ij})$ denote an upper triangular matrix with a one in the upper left-hand corner. Define

$$\text{vec}(A) = (a_{12} \ldots a_{1p}\, a_{22} \ldots a_{2p} \ldots a_{pp})' \tag{A1}$$

as the $Q = k(k+1)/2 - 1$ vector of elements from A. Define $\eta = (\theta', \text{vec}(A)')'$ as the $k + Q$ vector of parameters. The equations in (4) and (5) define the functional equation

$$T(F, \eta_0) = 0. \tag{A2}$$

The linearisation of this functional equation produces

$$\sqrt{n}(\hat{\eta} - \eta_0) \doteq -H^{-1}\sqrt{n}T(F_n, \eta_0),$$

where

$$T(F_n, \eta_0) = \left(n^{-1} \sum_{i=1}^{n} V_i', \text{vec}\left(n^{-1} \sum_{i=1}^{n} V_i V_i' - k^{-1} I\right)'\right)',$$

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

$V_i = A_0(X_i - \theta_0)/\|A_0(X_i - \theta_0)\|$ and H is the Hessian matrix of the functional defined in (A2). The central limit theorem applies to this approximation, yielding

$$\sqrt{n}(\hat{\eta} - \eta_0) \to N(0, H^{-1}\Sigma^*(H^{-1})'),$$

in distribution. Here

$$\Sigma^* = \begin{pmatrix} \Sigma_{11}^* & 0 \\ 0 & \Sigma_{22}^* \end{pmatrix}$$

is block diagonal with $\Sigma_{11}^* = \text{cov}(V)$ and $\Sigma_{22}^* = \text{cov}\{\text{vec}(VV' - k^{-1}I)\}$. The Hessian

$$H = \begin{pmatrix} H_{11} & 0 \\ 0 & H_{22} \end{pmatrix}$$

is also block diagonal with H_{11} described in expression (6). If we index the elements of vec(.) via the double subscripts described in (A1), the element of H_{22} in the jj'th row and ii'th column, where $j \leqslant j'$ and $i \leqslant i'$, is

$$E\left\{V_{[j']}\frac{X_{[i']} - \theta_{0i'}}{\|A_0(X - \theta_0)\|}(1 - 2V_{[j]}^2)\right\} \quad (i = j < j'),$$

$$E\left\{V_{[j]}\frac{X_{[i']} - \theta_{0i'}}{\|A_0(X - \theta_0)\|}(1 - 2V_{[j']}^2)\right\} \quad (j < j' = i),$$

$$2E\left\{V_{[j]}\frac{X_{[i']} - \theta_{0i'}}{\|A_0(X - \theta_0)\|}(1 - V_{[j]}V_{[i]})\right\} \quad (i = j = j'),$$

$$-2E\left\{\frac{X_{[i']} - \theta_{0i'}}{\|A_0(X - \theta_0)\|}V_{[j]}V_{[j']}V_{[i]}\right\} \quad (i \neq j, i \neq j'),$$

where $X_{[i]}$ and $V_{[i]}$ denote the ith components of the generic vectors X and V, respectively.

References

Adrover, J. G. (1998). Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution. *Ann. Statist.* **26**, 2301–20.

Andrews, D. F. & Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker.* New York: Springer-Verlag.

Bedall, K. F. & Zimmermann, H. (1979). Algorithm AS 143. The mediancentre. *Appl. Statist.* **28**, 325–8.

Brown, B. M. (1983). Statistical uses of the spatial median. *J. R. Statist. Soc.* B **45**, 25–30.

Chakraborty, B., Chaudhuri, P. & Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica* **8**, 767–84.

Donoho, D. L. & Huber, P. J. (1983). The notion of a breakdown point. In *A Festschrift for Erich L. Lehmann*, Ed. P. J. Bickel, K. A. Doksum and J. L. Hodges Jr., pp. 157–84. Belmont, CA: Wadsworth.

Duembgen, L. (1998). On Tyler's *M*-functional of scatter in high dimensions. *Ann. Inst. Statist. Math.* **50**, 471–91.

Gini, C. & Galvani, L. (1929). Di talune estensioni dei concetti di media ai caratteri qualitative. *Metron* **8**, 3–209.

Gower, J. S. (1974). Algorithm AS 78. The mediancentre. *Appl. Statist.* **23**, 466–70.

Haldane, J. B. S. (1948). Note on the median on a multivariate distribution. *Biometrika* **35**, 414–5.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Statist.* **42**, 1887–96.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

Hayford, J. F. (1902). What is the center of an area or the center of a population? *J. Am. Statist. Assoc.* **8**, 47–58.

Hettmansperger, T. P. & McKean, J. W. (1998). *Robust Nonparametric Statistical Methods.* London: Arnold.

Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1-norm and Related Methods*, Ed. Y. Dodge, pp. 217–30. Amsterdam: North-Holland.

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions

KENT, J. T. & TYLER, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19**, 2102–19.

LOPUHAA, H. P. & ROUSSEEUW, P. J. (1991). Breakdown properties of affine equivariant estimators of multivariate location of covariance matrices. *Ann. Statist.* **19**, 229–48.

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51–67.

MARONNA, R. A., STAHEL, W. A. & YOHAI, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Mult. Anal.* **42**, 141–61.

NIINIMAA, A. & OJA, H. (1995). On the influence functions of certain bivariate medians. *J. R. Statist. Soc.* B **57**, 565–74.

NIINIMAA, A., OJA, H. & TABLEMAN, M. (1990). The finite-sample breakdown point of the Oja bivariate median. *Statist. Prob. Lett.* **10**, 325–8.

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Prob. Lett.* **1**, 327–32.

RANDLES, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Am. Statist. Assoc.* **84**, 1045–50.

RANDLES, R. H. (2000). A simpler, affine invariant, multivariate, distribution-free sign test. *J. Am. Statist. Assoc.* **95**, 1263–8.

ROCKE, D. M. & WOODRUFF, D. L. (1993). Computation of robust estimates of multivariate location and shape. *Statist. Neer.* **47**, 27–42.

RONKAINEN, T., OJA, H. & ORPONEN, P. (2002). Computation of the multivariate Oja median. In *Developments in Robust Statist.*, Ed. R. Dutter, P. Filzmoser, U. Gather and P. Rousseeuw. To appear. Heidelberg: Springer-Verlag.

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

ROUSSEEUW, P. J. & YOHAI, V. (1984). Robust regression by means of S estimators. In *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics, **26**, pp. 256–72. New York: Springer.

SMALL, C. G. (1990). A survey of multidimensional medians. *Int. Statist. Rev.* **58**, 263–77.

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15**, 234–51.

TYLER, D. E. (1988). Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter. *SIAM J. Sci. Statist. Comp.* **9**, 354–62.

TYLER, D. E. (1991). Some issues in the robust estimation of multivariate location and scatter. In *Directions in Robust Statistics and Diagnostics Part II*, Ed. W. Stahel and S. Weisberg, pp. 327–36. New York: Springer-Verlag.

[*Received January* 2001. *Revised May* 2002]

This content downloaded from 134.84.192.102 on Fri, 05 Feb 2016 05:54:14 UTC
All use subject to JSTOR Terms and Conditions