J. R. Statist. Soc. B (1998)
60, *Part* 1, *pp.* 145–157

On an adaptive transformation–retransformation estimate of multivariate location

Biman Chakraborty† and Probal Chaudhuri
Indian Statistical Institute, Calcutta, India

[Received August 1995. Final revision December 1996]

Summary. An affine equivariant estimate of multivariate location based on an adaptive transformation and retransformation approach is studied. The work is primarily motivated by earlier work on different versions of the multivariate median and their properties. We explore an issue related to efficiency and equivariance that was originally raised by Bickel and subsequently investigated by Brown and Hettmansperger. Our estimate has better asymptotic performance than the vector of co-ordinatewise medians when the variables are substantially correlated. The finite sample performance of the estimate is investigated by using Monte Carlo simulations. Some examples are presented to demonstrate the effect of the adaptive transformation–retransformation strategy in the construction of multivariate location estimates for real data.

Keywords: Affine transformation; Asymptotic efficiency; Equivariance; Generalized variance; Multivariate median

1. Introduction

Various versions of the multivariate median and their statistical properties have been extensively investigated (see Small (1990) and Chaudhuri (1992) for two recent detailed reviews). Bickel (1964) (see also Barnett (1976) and Babu and Rao (1988)) investigated the vector of medians, which is not equivariant under rotation and arbitrary affine transformation of the data, and compared it with the affine equivariant vector of means. One of Bickel's main conclusions was that, despite some very encouraging robustness as well as efficiency properties, the performance of this vector of medians becomes very poor when the real-valued components of the data vector are highly correlated. He expressed a strong suspicion that this pathological behaviour may be partly due to the lack of affine equivariance. A similar feeling has been expressed by Brown and Hettmansperger (1987), who discussed the issue in detail and recommended some affine equivariant procedures (see also Brown and Hettmansperger (1989)). However, they did not attempt to dig very deeply into this issue of a possible connection between affine equivariance of a multivariate location estimate and its asymptotic efficiency when the real-valued components of multivariate observations are substantially correlated.

In many situations, non-equivariant versions of the multivariate median will not be a very sensible location estimate for reasons that arise from simple and natural geometric considerations. The problem of locating the 'geographical centre' of the population in a country, which has been extensively discussed by Small (1990) and Chaudhuri (1996) in connection with the population of the USA, is an excellent example making a rather

†*Address for correspondence*: Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India.
E-mail: res9421@isical.ernet.in

 1369–7412/98/60145

convincing case for the affine equivariant multivariate median. It is necessary to have the affine equivariance of such a location estimate because, if the estimate lacks affine equivariance, we shall obtain different centres of the same population just by choosing different co-ordinate systems, i.e. rotations of the map of the country will lead to different geographical centres of its population, which is not at all desirable.

Several affine equivariant versions of the multivariate median have been proposed (see for example Tukey (1975), Oja (1983) and Liu (1990)). But all are computationally quite intensive especially when the dimension of the data vector is large. This is primarily because each is defined as the solution of a complex minimization problem and cannot be expressed as a simple function of the data in a closed form. Moreover, their sampling distributions (even asymptotic distributions) are typically not easy to derive, and it is often very difficult to estimate their sampling variations from the data. Though not affine equivariant, the vector of medians, however, is very easy to compute as it is based on several univariate medians, and for the same reason its sampling distribution and related matters are fairly easy to work out. Chakraborty and Chaudhuri (1996) proposed a data-driven transformation and retransformation strategy for creating an affine equivariant version of the multivariate median from the non-equivariant vector of univariate medians. The purpose of this paper is to investigate in detail the properties of this multivariate location estimate when the transformation used in the construction of the estimate is chosen in an adaptive data-based way. Also, we shall explore the intriguing connection between affine equivariance and asymptotic efficiency of this location estimate in the presence of correlation between the variables observed.

In Section 2, we briefly describe our adaptive transformation and retransformation procedure, discuss some of its main features and demonstrate its usefulness in locating the geographical centre of the population of a country. In Section 3, we present some results related to the asymptotic properties of the estimate proposed, and we show that the estimate is always at least as efficient as the vector of medians and performs significantly better when there are high correlations between the variables in the data. We also present some simulation results for small samples drawn from some standard bivariate probability distributions to demonstrate the performance of our estimate in finite sample situations. In Section 4, we apply our techniques to some real data sets. There we estimate the generalized variance of the proposed estimate of multivariate location by using the bootstrap method and observe that this adaptive equivariant estimate outperforms the vector of medians in many cases, though not always. On the basis of this critical observation, we suggest a rule for deciding when we shall benefit by using the adaptive equivariant estimate and when the non-equivariant vector of medians will suffice. Section 5 concludes the paper with some brief remarks on the issues that have transpired in the course of our investigation. All the proofs are presented in Appendix A.

2. Adaptive transformation–retransformation estimate

We begin by introducing some notation following Chakraborty and Chaudhuri (1996). Consider data points $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$. Unless specified otherwise, all vectors in this paper will be column vectors, and the superscript T will be used to denote the transpose of a vector or a matrix. Define

$$S_n = \{\alpha | \alpha \subseteq \{1, 2, \ldots, n\} \text{ and } |\alpha| = d+1\},$$

which is the collection of all subsets of size $d+1$ of $\{1, 2, \ldots, n\}$. For a fixed $\alpha = \{i_0, i_1, \ldots, i_d\} \in S_n$, let $\mathbf{X}(\alpha, i_0)$ be the $d \times d$ matrix whose columns are the random vectors

$(X_i - X_{i_0})$ with $i \in \alpha$ and $i \neq i_0$. We assume here that $1 \leqslant i_1 < i_2 < \ldots < i_d \leqslant n$ and $i_0 \neq i_k$ for $k = 1, 2, \ldots, d$. If the X_i are independent and identically distributed (IID) observations with a common probability distribution that happens to be absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, $\mathbf{X}(\alpha, i_0)$ must be an invertible matrix with probability 1. Define, for each $i \notin \alpha$, $Y_i^{(\alpha,i_0)} = \mathbf{X}(\alpha, i_0)^{-1} X_i$, and set $\hat{\phi}_n^{(\alpha,i_0)}$ to be the vector of medians of the co-ordinates of the d-dimensional transformed observations $Y_i^{(\alpha,i_0)}$. Then the multivariate median $\hat{\theta}_n^{(\alpha,i_0)}$ for the original data is defined by retransforming $\hat{\phi}_n^{(\alpha,i_0)}$ as $\hat{\theta}_n^{(\alpha,i_0)} = \mathbf{X}(\alpha, i_0)\hat{\phi}_n^{(\alpha,i_0)}$. The asymptotic behaviour of the affine equivariant location estimate $\hat{\theta}_n^{(\alpha,i_0)}$ has been worked out in detail in section 3 in Chakraborty and Chaudhuri (1996), and we shall only sketch the main results here.

Suppose that the underlying distribution of the X_i is absolutely continuous with a common density $h(\mathbf{x})$. Then we know from the discussion at the beginning of section 3 in Chakraborty and Chaudhuri (1996) that, if $h(\mathbf{x})$ is such that any real-valued linear function of X_i has a differentiable and positive density, $\hat{\theta}_n^{(\alpha,i_0)}$ is an $n^{1/2}$-consistent and asymptotically normal location estimate. Further, this limiting multivariate normal distribution takes an interesting form when h is elliptically symmetric, i.e.

$$h(\mathbf{x}) = \det(\Sigma)^{-1/2} f\{(\mathbf{x} - \theta)^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \theta)\},$$

where $\theta \in \mathbb{R}^d$ is the location of symmetry, Σ is a $d \times d$ positive definite matrix and $f(|\mathbf{x}|)$ is a spherically symmetric density on $\mathbb{R}^d$. Let us write

$$\{\Sigma^{-1/2}\, \mathbf{X}(\alpha, i_0)\}^{-1} = R(\alpha, i_0)\, J(\alpha, i_0),$$

where $R(\alpha, i_0)$ is a diagonal matrix with positive diagonal entries and $J(\alpha, i_0)$ is a matrix whose rows are of unit length, i.e. the rows of $J(\alpha, i_0)$ are obtained by normalizing the rows of $\{\Sigma^{-1/2}\, \mathbf{X}(\alpha, i_0)\}^{-1}$, and the diagonal elements of $R(\alpha, i_0)$ are the lengths of those rows. Then it follows from theorem 3.1 in Chakraborty and Chaudhuri (1996) that, if the univariate marginals of f are differentiable and positive at 0, $\hat{\theta}_n^{(\alpha,i_0)}$ is an $n^{1/2}$-consistent and asymptotically normal estimate of θ, and its conditional asymptotic generalized variance given the X_i for $i \in \alpha$ is

$$(c/n)^d \det(\Sigma) \det\{D(\alpha, i_0)\} \det\{J(\alpha, i_0)\}^{-2}.$$

Here $c = \{2\, g(0)\}^{-2}$, g being any univariate marginal of the spherically symmetric density f, and $D(\alpha, i_0)$ is the $d \times d$ matrix whose (i, j)th element is $(2/\pi) \sin^{-1}(\gamma_{ij})$, γ_{ij} being the inner product of the ith and the jth row of $J(\alpha, i_0)$.

Consider next the symmetric positive definite matrix

$$V(\alpha, i_0) = J(\alpha, i_0)^{-1}\, D(\alpha, i_0)\, \{J(\alpha, i_0)^{\mathrm{T}}\}^{-1}.$$

It was established in theorem 3.2 in Chakraborty and Chaudhuri (1996) that $\det\{V(\alpha, i_0)\} = v(\alpha, i_0)$ (say) $\geqslant 1$. Our adaptive procedure to select the best subset $\alpha \in S_n$ and $i_0 \in \alpha$ can be described as follows. First obtain some consistent estimate of the scale matrix Σ, say $\hat{\Sigma}$, that is equivariant under non-singular linear transformation of the data. Then normalize each data point X_i by multiplying by $\hat{\Sigma}^{-1/2}$ to define $Z_i = \hat{\Sigma}^{-1/2} X_i$ for $1 \leqslant i \leqslant n$. For each $\alpha \in S_n$ and $i_0 \in \alpha$, compute $\hat{J}(\alpha, i_0)$, $\hat{D}(\alpha, i_0)$ and $\hat{V}(\alpha, i_0)$ on the basis of the Z_i (instead of the X_i) as described before. Then minimize $\det\{\hat{V}(\alpha, i_0)\}$ $(= \hat{v}(\alpha, i_0)$, say) over all possible choices of $\alpha \in S_n$ and $i_0 \in \alpha$, and suppose that $\hat{\alpha}$ and $\hat{i}_0$ are some minimizers of this estimated conditional generalized variance. Form $\mathbf{X}(\hat{\alpha}, \hat{i}_0)$, and use it to compute the adaptive transformation–retransformation estimate $\hat{\theta}_n^{(\hat{\alpha},\hat{i}_0)}$ from the original observations X_i. The term 'adaptive' is being used here to indicate the data-based selection of α as well as i_0 that is required to construct the transformation matrix $\mathbf{X}(\alpha, i_0)$.

As an illustration of the methodology, let us now consider the following example where we locate the geographical centre of the Indian population by using the transformation–retransformation median computed from decennial census data. This example will demonstrate the usefulness of this affine equivariant location estimate as a multivariate descriptive statistic before we start exploring its asymptotic efficiency and related matters in the following sections.

2.1. Example 1

To estimate the geographical centre of a population distribution, earlier statisticians used the centroid (i.e. the usual multivariate mean) but observed that the centroid may be highly sensitive to the influence of probability masses at the extremes (see for example Small (1990) and Chaudhuri (1996)). In other words an event like a death or a birth at the periphery of the country tends to have more influence on the centroid of the population than does a similar event in the central part of the country. This motivates the use of a median-like measure of the centre of a population. For India, we have used the data obtained in census years during the period 1872 to 1971 and considered only the populations of type I towns (as classified in 1971), which cover nearly 80% of the population. The rest of the population is scattered in smaller towns and villages, which have an insignificant effect on the estimation of the centre of the population, and by ignoring them we have substantially reduced the time required for the compilation of the data and subsequent numerical computation. As the radius of the earth is very large compared with the size of India, we have ignored the effect of the curvature of the earth in this example, and the population is regarded as living on an essentially flat surface in which the lines of latitude and those of longitude are assumed to be orthogonal straight lines. The geographical centres of population located by our transformation–retransformation median are given in Fig. 1.

3. Asymptotic results

In this section, we shall discuss the asymptotic performance of the adaptive transformation–retransformation estimate and establish some efficiency results. Suppose that $\alpha^* \in S_n$ and $i_0^* \in \alpha^*$ minimize $\det\{V(\alpha, i_0)\} = \upsilon(\alpha, i_0)$, and recall that $X_1, X_2, \ldots, X_n$ are IID observations with a common density h on $\mathbb{R}^d$, which need not be elliptically symmetric for the time being.

Theorem 1. Assume that h satisfies

$$\int_{\mathbb{R}^d} h(\mathbf{y})^{d+1} \, \mathrm{d}\mathbf{y} < \infty.$$

Then $\upsilon(\alpha^*, i_0^*)$ converges to 1 *in probability* as $n \to \infty$.

Clearly, the integrability condition imposed on h in this theorem will hold if h happens to be a bounded density on $\mathbb{R}^d$. In the presence of elliptic symmetry with

$$h(\mathbf{x}) = \det(\Sigma)^{-1/2} f\{(\mathbf{x} - \theta)^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \theta)\},$$

this condition translates into an integrability condition on f, which is again trivially satisfied for any bounded spherically symmetric density f on $\mathbb{R}^d$. This theorem implies that, when the

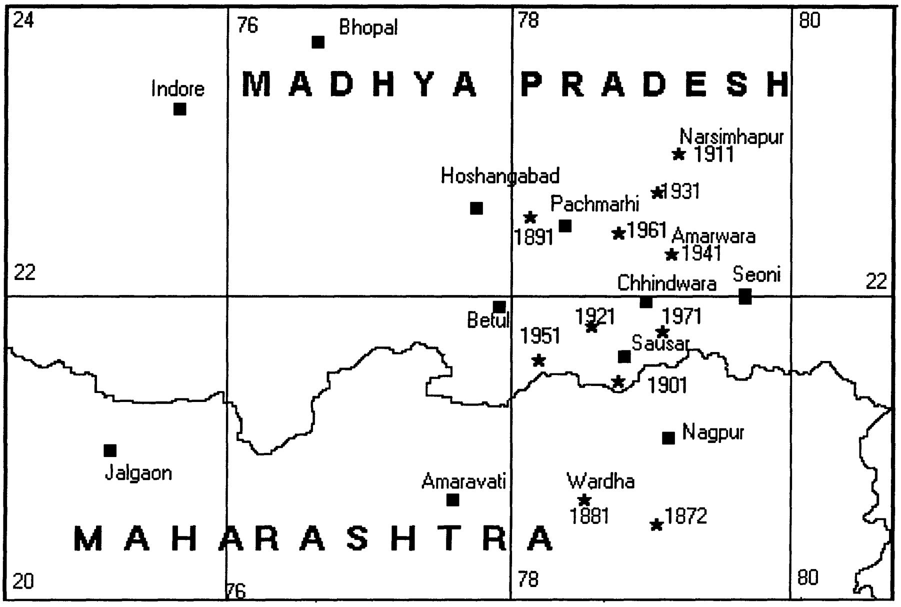

Fig. 1. Geographical centres (⋆) of the Indian population during 1872–1981

scale matrix Σ is known and the adaptive selection of α^* and S_n is done using that known Σ, the conditional generalized variance of the resulting transformation–retransformation estimate tends to the lower bound established in theorem 3.2 in Chakraborty and Chaudhuri (1996) (see our discussion in Section 2). However, in practice Σ is unknown, and we shall estimate it by a consistent and affine equivariant estimate $\hat{\Sigma}$ when we minimize $\hat{v}(\alpha, i_0)$ to obtain $\hat{\alpha}$ and $\hat{i}_0$. The next theorem says that the difference between $v(\hat{\alpha}, \hat{i}_0)$ and $v(\alpha^*, i_0^*)$ is asymptotically negligible.

Theorem 2. Under the condition assumed in theorem 1, $v(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)$ converges *in probability* to 0 as $n \to \infty$.

It follows from theorems 1 and 2 that both of $v(\alpha^*, i_0^*)$ and $v(\hat{\alpha}, \hat{i}_0)$ converge to 1, which is the lower bound discussed in section 2 following theorem 3.2 in Chakraborty and Chaudhuri (1996). Recall from this discussion that the asymptotic generalized variance of $\hat{\theta}_n^{(\alpha,i_0)}$ is $(c/n)^d \det(\Sigma)\, v(\alpha, i_0)$. Consequently, it now follows from theorems 1 and 2 that the adaptive selection of $\alpha \in S_n$ and $i_0 \in \alpha$ will produce an estimate with asymptotic generalized variance $(c/n)^d \det(\Sigma)$. As noted by Bickel (1964) and Babu and Rao (1988), the asymptotic generalized variance of the vector of medians is $(c/n)^d \det(\Gamma)$, where the (i, j)th element of Γ is

$$(\sigma_{ii}\sigma_{jj})^{1/2}(2/\pi)\sin^{-1}(\rho_{ij}),$$

$\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$, σ_{ij} is the (i, j)th element of Σ and c is as defined earlier. Following the line of arguments used in the proof of theorem 3.2 in Chakraborty and Chaudhuri (1996), it is easy to see that $\det(\Gamma) \geqslant \det(\Sigma)$, and equality holds only if Σ is a diagonal matrix. If the asymptotic efficiency of two competing estimates of a d-dimensional location parameter is now defined as the dth root of the ratio of their asymptotic generalized variances, the efficiency of our adaptive equivariant estimate compared with the non-equivariant vector of medians is always greater than or equal to 1. Further, the asymptotic efficiency of our estimate compared with the usual vector of means is the same as the efficiency of the sample median compared with the sample mean in the univariate problem, and it may be greater or smaller than 1 depending on the nature of the tail of the univariate marginal g of the d-variate spherically symmetric density f. These critical observations enable us to sense the subtle and intriguing connection between affine equivariance and asymptotic efficiency of multivariate versions of the median when there are correlations between the observed variables. They also provide an understanding of some related issues raised and discussed by Bickel (1964) and Brown and Hettmansperger (1987), which we have mentioned at the beginning of Section 1.

We close this section by presenting some simulation results to demonstrate the performance of the adaptive equivariant estimate in small samples. We have generated observations from the bivariate normal (i.e. $h(x, y) = (2\pi)^{-1} \exp\{-(x^2 + y^2)/2\}$) and Laplace (i.e. $h(x, y) = (2\pi)^{-1} \exp\{-\sqrt{(x^2 + y^2)}\}$) distributions with

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and $\theta = (0, 0)^{\mathrm{T}}$. We have used a set of five different values of ρ and two sample sizes, namely 20 and 30. Our adaptive equivariant estimate was compared with the non-equivariant vector of medians, and for the efficiency computation the estimates of their generalized variances were computed on the basis of 2000 Monte Carlo replications. The efficiency is taken to be the square root of the ratio of the generalized variances of the two competing bivariate location estimates.

It is apparent from Tables 1 and 2 that even with small sample sizes there is a gain in efficiency when the adaptive equivariant estimate is used instead of the non-equivariant vector of univariate medians if the correlation between the variables is high. As ρ increases, the efficiency increases, and there is also an increase in efficiency with an increase in the sample size. In small samples, the gain in efficiency for the adaptive equivariant estimate seems to be more in the bivariate normal case than in the bivariate Laplace case.

Table 1. Efficiency figures for the bivariate normal distribution example

Sample size	*Results for the following values of ρ:*				
	0.75	*0.80*	*0.85*	*0.90*	*0.95*
20	1.1039	1.1876	1.2657	1.3702	1.6202
30	1.1447	1.2637	1.3031	1.3882	1.6849

Table 2. Efficiency figures for the bivariate Laplace distribution example

Sample size	*Results for the following values of ρ:*				
	0.75	*0.80*	*0.85*	*0.90*	*0.95*
20	1.0679	1.1035	1.1611	1.2533	1.4819
30	1.0746	1.1659	1.2314	1.4326	1.7864

4. Some real examples

In this section, we shall consider two real data sets and explore the effect of the adaptive transformation and retransformation strategy on their analysis. In both examples we shall estimate the generalized variances of the location estimates by the bootstrap method (see for example Efron (1982)). One of the primary motivations behind considering the transformation–retransformation estimate is that, once we have the desired transformation matrix $\mathbf{X}(\alpha, i_0)$, it is quite easy to compute the estimate as it involves only determining the vector of coordinatewise medians of the transformed observations $\mathbf{X}(\alpha, i_0)^{-1}X_i$ and then retransforming that vector of univariate medians. As a consequence, we can conveniently estimate the conditional generalized variance of the transformation–retransformation estimate by using the bootstrap method once $\alpha \in S_n$ and $i_0 \in \alpha$ are fixed and the transformation matrix is formed. In each case considered here, we used 10000 bootstrap replications to estimate the generalized variance, and it took only a negligible amount of time on a 486 personal computer equipped with a standard Fortran compiler. We note here that the sampling variation of any other affine equivariant multivariate median that has been proposed (e.g. Tukey (1975), Oja (1983) and Liu (1990)) is extremely difficult to estimate from the data. It is virtually impossible to use the bootstrap or other resampling techniques for any of them in practice owing to the complex computational problems associated with each of them in the case of high or even moderately high dimensional data.

4.1. Example 2

Our second example deals with the famous iris data analysed by R. A. Fisher and many eminent statisticians by assuming multivariate normality. We have applied our technique of adaptive transformation and retransformation to all three different species considered in this data set, namely *Iris Setosa*, *Iris Versicolour* and *Iris Virginica*. Each data point in the set is four dimensional with variables sepal length, sepal width, petal length and petal width, and there are 50 observations for each species. Table 3 gives the adaptive transformation–retransformation medians and their estimated root-mean-squared errors (RMSEs) for these variables separately for the three different species.

The estimated correlation matrices of the sample medians for the three iris species are

$$\begin{pmatrix} 1.0 & 0.81 & 0.33 & 0.25 \\ & 1.0 & 0.22 & 0.27 \\ & & 1.0 & 0.31 \\ & & & 1.0 \end{pmatrix}, \quad \begin{pmatrix} 1.0 & 0.50 & 0.75 & 0.24 \\ & 1.0 & 0.61 & 0.72 \\ & & 1.0 & 0.53 \\ & & & 1.0 \end{pmatrix}, \quad \begin{pmatrix} 1.0 & 0.78 & 0.72 & 0.52 \\ & 1.0 & 0.79 & 0.74 \\ & & 1.0 & 0.84 \\ & & & 1.0 \end{pmatrix}$$

In addition to the adaptive equivariant estimate, we have computed the non-equivariant vector of medians and estimated the generalized variances for both of them in each species to

Table 3. Transformation–retransformation medians and their estimated RMSEs for the iris data

Species	*Sepal length (cm)*	*Sepal width (cm)*	*Petal length (cm)*	*Petal width (cm)*
Setosa	4.99	3.39	1.46	0.23
	(0.0690)	(0.0704)	(0.0285)	(0.0161)
Virginica	6.4456	2.9658	5.4039	2.0434
	(0.1264)	(0.0534)	(0.0769)	(0.0640)
Versicolour	6.0355	2.8285	4.3511	1.3482
	(0.1319)	(0.0549)	(0.0973)	(0.0475)

make a comparison. Interestingly, the equivariant estimate turns out to be more efficient than the non-equivariant estimate for *Iris Versicolour* and *Iris Virginica* (estimated efficiencies being 1.9158 and 1.8259 respectively in the two cases), whereas it turns out to be less efficient in the case of *Iris Setosa* (estimated efficiency being only 0.8522).

4.2. Example 3

The data set used in the third example was originally obtained from the laboratory of Dr James S. Elliot, of the Urology Section, Veterans' Administration Medical Center, Palo Alto, California, and the Division of Urology, Stanford University School of Medicine, Stanford, California, and it is reported in Andrews and Herzberg (1985). We considered four physical characteristics of 33 urine specimens with calcium oxalate crystals. These variables are specific gravity (i.e. the density of urine relative to water), pH (i.e. the negative logarithm of the hydrogen ion concentration), osmolarity (which is proportional to the concentration of molecules in the solution) and conductivity (which is proportional to the concentration of charged ions in the solution). As we would expect, the correlations between these variables are fairly high and the estimated efficiency of the adaptive equivariant estimate compared with the non-equivariant vector of medians turns out to be 2.2870, i.e. the transformation–retransformation strategy significantly reduces the sampling variation in the location estimate in this case. The transformation and retransformation medians and their estimated RMSEs and correlation matrix are presented in Table 4.

It is clear from the preceding two examples that we sometimes (though not always) gain by using the adaptive equivariant estimate. Our analysis enables us to choose between the equivariant transformation–retransformation median and the non-equivariant vector of usual medians by using a simple and convenient rule after the sampling variations of the two multivariate location estimates have been estimated from the data. This leads to a way of dealing with the equivariance and efficiency problems in real data analysis.

5. Concluding remarks

Remark 1. Once the matrix $\mathbf{X}(\hat{\alpha}, \hat{i}_0)$ has been formed, the computation of $\hat{\theta}_n^{(\hat{\alpha},\hat{i}_0)}$ is straightforward as it does not require any further optimization or iterative computation. But the selection of the optimal $(\hat{\alpha}, \hat{i}_0)$ may require a search over $\binom{n}{d+1}$ possible subsets α, and this number grows very fast with n and d. We can reduce the amount of computation involved for searching the optimal (α, i_0) by stopping whenever $\hat{v}(\alpha, i_0)$ is sufficiently close to 1 because we know from theorem 3.2 of Chakraborty and Chaudhuri (1996) that the lower bound for

Table 4. Transformation–retransformation medians, their estimated RMSEs and correlations for the urine data

Variable	*Median*	*Correlation matrix*			
Specific gravity	1.0222 (0.0015)	1.00	−0.1161	0.9207	0.5223
pH	5.8718 (0.1253)		1.00	−0.2217	−0.4135
Osmolarity	730.1650 (55.3338)			1.00	0.7599
Conductivity (mΩ^{-1})	21.6264 (1.7926)				1.00

$\upsilon(\alpha, i_0)$ is 1 (see our discussion in Section 2). We have observed that this approximation makes the algorithm very fast without making any serious change in the sampling variation or any significant loss of efficiency of the resulting estimate. In all the examples that we have considered, it performed satisfactorily. An alternative approach would be to make a random search over different subsets α and different indices $i_0 \in \alpha$ and stopping when $\upsilon(\alpha, i_0)$ stabilizes in some appropriate sense. Approaches that are similar to this have been considered in computing the least median of squares estimates (see for example Rousseeuw and Leroy (1987)).

Remark 2. A version of the multivariate median, which is popularly known as the 'spatial median' (see for example Haldane (1948), Gower (1974), Brown (1983), Small (1990) and Chaudhuri (1992, 1996)), has received considerable attention. Though equivariant under rotation or other forms of orthogonal transformation of the data, the spatial median is not equivariant under an arbitrary scale change of different real-valued components of multivariate observations. This lack of scale equivariance makes it an inappropriate location estimate for data sets (e.g. those considered in examples 2 and 3), where the variables have widely different scales. It is not meaningful to compute the spatial median when different real-valued components of a multivariate data set are measured in different units. The foremost motivation behind considering the adaptive transformation–retransformation strategy is to come up with an affine equivariant version of the multivariate median that will be reasonably easy to compute even for high dimensional data. At the same time, it is very much desirable in practice that we have a convenient and computationally feasible way of estimating the sampling variation of our proposed location estimate. It has been amply demonstrated in Section 4 that we can comfortably use the bootstrap method to estimate the sampling variation of the adaptive transformation and retransformation median in finite sample situations involving high dimensional data. All these make our multivariate median quite attractive for potential practical applications.

Remark 3. As we have discussed in detail in Section 3, the concern about poor efficiency of the non-equivariant vector of univariate medians raised by Bickel (1964) and Brown and Hettmansperger (1987) can be settled by using our adaptive transformation and retransformation strategy. Asymptotically our equivariant estimate outperforms the non-equivariant vector of medians as well as the affine equivariant vector of means in the presence of correlation between the variables if the underlying distribution is elliptically symmetric with univariate marginals having heavy tails. Our simulation results amply indicate a gain in the efficiency over the vector of co-ordinatewise medians even in finite sample situations for standard elliptically symmetric distributions when the amount of correlations between the variables in the data is significant.

Remark 4. When the underlying distribution deviates significantly from being elliptically symmetric, instead of minimizing $\upsilon(\alpha, i_0)$, we can try to estimate the generalized variance of the transformation–retransformation median for a fixed α and i_0 by using some resampling technique, and then to minimize that estimated variance with respect to $\alpha \in S_n$ and $i_0 \in \alpha$. However, such an approach will be computationally quite intensive, and we shall not discuss it here. We conclude by pointing out that this adaptive transformation and retransformation strategy is essentially a way of finding an appropriate 'data-driven co-ordinate system' (see Chaudhuri and Sengupta (1993)) so that data points can be expressed in terms of that co-ordinate system before analysis and the computation of descriptive statistics (such as the median) to enable us to make efficient statistical inference.

Acknowledgements

The research presented here has been partially supported by a grant from the Indian Statistical Institute. The authors are grateful to the two referees for their careful reading of earlier versions of the paper and several comments and suggestions.

Appendix A: Proofs

A.1. Proof of theorem 1

Assume without loss of generality that Σ is the d-dimensional identity matrix. Consider $\alpha = \{1, 2, \ldots, d+1\}$ and $i_0 = 1$. As the underlying distribution of the X_i are IID with density h, the joint probability density function of $X_1, \ldots, X_{d+1}$ can be written as $\Pi_{i=1}^{d+1} h(\mathbf{x}_i)$. Now we make the following transformation of variables:

$$Y_1 = X_2 - X_1, \ldots, Y_d = X_{d+1} - X_1, \ Y_{d+1} = X_1.$$

Then the joint density of $Y_1, \ldots, Y_{d+1}$ is given by

$$h(\mathbf{y}_{d+1}) \prod_{i=1}^{d} h(\mathbf{y}_i + \mathbf{y}_{d+1}).$$

Therefore, the joint density of $Y_1, \ldots, Y_d$ at the origin in $\mathbb{R}^{d\times d}$ is

$$\int_{\mathbb{R}^d} h(\mathbf{y})^{d+1} \, d\mathbf{y},$$

which is finite and positive by the condition assumed in the statement of the theorem. This condition further implies that the map

$$(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_d) \mapsto \int_{\mathbb{R}^d} h(\mathbf{y}) \prod_{i=1}^{d} h(\mathbf{y}_i + \mathbf{y}) \, d\mathbf{y}$$

from $\mathbb{R}^{d\times d}$ to $\mathbb{R}$ is everywhere continuous. Therefore the joint density of $Y_1, \ldots, Y_d$ must remain bounded away from 0 in a neighbourhood of $\mathbf{0} \in \mathbb{R}^{d\times d}$. Consequently the probability of the event that the columns of $\mathbf{X}(\alpha, i_0)$ will be nearly orthogonal (and hence $v(\alpha, i_0) = \det\{\mathbf{V}(\alpha, i_0)\}$ will be very close to 1) is bounded away from 0, i.e. we have for any $\epsilon > 0$

$$\mathrm{pr}[\det\{V(\alpha, i_0)\} = v(\alpha, i_0) < 1 + \epsilon] = p_\epsilon > 0.$$

Let $\alpha_1, \alpha_2, \ldots, \alpha_{k_n}$ be disjoint subsets of S_n and $i_{0,j} \in \alpha_j$ for $1 \leqslant j \leqslant k_n$ such that $k_n \to \infty$ as $n \to \infty$ (for example k_n may be equal to $n/(d+1)$). Then

$$\begin{aligned} \mathrm{pr}\{v(\alpha^*, i_0^*) \geqslant 1 + \epsilon\} &= \mathrm{pr}\{\forall \alpha \in S_n \text{ and } i_0 \in \alpha, \ v(\alpha, i_0) \geqslant 1 + \epsilon\} \\ &\leqslant \mathrm{pr}\{v(\alpha_1, i_{0,1}) \geqslant 1 + \epsilon, \ldots, v(\alpha_{k_n}, i_{0,k_n}) \geqslant 1 + \epsilon\} \\ &= (1 - p_\epsilon)^{k_n} \to 0 \text{ as } n \to \infty. \end{aligned}$$ □

A.2. Preliminary results for proof of theorem 2

To prove theorem 2, we shall prove some preliminary results first.

Lemma 1. $\sup_{\alpha\in S_n} \sup_{i_0\in\alpha} |\hat{J}(\alpha, i_0) - J(\alpha, i_0)|$ converges in probability to 0 as $n \to \infty$.

Proof. Let us write $\mathbf{X}(\alpha, i_0)^{-1}\Sigma^{1/2} = R(\alpha, i_0)\, J(\alpha, i_0)$ and similarly $\mathbf{X}(\alpha, i_0)^{-1}\hat{\Sigma}^{1/2} = \hat{R}(\alpha, i_0)\, \hat{J}(\alpha, i_0)$, where $\hat{\Sigma}$ is a consistent estimate of Σ. Clearly, the rows of $J(\alpha, i_0)$ and $\hat{J}(\alpha, i_0)$ are just the normalized rows of $\mathbf{X}(\alpha, i_0)^{-1}\Sigma^{1/2}$ and $\mathbf{X}(\alpha, i_0)^{-1}\hat{\Sigma}^{1/2}$ respectively. Let the jth row of $\mathbf{X}(\alpha, i_0)^{-1}$ be u_j^{T}. Then

$$\frac{u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}}{|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|}-\frac{u_j^{\mathrm{T}}\Sigma^{1/2}}{|u_j^{\mathrm{T}}\Sigma^{1/2}|}=\frac{u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|u_j^{\mathrm{T}}\Sigma^{1/2}|-u_j^{\mathrm{T}}\Sigma^{1/2}|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|}{|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|\,|u_j^{\mathrm{T}}\Sigma^{1/2}|}$$

$$=\frac{u_j^{\mathrm{T}}(\hat{\Sigma}^{1/2}-\Sigma^{1/2})|u_j^{\mathrm{T}}\Sigma^{1/2}|+u_j^{\mathrm{T}}\Sigma^{1/2}(|u_j^{\mathrm{T}}\Sigma^{1/2}|-|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|)}{|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|\,|u_j^{\mathrm{T}}\Sigma^{1/2}|}.$$

Now, since $\hat{\Sigma}\xrightarrow{\mathrm{p}}\Sigma$ (a positive definite matrix) as $n\rightarrow\infty$, for sufficiently large n and any $d\times 1$ vector u, we must have

$$\frac{u^{\mathrm{T}}\hat{\Sigma}u}{u^{\mathrm{T}}u}\geqslant c^2$$

for some $c>0$. Then

$$\left|\frac{u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}}{|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|}-\frac{u_j^{\mathrm{T}}\Sigma^{1/2}}{|u_j^{\mathrm{T}}\Sigma^{1/2}|}\right|\leqslant\frac{2|\hat{\Sigma}^{1/2}-\Sigma^{1/2}|}{c}.$$

Therefore,

$$\sup_{\alpha\in S_n}\sup_{i_0\in\alpha}\sup_{j}\left|\frac{u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}}{|u_j^{\mathrm{T}}\hat{\Sigma}^{1/2}|}-\frac{u_j^{\mathrm{T}}\Sigma^{1/2}}{|u_j^{\mathrm{T}}\Sigma^{1/2}|}\right|\leqslant\frac{2|\hat{\Sigma}^{1/2}-\Sigma^{1/2}|}{c},$$

i.e. we must have

$$\sup_{\alpha\in S_n}\sup_{i_0\in\alpha}|\hat{J}(\alpha,i_0)-J(\alpha,i_0)|\leqslant c^*|\hat{\Sigma}^{1/2}-\Sigma^{1/2}|,$$

for some positive constant c^*. The proof is now complete in view of the fact that $\hat{\Sigma}$ is a consistent estimate of Σ. □

Lemma 2. $\sup_{\alpha\in S_n}\sup_{i_0\in\alpha}|\hat{J}(\alpha,i_0)\hat{J}(\alpha,i_0)^{\mathrm{T}}-J(\alpha,i_0)J(\alpha,i_0)^{\mathrm{T}}|$ converges in probability to 0 as $n\rightarrow\infty$.

Proof. First observe that

$$\begin{aligned}|\hat{J}(\alpha,i_0)\,\hat{J}(\alpha,i_0)^{\mathrm{T}}-J(\alpha,i_0)\,J(\alpha,i_0)^{\mathrm{T}}|&=|\hat{J}(\alpha,i_0)\,\hat{J}(\alpha,i_0)^{\mathrm{T}}-J(\alpha,i_0)\,\hat{J}(\alpha,i_0)^{\mathrm{T}}\\&\quad+J(\alpha,i_0)\,\hat{J}(\alpha,i_0)^{\mathrm{T}}-J(\alpha,i_0)\,J(\alpha,i_0)^{\mathrm{T}}|\\&\leqslant|\hat{J}(\alpha,i_0)||\hat{J}(\alpha,i_0)-J(\alpha,i_0)|+|J(\alpha,i_0)||\hat{J}(\alpha,i_0)-J(\alpha,i_0)|\\&\leqslant c'|\hat{J}(\alpha,i_0)-J(\alpha,i_0)|,\end{aligned}$$

where c' is some positive constant. The last inequality follows from the fact that the rows of $J(\alpha,i_0)$ and $\hat{J}(\alpha,i_0)$ are of unit length. The result now follows from lemma 1. □

Lemma 3. For $M>0$, define $K_M^n=\{(\alpha,i_0)\colon\alpha\in S_n,\ i_0\in\alpha\text{ and }v(\alpha,i_0)\leqslant M\}$. Then

$$\sup_{(\alpha,i_0)\in K_M^n}|\hat{v}(\alpha,i_0)-v(\alpha,i_0)|$$

converges in probability to 0 as $n\rightarrow\infty$.

Proof. From lemma 2, it is easy to see that

$$\sup_{\alpha\in S_n}\sup_{i_0\in\alpha}|\hat{D}(\alpha,i_0)-D(\alpha,i_0)|\xrightarrow{\mathrm{p}}0,$$

$$\sup_{\alpha\in S_n}\sup_{i_0\in\alpha}|\det\{\hat{J}(\alpha,i_0)\}^2-\det\{J(\alpha,i_0)\}^2|\xrightarrow{\mathrm{p}}0$$

and

$$\sup_{\alpha\in S_n}\sup_{i_0\in\alpha} |\det\{\hat{D}(\alpha, i_0)\} - \det\{D(\alpha, i_0)\}| \xrightarrow{P} 0 \qquad \text{as } n\to\infty.$$

Next, note that there exists $\delta > 0$ such that, for any $(\alpha, i_0) \in K_M^n$, $\det\{J(\alpha, i_0)\}^2 > \delta$. The existence of such a δ follows from some routine analysis using some of the arguments in the proof of theorem 3.2 in Chakraborty and Chaudhuri (1996). So, for sufficiently large n, with probability tending to 1, we have $\det\{\hat{J}(\alpha, i_0)\}^2 > \delta$. Therefore, for $(\alpha, i_0) \in K_M^n$,

$$\begin{aligned}|\hat{v}(\alpha, i_0) - v(\alpha, i_0)| &\leqslant \frac{|\det\{\hat{D}(\alpha, i_0)\} - \det\{D(\alpha, i_0)\}|}{\det\{\hat{J}(\alpha, i_0)\}^2}\\ &\quad + \frac{|\det\{D(\alpha, i_0)\}|\,|\det\{\hat{J}(\alpha, i_0)\}^2 - \det\{J(\alpha, i_0)\}^2|}{\det\{\hat{J}(\alpha, i_0)\}^2 \det\{J(\alpha, i_0)\}^2}\\ &\leqslant \frac{|\det\{\hat{D}(\alpha, i_0)\} - \det\{D(\alpha, i_0)\}| + |\det\{\hat{J}(\alpha, i_0)\}^2 - \det\{J(\alpha, i_0)\}^2|}{\delta^2}.\end{aligned}$$

Hence, we have the result □

A.3. Proof of theorem 2

From theorem 1, we have that the α^* and the i_0^* which minimize $v(\alpha, i_0)$ are in the set K_M^n, and hence in view of lemma 3 $(\hat{\alpha}, \hat{i}_0)$ will be in K_M^n with probability tending to 1 as $n \to \infty$ if $M > 0$ is chosen to be suitably large.

Next, since $\hat{\alpha}$ and $\hat{i}_0$ minimize $\hat{v}(\alpha, i_0)$, and α^* and i_0^* minimize $v(\alpha, i_0)$, it follows by some straightforward analysis that $|\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\hat{\alpha}, \hat{i}_0)| < \epsilon$ and $|\hat{v}(\alpha^*, i_0^*) - v(\alpha^*, i_0^*)| < \epsilon$ will imply that $|\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)| < \epsilon$. Hence

$$\operatorname{pr}\{|\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)| > \epsilon\} \leqslant \operatorname{pr}\{|\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\hat{\alpha}, \hat{i}_0)| > \epsilon\} + \operatorname{pr}\{|\hat{v}(\alpha^*, i_0^*) - v(\alpha^*, i_0^*)| > \epsilon\}.$$

At this point, it follows from lemma 3 that $\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)$ converges in probability to 0. The proof is now complete after observing the inequality

$$|v(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)| \leqslant |v(\hat{\alpha}, \hat{i}_0) - \hat{v}(\hat{\alpha}, \hat{i}_0)| + |\hat{v}(\hat{\alpha}, \hat{i}_0) - v(\alpha^*, i_0^*)|$$

and using lemma 3. □

References

Andrews, D. F. and Herzberg, A. M. (1985) *Data: a Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer.
Babu, G. J. and Rao, C. R. (1988) Joint asymptotic distribution of marginal quantile functions in sample from a multivariate population. *J. Multiv. Anal.*, **27**, 15–23.
Barnett, V. (1976) The ordering of multivariate data (with discussion). *J. R. Statist. Soc.* A, **139**, 318–354.
Bickel, P. J. (1964) On some alternative estimates for shift in the *P*-variate one sample problem. *Ann. Math. Statist.*, **35**, 1079–1090.
Brown, B. M. (1983) Statistical uses of the spatial median. *J. R. Statist. Soc.* B, **45**, 25–30.
Brown, B. M. and Hettmansperger, T. P. (1987) Affine invariant rank methods in the bivariate location model. *J. R. Statist. Soc.* B, **49**, 301–310.
———(1989) An affine invariant bivariate version of the sign test. *J. R. Statist. Soc.* B, **51**, 117–125.
Chakraborty, B. and Chaudhuri, P. (1996) On a transformation and re-transformation technique for constructing affine equivariant multivariate median. *Proc. Am. Math. Soc.*, **124**, 2539–2547.
Chaudhuri, P. (1992) Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Ann. Statist.*, **20**, 897–916.
———(1996) On a geometric notion of quantiles for multivariate data. *J. Am. Statist. Ass.*, **91**, 862–872.
Chaudhuri, P. and Sengupta, D. (1993) Sign tests in multidimension: inference based on the geometry of the data cloud. *J. Am. Statist. Ass.*, **88**, 1363–1370.
Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.
Gower, J. C. (1974) Algorithm AS 78: The mediancenter. *Appl. Statist.*, **23**, 466–470.

Haldane, J. B. S. (1948) Note on the median of a multivariate distribution. *Biometrika*, **35**, 414–415.
Liu, R. Y. (1990) On a notion of data depth based on random simplices. *Ann. Statist.*, **18**, 405–414.
Oja, H. (1983) Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, **1**, 327–332.
Rousseeuw, P. J. and Leroy, A. (1987) *Robust Regression and Outlier Detection.* New York: Wiley.
Small, C. G. (1990) A survey of multidimensional medians. *Int. Statist. Rev.*, **58**, 263–277.
Tukey, J. W. (1975) Mathematics and picturing data. In *Proc. Int. Congr. Mathematicians, Vancouver 1974* (ed. R. D. James), vol. 2, pp. 523–531. Vancouver: Canadian Mathematical Congress.