ON AFFINE EQUIVARIANT MULTIVARIATE QUANTILES

Biman Chakraborty
Department of Statistics and Applied Probability
The National University of Singapore
10, Kent Ridge Crescent, Singapore 119260
Republic of Singapore

Running Head: **Multivariate Quantiles**

Abstract

An extension of univariate quantiles in the multivariate set-up has been proposed and studied. The proposed approach is affine equivariant, and it is based on an adaptive transformation retransformation procedure. Bahadur type linear representations of the proposed quantiles are established and consequently asymptotic distributions are also derived. As applications of these multivariate quantiles, we develop some affine equivariant quantile contour plots which can be used to study the geometry of the data cloud as well as the underlying probability distribution and to detect outliers. These quantiles can also be used to construct affine invariant versions of multivariate Q-Q plots which are useful in checking how well a given multivariate probability distribution fits the data and for comparing the distributions of two data sets. We illustrate these applications with some simulated and real data sets. We also indicate a way of extending the notion of univariate L-estimates and trimmed means in the multivariate set-up using these affine equivariant quantiles.

Keywords and Phrases: Bahadur representation, L estimates, multivariate ranks, Q-Q plots, quantile contour plots, transformation-retransformation.

1 Introduction

The problem of finding out suitable analogs of quantiles for multivariate data has a long history in statistics. Univariate quantiles are quite popular for their usefulness in constructing useful descriptive statistics like the median, the inter-quartile range and various measures of skewness and kurtosis. They are also used in constructing robust L-estimates of location. As there is no inherent ordering in multidimension, extending the notion of quantiles poses a big problem. In a classic paper, Barnett (1976) reviewed different possible techniques for ordering multivariate data (see also Plackett, 1976 and Reiss, 1989). Brown and Hettmansperger (1987, 1989) have proposed a notion of bivariate quantiles based on Oja's simplicial median (see Oja, 1983). Eddy (1983, 1985) proposed an interesting approach to define multivariate quantiles using certain nested sequence of convex sets. Very recently, Chaudhuri (1996) and Koltchinskii (1997) proposed the notion of geometric or spatial quantile which generalizes the notion of spatial median that has been studied by earlier authors (see e.g. Brown 1983, Chaudhuri 1992). Chaudhuri (1996) indexed multivariate geometric quantiles, based on Euclidean distances, using the elements of d-dimensional open unit ball. The corresponding quantiles not only give the idea of 'extreme' or 'central' observations but also about their orientations in the data cloud. He also presented a Bahadur type representation for the geometric quantiles and indicated various ways of extending these quantiles to L-estimates, regression quantiles etc. Recently, Marden (1998) proposed some analogs of bivariate Q-Q plots based on geometric quantiles. These bivariate Q-Q plots can be used in comparing a sample to a given population distribution and they may reveal differences in location, scale and skewness, as well as outliers.

Whether the notion of multivariate quantiles would be based on some univariate concept of ordering or on some vector valued concept of ranks is a debatable issue. In many ways it seems to be a good idea to make use of the orientation information in any version of multivariate quantile. That is the only way in which one can talk about the 'high points' and the 'low points' in a multivariate data cloud. In a multivariate situation an observation may have 'high' values in some direction but 'low' values in some other direction. To capture these intrinsic geometric features of the multivariate data cloud, it seems reasonable to index the multivariate quantiles by some multivariate quantities, which will give us a way of measuring the closeness (or deviation) of a specific data point to (or from) the center of the data cloud as well as its spatial orientation with respect to the data cloud. Brown and Hettmansperger (1987, 1989) introduced a notion of bivariate quantile which is based on their definition of multivariate ranks derived from Oja's criterion function (cf. Oja, 1983). The problem with their approach is that the criterion function used by them is not 'self-normalized' in the sense that it is the gradient vector of Oja loss function based on simplicial volumes and is not bounded. For certain losses and distances, the gradient leads to 'self-normalized' orientation information. For instance, the gradient vector of the function $f(x_1, x_2, \ldots, x_d) = |x_1| + |x_2| + \cdots + |x_d|$ (i.e. the l_1-norm) is the coordinatewise sign vectors for which each coordinate is bounded by 1. If $f(x_1, x_2, \ldots, x_d) = (x_1^2 + x_2^2 + \cdots + x_d^2)^{1/2}$ (i.e. the l_2-norm), the gradient is a unit direction vector [see Möttönen and Oja (1995) and Möttönen, Oja and Tienari (1997) for the notion of the ranks of the data points constructed using such a gradient]. The advantage of using 'self-normalized' orientation information is that it becomes easy to interpret what is 'high' and what is 'low' in a multidimensional setting.

The problem with geometric quantiles (Chaudhuri 1996, Koltchinskii 1997) is that they are not equivariant under arbitrary affine transformations though they are equivariant under rotations of the data cloud. Due to lack of affine equivariance, these geometric quantiles do not lead to any sensible estimate when the different coordinate variables of the data-vectors are measured in different units or they have different degrees of statistical variations. In this paper we have used a transformation retransformation approach based on a 'data-driven coordinate system' (cf. Chaudhuri and Sengupta, 1993) to construct affine equivariant estimates of multivariate quantiles. In Section 2, we introduce the notion of l_p-quantiles and a proper indexing for them. Then with the help of transformation retransformation methodology, we extend l_p-quantiles to a family of affine equivariant multivariate quantiles and explore their basic properties with regard to uniqueness, existence and computation. In Section 3, we discuss asymptotic behavior of multivariate quantiles. We establish a Bahadur-type linear representation and use it to derive asymptotic distributions of sample quantiles. In the same section, we indicate a procedure to select a suitable 'data-driven coordinate system' and discuss a few interesting results related to that. In Section 4, we present some applications of our proposed quantiles. In particular, we discuss construction of quantile based contour plots for distributions and indicate a procedure for multivariate generalization of Q-Q plots and demonstrate with some simulation results and real data sets about how they can be used in comparing a multivariate sample to a given distribution. We also construct L-estimates and trimmed mean estimates for multivariate location based on these multivariate quantiles. All the proofs are relegated to the Appendix.

2 l_p-Quantiles And Transformation Retransformation

It is easy to see that given any α such that $0 < \alpha < 1$ and $u = 2\alpha - 1$, the sum $\sum_{i=1}^{n}\{|X_i - Q| + u(X_i - Q)\}$ is minimized when Q is the sample α-th quantile based on the real-valued observations X_i's. In this article, we generalize this concept to d-dimensional l_p spaces for $1 \leq p < \infty$. Define, the open unit ball $B_p^{(d)}$ in l_p space as $\{\boldsymbol{u} : \boldsymbol{u} \in \mathbb{R}^d, \|\boldsymbol{u}\|_p < 1\}$ where $\boldsymbol{u} = (u_1, \ldots, u_d)^T$ and $\|\boldsymbol{u}\|_p = (|u_1|^p + \cdots + |u_d|^p)^{1/p}$ and $\|\boldsymbol{u}\|_\infty = \max(|u_1|, \ldots, |u_d|)$. For $1 \leq p < \infty$, and for any $\boldsymbol{u} \in B_q^{(d)}$, $\boldsymbol{t} \in \mathbb{R}^d$, where $1/p + 1/q = 1$ with the convention that $q = \infty$ when $p = 1$, let us define

$$\Phi_p(\boldsymbol{u}, \boldsymbol{t}) = \|\boldsymbol{t}\|_p + \boldsymbol{u}^T\boldsymbol{t}. \tag{2.1}$$

Then the l_p-quantile $\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u})$ corresponding to $\boldsymbol{u}$ is defined as

$$\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) = \arg\min_{Q \in \mathbb{R}^d} \sum_{i=1}^{n} \Phi_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}). \tag{2.2}$$

Observe at this point that a vector $\boldsymbol{u}$ for which $\|\boldsymbol{u}\|_q$ is close to one corresponds to an extreme quantile whereas a vector $\boldsymbol{u}$ for which $\|\boldsymbol{u}\|_q$ is close to zero corresponds to a central quantile. Since the vector $\boldsymbol{u}$ has a direction in addition to its magnitude, this immediately leads to a notion of directional outlyingness of a point with respect to the center of a cloud of observations based on the geometry of the cloud. Kemperman (1987) introduced and extensively studied a notion of median in Banach spaces. Observe that the second term in the definition of $\Phi_p(\boldsymbol{u}, \boldsymbol{t})$ can be viewed as a real-valued linear functional with norm less than one. In other words, quantiles in a Banach space will be indexed by the elements of the open unit ball around the origin in the dual Banach space of real-valued linear functionals. This yields a generalisation of Kemperman's (1987) idea of median into a notion of quantiles in Banach spaces. It is also noteworthy that if we view the d-dimensional space $\mathbb{R}^d$ equipped with l_q-norm as the dual of the Banach space $\mathbb{R}^d$ equipped with l_p-norm where $1/p + 1/q = 1$, our index vector $\boldsymbol{u}$ is an element of the open unit ball in that dual space.

It is easy to observe that for $1 \leq p < \infty$, l_p-quantiles are not equivariant under arbitrary affine transformations of the data vectors and they are not even equivariant under orthogonal transformations unless $p = 2$ (for rotational equivariance in the case $p = 2$ see Chaudhuri 1996). Thus when the coordinate variables are measured in different units, or they have different degrees of statistical variation l_p-quantiles do not make much sense. This lack of affine equivariance makes l_p-quantiles very much dependent on the choice of the coordinate system, which is not at all desirable. To resolve the problem of lack of affine equivariance of the vector of coordinatewise medians, Chakraborty and Chaudhuri (1996) introduced a transformation retransformation methodology. Chakraborty, Chaudhuri and Oja (1998) adopted the same methodology to develop an affine equivariant modification of spatial median.

Let us now consider n data points $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ in $\mathbb{R}^d$, and assume that $n > d+1$. Let $\alpha = \{i_0, i_1, \ldots, i_d\}$ denote a subset of size $(d+1)$ of $\{1, 2, \ldots, n\}$. Consider the points $\boldsymbol{X}_{i_0}, \boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_d}$, which will form a 'data-driven coordinate system', where $\boldsymbol{X}_{i_0}$ will determine the origin and the lines joining that origin to the remaining d data points $\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_d}$ will form various coordinate axes. The $d \times d$ matrix $\mathbf{X}(\alpha)$ containing the columns $\boldsymbol{X}_{i_1} - \boldsymbol{X}_{i_0}, \ldots, \boldsymbol{X}_{i_d} - \boldsymbol{X}_{i_0}$ can be taken as the transformation matrix for transforming the remaining data points $\boldsymbol{X}_j$'s $1 \leq j \leq n$, $j \notin \alpha$ to express them in terms of the new coordinate system as $\boldsymbol{Y}_j^{(\alpha)} = \{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{X}_j$. If the $\boldsymbol{X}_j$'s are i.i.d. observations with a common probability distribution that happens to be absolutely continuous w.r.t the Lebesgue measure on $\mathbb{R}^d$, $\mathbf{X}(\alpha)$ must be an invertible matrix with probability one. To compute the $\boldsymbol{u}$-th l_p-quantile for $1 \le p < \infty$ and $\|\boldsymbol{u}\|_q < 1$ with $1/p + 1/q = 1$ define

$$\begin{aligned} \boldsymbol{v}(\alpha) &= \tfrac{\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}}{\|\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}\|_q}\|\boldsymbol{u}\|_q \quad \text{for } \boldsymbol{u} \neq 0 \\ &= 0 \quad \text{for } \boldsymbol{u} = 0 \end{aligned}$$

Let $\hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{u})$ be the $\boldsymbol{v}(\alpha)$-th l_p-quantile based on $\boldsymbol{Y}_j^{(\alpha)}$'s with $1 \le j \le n$, $j \notin \alpha$ as defined in (2.2). Then define the multivariate transformation retransformation (TR) l_p-quantile $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ for the original data by retransforming $\hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{u})$ as $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) = \{\mathbf{X}(\alpha)\}\hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{u})$. Note that as we transform the observations in the new coordinate system, we need to suitably modify the orientation of the index vector $\boldsymbol{u}$. In the new coordinate system, the vector $\boldsymbol{u}$ should be transformed to $\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}$, but it may not be in the open unit ball $B_q^{(d)}$. To preserve the l_q-norm of the vector $\boldsymbol{u}$, we rescale $\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}$ by multiplying it with $\|\boldsymbol{u}\|_q/\|\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}\|_q$. In the transformed coordinate system, we compute $\boldsymbol{v}(\alpha)$-th l_p-quantile based on transformed observations and then retransform it back to the original coordinate system. In other words,

$$\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) = \arg\min_{Q \in \mathbb{R}^d} \sum_{i \notin \alpha} \left[\|\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \boldsymbol{Q})\|_p + \{\boldsymbol{v}(\alpha)\}^T\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \boldsymbol{Q}) \right]. \tag{2.3}$$

We now state a Theorem demonstrating the equivariance of the TR l_p-quantile under arbitrary affine transformations of data vectors.

Theorem 2.1 *Let the d-dimensional random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be transformed to $\mathbf{A}\boldsymbol{X}_1 + \boldsymbol{b}, \mathbf{A}\boldsymbol{X}_2 + \boldsymbol{b}, \ldots, \mathbf{A}\boldsymbol{X}_n + \boldsymbol{b}$, where $\mathbf{A}$ is a $d \times d$ nonsingular matrix and $\boldsymbol{b}$ is a vector in $\mathbb{R}^d$. Then for $\boldsymbol{w} = (\|\boldsymbol{u}\|_q/\|\mathbf{A}\boldsymbol{u}\|_q)\mathbf{A}\boldsymbol{u}$,the $\boldsymbol{w}$-th TR l_p-quantile based on $\mathbf{A}\boldsymbol{X}_1 + \boldsymbol{b}, \ldots, \mathbf{A}\boldsymbol{X}_n + \boldsymbol{b}$ is given by $\mathbf{A}\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) + \boldsymbol{b}$, where $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ is the $\boldsymbol{u}$-th TR l_p-quantile based on $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$.*

It is easy to see that, if we take $p = 2$ and $\mathbf{A}$ happens to be an orthogonal matrix, then $\mathbf{A}\boldsymbol{u}$-th quantile based on $\mathbf{A}\boldsymbol{X}_1 + \boldsymbol{b}, \ldots, \mathbf{A}\boldsymbol{X}_n + \boldsymbol{b}$ will be given by $\mathbf{A}\hat{\boldsymbol{Q}}_n^{(\alpha,2)}(\boldsymbol{u}) + \boldsymbol{b}$ where $\hat{\boldsymbol{Q}}_n^{(\alpha,2)}(\boldsymbol{u})$ is the $\boldsymbol{u}$-th transformation retransformation geometric quantile based on $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ (cf. Fact 2.2.1 of Chaudhuri 1996).

It should be noted that general M-quantiles defined by Koltchinskii (1997) are not affine equivariant in nature and we can employ this transformation retransformation strategy to general M-quantiles also to make them affine equivariant. But we have decided to restrict ourselves to l_p-quantiles here mainly because in many practical situations l_1-quantiles and spatial (or l_2) quantiles turn out to be adequate to explore different statistically important geometric aspects of a multivariate data cloud, some of which we will see later. The mathematical treatment of l_p-quantiles is not much different from those of l_2-quantiles and for each $p \geq 1$ the l_p-norm leads to a notion of multidimensional symmetry and associated symmetric probability distributions will have contours that coincide with the balls defined by the l_p-norm. The existence and uniqueness of TR l_p-quantiles are given in the following Fact.

Fact 2.2 *Consider observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ in $\mathbb{R}^d$ and $\alpha = \{i_0, i_1, \ldots, i_d\} \subset \{1, 2, \ldots, n\}$ such that the matrix $\mathbf{X}(\alpha)$ as defined earlier is invertible. Then the TR l_p-quantile $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ exists for any given $\boldsymbol{u}$ in the open unit ball $B_q^{(d)}$, where $1/p + 1/q = 1$. Further, for $d \geq 2$ and $1 < p < \infty$, it will be unique if the $\boldsymbol{X}_i$'s, $i \notin \alpha$ are not all carried by a single straight line in $\mathbb{R}^d$.*

Efficient algorithms for computing spatial median have been extensively studied by Gower (1974) and Bedall and Zimmermann (1979). Chaudhuri (1996) suggested an algorithm to compute geometric quantiles which is a minor modification of Newton-Raphson algorithm for finding roots of multivariate equations. We now state a fact characterizing TR l_p-quantiles in terms of data points from which it is computed.

Fact 2.3 *Consider $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ in $\mathbb{R}^d$ and $\alpha = \{i_0, i_1, \ldots, i_d\} \subset \{1, 2, \ldots, n\}$ such that the matrix $\mathbf{X}(\alpha)$, as defined earlier, is invertible, and $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ is the $\boldsymbol{u}$-th TR l_p-quantile based on these observations. If $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) \neq \boldsymbol{X}_i$ for all $i \notin \alpha$, we have for $1 < p < \infty$ and $1/p + 1/q = 1$*

$$\sum_{i \notin \alpha} \frac{\nu[\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}))]}{\|\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}))\|_p^{p-1}} + (n - d - 1)\boldsymbol{v}(\alpha) = 0 \tag{2.4}$$

On the other hand, if $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) = \boldsymbol{X}_i$ *for some* $i \notin \alpha$*, we will have*

$$\left\| \sum_{i\notin\alpha, X_i \neq \hat{Q}_n^{(\alpha,p)}} \left\{ \frac{\nu[\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}))]}{\|\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}))\|_p^{p-1}} + \boldsymbol{v}(\alpha) \right\} \right\|_q$$

$$\leq (1 + \|\boldsymbol{v}(\alpha)\|_q)[\#\{i: \ \boldsymbol{X}_i = \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}), i \notin \alpha\}], \tag{2.5}$$

where $\nu[(x_1, x_2, \ldots, x_d)^T] = (sign(x_1)|x_1|^{p-1}, \ldots, sign(x_d)|x_d|^{p-1})^T$*,* $\boldsymbol{v}(\alpha)$ *as defined earlier, and* $\#$ *denotes the number of elements in a set.*

This fact implies that one can use iterative methods like Newton-Raphson type method to compute $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ for $1 < p < \infty$. For $p = 1$, l_1-quantiles are nothing but coordinatewise quantiles. Thus, after transformation, one has to compute coordinatewise quantiles of the transformed observations and then retransform it back. This shows the simplicity of the computation involved in TR l_p-quantiles once the transformation matrix is fixed. Both of Facts 2.2 and 2.3 follow from some minor modifications of some of the fundamental results in Kemperman (1987) and Chaudhuri (1996), and we will skip their proofs here.

3 Large Sample Properties: Main Results

Let us begin by introducing some notations. For any $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{u} \in B_q^{(d)}$, we will write $\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{x}) = \nu(\boldsymbol{x})/\|\boldsymbol{x}\|_p^{p-1} + \boldsymbol{u}$ for $\boldsymbol{x} \neq 0$ and $\boldsymbol{\varphi}_p(\boldsymbol{u}, 0) = \boldsymbol{u}$. Note that $\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{x})$ is the gradient or first order derivative of the function $\Phi_p(\boldsymbol{u}, \boldsymbol{x})$ w.r.t $\boldsymbol{x}$ when $\boldsymbol{x} \neq 0$. Let $\Psi_p(\boldsymbol{x})$ denote the $d \times d$ Hessian matrix or the second order derivative of $\Phi_p(\boldsymbol{u}, \boldsymbol{x})$ for $1 < p < \infty$. So for $\boldsymbol{x} \neq 0$,

$$\Psi_p(\boldsymbol{x}) = (p-1)\|\boldsymbol{x}\|_p^{1-p}[W_p(\boldsymbol{x}) - \frac{\nu(\boldsymbol{x})\{\nu(\boldsymbol{x})\}^T}{\|\boldsymbol{x}\|_p^p}],$$

where $W_p(\boldsymbol{x})$ is the diagonal matrix $diag(|x_1|^{p-2}, \ldots, |x_d|^{p-2})$. We will adopt the convention that $\Psi_p(0) = \mathbf{0}$ = the zero matrix. Note that when $p = 1$, $\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{x})$ becomes $(sign(x_1), \ldots, sign(x_d))^T + \boldsymbol{u}$ and $\Psi_p(\boldsymbol{x})$ is identically equal to $\mathbf{0}$.

3.1 Asymptotic Behavior of TR l_p-quantiles

Let us define $\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})$ as

$$\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u}) = \arg \min_{Q\in\mathbb{R}^d} E^{(\alpha)}[\Phi_p(\boldsymbol{v}(\alpha), \{\mathrm{X}(\alpha)\}^{-1}(\boldsymbol{X}-\boldsymbol{Q})) - \Phi_p(\boldsymbol{v}(\alpha), \{\mathrm{X}(\alpha)\}^{-1}\boldsymbol{X})]$$

where $E^{(\alpha)}$ denotes the conditional expectation given the $\boldsymbol{X}_i$'s for which $i \in \alpha$ and $\boldsymbol{v}(\alpha)$ is as defined in Section 2. In this Section, the observations $\boldsymbol{X}_i$'s will be assumed to be i.i.d. observations with a common probability distribution having density $h(\boldsymbol{x})$ on $\mathbb{R}^d$. Let us define

$$D_1^{(\alpha,p)}(\boldsymbol{Q}) = E^{(\alpha)}\{\Psi_p(\{\mathrm{X}(\alpha)\}^{-1}(\boldsymbol{X}-\boldsymbol{Q}))\},$$

and

$$D_2^{(\alpha,p)}(\boldsymbol{Q},\boldsymbol{u}) = E^{(\alpha)}[\{\boldsymbol{\varphi}_p(\boldsymbol{v}(\alpha), \{\mathrm{X}(\alpha)\}^{-1}(\boldsymbol{X}-\boldsymbol{Q}))\}^T\{\boldsymbol{\varphi}_p(\boldsymbol{v}(\alpha), \{\mathrm{X}(\alpha)\}^{-1}(\boldsymbol{X}-\boldsymbol{Q}))\}].$$

Theorem 3.1 (Bahadur type representation of TR l_1-quantiles) *Assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ is a sequence of i.i.d. observations with a common density $h(\boldsymbol{x})$. Fix $\alpha = \{i_0, i_1, \ldots, i_d\} \subset \{1, 2, \ldots, n\}$ and the matrix $\mathrm{X}(\alpha)$ and assume that the j-th marginal g_j of the density $f(\boldsymbol{y}) = |\det\{\mathrm{X}(\alpha)\}|h\{\mathrm{X}(\alpha)\boldsymbol{y}\}$ is differentiable and positive at $\boldsymbol{Q}_j^{\#}(\boldsymbol{u})$, where $\boldsymbol{Q}_j^{\#}(\boldsymbol{u})$ is the j-th element of $\{\mathrm{X}(\alpha)\}^{-1}\boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u})$ for $i = 1, \ldots, d$. Then for any $\boldsymbol{u} \in \mathbb{R}^d$ such that $\|\boldsymbol{u}\|_\infty < 1$, and given the $\boldsymbol{X}_i$'s with $i \in \alpha$, we have*

$$\begin{aligned} \hat{\boldsymbol{Q}}_n^{(\alpha,1)}(\boldsymbol{u}) - \boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u}) &= \\ n^{-1}\mathrm{X}(\alpha)\{D_f(\alpha)\}^{-1}\sum_{i\notin\alpha}\Big\{Sign[\{\mathrm{X}(\alpha)\}^{-1}\{\boldsymbol{X}_i - \boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u})\}] + \boldsymbol{v}(\alpha)\Big\} &+ R_n(\boldsymbol{u}), \end{aligned} \tag{3.1}$$

where $D_f(\alpha)$ is the diagonal matrix $diag(2g_1\{\boldsymbol{Q}_1^{\#}(\boldsymbol{u})\}, \ldots, 2g_d\{\boldsymbol{Q}_d^{\#}(\boldsymbol{u})\})$, Sign denotes the vector of coordinatewise signs, and as $n \to \infty$, the remainder term $R_n(\boldsymbol{u})$ is almost surely $O(n^{-3/4}(\log n)^{3/4})$.

Theorem 3.2 (Bahadur type representation of TR l_p-quantiles, $1 < p < \infty$) *Assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ is a sequence of i.i.d. observations with a common density $h(\boldsymbol{x})$ which is bounded on every compact subset of $\mathbb{R}^d$ with $d \geq 2$. Fix $\alpha = \{i_0, i_1, \ldots, i_d\} \subset \{1, 2, \ldots, n\}$ and the matrix $\mathrm{X}(\alpha)$. Then for any fixed*

$\boldsymbol{u} \in B_q^{(d)}$, *where* $1 < p < \infty$ *and* $1/p + 1/q = 1$, *the expectation defining the matrix* $D_1^{(\alpha,p)}[\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})]$ *will exist as a finite and invertible matrix, and given the* $\boldsymbol{X}_i$*'s with* $i \in \alpha$, *we have*

$$\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u}) =$$

$$n^{-1}\mathbf{X}(\alpha)[D_1^{(\alpha,p)}(\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u}))]^{-1}\sum_{i \notin \alpha}\boldsymbol{\varphi}_p[\boldsymbol{v}(\alpha), \{\mathbf{X}(\alpha)\}^{-1}\{\boldsymbol{X}_i - \boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}] + R_n(\boldsymbol{u}), \tag{3.2}$$

where as $n \to \infty$, $R_n(\boldsymbol{u})$ *is almost surely* $O(\log n/n)$ *if* $d \geq 3$, *and when* $d = 2$, $R_n(\boldsymbol{u})$ *is almost surely* $o(n^{-\beta})$ *for any fixed* β *such that* $0 < \beta < 1$.

It will be appropriate to note here that Chaudhuri (1996) established a Bahadur type representation of nonequivariant l_2-quantiles and Koltchinskii (1997) considered general non-equivariant multivariate M-estimators and obtained their asymptotic distributions using empirical processes. The following Corollaries are easy consequence of the above two Theorems.

Corollary 3.3 *Under the assumptions of Theorem 3.1, for any fixed* $\boldsymbol{u} \in \mathbb{R}^d$ *such that* $\|\boldsymbol{u}\|_\infty < 1$, *the conditional distribution of* $\sqrt{n}\{\hat{\boldsymbol{Q}}_n^{(\alpha,1)}(\boldsymbol{u}) - \boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u})\}$ *given the* $\boldsymbol{X}_i$*'s with* $i \in \alpha$ *converges weakly to a d-dimensional normal distribution with zero mean and dispersion matrix*

$$\{\mathbf{X}(\alpha)\}\{D_f(\alpha)\}^{-1}[D_2^{(\alpha,1)}(\boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u}), \boldsymbol{u})]\{D_f(\alpha)\}^{-1}\{\mathbf{X}(\alpha)\}^T$$

as $n \to \infty$.

Corollary 3.4 *Under the assumptions of Theorem 3.2, for* $1 < p < \infty$ *and for any* $\boldsymbol{u} \in B_q^{(d)}$ *where* $1/p + 1/q = 1$, *the conditional distribution of* $\sqrt{n}\{\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}$ *given the* $\boldsymbol{X}_i$*'s with* $i \in \alpha$ *converges weakly to a d-dimensional normal distribution with zero mean and dispersion matrix*

$$\{\mathbf{X}(\alpha)\}[D_1^{(\alpha,p)}\{\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}]^{-1}[D_2^{(\alpha,p)}\{\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u}), \boldsymbol{u}\}][D_1^{(\alpha,p)}\{\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}]^{-1}\{\mathbf{X}(\alpha)\}^T$$

as $n \to \infty$.

It may be useful to note here that the above corollaries can be used to construct large-sample confidence ellipsoids for $\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})$, provided that we can construct a reasonable estimate of the limiting dispersion matrix of $\sqrt{n}\{\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})-\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}$ from the data. For co-ordinatewise sample quantile vector, Babu and Rao (1988) discussed a consistent method of estimating the asymptotic variance covariance matrix. We can apply the same method to the transformed observations to estimate $\{D_f(\alpha)\}^{-1}[D_2^{(\alpha,1)}(\boldsymbol{Q}^{(\alpha,1)}(\boldsymbol{u}),\boldsymbol{u})]\{D_f(\alpha)\}^{-1}$ when $p=1$. For $p=2$, Chaudhuri (1996) discussed a simple but $n^{-1/2}$ consistent estimate of the asymptotic dispersion matrix of the geometric quantiles. The same methodology can be extended very easily to estimate the asymptotic dispersion matrix of $\sqrt{n}\{\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})-\boldsymbol{Q}^{(\alpha,p)}(\boldsymbol{u})\}$ for $1<p<\infty$.

3.2 Selection of α

The asymptotic normal distribution of $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ established in the preceding section and the form of the associated dispersion matrix clearly indicates that the performance of the TR l_p-quantiles will depend upon the choice of the transformation matrix $\mathbf{X}(\alpha)$. Hence it is important to select a suitable subset of indices α. Before we state any formal method for selecting the transformation matrix $\mathbf{X}(\alpha)$, the following Facts, which are direct consequences of the main results in Chakraborty and Chaudhuri (1996) and Chakraborty, Chaudhuri and Oja (1998), in the special cases of l_1 and l_2 TR medians will provide some valuable insights into the problem. Let us assume that $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n,\ldots$ are independent and identically distributed random variables with a common elliptically symmetric density $|\det(\Sigma)|^{-1/2}f\{(\boldsymbol{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})\}$ where Σ is a $d\times d$ positive definite matrix, $\boldsymbol{\theta}\in\mathbb{R}^d$ and $f(\boldsymbol{x}^T\boldsymbol{x})$ is a density in $\mathbb{R}^d$.

Fact 3.5 *Assume that the density function f, as defined above, is such that any univariate marginal g of the spherically symmetric density $f(\boldsymbol{x}^T\boldsymbol{x})$ is differentiable and positive at zero. Then for any given subset α of $\{1,2,\ldots,n\}$ with size $d+1$ and given the $\boldsymbol{X}_i$'s with $i\in\alpha$, the conditional asymptotic distribution of $\sqrt{n}\left\{\hat{\boldsymbol{Q}}_n^{(\alpha,1)}(\mathbf{0})-\boldsymbol{\theta}\right\}$ is d-dimensional normal with zero mean and dispersion matrix*

$$\mathbf{V}(\alpha)=c\Sigma^{1/2}\{\mathbf{J}(\alpha)\}^{-1}\{\mathbf{D}(\alpha)\}\{[\mathbf{J}(\alpha)]^T\}^{-1}\Sigma^{1/2}.$$

Here $c = \{2g(0)\}^{-2}$, $\{\Sigma^{-1/2}\mathbf{X}(\alpha)\}^{-1} = \mathbf{R}(\alpha)\mathbf{J}(\alpha)$ with the rows of $\mathbf{J}(\alpha)$ having unit length and $\mathbf{R}(\alpha)$ is a diagonal matrix, and $\mathbf{D}(\alpha)$ 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 i-th and the j-th row of $\mathbf{J}(\alpha)$. Further, $\det\{\mathbf{V}(\alpha)\} \geq \det\{c\Sigma\}$ and equality holds if $\mathbf{J}(\alpha) = \mathbf{I}_d$.

Fact 3.6 *For any given subset α of $\{1, \ldots, n\}$ with size $d+1$ and given the $\boldsymbol{X}_i$'s with $i \in \alpha$, the conditional asymptotic distribution of $\sqrt{n}\left\{\hat{\boldsymbol{Q}}_n^{(\alpha,2)}(\mathbf{0}) - \boldsymbol{\theta}\right\}$ is d-variate normal with zero mean and variance covariance matrix $\Delta\{f, \Sigma, \mathbf{X}(\alpha)\}$ that depends on f, Σ and the transformation matrix $\mathbf{X}(\alpha)$. Here the positive definite matrix Δ is such that the difference $\Delta\{f, \Sigma, \mathbf{A}\} - \Delta\{f, \Sigma, \mathbf{B}\}$ is non-negative definite for any f, Σ and any two $d \times d$ invertible matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{B}^T\Sigma^{-1}\mathbf{B} = \lambda\mathbf{I}_d$, where $\lambda > 0$ is a constant and $\mathbf{I}_d$ is the $d \times d$ identity matrix.*

The main message communicated by these facts is that for $\boldsymbol{u} = \mathbf{0}$ (i.e. in the case of multivariate median) and $p = 1$ or 2, we need to choose $\mathbf{X}(\alpha)$ in such a way that $\{\mathbf{X}(\alpha)\}^T\Sigma^{-1}\mathbf{X}(\alpha)$ becomes as close as possible to a matrix of the form $\lambda\mathbf{I}_d$, which is a diagonal matrix with all diagonal entries equal. In other words, the coordinate system represented by the matrix $\Sigma^{-1/2}\mathbf{X}(\alpha)$ should be as orthonormal as possible. It also implies that when $\{\mathbf{X}(\alpha)\}^T\Sigma^{-1}\mathbf{X}(\alpha)$ is chosen to be close to a diagonal matrix with all diagonal entries equal, the asymptotic efficiency of the estimate $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\mathbf{0})$ becomes close to that of the l_p-median under spherically symmetric models, and it will be more efficient than l_p median in elliptically symmetric models for $p = 1$ or 2.

Keeping in mind the fact that the above selection procedure provides "the most efficient transformation" for the multivariate median problem, we propose to select the transformation matrix $\mathbf{X}(\alpha)$ in such a way that $\{\mathbf{X}(\alpha)\}^T\Sigma^{-1}\mathbf{X}(\alpha)$ becomes as close as possible to a diagonal matrix with all diagonal entries equal. Here Σ is the scatter matrix associated with the underlying distribution of the $\boldsymbol{X}_i$'s which may not necessarily be elliptically symmetric. If the second moments of the underlying distribution exist, Σ can be taken to be the variance covariance matrix of that distribution. Since Σ will be an unknown parameter in practice, we have to estimate that from the data, and we will need an affine equivariant estimate (say $\hat{\Sigma}$). After obtaining $\hat{\Sigma}$, we will try to choose $\mathbf{X}(\alpha)$ in such a way that the eigen values of the positive definite matrix $\{\mathbf{X}(\alpha)\}^T\hat{\Sigma}^{-1}\mathbf{X}(\alpha)$ becomes as equal as possible. To achieve this, our strategy will be to minimize the ratio between the arithmetic mean and the geometric mean of the eigenvalues. Since the arithmetic mean and the geometric mean of the eigenvalues of a symmetric matrix can be obtained from its trace and the determinant respectively, we do not need to compute individual eigenvalues. Instead of minimizing the ratio over all possible subsets α with size $d+1$ of $\{1,\ldots,n\}$, one can substantially reduce the amount of computation by stopping the search for optimal $\mathbf{X}(\alpha)$ as soon as the ratio becomes smaller than $1+\epsilon$, where ϵ is a preassigned small positive number. In our simulations and data analysis, we did not observe such an approach to cause any significant change in the statistical performance of the procedures though there was considerable gain in the speed of computation. The algorithm leads to stable estimates with small sample sizes even when the dimension is large.

Define now $\mathbf{X}^*(\alpha) = |\det(\mathbf{X}(\alpha))|^{-1/d}\mathbf{X}(\alpha)$ and $\hat{\Sigma}^* = \{\det(\hat{\Sigma})\}^{-1/d}\hat{\Sigma}$ where $\hat{\Sigma}$ is a positive definite matrix computed from the data. Note that, the absolute values of the determinants of the newly defined matrices $\mathbf{X}^*(\alpha)$ and $\hat{\Sigma}^*$ are both equal to 1, and the operation can be viewed as a way of normalizing matrices. Then to select the optimal α according to the above mentioned criteria, we only have to minimize the trace of $\{\mathbf{X}^*(\alpha)\}^T\hat{\Sigma}^{*-1}\mathbf{X}^*(\alpha)$. Suppose that for the subset of indices $\hat{\alpha}$, the trace of $\{\mathbf{X}^*(\alpha)\}^T\hat{\Sigma}^{*-1}\mathbf{X}^*(\alpha)$ is minimized, that is, $\hat{\alpha}$ is our optimal subset of indices used to construct the optimal transformation matrix.

Theorem 3.7 *Assume that, the random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are independent and identically distributed with a common density $h(\boldsymbol{x})$ which satisfies*

$$\int_{\mathbb{R}^d} \{h(\boldsymbol{x})\}^{d+1}\, d\boldsymbol{x} < \infty.$$

Further assume that $\hat{\Sigma}^*$ converges in probability to a positive definite matrix Σ^*. Then $\det(\Sigma^*) = 1$, and trace$[\{\mathbf{X}^*(\hat{\alpha})\}^T\Sigma^{*-1}\mathbf{X}^*(\hat{\alpha})]/d$ 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 case of elliptic symmetry with $h(\boldsymbol{x}) = \{\det(\Sigma)\}^{-1/2} f\{(\boldsymbol{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})\}$, this condition translates into an integrability condition on f, which is again trivially satisfied for any bounded spherically symmetric density f on $\mathbb{R}^d$. It is interesting to note that if the second moments of the distribution of $\boldsymbol{X}_i$'s exist, $\hat{\Sigma}$ can taken to be the usual sample variance covariance matrix and $\hat{\Sigma}^*$ will converge in probability to Σ^* where Σ^* is the normalized version of the variance covariance matrix of the distribution. In the case of elliptically symmetric distributions, one can use any consistent affine equivariant estimate of the associated scale matrix Σ upto a scalar multiple.

As an alternative affine equivariant modification of spatial median, Isogai (1985) and Rao (1988) suggested spatial median based on observations transformed by the square root of the usual variance covariance matrix. While transforming the data points by the square root of the sample variance covariance matrix is a popular approach, the resulting coordinate system does not have any simple geometric interpretation. Further, such a transformation cannot lead to an affine equivariant modification of multivariate location estimates which are obtained by minimizing the general l_p distances for a p different from 2, (see Chakraborty and Chaudhuri 1996), and the limitation of that approach is primarily due to the fact that there does not exist a way to extract an affine equivariant square root of the sample variance covariance matrix. On the other hand, observe that in a sense our selection procedure gives an "affine equivariant estimate" of the matrix $\Sigma^{1/2}$ which is further justified from our next result.

Corollary 3.8 *Under the conditions assumed in Theorem 3.7, the matrix* $\mathbf{X}^*(\hat{\alpha})\{\mathbf{X}^*(\hat{\alpha})\}^T$ *converges in probability to the matrix* Σ^* *as* $n \to \infty$.

Our results hold for any consistent and affine equivariant estimate of Σ (or Σ^*) and one can use robust estimates of scale as discussed by Davies (1987), which however are computationally quite intensive. Note that, this 'data-driven coordinate system' is a widely applicable tool for converting non-equivariant (or non-invariant) procedures into equivariant (or invariant) procedures, which is not limited to only l_p-quantiles. Besides, it has a very nice and intuitively meaningful geometric interpretation, and an attractive feature of this data-based transformation retransformation strategy is the clean and elegant mathematical theory associated with the approach.

4 Applications

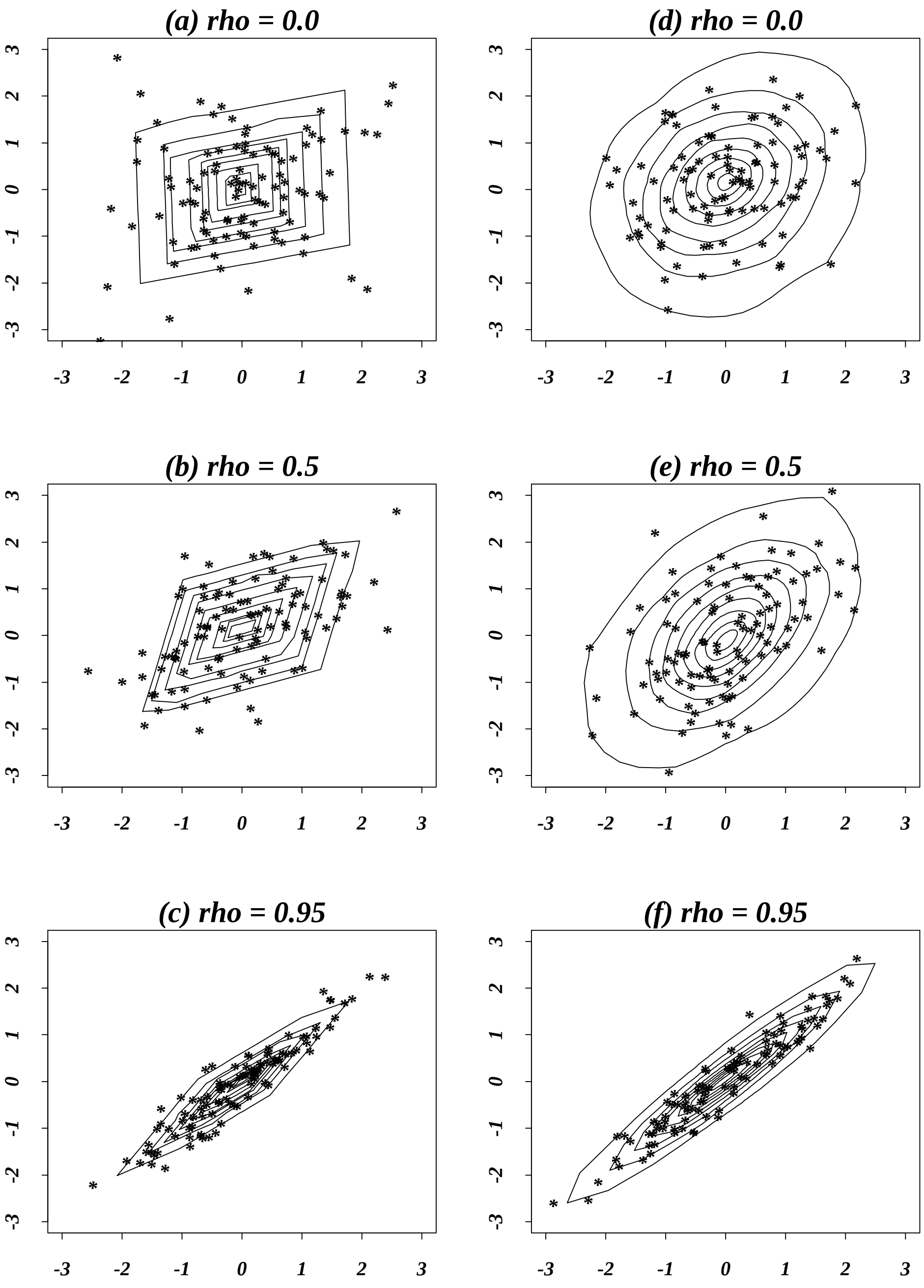
4.1 Quantile Contour Plots

In the univariate set-up the quantiles uniquely determine the population distribution, and the sample quantiles provide a fair idea about the shape of the distribution. While exploring a multivariate data cloud, one may be interested to find out quantile contours, which join the quantiles for which the length of the index vector $\boldsymbol{u}$ is a constant, to get ideas about the shape of the underlying population distribution. Thus quantile contours can be described by the sets $\{\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) : \|\boldsymbol{u}\|_q = r\}$ where $0 < r < 1$. For $r = 0$, it comprises of only one point – the TR l_p median. In principle, quantile contours can be constructed for any dimension $d \geq 2$, but for practical purposes, it is easier to visualize things only for bivariate data.

It is interesting to note that, for the optimal selection of the transformation matrix $\mathbf{X}(\alpha)$, the population quantile contours corresponding to $p = 2$ are nothing but the level sets of the probability density function (or, probability density contours) when the underlying distribution is elliptically symmetric with density $\{\det(\Sigma)\}^{-1/2} f\{(\boldsymbol{x} - \boldsymbol{\theta})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\theta})\}$. The optimal selection of $\mathbf{X}(\alpha)$ provides an estimate of the matrix $\Sigma^{1/2}$ upto a scalar multiple and premultiplying the observations by $\{\mathbf{X}(\alpha)\}^{-1}$ makes the data spherical. As probability density contours characterize a distribution, the affine equivariant TR l_2-quantile contour plots can be used to measure the closeness of the data to a specific elliptically symmetric probability distribution. Even when the underlying probability distribution is not elliptically symmetric, Koltchinskii (1997) observed that the spatial quantile process uniquely determines the population distribution. Vector of coordinatewise quantiles determine the marginals of the joint multivariate distribution. However, marginals do not uniquely determine the joint distribution. Thus TR l_1-quantile contour plots cannot be used as a tool for measuring proximity to a multivariate distribution. Nevertheless, they can provide some insights into the geometry of the multivariate data cloud and help in identifying possible outliers.

To illustrate quantile contour plots, we simulated 100 observations from bivariate normal populations with zero means, unit standard deviations and varying

Figure 1: Quantile contour plot for bivariate normal



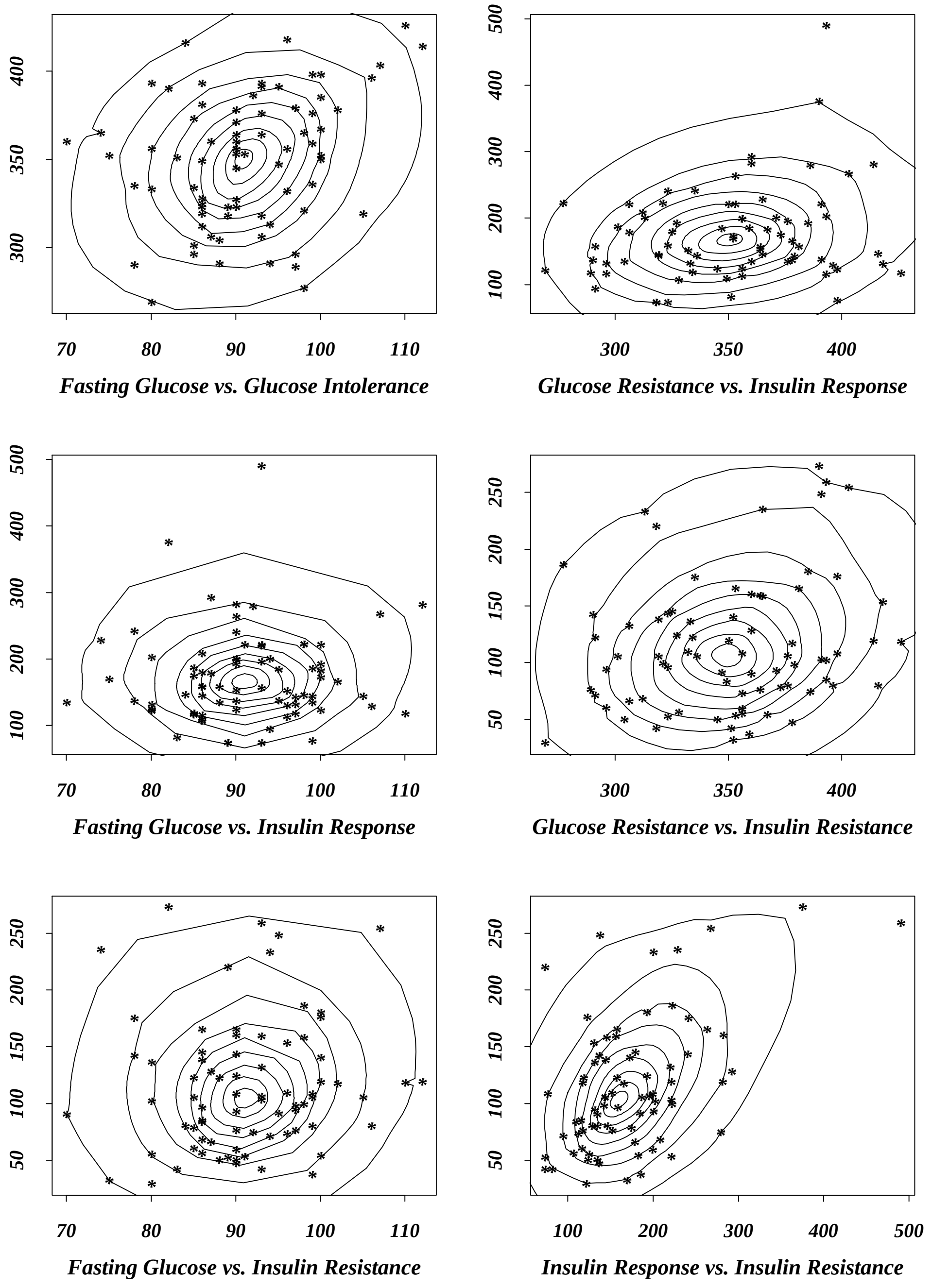
correlation coefficients $\rho = 0.0, 0.5$ and 0.95. In Figure 1 (a), (b) and (c), we have plotted TR l_1-quantiles for $r = 0.1, 0.2, \ldots, 0.9$. To construct quantile contours, for each r, we have taken 32 values of $\boldsymbol{u}$ such that $\|\boldsymbol{u}\|_\infty = r$ and joined the corresponding quantiles. In Figures 1 (d), (e) and (f) we have similarly plotted TR l_2-quantiles for $r = 0.1, 0.2, \ldots, 0.9$. For each r, we have computed quantiles corresponding to $\boldsymbol{u} = (r\cos\theta, r\sin\theta)^T$ where $\theta = \pi k/16$, $k = 0, 1, \ldots, 31$ and joined them. We notice that as the TR quantiles are affine equivariant, quantile contours nicely capture the shift of the distribution from spherical symmetry to elliptical symmetry. The regions enclosed by quantile contours can be viewed as multivariate analogs of box and whisker plots used for univariate data.

Another interesting application of these quantile contours is in detecting outliers in the multivariate data. In multidimension, it is really difficult to detect the outliers. Here we suggest a simple procedure. We compute the quantile contour for some r close to 1 (the choice of r depends on the problem and the user's preference), and if a particular observation lies outside this contour, then we will call it an outlier. We demonstrate the methodology in a real data set. Reaven and Miller (1979) examined the relationship between chemical, subclinical and overt nonketotic diabetes in 145 non-obese adult subjects. The three primary variables used in the analysis are glucose intolerance, insulin response to oral glucose and insulin resistance. In addition, the relative weight and fasting plasma glucose were also measured for each individual in the study conducted at the Stanford Clinical Research Center. We have taken only 48 overt nonketotic diabetic patients and in Figure 2 we have shown the TR l_2-quantile contours by taking two variables at a time and $r = 0.0, 0.1, \ldots, 0.9$. These quantile contours clearly reveal that there are some outliers in the data set. Note that affine equivariance of the quantiles is crucial in outlier detection as the outlyingness of a data point should not be judged differently in different coordinate systems.

4.2 Multivariate Ranks

In univariate set-up, the concept of ranks and quantiles are closely related. Jan and Randles (1994) and Möttönen and Oja (1995) considered some notions of multi-

Figure 2: Quantile contour plots for Overt diabetic patients



variate ranks which are closely related to geometric quantiles (or l_2-quantiles). Chaudhuri (1996) suggested the d-dimensional direction vector $n^{-1} \sum_{X_i \neq y} \|\boldsymbol{X}_i - \boldsymbol{y}\|_2^{-1}(\boldsymbol{X}_i - \boldsymbol{y})$ as the multivariate rank of $\boldsymbol{y} \in \mathbb{R}^d$. We may define affine invariant notions of multivariate ranks based on our transformation retransformation approach as follows. Consider the d-dimensional direction vector based on l_2-norm $n^{-1} \sum_{X_i \neq y, i \notin \alpha} \|\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \boldsymbol{y})\|_2^{-1}\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \boldsymbol{y})$ or alternatively based on l_1-norm $n^{-1} \sum_{X_i \neq y, i \notin \alpha} Sign[\{\mathbf{X}(\alpha)\}^{-1}(\boldsymbol{X}_i - \boldsymbol{y})]$, which can be viewed as descriptive statistics that determine the geometric position of the point $\boldsymbol{y} \in \mathbb{R}^d$ with respect to the data cloud formed by the observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ and these lead to vector valued concepts of multivariate centered ranks corresponding to TR l_2 and l_1-quantiles. Similarly, from the gradient vectors of the other l_p-norms, one can construct different versions of multivariate ranks. However, it is rather easy to interpret and geometrically visualize things for $p = 1$ and $p = 2$. Observe that the multivariate rank vectors associated with TR l_p-quantiles lie inside the unit ball $B_q^{(d)}$ where as usual $1/p + 1/q = 1$. There are some attempts to construct ranks as univariate quantities based on different data depth concepts like Tukey's half-space depth (Tukey, 1975) and Liu's simplicial depth (Liu, 1990), but they fail to take into account the orientation of a point in the data cloud. Univariate concepts of ranks can distinguish between 'extreme' points and 'central' points but they do not provide the information whether the 'extreme' observations are 'low' or 'high' observations with respect to some specific directions. For these limitations multivariate notions of ranks are often preferred over univariate notions. Based on these affine invariant multivariate ranks one can construct different rank related methodologies in multidimension extending univariate rank based methodologies.

These affine invariant notions of multivariate ranks may be used to construct tests for multivariate location. Chakraborty and Chaudhuri (1999) discussed affine invariant rank tests in one sample and two sample location problems based on the l_1-ranks. Similar methods can be constructed for other l_p-ranks also. This vector-valued notion of multivariate ranks is also useful to generalise rank regression in multi-response linear models. Chakraborty and Chaudhuri (1997) considered a few specific cases of multivariate rank regression using l_1-ranks. In the next subsection, we will use l_p-rank vectors to construct multivariate Q-Q plots. In fact, with the help of this notion of multivariate ranks, one can generalise most of the rank based methodologies from the univariate set-up to the multidimension.

4.3 Multivariate Q-Q Plots

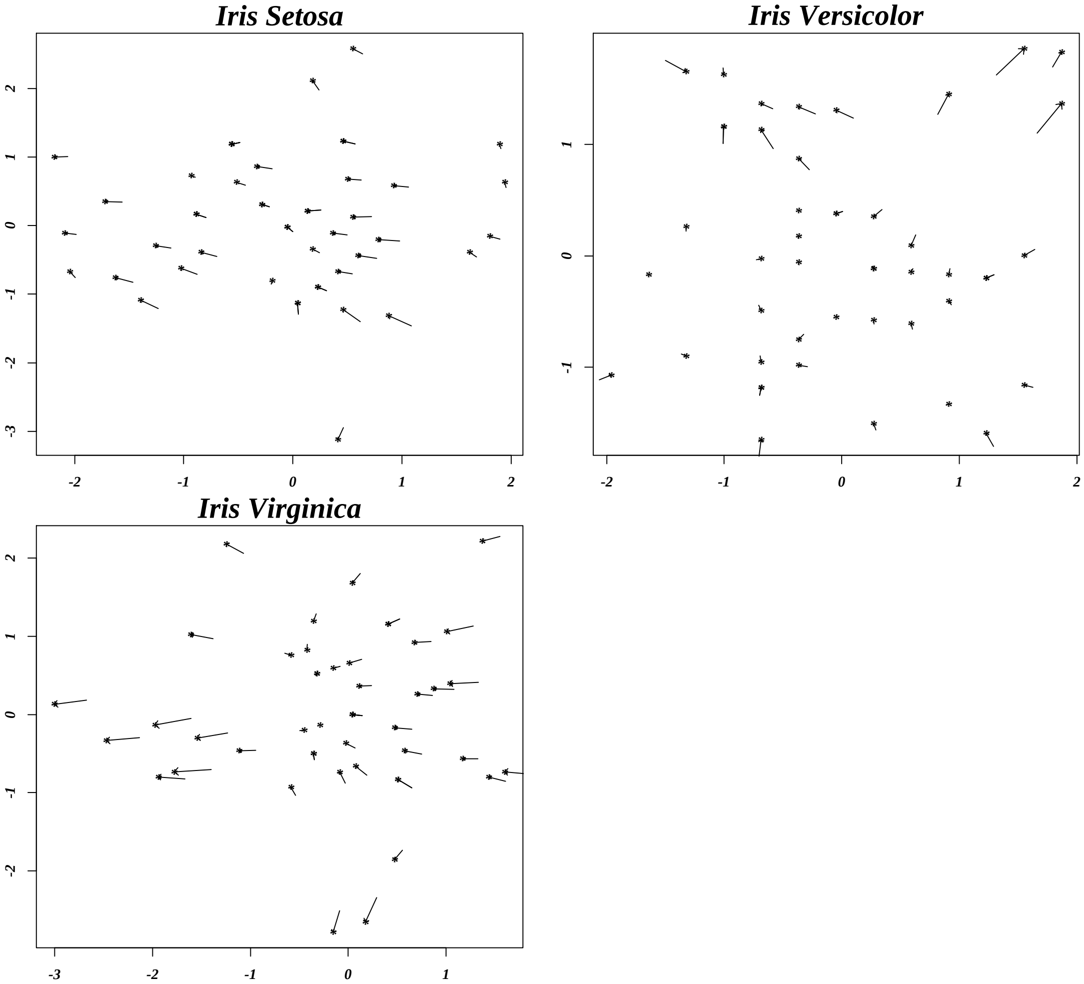
Q-Q plots are popular and useful diagnostic tools in univariate data analysis. With their help, it is possible to assess graphically the closeness of a sample to a particular univariate distribution or the closeness between two independent samples. The idea behind Q-Q plots is to compute and plot a finite number of quantiles from the sample and corresponding quantiles from the comparing probability distribution or from the comparing sample. Now we can generalize this concept to define multivariate Q-Q plots. At first, we transform the data points by $\{\hat{\lambda}\mathbf{X}(\alpha)\}^{-1}$ where $\hat{\lambda}^2 = trace[\{\mathbf{X}(\alpha)\}^{-1}\hat{\Sigma}(\{\mathbf{X}(\alpha)\}^{T})^{-1}]/d$. After that l_p-ranks of the transformed observations are computed as discussed earlier. As we know that these l_p-rank vectors lie in $B_q^{(d)}$, and one can compute corresponding l_p-quantiles of the comparing probability distribution with scatter matrix $\mathbf{I}_d$ and location parameter $\mathbf{0}$ by taking the rank of the observation as the index vector $\boldsymbol{u}$. Then we plot the arrows from the l_p-quantiles of the given distribution to the transformed observations. We have noted earlier that a proper selection of the transformation matrix $\mathbf{X}(\alpha)$ leads to an estimate of the scatter matrix Σ and $\hat{\lambda}^{-2}\{\mathbf{X}(\alpha)\}^{-1}\Sigma\{[\mathbf{X}(\alpha)]^{T}\}^{-1}$ is expected to be close to a d-dimensional identity matrix. If these arrows are very small in length and randomly oriented, then one may conclude that the sample does not deviate much from the chosen probability distribution. But if most of the arrows are directed towards a particular direction then the sample is more skewed in that direction, and if in general arrow lengths are large then the sample obviously does not conform with the given distribution. Marden (1998) also used a similar technique to define bivariate Q-Q plots but they are not affine invariant in nature and thus the presence of high correlations among the coordinate variables will often lead to inappropriate inference. It would be wise to note here that the TR Q-Q plots can be constructed for any dimension d in principle, but in practice it is not possible to visualize the plots for dimensions $d > 3$. So, for dimensions greater than

3, one suggestion is to make a Q-Q plot matrix, taking all pairs of variables. But, definitely, we lose some features of the multivariate data cloud with this suggestion.

The TR Q-Q plots, which are affine invariant, can be used to construct tests of goodness of fit to a given multivariate distribution. There is no known good way of testing in practice whether the observed data is from a specified multivariate distribution or not. Both of the well-known χ^2-goodness of fit test and Kolmogorov-Smirnov test have serious practical limitations and are not very useful for multivariate problems. We suggest the following test procedure. At first, we make the data spherical by transforming the observations by $\{\hat{\lambda}\mathbf{X}(\alpha)\}^{-1}$ and then subtract the l_p-median of the transformed observations from them. Let us call these observations $\boldsymbol{Z}_i^{(\alpha,p)}$'s for $i \notin \alpha$. Thus, we have transformed observations with location parameter zero and identity as the scale matrix. Then we compute l_p-ranks of each of these transformed observations and corresponding l_p-quantiles of the population distribution (say, $\boldsymbol{Q}_i^{(\alpha,p)}$'s). In the case $p = 2$, these population l_p-quantiles can be computed using the formula given in Möttönen, Oja and Tienari (1997) for spherically symmetric distributions. After that, let us consider the statistic $T_n^{(\alpha,p)} = \sum_{i\notin\alpha} \|\boldsymbol{Z}_i^{(\alpha,p)} - \boldsymbol{Q}_i^{(\alpha,p)}\|$, which is nothing but the sum of the length of "directed arrows" as discussed earlier. If the observed data is close to the given distribution, the statistic $T_n^{(\alpha,p)}$ should be close to zero. Thus $T_n^{(\alpha,p)}$ can be used as a test statistic for testing goodness of fit of a multivariate distribution. As an illustration, we have constructed bivariate affine invariant Q-Q plots for Iris Setosa, Iris Virginica and Iris Versicolor of the famous Fisher's iris data using TR l_2-quantiles. We have considered only two variables sepal length and sepal width for the demonstration purpose and compared them with bivariate normal distributions in Figure 3, where the plots indicate fairly good fits. Möttönen, Oja and Tienari (1997) provided a result for computing geometric quantiles of the bivariate normal distribution, and we have used that for our calculations.

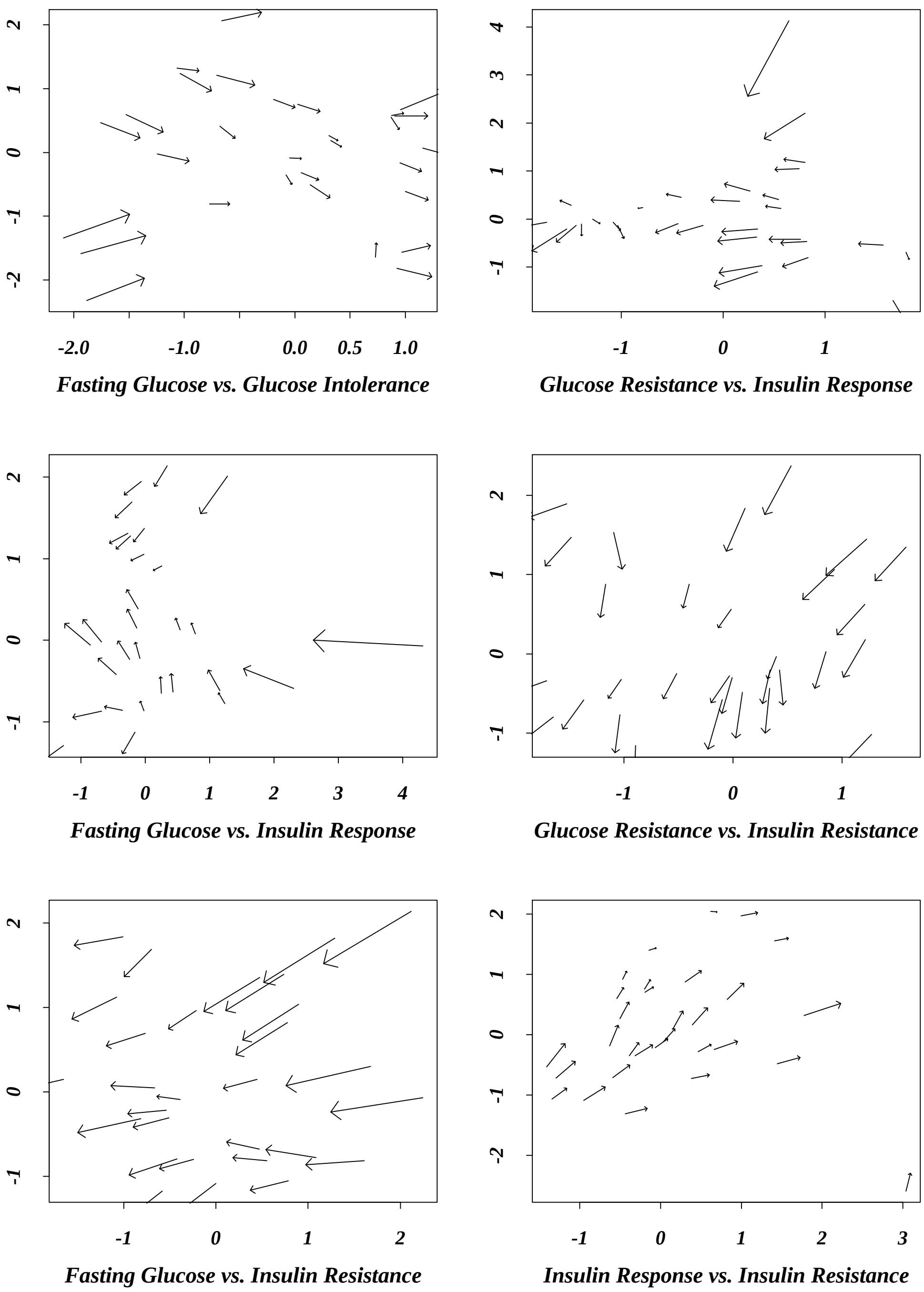
As discussed earlier, we can also construct tests of equality of the underlying distribution of two multivariate samples in a similar fashion. Here we compute transformed observations for both the samples and based on the ranks of the observations of one sample, we compute the quantiles of the other sample and draw directed arrows. Sum of these directed arrows provides us a test statistic for testing

Figure 3: Bivariate Q-Q plot for Iris data



equality of the underlying distribution of two samples. To illustrate the comparison between two samples using Q-Q plots, we again used the blood sugar data example which we have used earlier to demonstrate quantile contour plots. In Figure 4, we construct Q-Q plots for comparing normal patients with overt nonketotic patients by computing multivariate affine invariant l_2-ranks of the first sample and corresponding geometric quantiles of the transformed observations of the second sample. We have taken two variable at a time. From these Q-Q plots, it is quite apparent that the underlying distributions of the normal patients and overt diabetic patients are quite different. Large arrow lengths in all the plots suggest that there are possibly differences in locations and scales of the distributions and also

Figure 4: Q-Q plots for comparing Overt diabetic patients and normal patients



the arrows are oriented towards a common direction indicating possible differences in the shapes of the distributions.

4.4 L-Estimates

In the univariate set-up, linear combinations of order statistics or L-estimators have played an extremely important role in the development of robust methods for the one sample location problem. Serfling (1980) gave a detailed account of various important univariate descriptive statistics (e.g. trimmed mean, inter-quartile range etc.) by formulating them as L-statistics and derived their asymptotic properties. It is possible to extend the concept of L-estimators of univariate location to a multivariate set-up using TR l_p-quantiles in a natural way. To construct L-estimators we have to form suitable weighted averages of $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$'s as $\boldsymbol{u}$ varies over an appropriate subset of $B_q^{(d)}$. One has to keep in mind that a $\boldsymbol{u}$ with $\|\boldsymbol{u}\|_q$ close to zero corresponds to a central quantile and for a $\boldsymbol{u}$ with $\|\boldsymbol{u}\|_q$ close to one corresponds to an extreme quantile.

Suppose that μ is an appropriately chosen probability measure on $B_q^{(d)}$ supported on a subset S of $B_q^{(d)}$. Then an L-estimate of multivariate location will have the form $\int_S \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})\mu(d\boldsymbol{u})$. Specifically, if we consider $J(\boldsymbol{u})$, a bounded, real-valued continuous function defined on $B_q^{(d)}$, we may define L-estimate corresponding to the function J as,

$$\hat{\boldsymbol{\theta}}_J^{(\alpha,p)} = \int_{B_q^{(d)}} J(\boldsymbol{u})\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})\, d\boldsymbol{u} \tag{4.1}$$

By considering different forms of the function $J(\boldsymbol{u})$, one can construct various interesting descriptive statistics of the multivariate data cloud. One can define analogs of trimmed mean or inter-quartile range for a multivariate set-up. In the above set-up, if we consider S to be the l_q-ball with center at the origin and radius r, where r is a constant such that $0 < r < 1$, (i.e. is $S = \{\boldsymbol{u}|\boldsymbol{u} \in \mathbb{R}^d, \|\boldsymbol{u}\|_q \leq r\}$) and the probability measure μ is chosen to be the uniform probability measure on S, $\int_S \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})\mu(d\boldsymbol{u})$ will be a typical definition of trimmed mean by taking $J(\boldsymbol{u}) = (\lambda(S))^{-1}\mathbf{1}_{\{S\}}(\boldsymbol{u})$, where S is the l_q-ball of radius r as defined above and $\lambda(S)$ is the Lebesgue measure of the set S. Thus the r-trimmed multivariate mean is given by

$$\hat{\boldsymbol{\theta}}_{(r)}^{(\alpha,p)} = \frac{1}{\lambda(S)} \int_S \hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})\, d\boldsymbol{u}. \tag{4.2}$$

As the transformation retransformation l_p-quantiles are equivariant under arbitrary affine transformations, the L-estimators $\hat{\boldsymbol{\theta}}_J^{(\alpha,p)}$ or the trimmed multivariate mean $\hat{\boldsymbol{\theta}}_{(r)}^{(\alpha,p)}$ are also affine equivariant. Some recent attempts to construct and study various versions of trimmed mean estimate of multivariate location using different ideas can be found in Donoho and Gasko (1992), Gordaliza (1991) and Nolan (1992). Recently, Koltchinskii (1997) showed that the geometric quantile process converges asymptotically to a Gaussian process under some suitable conditions. Using that result, he proved the asymptotic normality of the L-estimates based on non-equivariant geometric quantiles. We can establish similar results for our TR l_p-quantile processes and derive asymptotic normality of affine equivariant L-estimates based on them.

5 Appendix : The Proofs

Proof of Theorem 2.1: As d-dimensional random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are transformed to $\mathbf{A}\boldsymbol{X}_1+\boldsymbol{b}, \ldots, \mathbf{A}\boldsymbol{X}_n+\boldsymbol{b}$, where $\mathbf{A}$ is a $d\times d$ nonsingular matrix and $\boldsymbol{b}$ is $d\times 1$ vector, the transformation matrix $\mathbf{X}(\alpha)$ gets transformed to $\mathbf{AX}(\alpha)$. For $\boldsymbol{u} \in B_q^{(d)}$, define $\boldsymbol{w} = (\|\boldsymbol{u}\|_q/\|\mathbf{A}\boldsymbol{u}\|_q)\mathbf{A}\boldsymbol{u}$. Note that the index vector $\boldsymbol{v}(\alpha)$ based on original observations and corresponding to $\boldsymbol{u}$ is defined as $[\|\boldsymbol{u}\|_q/\|\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}\|_q]\{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{u}$ and that based on transformed observations and corresponding to $\boldsymbol{w}$ is given by

$$\boldsymbol{v}^*(\alpha) = \frac{\{\mathbf{AX}(\alpha)\}^{-1}\boldsymbol{w}}{\|\{\mathbf{AX}(\alpha)\}^{-1}\boldsymbol{w}\|_q}\|\boldsymbol{w}\|_q = \frac{\{\mathbf{AX}(\alpha)\}^{-1}\mathbf{A}\boldsymbol{u}}{\|\{\mathbf{AX}(\alpha)\}^{-1}\mathbf{A}\boldsymbol{u}\|_q}\|\boldsymbol{u}\|_q = \boldsymbol{v}(\alpha).$$

Also, note that the $\boldsymbol{Y}_i^{(\alpha)}$'s will be transformed to $\boldsymbol{Z}_i^{(\alpha)} = \boldsymbol{Y}_i^{(\alpha)} + \{\mathbf{AX}(\alpha)\}^{-1}\boldsymbol{b}$, and the $\boldsymbol{v}(\alpha)$-th l_p-quantile is equivariant under a location shift of the data points. Hence, the $v(\alpha)$-th l_p-quantile based on $\boldsymbol{Z}_i^{(\alpha)}$'s is transformed to $\hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{w}) = \hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{u}) + \{\mathbf{AX}(\alpha)\}^{-1}\boldsymbol{b}$. Consequently $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{w})$, the $\boldsymbol{w}$-th TR l_p-quantile based on transformed observations, which is defined as $\{\mathbf{AX}(\alpha)\}\hat{\boldsymbol{R}}_n^{(\alpha,p)}(\boldsymbol{w})$, will be equal to $\mathbf{A}\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) + \boldsymbol{b}$. Thus the $\boldsymbol{w}$-th TR l_p-quantile based on transformed observations $(\mathbf{A}\boldsymbol{X}_i + \boldsymbol{b})$ is $\mathbf{A}\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u}) + \boldsymbol{b}$, where $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$ is the $\boldsymbol{u}$-th TR l_p-quantile based on original $\boldsymbol{X}_i$'s. $\square$

Before we prove Theorem 3.1, we state an asymptotic representation of $\hat{\boldsymbol{Q}}_n^{(1)}(\boldsymbol{u})$, which is the non-equivariant vector of coordinatewise sample quantiles. Consider $\boldsymbol{Q}^{(1)}(\boldsymbol{u})$ as the vector of marginal quantiles of the population distribution function F.

Lemma 5.1 *Let the j-th marginal distribution F_j of F be twice differentiable and $f_j(Q_j^{(1)}(\boldsymbol{u})) > 0$ where f_j is the j-th marginal density for $1 \leq j \leq d$ and $\boldsymbol{Q}^{(1)}(\boldsymbol{u}) = (Q_1^{(1)}(\boldsymbol{u}), \ldots, Q_d^{(1)}(\boldsymbol{u}))^T$. Then*

$$\hat{\boldsymbol{Q}}_n^{(1)}(\boldsymbol{u}) - \boldsymbol{Q}^{(1)}(\boldsymbol{u}) = n^{-1} D_f^{-1} \sum_{i=1}^{n} [Sign(\boldsymbol{X}_i - \boldsymbol{Q}^{(1)}(\boldsymbol{u})) + \boldsymbol{u}] + R_n(\boldsymbol{u}), \tag{5.1}$$

where D_f is the diagonal matrix $diag(2f_1(Q_1^{(1)}(\boldsymbol{u})), \ldots, 2f_d(Q_d^{(1)}(\boldsymbol{u})))$ and as $n \to \infty$, the remainder term $R_n(\boldsymbol{u})$ is almost surely $O(n^{-3/4}(\log n)^{3/4})$.

The above Lemma follows almost directly from Bahadur (1966) representation for sample quantiles in the univariate case and thus we omit the proof of this Lemma. For the vector of marginal quantiles, the asymptotic normality can be derived with weaker conditions and it is studied in detail by Babu and Rao (1988).

Proof of Theorem 3.1: Define the transformed observation $\boldsymbol{Z}_i^{(\alpha)}$ as $\boldsymbol{Z}_i^{(\alpha)} = \{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{X}_i$. Then, given the $\boldsymbol{X}_i$'s for which $i \in \alpha$, the transformed observations $\boldsymbol{Z}_i^{(\alpha)}$'s with $i \notin \alpha$ are conditionally i.i.d. random vectors with common density $|\det\{\mathbf{X}(\alpha)\}| h\{\mathbf{X}(\alpha)\boldsymbol{z}\}$. The conditions of Theorem 3.1 implies that the conditions of Lemma 5.1 holds for the density of transformed data. Thus using Lemma 5.1 for the coordinatewise quantiles of transformed observations $\boldsymbol{Z}_i^{(\alpha)}$'s for $1 \leq i \leq n$, $i \notin \alpha$, we have the representation in (3.1) for the TR l_1-quantile $\hat{\boldsymbol{Q}}_n^{(\alpha,1)}(\boldsymbol{u})$. $\square$

Before we prove Theorem 3.2, let us prove a Lemma on asymptotic representation of non-equivariant l_p-quantiles $\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u})$ for $1 < p < \infty$. Let $\boldsymbol{Q}^{(p)}(\boldsymbol{u})$ be the population l_p quantile and define the matrices $D_1^{(p)}(\boldsymbol{Q}) = E\{\Psi_p(\boldsymbol{X} - \boldsymbol{Q})\}$ and $D_2^{(p)}(\boldsymbol{Q}, \boldsymbol{u}) = E\{[\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X} - \boldsymbol{Q})][\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X} - \boldsymbol{Q})]^T\}$. Note that, $D_1^{(p)}(\boldsymbol{Q})$ will be positive definite unless the distribution of $\boldsymbol{X}$ is completely supported on a straight line in $\mathbb{R}^d$, and the expectation defining $D_1^{(p)}(\boldsymbol{Q})$ will exist finitely for $d \geq 2$ whenever $\boldsymbol{X}$ has a density that is bounded on compact subsets of $\mathbb{R}^d$. These facts can be verified directly.

Lemma 5.2 *Assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ is a sequence of independent and identically distributed random vectors in $\mathbb{R}^d$ such that their common density is bounded on every bounded subset of $\mathbb{R}^d$. Then for any fixed $\boldsymbol{u} \in B_q^{(d)}$, where $1 < p < \infty$ and $1/p + 1/q = 1$, we have the following Bahadur type representation for the $\boldsymbol{u}$-th l_p-quantile:*

$$\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(p)}(\boldsymbol{u}) = n^{-1}[D_1^{(p)}(\boldsymbol{Q}(\boldsymbol{u}))]^{-1} \sum_{i=1}^{n} \boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}^{(p)}(\boldsymbol{u})) + R_n(\boldsymbol{u}), \quad (5.2)$$

where as $n \to \infty$, $R_n(\boldsymbol{u})$ is almost surely $O(logn/n)$ if $d \geq 3$ and when $d = 2$, $R_n(\boldsymbol{u})$ is almost surely $o(n^{-\beta})$ for any fixed β such that $0 < \beta < 1$.

Proof: We present the proof of the lemma following arguments similar to those used to prove the main results in Chaudhuri (1992, 1996) with suitable modifications. We split the proof in several parts to expose the key ideas. Koltchinskii (1997) obtained a similar representation theorem but with slower rate of convergence for the remainder term $R_n(\boldsymbol{u})$. It follows form his result that there exists a constant $K_1 > 0$ such that we have almost surely $\|\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(p)}(\boldsymbol{u})\|_p \leq K_1$ for all n sufficiently large.

Now observe that $\sum_{i=1}^{n} \boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}))$ is bounded [cf. Kemperman (1987), Chaudhuri(1996)] with the convention $\boldsymbol{\varphi}_p(\boldsymbol{u}, 0) = \boldsymbol{u}$. Consequently, an easy extension of Proposition 5.6 of Chaudhuri (1992) implies the existence of a constant $K_2 > 0$ such that almost surely $\|\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(p)}(\boldsymbol{u})\|_p \leq K_2 n^{-1/2}(\log n)^{1/2}$ for all n sufficiently large. Recall here that $\boldsymbol{Q}^{(p)}(\boldsymbol{u})$ satisfies $E[\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{Q}^{(p)}(\boldsymbol{u}))] = \mathbf{0}$, and lemmas 5.3 and 5.4 of Chaudhuri (1992) can be suitably modified to imply that the magnitude of the d-dimensional vector $\sum_{i=1}^{n} \boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q})$ will explode to infinity almost surely as $n \to \infty$, unless $\boldsymbol{Q}$ lies inside a ball in $\mathbb{R}^d$ with center at $\boldsymbol{Q}^{(p)}(\boldsymbol{u})$ and radius of the order $O(n^{-1/2}[\log n]^{1/2})$.

Let B_n be the subset of $\mathbb{R}^d$ defined as

$$B_n = \{(v_1, \ldots, v_d) | n^4 v_i = \text{ an integer and } |v_i| \leq K_2 n^{-1/2}(\log n)^{1/2} \text{ for } 1 \leq i \leq d\}.$$

For $\boldsymbol{Q} \in \mathbb{R}^d$, define

$$\Delta(\boldsymbol{Q}) = E\{\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_1 - \boldsymbol{Q})\} + \{D_1^{(p)}(\boldsymbol{Q}^{(p)}(\boldsymbol{u}))\}\{\boldsymbol{Q} - \boldsymbol{Q}^{(p)}(\boldsymbol{u})\}$$

and for $\boldsymbol{Q} \in B_n$, define

$$\begin{aligned} &\Lambda_n(\boldsymbol{Q}^{(p)}(\boldsymbol{u}), \boldsymbol{Q} + \boldsymbol{Q}^{(p)}(\boldsymbol{u})) \\ = \;& n^{-1}\sum_{i=1}^{n}\{\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}^{(p)}(\boldsymbol{u})) - \boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}^{(p)}(\boldsymbol{u}) - \boldsymbol{Q})\} \\ &+E\{\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_1 - \boldsymbol{Q}^{(p)}(\boldsymbol{u}) - \boldsymbol{Q})\}. \end{aligned}$$

Consider a sample sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ such that, for all n sufficiently large, we have $\|\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}^{(p)}(\boldsymbol{u})\|_p \leq K_4(\log n/n)^{1/2}$, and $\|\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}_n^*\|_p \leq K_3(\log n/n)$ for some $K_3 > 0$ and $\boldsymbol{Q}_n^*$ is a point in $\mathbb{R}^d$ such that $\boldsymbol{Q}_n^* - \boldsymbol{Q}^{(p)}(\boldsymbol{u}) \in B_n$, and $\boldsymbol{Q}_n^*$ is closest to $\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u})$ in l_p-norm. If there are several choices for such a $\boldsymbol{Q}_n^*$, we can choose any one of them. It is quite easy to verify (see the proof of Proposition 5.6 in Chaudhuri 1992) that the collection of all sample sequences satisfying these requirements will form a set of probability one. Now, we can write

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}^{(p)}(\boldsymbol{u})) \\ = \;&\Lambda_n\{\boldsymbol{Q}^{(p)}(\boldsymbol{u}), \boldsymbol{Q}_n^*\} + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}_n^*) - \Delta(\boldsymbol{Q}_n^*) + D_1^{(p)}(\boldsymbol{Q}^{(p)}(\boldsymbol{u}))\{\boldsymbol{Q}_n^* - \boldsymbol{Q}^{(p)}(\boldsymbol{u})\}. \end{aligned}$$

Some minor modifications of the arguments used in the proof of Fact 5.8 and Lemma 5.9 in Chaudhuri (1992) implies that for a fixed constant $M^* > 0$,

$$\sup_{\|Q - Q^{(p)}(u)\|_p \leq M^*(\log n/n)^{1/2}} \|\Delta(\boldsymbol{Q})\|_q = O(\log n/n) \tag{5.3}$$

as $n \to \infty$, for $d \geq 3$. On the other hand for $d = 2$, we have

$$\sup_{\|Q - Q^{(p)}(u)\|_p \leq M^*(\log n/n)^{1/2}} \|\Delta(\boldsymbol{Q})\|_q = o(n^{-\omega}) \tag{5.4}$$

as $n \to \infty$ for any constant ω such that $1/2 < \omega < 1$. We also have that for $d \geq 3$ there is a constant $K_5 > 0$ such that $\max_{Q \in B_n} \|\Lambda_n(\boldsymbol{Q}^{(p)}(\boldsymbol{u}), \boldsymbol{Q} + \boldsymbol{Q}^{(p)}(\boldsymbol{u}))\|_q \leq K_5(\log n/n)$ almost surely for all n sufficiently large. Also, if $d = 2$, we have

$\max_{Q \in B_n} \|\Lambda_n(\boldsymbol{Q}^{(p)}(\boldsymbol{u}), \boldsymbol{Q} + \boldsymbol{Q}^{(p)}(\boldsymbol{u}))\|_q = o(n^{-\omega})$ almost surely as $n \to \infty$, where ω is any constant satisfying $0 < \omega < 1$. On the other hand, it is quite easy to verify (cf. the inequality (6) in the proof of proposition 5.6 in Chaudhuri 1992) that $n^{-1} \sum_{i=1}^{n} \boldsymbol{\varphi}_p(\boldsymbol{u}, \boldsymbol{X}_i - \boldsymbol{Q}_n^*) = O(n^{-1} \log n)$ almost surely as $n \to \infty$.

The proof of the lemma is now complete using the positive definiteness of the matrix $D_1^{(p)}[\boldsymbol{Q}^{(p)}(\boldsymbol{u})]$ together with the fact that $\|\hat{\boldsymbol{Q}}_n^{(p)}(\boldsymbol{u}) - \boldsymbol{Q}_n^*\|_p$ is $O(n^{-4})$ as $n \to \infty$ along our chosen sample sequence. □

Proof of Theorem 3.2: Define $\boldsymbol{Y}_i^{(\alpha)} = \{\mathbf{X}(\alpha)\}^{-1}\boldsymbol{X}_i$, for $1 \le i \le n$, $i \notin \alpha$. Then, given the $\boldsymbol{X}_j$'s for which $j \in \alpha$, the transformed observations $\boldsymbol{Y}_i^{(\alpha)}$'s with $i \notin \alpha$ are conditionally i.i.d random vectors with common density $|\det\{\mathbf{X}(\alpha)\}| h\{\mathbf{X}(\alpha)\boldsymbol{y}\}$. As the density h is bounded on every bounded subset of $\mathbb{R}^d$, the conditions in Lemma 5.2 hold for transformed observations. Using Lemma 5.2 for representation of l_p-quantiles of transformed observations $\boldsymbol{Y}_i^{(\alpha)}$'s, we have the representation in (3.2) for TR l_p-quantile $\hat{\boldsymbol{Q}}_n^{(\alpha,p)}(\boldsymbol{u})$. □

Lemma 5.3 *Assume that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ are independent and identically distributed random variables with a common density $h(\boldsymbol{x})$ such that $\int_{\mathbb{R}^d}\{h(\boldsymbol{x})\}^{d+1}\, d\boldsymbol{x} < \infty$. Let Γ be a positive definite matrix with determinant equal to 1 and $\bar{\alpha}$ minimizes $t(\alpha) = trace[\{\mathbf{X}^*(\alpha)\}^T \Gamma^{-1} \mathbf{X}^*(\alpha)]/d$. Then $t(\bar{\alpha})$ converges in probability to 1 as $n \longrightarrow \infty$.*

Proof: Let $\mathbf{A}$ be a $d \times d$ positive definite matrix such that $\Gamma = \mathbf{A}\mathbf{A}^T$. Consider $\alpha = \{1, 2, \ldots, d+1\}$. As the underlying distribution of the $\boldsymbol{X}_i$'s are independent and identically distributed with a common density h, the joint probability density function of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{d+1}$ can be written as $\prod_{i=1}^{d+1} h(\boldsymbol{x}_i)$. Now we make the following transformation of variables:

$$\boldsymbol{Y}_1 = \mathbf{A}^{-1}(\boldsymbol{X}_2 - \boldsymbol{X}_1), \ldots, \boldsymbol{Y}_d = \mathbf{A}^{-1}(\boldsymbol{X}_{d+1} - \boldsymbol{X}_1), \boldsymbol{Y}_{d+1} = \mathbf{A}^{-1}\boldsymbol{X}_1.$$

Then the joint density of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{d+1}$ is given by

$$h(\mathbf{A}\boldsymbol{y}_{d+1}) \prod_{i=1}^{d} h\{\mathbf{A}(\boldsymbol{y}_i + \boldsymbol{y}_{d+1})\}. \tag{5.5}$$

Therefore, the joint density of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_d$ at the origin in $\mathbb{R}^{d\times d}$ is

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

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

$$(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_d) \mapsto \int_{\mathbb{R}^d} h(\mathbf{A}\boldsymbol{y}) \prod_{i=1}^{d} h\{\mathbf{A}(\boldsymbol{y}_i + \boldsymbol{y})\}\, d\boldsymbol{y} \tag{5.6}$$

from $\mathbb{R}^{d\times d}$ to $\mathbb{R}$ is everywhere continuous. Therefore the joint density of $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_d$ must remain bounded away from zero in a neighbourhood of $\mathbf{0} \in \mathbb{R}^{d\times d}$. Consequently the probability of the event that the columns of $\mathbf{AX}(\alpha)$ will be nearly orthogonal and of nearly same length (and hence $\{\mathbf{X}^*(\alpha)\}^T\Gamma^{-1}\mathbf{X}^*(\alpha)$ will be very close to $\mathbf{I}_d$) is bounded away from zero. In other words, we have for any $\epsilon > 0$,

$$P[\|\{\mathbf{X}^*(\alpha)\}^T\Gamma^{-1}\mathbf{X}^*(\alpha) - \mathbf{I}_d\|_1 < \epsilon] = p_\epsilon > 0. \tag{5.7}$$

Let $\alpha_1, \alpha_2, \ldots, \alpha_{k_n}$ be disjoints subsets of $\{1, 2, \ldots, n\}$ each with size $d+1$ such that $k_n \to \infty$ as $n \to \infty$ [e.g. k_n may be equal to $n/(d+1)$]. Then
$P[|trace[\{\mathbf{X}^*(\bar{\alpha})\}^T\Gamma^{-1}\mathbf{X}^*(\bar{\alpha}) - \mathbf{I}_d]| > \epsilon]$

$$\begin{aligned}
&\leq P[|trace[\{\mathbf{X}^*(\alpha_j)\}^T\Gamma^{-1}\mathbf{X}^*(\alpha_j) - \mathbf{I}_d]| > \epsilon, \text{ for } 1 \leq j \leq k_n\} \\
&\leq (1 - p_\epsilon)^{k_n} \longrightarrow 0 \text{ as } n \to \infty
\end{aligned}$$

Hence, the result follows. □

Proof of Theorem 3.7: For $M > 1$, define $K_M^n = \{\alpha = \{i_0, i_1, \ldots, i_d\} : \ t(\alpha) \equiv trace[\{\mathbf{X}^*(\alpha)\}^T\Sigma^{*-1}\mathbf{X}^*(\alpha)]/d \leq M\}$. Then it is easy to see that there exists some $M_1 > 0$ such that for any $\alpha \in K_M^n$, $trace[\mathbf{X}^*(\alpha)\{\mathbf{X}^*(\alpha)\}^T] \leq M_1$. Observe that, for any $\alpha \in K_M^n$

$$\begin{aligned}
\|\{\mathbf{X}^*(\alpha)\}^T\hat{\Sigma}^{*-1}\mathbf{X}^*(\alpha) - \{\mathbf{X}^*(\alpha)\}^T\Sigma^{*-1}\mathbf{X}^*(\alpha)\|_2 &\leq \|\mathbf{X}^*(\alpha)\|_2^2\|\hat{\Sigma}^{*-1} - \Sigma^{*-1}\|_2 \\
&\leq M_1\|\hat{\Sigma}^{*-1} - \Sigma^{*-1}\|_2.
\end{aligned}$$

Now, since $\hat{\Sigma}^* \xrightarrow{p} \Sigma^*$ as $n \to \infty$, where Σ^* is a positive definite matrix with determinant equal to 1, we have

$$\sup_{\alpha \in K_M^n} \|\{\mathbf{X}^*(\alpha)\}^T\hat{\Sigma}^{*-1}\mathbf{X}^*(\alpha) - \{\mathbf{X}^*(\alpha)\}^T\Sigma^{*-1}\mathbf{X}^*(\alpha)\|_2 \xrightarrow{p} 0 \tag{5.8}$$

as $n \to \infty$. Since $\bar{\alpha}$ minimizes $t(\alpha)$, by taking Σ^* as Γ in Lemma 5.3, we must have with large probability $\bar{\alpha} \in K_M^n$ for all sufficiently large n. Therefore using the continuity of the trace function of a matrix, we have

$$|\hat{t}(\bar{\alpha}) - t(\bar{\alpha})| \xrightarrow{p} 0, \tag{5.9}$$

where $\hat{t}(\alpha) = trace[\{\mathbf{X}^*(\alpha)\}^T \hat{\Sigma}^{*-1} \mathbf{X}^*(\alpha)]/d$. Thus, for all sufficiently large n, $\hat{t}(\bar{\alpha}) \leq M_2$ with large probability for some $M_2 > 0$. In other words, for $\hat{\alpha}$ which minimizes $\hat{t}(\alpha)$, we have $\hat{t}(\hat{\alpha}) \leq M_2$. Therefore $trace[\mathbf{X}^*(\hat{\alpha})\{\mathbf{X}^*(\hat{\alpha})\}^T]/d$ is also bounded in probability as $n \to \infty$. This in turn ensures that $|\hat{t}(\hat{\alpha}) - t(\hat{\alpha})| \xrightarrow{p} 0$ as $n \to \infty$.

Next, since $\hat{\alpha}$ minimizes $\hat{t}(\alpha)$ and $\bar{\alpha}$ minimizes $t(\alpha)$, it follows by some straightforward analysis that $|\hat{t}(\hat{\alpha}) - t(\hat{\alpha})| < \epsilon$ and $|\hat{t}(\bar{\alpha}) - t(\bar{\alpha})| < \epsilon$ will imply that $|\hat{t}(\hat{\alpha}) - t(\bar{\alpha})| < \epsilon$. Hence, we have

$$P[|\hat{t}(\hat{\alpha}) - t(\bar{\alpha})| > \epsilon] \leq P[|\hat{t}(\hat{\alpha}) - t(\hat{\alpha})| > \epsilon] + P[|\hat{t}(\bar{\alpha}) - t(\bar{\alpha})| > \epsilon],$$

and consequently $|\hat{t}(\hat{\alpha}) - t(\bar{\alpha})| \xrightarrow{p} 0$ as $n \to \infty$. Finally, since

$$|t(\hat{\alpha}) - t(\bar{\alpha})| \leq |t(\hat{\alpha}) - \hat{t}(\hat{\alpha})| + |\hat{t}(\hat{\alpha}) - t(\bar{\alpha})|,$$

it follows from Lemma 5.3 that $t(\hat{\alpha})$ converges in probability to 1. $\square$

Lemma 5.4 *Let $\{\mathbf{A}_n\}$ be a sequence of $d \times d$ random positive definite matrices such that $\det(\mathbf{A}_n) = 1$ for all $n \geq 1$ and $trace(\mathbf{A}_n) \xrightarrow{p} d$ as $n \to \infty$. Then $\mathbf{A}_n \xrightarrow{p} \mathbf{I}_d$ as $n \to \infty$.*

Proof : Let the eigenvalues of the positive definite matrix $\mathbf{A}_n$ be $\lambda_{1:n} \leq \lambda_{2:n} \leq \cdots \leq \lambda_{d:n}$. Then, if we show that $\lambda_{1:n} \xrightarrow{p} 1$ and $\lambda_{d:n} \xrightarrow{p} 1$ as $n \to \infty$, the proof of the Lemma will be complete. If possible, suppose that $\lambda_{d:n}$ does not converge in probability to 1 as $n \to \infty$. Then there exists some $\epsilon > 0$ and $\delta > 0$ such that for infinitely many values of $n \geq 1$, we will have

$$P[\lambda_{d:n} > 1 + \epsilon] > \delta.$$

Define $\mu_n = (\lambda_{1:n} + \cdots + \lambda_{d-1:n})/(d-1)$, i.e. the average of the eigenvalues excluding the maximum one. Then, as the product of all the eigenvalues is 1, we have by A.M.–G.M. inequality $\mu_n \geq \lambda_{d:n}^{-1/(d-1)}$. Thus we have

$$trace(\mathbf{A}_n)/d = \frac{\lambda_{d:n} + (d-1)\mu_n}{d} \geq \frac{\lambda_{d:n} + (d-1)\lambda_{d:n}^{-1/(d-1)}}{d} > 1 + \epsilon_1$$

for some $\epsilon_1 > 0$ whenever $\lambda_{d:n} > 1 + \epsilon$. Here ϵ_1 depends on ϵ and d only. Therefore

$$P[trace(\mathbf{A}_n)/d > 1 + \epsilon_1] \geq P[\lambda_{d:n} > 1 + \epsilon] > \delta,$$

which contradicts the fact that $trace(\mathbf{A}_n)$ converges in probability to d as $n \to \infty$. Hence, we must have $\lambda_{d:n} \xrightarrow{p} 1$ as $n \to \infty$.

As the maximum eigenvalue $\lambda_{d:n}$ converges to 1 and the determinant of the matrix $\mathbf{A}_n$ is 1, all other eigenvalues including the minimum one must converge to 1 in probability as $n \to \infty$. □

Proof of Corollary 3.8: Theorem 3.7 implies that $trace[\{\mathbf{X}^*(\hat{\alpha})\}^T \Sigma^{*-1} \mathbf{X}^*(\hat{\alpha})]$ tends to d in probability as $n \to \infty$. Hence, $\|\mathbf{X}^*(\hat{\alpha})\|_2$ must remain bounded in probability as $n \to \infty$. Also, since $\det[\{\mathbf{X}^*(\hat{\alpha})\}^T \Sigma^{*-1} \mathbf{X}^*(\hat{\alpha})] = 1$, Theorem 3.7 and Lemma 5.4 imply that

$$\{\mathbf{X}^*(\hat{\alpha})\}^T \Sigma^{*-1} \mathbf{X}^*(\hat{\alpha}) \xrightarrow{p} \mathbf{I}_d$$

as $n \to \infty$. The proof is now complete by observing the fact

$$\|\mathbf{X}^*(\hat{\alpha})\{\mathbf{X}^*(\hat{\alpha})\}^T - \Sigma^*\|_2 \leq \|\mathbf{X}^*(\hat{\alpha})\|_2^2 \|\{\mathbf{X}^*(\hat{\alpha})\}^{-1} \Sigma^* \{[\mathbf{X}^*(\hat{\alpha})]^T\}^{-1} - \mathbf{I}_d\|_2.$$

□

Acknowledgements

The author is grateful to Probal Chaudhuri for helpful suggestions and comments. Thanks are also due to two anonymous referees for a very careful reading of an earlier version of the manuscript and raising several interesting points to improve the paper.

References

BABU, G. J. AND RAO, C. R. (1988) Joint asymptotic distribution of marginal quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis*, **27**, 15–23.

BAHADUR, R.R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, **37**, 577–580.

BARNETT, V. (1976). The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society, Series A*, **139**, 318–354.

BEDALL, F.K. AND ZIMMERMANN, H. (1979). AS143: The mediancenter, *Applied Statistics*, **28**, 325–328.

BROWN, B. M. (1983). Statistical use of the spatial median. *Journal of the Royal Statistical Society, Series B*, **45**, 25−30.

BROWN, B.M. AND HETTMANSPERGER, T.P. (1987). Affine invariant rank methods in the bivariate location model. *Journal of the Royal Statistical Society, Series B*, **49**, 301–310.

BROWN, B.M. AND HETTMANSPERGER, T.P. (1989). An affine invariant bivariate version of the sign test. *Journal of the Royal Statistical Society, Series B*, **51**, 117–125.

CHAKRABORTY, B. AND CHAUDHURI, P. (1996). On a transformation and re-transformation technique for constructing affine equivariant multivariate median. *Proceedings of the American Mathematical Society*, **124**, 2539–2547.

CHAKRABORTY, B. AND CHAUDHURI, P. (1997). On multivariate rank regression. *L_1-Statistical Procedures and Related Topics*, ed. Y. Dodge, IMS Lecture Notes-Monograph Series, Volume 31, pp. 399–414.

CHAKRABORTY, B., AND CHAUDHURI, P. (1998) On an adaptive transformation retransformation estimate of multivariate location. *Journal of the Royal Statistical Society, Series B*, **60**, 145–157.

CHAKRABORTY, B. AND CHAUDHURI, P. (1999). On affine equivariant sign and rank tests in one and two sample multivariate problems. *Multivariate Analysis, Design of Experiments and Survey Sampling*, ed. S. Ghosh. Marcel Dekker Inc., New York.

CHAKRABORTY, B., CHAUDHURI, P. AND OJA, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica*, **8**, 767–784.

CHAUDHURI, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *The Annals of Statistics*, **20**, 897–916.

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, **91**, 862–872.

CHAUDHURI, P. AND SENGUPTA, D. (1993). Sign tests in multidimension : Inference based on the geometry of the data cloud. *Journal of the American Statistical Association*, **88**, 1363–1370.

DAVIES, P.L. (1987), Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, **15**, 1269–1292.

DONOHO, D.L. AND GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, **20**, 1803–1827.

EDDY, W.F. (1983). Set valued ordering of bivariate data. *Stochastic Geometry, Geometric Statistics, and Stereology*, eds. R.V. Ambartsumian and W. Weil, Leipzig: Tuebner, pp. 79–90.

EDDY, W.F. (1985). Ordering of multivariate data. *Computer Science and Statistics: The Interface*, ed. L. Billard, Amsterdam: North-Holland, pp. 25–30.

GOWER, J. C. (1974). The mediancenter. *Journal of the Royal Statistical Society, Series C*, **23**, 466–470.

Gordaliza, A. (1991). On the breakdown point of multivariate location estimators based on trimming procedures. *Statistics & Probability Letters*, **11**, 387–394.

Hettmansperger, T. P., Nyblom, J. and Oja, H. (1994). Affine invariant multivariate one sample sign tests. *Journal of the Royal Statistical Society, Series B*, **56**, 221−234.

Isogai, T. (1985). Some extension of Haldane's multivariate median and its application. *Annals of the Institute of Statistical Mathematics.* **37**, 289–301.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. *Statistical Data Analysis Based on L_1 norm and Related Methods*, ed. Y. Dodge, Amsterdam: North-Holland, pp. 217–230.

Koltchinskii, V. (1997). M-Estimation, convexity and quantiles, *Annals of Statistics*, **25**, 435–477.

Liu, R.Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, **18**, 405–414.

Marden, J. (1998). Bivariate Q-Q plots and spider webplots. *Statistica Sinica*, **8**, 813–826.

Möttönen, J., and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, **5**, 201−213.

Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, **25**, 542–552.

Nolan, D.(1992). Asymptotics for multivariate trimming. *Stochastic Processes and Their Applications*, **42**, 157–169.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, **1**, 327–332.

Plackett, R. L. (1976). Comment on "The ordering of multivariate data", by V. Barnett. *Journal of the Royal Statistical Society, Series A*, **139**, 344–346.

JAN, S.L. AND RANDLES, R.H. (1994). A multivariate signed sum test for the one sample location problem, *Journal of Nonparametric Statistics*, **4**, 49–63.

RAO, C. R. (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā, Series A*, **50**, 289–313.

REAVEN, G.M. AND MILLER, R.G. (1979). An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologia*, **16**, 17–24.

REISS, R.D. (1989). *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics*, New York: Springer.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

TUKEY, J.W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974*, ed. R.D. James, Canadian mathematical Congress, vol. 2, 523–531.