Aust. N. Z. J. Stat. **45**(3), 2003, 343–352

A MULTIVARIATE PARALLELOGRAM AND ITS APPLICATION TO MULTIVARIATE TRIMMED MEANS

JYH-JEN HORNG SHIAU[1]* AND LIN-AN CHEN[1]

National Chiao Tung University

Summary

This paper introduces a multivariate parallelogram that can play the role of the univariate quantile in the location model, and uses it to define a multivariate trimmed mean. It assesses the asymptotic efficiency of the proposed multivariate trimmed mean by its asymptotic variance and by Monte Carlo simulation.

Key words: multivariate parallelogram; quantile; trimmed mean.

1. Introduction

Let $y_1, \ldots, y_n$ denote a random sample from a univariate population with distribution function F, and let $\hat{F}$ be the empirical distribution function obtained from this sample. Let $Q(\alpha_1, \alpha_2) = (F^{-1}(\alpha_1), F^{-1}(\alpha_2))$ denote the (α_1, α_2)-quantile interval of F and let $\hat{Q}(\alpha_1, \alpha_2) = (\hat{F}^{-1}(\alpha_1), \hat{F}^{-1}(\alpha_2))$ denote the corresponding sample quantile interval, where F^{-1} and $\hat{F}^{-1}$ are the inverse functions of F and $\hat{F}$, respectively. The sample quantile interval plays a very important role in statistical inference. For example, as a region with a particular coverage probability, the interval is a natural estimator for scale parameters such as the range and interquartile range. With this property, the quantile interval can be used in industrial applications to define a process capability index for process capability assessment, especially for non-normal processes. Also, this interval is routinely used in classifying the observations of a sample into good or bad observations in robust mean estimation, such as for the trimmed mean and Winsorized mean.

Analogues have been proposed for quantiles or order statistics in high dimensions. It is well known that the univariate quantile can be obtained by solving a minimization problem. Breckling & Chambers (1988) and Koltchinskii (1997) generalized the minimization problem for the multivariate case and then defined a multivariate quantile as the minimizer of the problem. Chaudhuri (1996) considered a geometric quantile that uses the geometry of multivariate data clouds. Chakraborty (2001) used a transformation-retransformation technique to introduce a multivariate quantile. However, these approaches do not have obvious settings for defining multivariate regions suitable for constructing descriptive statistics because they lack a natural ordering in multi-dimensional data.

Received November 2001; revised July 2002; accepted September 2002.
* Author to whom correspondence should be addressed.
[1] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. e-mail: jyhjen@stat.nctu.edu.tw
Acknowledgments. The authors thank the Managing Editor, the Technical Editor, an associate editor and a referee for their valuable comments, which greatly improved the quality of the paper. This research work was partially supported by the National Science Council of the Republic of China Grant Nos NSC90-2118-M-009-008 and NSC90-2118-M-009-016.

In contrast to the above approaches, Chen & Welsh (2002) proposed for the bivariate random vector $\boldsymbol{y}$ a bivariate quantile that partitions $\mathbb{R}^2$ into two-dimensional intervals of specified probabilities. In this sense, their approach is a natural extension of the univariate quantile. They first applied an appropriate transformation to $\boldsymbol{y}$, attempting to make the two coordinate components of the transformed random vector $\boldsymbol{x}$ uncorrelated. Denote $\boldsymbol{x} = (x_1, x_2)$. Then the partition takes the following two steps. (i) Find q_1 such that $\Pr(x_1 \leq q_1) = \alpha_1$. More specifically, the line $x_1 = q_1$ divides $\mathbb{R}^2$ into two sets, of which one set has coverage probability α_1 and the other has coverage probability $1 - \alpha_1$. (ii) Find q_2 such that $\Pr(x_1 \leq q_1, x_2 \leq q_2) = \alpha_2$. Thus, the line $x_2 = q_2$ divides the set with coverage probability α_1 into two subsets such that one has coverage probability α_2 and the other has coverage probability $\alpha_1 - \alpha_2$. Then (q_1, q_2) is called the (α_1, α_2)-th bivariate quantile of $\boldsymbol{x}$ in the transformed space. Finally, the bivariate quantile of $\boldsymbol{y}$ is obtained by back-transforming the bivariate quantile of $\boldsymbol{x}$. Let $(\hat{q}_1, \hat{q}_2)$ denote the corresponding sample quantile obtained from the data. Note that the distribution of $\hat{q}_2$ depends on the distribution of $\hat{q}_1$, which makes the asymptotic properties of $(\hat{q}_1, \hat{q}_2)$ quite complicated. This approach can be extended to higher dimensional data. However, it would be too complicated to study the asymptotic distribution of the sample multivariate quantile defined in this way, and difficult to use it to make statistical inference from data.

There is some work in the literature on multivariate median estimation. Oja (1983) defined the multivariate simplex median by minimizing the sum of volumes of simplices with vertices on the observations; Liu (1988, 1990) introduced the simplicial depth median by maximizing an empirical simplicial depth function. Small (1990) gives an excellent review of these papers.

Chaudhuri (1996) noted that most authors introduced descriptive statistics by merely generalizing the univariate statistics to the multivariate setup, with no clear population analogues for these multivariate descriptive statistics. In other words, descriptive statistics were being defined without the target population parameters to be estimated. Although the approach of Chen & Welsh (2002) defines both the multivariate population parameters and their corresponding estimators, it is worth developing alternatives that are easier to use in theoretical study and practical applications. The major purposes of this paper are to define a multivariate parallelogram region as a counterpart of the univariate quantile interval, and to propose a statistic to estimate it. The approach used here considers the same variable transformation as that in Chen & Welsh (2002), but it defines the multivariate quantile through the univariate quantile of each coordinate of the transformed variable. This avoids the difficulty that we encountered in studying the complicated asymptotic distribution of the Chen & Welsh multivariate quantile.

Inspired by an idea that Huber (1973, 1981) used when constructing a location-scale equivariant studentized M-estimator for location, we introduce multivariate quantile points and use them to construct a multivariate parallelogram. With sample multivariate parallelograms, many multivariate descriptive statistics, such as multivariate versions of scale estimators, process capability indices, and trimmed means, are easy to construct. In this paper, we study the large-sample properties of the sample multivariate quantile points and the trimmed means constructed by this parallelogram. We compute asymptotic generalized variances of the proposed multivariate trimmed mean and the Cramér–Rao lower bounds for various multivariate contaminated normal distributions. The study reveals that the proposed trimmed mean is quite efficient.

Section 2 introduces the multivariate parallelogram and the multivariate quantile points, and gives a large sample representation of the multivariate quantile points. Section 3 presents a multivariate trimmed mean and its large sample representation. Section 4 gives a comparative study of multivariate trimmed means for two different coordinate transformations and the sample mean based on the asymptotic efficiency and Monte Carlo simulations. Proofs of the theorems in this paper are given in Shiau & Chen (2002), which is available at http://www.stat.nctu.edu.tw/TechnicalReports/jyhjen/MultiTrimMean.ps

2. Multivariate parallelograms

Consider the multivariate location model

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{v}\,,$$

where $\boldsymbol{y}, \boldsymbol{\mu}$, and $\boldsymbol{v}$ are $p \times 1$ vectors with $\mathrm{E}(\boldsymbol{v}) = 0$ and $\mathrm{V}(\boldsymbol{y}) = \mathrm{V}(\boldsymbol{v}) = \boldsymbol{\Sigma}$. We want to find a subset of the sample space of $\boldsymbol{y}$ with a fixed (either exact or approximate) coverage probability. Apparently, there can be many choices for the shape of this subset. Popular ones include the cube, ellipsoid, parallelogram, trapezoid, etc. If we further require this subset to have the smallest volume, then we can expect different shapes for different distributions. For example, our investigation shows that the ellipsoid is better than the parallelogram when the distribution is bivariate normal while the result is opposite when the distribution is bivariate exponential or chi-squared.

Since $\boldsymbol{\Sigma}$ is symmetric positive definite, there exists a non-singular matrix $\boldsymbol{D} = \boldsymbol{\Sigma}^{1/2}$, such that $\boldsymbol{\Sigma} = \boldsymbol{D}\boldsymbol{D}^{\top}$. The transformed random vector $\boldsymbol{x} = \boldsymbol{D}^{-1}\boldsymbol{y}$ has the model

$$\boldsymbol{x} = \boldsymbol{D}^{-1}\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

with $\boldsymbol{\epsilon} = \boldsymbol{D}^{-1}\boldsymbol{v}$. Denote $\boldsymbol{x} = (x_1, \ldots, x_p)$. Since $\mathrm{V}(\boldsymbol{x}) = \boldsymbol{I}$, $x_1, \ldots, x_p$ are uncorrelated. It is then natural to consider a region formed by the Cartesian product of the p marginal quantile intervals. The multivariate parallelogram is thus obtained by transforming this region back to the $\boldsymbol{y}$-space. An advantage of choosing the parallelogram as the shape of the region is that asymptotic study of the proposed statistics based on the sample parallelogram is relatively simple compared to other shapes.

Definition 1. *For $j = 1, \ldots, p$ and $0 < \alpha_j < 1$, let $\xi_j = F_j^{-1}(\alpha_j)$ denote the univariate α_j-th quantile of the random variable x_j.*

(a) *Define the $(\alpha_1, \ldots, \alpha_p)$-th multivariate quantile point by $\boldsymbol{q}(\boldsymbol{\alpha}) = \boldsymbol{D}\boldsymbol{\xi}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)$. Denote $\boldsymbol{q}(\alpha\boldsymbol{1})$ by $\boldsymbol{q}(\alpha)$ when $\alpha_1 = \cdots = \alpha_p = \alpha$, where $\boldsymbol{1} = (1, \ldots, 1)$.*

(b) *Let $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{p1})$ and $\boldsymbol{\alpha}_2 = (\alpha_{12}, \ldots, \alpha_{p2})$ where $\alpha_{j1} < \alpha_{j2}$. Define the multivariate quantile set by*

$$Q(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \{\boldsymbol{q}(\alpha_{1a_1}, \ldots, \alpha_{pa_p}) \colon a_j = 1, 2,\, j = 1, \ldots, p\}\,,$$

which contains 2^p quantile points. Define $\xi_{jk} = F_j^{-1}(\alpha_{jk})$. The region

$$R(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \{\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \colon \xi_{j1} \le x_j \le \xi_{j2},\ j = 1, 2, \ldots, p\}$$

is called the parallel multivariate quantile region.

Suppose that we have a random sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ obeying the following multivariate location model:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{v}_i \qquad (i = 1, \ldots, n),$$

where $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent and identically distributed error vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Let $\hat{\boldsymbol{\Sigma}}$ denote an estimator of the variance–covariance matrix $\boldsymbol{\Sigma}$. Then $\hat{\boldsymbol{D}} = \hat{\boldsymbol{\Sigma}}^{1/2}$. Let $\boldsymbol{x}_i = \hat{\boldsymbol{D}}^{-1}\boldsymbol{y}_i, i = 1, \ldots, n,$ denote the transformed random vectors. Denote $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})$. Let $\hat{\xi}_{j1}$ and $\hat{\xi}_{j2}$ denote the α_{j1}-th and α_{j2}-th sample quantiles, respectively, based on the transformed observations $x_{j1}, \ldots, x_{jn}$.

Definition 2. *Define an estimator of the multivariate quantile point $\boldsymbol{q}(\boldsymbol{\alpha})$ by*

$$\hat{\boldsymbol{q}}(\boldsymbol{\alpha}) = \hat{\boldsymbol{D}}\hat{\boldsymbol{\xi}},$$

where $\hat{\boldsymbol{\xi}} = (\hat{\xi}_1, \ldots, \hat{\xi}_p)$. Consequently, an estimator of the multivariate quantile set is

$$\hat{Q}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \{\hat{\boldsymbol{q}}(\alpha_{1a_1}, \ldots, \alpha_{pa_p})\colon a_j = 1, 2, j = 1, \ldots, p\}.$$

The parallel multivariate quantile region is estimated by

$$\hat{R}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \{\boldsymbol{y} = \hat{\boldsymbol{D}}\boldsymbol{x}\colon \hat{\xi}_{j1} \le x_j \le \hat{\xi}_{j2}, j = 1, \ldots, p\}. \tag{1}$$

Theorem 1 states that the estimators of the quantile points and the parallelogram defined above are of desired equivariance properties. First, to simplify the notation, we fix the quantile levels $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ and $\boldsymbol{\alpha}_2 = 1 - \boldsymbol{\alpha}_1 = (1-\alpha_1, 1-\alpha_2, \ldots, 1-\alpha_p)$, and then suppress them from the notation of the statistics under study. Re-denote the estimated multivariate quantile points obtained from the samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ and $\boldsymbol{A}\boldsymbol{y}_1 + \boldsymbol{c}, \ldots, \boldsymbol{A}\boldsymbol{y}_n + \boldsymbol{c}$ by $\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_1}(\boldsymbol{y})$ and $\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_1}(\boldsymbol{A}\boldsymbol{y}+\boldsymbol{c})$, respectively. We also re-denote $\boldsymbol{D}$ by $\boldsymbol{D}(\boldsymbol{y})$ and $\hat{\boldsymbol{\xi}}_j$ by $\hat{\boldsymbol{\xi}}_j(\alpha_j, \boldsymbol{y})$ to indicate which dataset the statistics are based on. For the moment, we let $\hat{\boldsymbol{\Sigma}}$ be more general than the sample covariance matrix. Let $\hat{\boldsymbol{\Sigma}}(\boldsymbol{y})$ denote a $p \times p$ matrix representing a statistic obtained from the random sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$.

Theorem 1. *Let $\boldsymbol{A}$ denote a $p \times p$ non-singular matrix and $\boldsymbol{c}$ a $p \times 1$ vector. Suppose that the statistic $\hat{\boldsymbol{D}}$ satisfies $\hat{\boldsymbol{D}}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = \boldsymbol{A}\hat{\boldsymbol{D}}(\boldsymbol{y})$ and we denote the estimated parallel multivariate quantile region of (1) by $\hat{R}(\boldsymbol{\alpha}, \boldsymbol{y})$. Then*

(a) $\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_1}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = \boldsymbol{A}\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_1}(\boldsymbol{y}) + \boldsymbol{c}$ *and*

(b) $\hat{R}(\boldsymbol{\alpha}_1, \boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = \boldsymbol{A}\hat{R}(\boldsymbol{\alpha}_1, \boldsymbol{y}) + \boldsymbol{c}$.

On the other hand, suppose that the statistic $\hat{\boldsymbol{D}}$ satisfies $\hat{\boldsymbol{D}}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = -\boldsymbol{A}\hat{\boldsymbol{D}}(\boldsymbol{y})$. Then

(c) $\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_1}(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = \boldsymbol{A}\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}_2}(\boldsymbol{y}) + \boldsymbol{c}$ *and*

(d) $\hat{R}(\alpha_1, \boldsymbol{A}\boldsymbol{y} + \boldsymbol{c}) = \boldsymbol{A}\hat{R}(\boldsymbol{\alpha}_2, \boldsymbol{y}) + \boldsymbol{c}$.

For the large sample study, the following conditions are assumed for random vector $\boldsymbol{v}$ and sample covariance matrix $\hat{\boldsymbol{\Sigma}}$. For $j = 1, \ldots p$, let g_j and G_j denote the probability density function (pdf) and the cumulative distribution function (cdf) of the transformed error vector $\boldsymbol{D}^{-1}\boldsymbol{v}$, respectively. Let $\boldsymbol{\sigma}_j^{\mathsf{T}}$ denote the jth row of $\boldsymbol{D}^{-1}$. Let $\boldsymbol{u}, \boldsymbol{\delta} \in \mathbb{R}^p$.

(i) The pdf g_j and its derivative are both bounded and bounded away from 0 in a neighbourhood of $G_j^{-1}(\alpha_j)$ for $\alpha_j \in (0, 1)$, $j = 1, 2, \ldots, p$.

(ii) $n^{1/2}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) = \boldsymbol{O}_p(1)$.

(iii) There exists $\theta > 0$ such that the pdf of $\boldsymbol{v}^\top(\boldsymbol{\sigma}_j + \boldsymbol{\delta})$ is uniformly bounded in a neighbourhood of $G_j^{-1}(\alpha)$ for $\|\boldsymbol{\delta}\| \le \theta$, and the pdf of $\boldsymbol{v}^\top(\boldsymbol{\sigma}_j + \boldsymbol{\delta})(\boldsymbol{v}^\top\boldsymbol{u})(\boldsymbol{v}^\top\boldsymbol{\sigma}_j)$ is uniformly bounded away from 0 for $\|\boldsymbol{u}\| = 1$ and $\|\boldsymbol{\delta}\| \le \theta$.

(iv) $\mathrm{E}\big((\boldsymbol{v}^\top\boldsymbol{\sigma}_j)^2\|\boldsymbol{v}\|\big) < \infty$.

A representation of the multivariate quantile point is stated in Theorem 2.

Theorem 2. *Under conditions (i)–(iv),*

$$n^{1/2}(\hat{\boldsymbol{q}}_{\boldsymbol{\alpha}} - \boldsymbol{q}_{\boldsymbol{\alpha}}) = n^{-1/2}\boldsymbol{D}\sum_{i=1}^{n}\boldsymbol{u}_i + n^{1/2}(\hat{\boldsymbol{D}} - \boldsymbol{D})\boldsymbol{\xi} + \boldsymbol{D}n^{1/2}\big((\hat{\boldsymbol{D}}^{-1} - \boldsymbol{D}^{-1})\boldsymbol{\mu} + \boldsymbol{v}\big) + \boldsymbol{o}_p(1)\,,$$

where $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ip})$ *with* $u_{ij} = f_j^{-1}(\xi_j)(\alpha_j - \mathrm{I}(x_{ji} \le \xi_j))$; $\boldsymbol{v} = (v_1, \ldots, v_p)$ *with* $\boldsymbol{v}_j = (\boldsymbol{s}_j - \boldsymbol{\sigma}_j)^\top\mathrm{E}(\boldsymbol{v} \mid \boldsymbol{x}_j = \boldsymbol{\xi}_j)$; f_j *is the pdf of* $\boldsymbol{x}_j$; $\boldsymbol{s}_j^\top$ *and* $\boldsymbol{\sigma}_j^\top$ *are the* j*th row of* $\hat{\boldsymbol{D}}^{-1}$ *and* $\boldsymbol{D}^{-1}$, *respectively; and* I *denotes the indicator function.*

The asymptotic distribution of the multivariate quantile point completely relies on the asymptotic property of the estimator of the scale matrix $\boldsymbol{D}$.

Having defined the multivariate quantile point and parallel multivariate quantile region, we now introduce a simple multivariate median.

Definition 3. *Define a multivariate median by* $\boldsymbol{q}(0.5)$ *and let* $\hat{\boldsymbol{q}}(0.5)$ *denote its estimator.*

Corollary 3. *Under conditions (i)–(iv),* $\hat{\boldsymbol{q}}(0.5)$ *has the following representation:*

$$\begin{aligned} &n^{1/2}\big(\hat{\boldsymbol{q}}(0.5) - \boldsymbol{q}(0.5)\big) \\ &\quad = n^{-1/2}\tfrac{1}{2}\boldsymbol{D}\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_i + \boldsymbol{D}n^{1/2}(\hat{\boldsymbol{D}}^{-1} - \boldsymbol{D}^{-1})\boldsymbol{\mu} + n^{1/2}(\hat{\boldsymbol{D}} - \boldsymbol{D})\tilde{\boldsymbol{\xi}} + \boldsymbol{D}n^{1/2}\tilde{\boldsymbol{v}} + \boldsymbol{o}_p(1)\,, \end{aligned}$$

where $\tilde{\boldsymbol{u}}_i = (\tilde{u}_{i1}, \ldots, \tilde{u}_{ip})$, $\tilde{u}_{ij} = f_j^{-1}(\tilde{\xi}_j)\,\mathrm{sgn}(x_{ji} \le \tilde{\xi}_j)$, $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_p)$, $\tilde{\xi}_j = F_j^{-1}(0.5)$, *and* $\tilde{\boldsymbol{v}} = (\tilde{v}_1, \ldots, \tilde{v}_p)$, $\tilde{v}_j = (\boldsymbol{s}_j - \boldsymbol{\sigma}_j)^\top\mathrm{E}(\boldsymbol{v} \mid \boldsymbol{x}_j = \tilde{\boldsymbol{\xi}}_j)$. *Here* $\mathrm{sgn}(A) = 1$, *if condition* A *holds, and* -1 *otherwise.*

Compared to the multivariate medians defined by Liu (1988) and Oja (1983), this definition is much simpler, and the estimate is easier to compute.

3. Multivariate trimmed means by parallelogram

The simplest way to construct a robust multivariate estimator is to take a robust estimator for each coordinate. The multivariate trimmed mean of this type has been studied by Gnanadesikan & Kettenring (1972). Unfortunately, this approach does not consider all variables simultaneously so that the estimators thus constructed do not have the equivariance property. We propose a multivariate trimmed mean based on the studentized observations $\boldsymbol{x}_i = \hat{\boldsymbol{D}}^{-1}\boldsymbol{y}_i, i = 1, \ldots, n$. Recall that $\boldsymbol{s}_j^\top$ is the jth row of $\hat{\boldsymbol{D}}^{-1}$, $j = 1, \ldots, p$.

Definition 4. *The multivariate trimmed mean is* $\hat{\boldsymbol{\mu}}_t = \hat{\boldsymbol{D}}\hat{\boldsymbol{m}}$ *with* $\hat{\boldsymbol{m}} = (\hat{m}_1, \ldots, \hat{m}_p)$, *where*

$$\hat{m}_j = \frac{\sum_{i=1}^{n} x_{ji}\,\mathrm{I}(\hat{\xi}_{j1} \le x_{ji} \le \hat{\xi}_{j2})}{\sum_{i=1}^{n}\mathrm{I}(\hat{\xi}_{j1} \le x_{ji} \le \hat{\xi}_{j2})}\,.$$

Denote the multivariate trimmed mean by $\hat{\boldsymbol{\mu}}_t(\alpha)$ *for simplicity if* $\alpha = \alpha_{11} = \cdots = \alpha_{p1} = 1 - \alpha_{12} = \cdots = 1 - \alpha_{p2}$.

Theorem 4. *Suppose that the sample scale matrix $\hat{\boldsymbol{D}}$ is such that $\hat{\boldsymbol{D}}(\boldsymbol{A}\boldsymbol{y}+\boldsymbol{c}) = \boldsymbol{A}\hat{\boldsymbol{D}}(\boldsymbol{y})$. Here we denote $\hat{\boldsymbol{\mu}}_t$ by $\hat{\boldsymbol{\mu}}_t(\boldsymbol{y})$. Then $\hat{\boldsymbol{\mu}}_t(\boldsymbol{A}\boldsymbol{y}+\boldsymbol{c}) = \boldsymbol{A}\hat{\boldsymbol{\mu}}_t(\boldsymbol{y})+\boldsymbol{c}$.*

Let $\boldsymbol{\sigma}_j^{\top}$ denote the jth row of $\boldsymbol{D}^{-1}$. Let G_j and g_j denote the cdf and pdf of $\epsilon_j = \boldsymbol{\sigma}_j^{\top}\boldsymbol{v}$, respectively. Denote $\eta_j = G_j^{-1}(\alpha_j)$ and $\eta_{jk} = G_j^{-1}(\alpha_{jk})$. It is seen that $\xi_j = \boldsymbol{\sigma}_j^{\top}\boldsymbol{\mu} + \eta_j$. Denote $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_p)$ with $\delta_j = \int_{\eta_{j1}}^{\eta_{j2}} \epsilon_j g_j(\epsilon_j) d\epsilon_j$, $\phi(\epsilon_j) = \epsilon_j \,\mathrm{I}\,(\eta_{j1} \le \epsilon_j \le \eta_{j2}) - \delta_j + \eta_{j1}(\mathrm{I}\,(\epsilon_j < \eta_{j1}) - \alpha_{j1}) + \eta_{j2}(\mathrm{I}\,(\epsilon_j > \eta_{j2}) - (1-\alpha_{j2}))$, $\boldsymbol{E}_{vj} = \mathrm{E}(\boldsymbol{v}^{\top}\mathrm{I}\,(\eta_{j1} \le \epsilon_j \le \eta_{j2}))$, and $\boldsymbol{H} = \mathrm{diag}(h_1, \ldots, h_p)$, where $h_j = 1/(\alpha_{j2} - \alpha_{j1})$.

Theorem 5. *Under conditions (i)–(iv),*

$$n^{1/2}\big(\hat{\boldsymbol{\mu}}_t - (\boldsymbol{\mu} + \hat{\boldsymbol{D}}\boldsymbol{H}\boldsymbol{\delta})\big) = \boldsymbol{D}\boldsymbol{H}n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{\phi}_i + \boldsymbol{\omega}) + \boldsymbol{o}_p(1)\,,$$

where $\boldsymbol{\phi}_i = (\phi(\epsilon_{1i}), \ldots, \phi(\epsilon_{pi}))$ and $\boldsymbol{\omega} = (\boldsymbol{E}_{v1}(\boldsymbol{s}_1 - \boldsymbol{\sigma}_1), \ldots, \boldsymbol{E}_{vp}(\boldsymbol{s}_p - \boldsymbol{\sigma}_p))$.

Corollary 6. *Suppose that $\boldsymbol{E}_{vj} = \boldsymbol{0}$ and G_j are all symmetric about zeros. Denote $\tilde{\eta}_{j1} = G_j^{-1}(\alpha)$ and $\tilde{\eta}_{j2} = G_j^{-1}(1-\alpha)$. Then*

$$n^{1/2}\big(\hat{\boldsymbol{\mu}}_t(\alpha) - \boldsymbol{\mu}\big) = (1-2\alpha)^{-1}\boldsymbol{D}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\phi}_{0i} + \boldsymbol{o}_p(1)\,,$$

where $\boldsymbol{\phi}_{0i} = (\phi_0(\epsilon_{1i}), \ldots, \phi_0(\epsilon_{pi}))$ with

$$\phi_0(\epsilon_j) = \begin{cases} \tilde{\eta}_{j1} & \epsilon_j < \tilde{\eta}_{j1}\,, \\ \epsilon_j & \tilde{\eta}_{j1} \le \epsilon_j \le \tilde{\eta}_{j2}\,, \\ \tilde{\eta}_{j2} & \epsilon_j > \tilde{\eta}_{j2}\,, \end{cases}$$

for $j = 1, \ldots, p$, and $n^{1/2}(\hat{\boldsymbol{\mu}}_t(\boldsymbol{\alpha}) - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{K})$, where $\boldsymbol{K} = (1-2\alpha)^{-2}\boldsymbol{D}\boldsymbol{T}\boldsymbol{D}^{\top}$, $\boldsymbol{T} = [\tau_{jk}]$, with

$$\tau_{jj} = \int_{\tilde{\eta}_{j1}}^{\tilde{\eta}_{j2}} \epsilon^2 g_j(\epsilon)d\epsilon + 2\alpha(\tilde{\eta}_{j2})^2\,, \quad \text{for } j = 1, \ldots, p\,,$$

and

$$\begin{aligned}\tau_{jk} = {} & \tilde{\eta}_{j1}\tilde{\eta}_{k1}\Pr(\epsilon_j < \tilde{\eta}_{j1},\ \epsilon_k < \tilde{\eta}_{k1}) + \tilde{\eta}_{j1}\int_{\tilde{\eta}_{k1}}^{\tilde{\eta}_{k2}}\int_{-\infty}^{\tilde{\eta}_{j1}} \epsilon_k g_{jk}(\epsilon_j, \epsilon_k)\,d\epsilon_j\,d\epsilon_k \\ & + \tilde{\eta}_{j1}\tilde{\eta}_{k2}\Pr(\epsilon_j < \tilde{\eta}_{j1},\ \epsilon_k > \tilde{\eta}_{k2}) + \tilde{\eta}_{k1}\int_{-\infty}^{\tilde{\eta}_{k1}}\int_{\tilde{\eta}_{j1}}^{\tilde{\eta}_{j2}} \epsilon_j g_{jk}(\epsilon_j, \epsilon_k)\,d\epsilon_j\,d\epsilon_k \\ & + \int_{\tilde{\eta}_{k1}}^{\tilde{\eta}_{k2}}\int_{\tilde{\eta}_{j1}}^{\tilde{\eta}_{j2}} \epsilon_j\epsilon_k g_{jk}(\epsilon_j, \epsilon_k)\,d\epsilon_j\,d\epsilon_k + \tilde{\eta}_{k2}\int_{\tilde{\eta}_{k2}}^{\infty}\int_{\tilde{\eta}_{j1}}^{\tilde{\eta}_{j2}} \epsilon_j g_{jk}(\epsilon_j, \epsilon_k)\,d\epsilon_j\,d\epsilon_k \\ & + \tilde{\eta}_{j2}\tilde{\eta}_{k1}\Pr(\epsilon_j > \tilde{\eta}_{j2},\ \epsilon_k < \tilde{\eta}_{k1}) + \tilde{\eta}_{j2}\int_{\tilde{\eta}_{k1}}^{\tilde{\eta}_{k2}}\int_{\tilde{\eta}_{j2}}^{\infty} \epsilon_k g_{jk}(\epsilon_j, \epsilon_k)\,d\epsilon_j\,d\epsilon_k \\ & + \tilde{\eta}_{j2}\tilde{\eta}_{k2}\Pr(\epsilon_j > \tilde{\eta}_{j2},\ \epsilon_k > \tilde{\eta}_{k2})\,,\end{aligned}$$

for $j \ne k$, $\tau_{jk} = \tau_{kj}$, $j, k = 1, \ldots, p$, where g_{jk} denotes the joint pdf of ϵ_j and ϵ_k.

TABLE 1

Asymptotic generalized variances of estimators and Cramér–Rao lower bounds ($\delta = 0.1$)

Estimate	$\rho = 0.2$			$\rho = 0.5$		
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
$\bar{\boldsymbol{y}}$	1.287	3.394	10.89	1.219	3.368	10.89
$\hat{\boldsymbol{\mu}}_t$						
$\alpha = 0.1$	1.202	1.364	1.438	1.138	1.314	1.364
$\alpha = 0.2$	1.282	1.387	1.432	1.199	1.318	1.356
$\alpha = 0.3$	1.359	1.495	1.531	1.299	1.437	1.453
$\hat{\boldsymbol{\mu}}_{tt}$						
$\alpha = 0.1$	1.203	1.367	1.441	1.153	1.302	1.365
$\alpha = 0.2$	1.270	1.398	1.437	1.226	1.332	1.377
$\alpha = 0.3$	1.388	1.483	1.519	1.332	1.414	1.453
C–R	1.152	1.152	1.115	1.017	1.013	0.983

4. Asymptotic efficiency and Monte Carlo study for the trimmed means

From Theorem 5, the asymptotic efficiency of the multivariate trimmed mean relies on the performance of the estimator $\hat{\boldsymbol{\Sigma}}$ since the covariance matrix $\boldsymbol{\Sigma}$ affects $\boldsymbol{D}$ and $\boldsymbol{\phi}$. It is well known that the sample covariance matrix is efficient as an estimator of the covariance matrix for normal distributions, but becomes less efficient when the error vector $\boldsymbol{\epsilon}$ departs from normal distributions. It is then quite natural to suspect that using $\hat{\boldsymbol{D}}$ may reduce the efficiency of the multivariate trimmed mean when $\boldsymbol{\epsilon}$ departs from normal distributions. We thus consider a 'robustified' version of the multivariate trimmed mean.

Denote the multivariate trimmed mean based on the ordinary sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ by $\hat{\boldsymbol{\mu}}_t$. Let z_α denote the truncated variable obtained by restricting the random variable z on the interval $[F_z^{-1}(\alpha), F_z^{-1}(1-\alpha)]$, where F_z^{-1} is the population quantile function of z. For simplicity, consider the bivariate case. Denote the pdf of $\boldsymbol{v} = (v_1, v_2)$ by h and the marginal pdfs of v_1 and v_2 by h_1 and h_2, respectively. Similar to the trimmed mean for the location parameter, a robust scale matrix can be defined based on the truncated variables $v_{1\alpha}$ and $v_{2\alpha}$. Define the trimmed covariance matrix by $\boldsymbol{\Sigma}_\alpha = \mathrm{V}(\boldsymbol{v}) = [\varphi_{ij}]$ with

$$\varphi_{ii} = \frac{1}{1-2\alpha}\int_{C_i} v_i^2 h_i(v_i)\,dv_i, \qquad \varphi_{12} = \frac{1}{a(\alpha)}\int_{C_2}\int_{C_1} v_1 v_2 h(v_1, v_2)\,dv_1\,dv_2\,,$$

where $a(\alpha) = \Pr(v_1 \in C_1, v_2 \in C_2)$, $C_j = [H_j^{-1}(\alpha), H_j^{-1}(1-\alpha)]$, $j = 1, 2$, and H_j^{-1} is the population quantile function of the variable v_j. Let $\hat{\boldsymbol{\Sigma}}_\alpha$ be an estimator of $\boldsymbol{\Sigma}_\alpha$ satisfying assumption (ii). Denote the robustified bivariate trimmed mean based on $\hat{\boldsymbol{\Sigma}}_\alpha$ by $\hat{\boldsymbol{\mu}}_{tt}$.

To compare the sample mean $\bar{\boldsymbol{y}}$, the trimmed mean $\hat{\boldsymbol{\mu}}_t$, and the robustified trimmed mean $\hat{\boldsymbol{\mu}}_{tt}$, we consider the error vector

$$\boldsymbol{v} \stackrel{\mathrm{d}}{=} \begin{cases} \mathrm{N}_2(\boldsymbol{0}, \boldsymbol{R}) & \text{with probability } 1-\delta\,, \\ \mathrm{N}_2(\boldsymbol{0}, \sigma^2\boldsymbol{I}) & \text{with probability } \delta\,, \end{cases} \quad \text{where} \quad \boldsymbol{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{2}$$

for $0 \le \delta \le 1$. When $\delta = 0$, $\boldsymbol{v}$ has a bivariate normal distribution. Define the generalized variance as the determinant of the covariance matrix. Consider $\delta = 0.1$, $\rho = 0.2, 0.5$, and $\sigma = 3, 5, 10$. Table 1 gives the square-root of the asymptotic generalized variances of $\bar{\boldsymbol{y}}$, $\hat{\boldsymbol{\mu}}_t$, and $\hat{\boldsymbol{\mu}}_{tt}$ for the cases $\alpha = 0.1, 0.2, 0.3$. The Cramér–Rao lower bounds (C-R) for the distributions under study are also included.

TABLE 2

Asymptotic generalized variances of the two trimmed means ($\delta = 0.3,\ \rho = 0.5,\ \sigma = 10$)

	$\hat{\boldsymbol{\mu}}_t$			$\hat{\boldsymbol{\mu}}_{tt}$		
	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
$\rho = 0.5$ $\sigma = 10$	1.263	1.299	1.421	1.328	1.311	1.432

The following are some observations from Table 1.

(a) Relatively, the sample mean $\bar{\boldsymbol{y}}$ has larger asymptotic generalized variances than the two trimmed means. This confirms that $\bar{\boldsymbol{y}}$ is quite sensitive to the outliers. It also shows that the two trimmed means are fairly robust, as expected.

(b) The two trimmed means have nearly the same efficiency.

(c) When the correlation coefficient ρ gets larger, the asymptotic generalized variances of the two trimmed means get smaller.

(d) With $\delta = 0.1$, which means that approximately 10% of observations are drawn from a distribution with larger variance, a 10% trimming percentage seems to be reasonable for both of the trimmed means under study. A similar observation has been made for the univariate trimmed means under the location and linear regression models (Ruppert & Carroll, 1980; Chen & Chiang, 1996).

The two trimmed means are quite competitive in the above study. Would the trimmed mean based on the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ be relatively less efficient than the one based on the trimmed covariance matrix $\hat{\boldsymbol{\Sigma}}_\alpha$, if the error distribution had more outliers? We compute the asymptotic generalized variances of these two trimmed means for the error distribution of (2) with $\delta = 0.3$ and $\sigma = 10$. Table 2 lists the results.

Surprisingly, the generalized variances are all smaller for $\hat{\boldsymbol{\mu}}_t$ than for $\hat{\boldsymbol{\mu}}_{tt}$, which indicates that the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ is a better choice.

To study the trimmed means under the multivariate model with asymmetric error distributions, we perform a Monte Carlo simulation for the multivariate location model $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{v}$. Let $\boldsymbol{z} = (z_1, z_2)$ denote a vector of two independent exponential random variables with mean 1. Assume that the error vector $\boldsymbol{v} = (v_1, v_2)$ has the following mixed distribution:

$$\boldsymbol{v} \overset{\mathrm{d}}{=} \begin{cases} \boldsymbol{G}\boldsymbol{z} & \text{with probability } 1 - \delta\,, \\ \sigma \boldsymbol{z} & \text{with probability } \delta\,, \end{cases} \qquad \text{where} \quad \boldsymbol{G} = \begin{bmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{bmatrix}.$$

This design ensures that $\boldsymbol{v}$ has a zero-mean asymmetric distribution, and has either a covariance matrix $\boldsymbol{R}$ with probability $(1-\delta)$ or a covariance matrix $\sigma^2\boldsymbol{I}$ with probability δ. Note that large values of σ may produce outliers.

In this study, we let $\boldsymbol{\mu} = (1, 1)$ and consider the cases $\delta = 0.1, 0.2,\ \rho = 0.2, 0.5, 0.8,$ and $\sigma = 2, 5, 10$. The sample size is $n = 30$. For each case, we simulate 1000 sets of data from the above mixture model. For $i = 1, \dots, 1000,$ let $\hat{\boldsymbol{\mu}}_i$ stand for the estimate of the ith replicate for the mean estimator $\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}}$ can be any of the three mean estimators under study. With 1000 replicates, we compute the averaged mean squared error (AMSE) of the mean estimators defined by

$$\mathrm{AMSE}(\hat{\boldsymbol{\mu}}) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu})^{\mathsf{T}}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu})\,,$$

for $\hat{\boldsymbol{\mu}}$ being $\bar{\boldsymbol{y}},\ \hat{\boldsymbol{\mu}}_t$, and $\hat{\boldsymbol{\mu}}_{tt}$, respectively. Tables 3, 4, and 5 give the results.

TABLE 3

AMSE *of the three mean estimators* ($\rho = 0.2$)

Estimate	$\delta = 0.1$			$\delta = 0.2$		
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
$\bar{\boldsymbol{y}}$	86.50	237.2	697.7	111.8	376.1	1370
$\hat{\boldsymbol{\mu}}_t$						
$\alpha = 0.1$	0.059	0.074	0.111	0.069	0.116	0.311
$\alpha = 0.2$	0.088	0.098	0.108	0.099	0.112	0.144
$\alpha = 0.3$	0.120	0.132	0.144	0.132	0.146	0.161
$\hat{\boldsymbol{\mu}}_{tt}$						
$\alpha = 0.1$	0.105	0.128	0.191	0.121	0.228	0.586
$\alpha = 0.2$	0.165	0.173	0.184	0.180	0.201	0.251
$\alpha = 0.3$	0.220	0.232	0.247	0.239	0.253	0.265

TABLE 4

AMSE *of the three mean estimators* ($\rho = 0.5$)

Estimate	$\delta = 0.1$			$\delta = 0.2$		
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
$\bar{\boldsymbol{y}}$	86.89	211.2	765.6	118.9	381.3	1422
$\hat{\boldsymbol{\mu}}_t$						
$\alpha = 0.1$	0.073	0.083	0.119	0.085	0.132	0.302
$\alpha = 0.2$	0.112	0.117	0.124	0.124	0.139	0.163
$\alpha = 0.3$	0.151	0.160	0.164	0.163	0.182	0.196
$\hat{\boldsymbol{\mu}}_{tt}$						
$\alpha = 0.1$	0.102	0.123	0.189	0.128	0.231	0.563
$\alpha = 0.2$	0.160	0.162	0.170	0.184	0.204	0.249
$\alpha = 0.3$	0.214	0.214	0.223	0.236	0.252	0.263

TABLE 5

AMSE *of the three mean estimators* ($\rho = 0.8$)

Estimate	$\delta = 0.1$			$\delta = 0.2$		
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
$\bar{\boldsymbol{y}}$	85.43	218.3	723.9	101.3	377.7	1320
$\hat{\boldsymbol{\mu}}_t$						
$\alpha = 0.1$	0.083	0.100	0.125	0.092	0.150	0.317
$\alpha = 0.2$	0.130	0.142	0.139	0.142	0.169	0.188
$\alpha = 0.3$	0.176	0.194	0.193	0.191	0.218	0.226
$\hat{\boldsymbol{\mu}}_{tt}$						
$\alpha = 0.1$	0.097	0.121	0.186	0.118	0.226	0.530
$\alpha = 0.2$	0.144	0.156	0.157	0.174	0.208	0.241
$\alpha = 0.3$	0.187	0.203	0.201	0.227	0.250	0.251

We observe the following for the asymmetric distribution from Tables 3, 4, and 5.

(a) Both the trimmed means perform better than the sample mean.

(b) Again, $\hat{\boldsymbol{\mu}}_t$ performs better than $\hat{\boldsymbol{\mu}}_{tt}$. It seems that the robustified trimmed mean does not benefit from trimming the covariance matrix. This is somewhat surprising. The trimmed mean using the ordinary sample covariance matrix attains good efficiency.

References

BRECKLING, J. & CHAMBERS, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

CHAKRABORTY, B. (2001). On affine equivariant multivariate quantiles. *Ann. Inst. Statist. Math.* **53**, 380–403.

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91**, 862–872.

CHEN, L-A. & CHIANG, Y-C. (1996). Symmetric quantile and trimmed means for location and linear regression model. *J. Nonparametr. Statist.* **7**, 171–185.

CHEN, L-A. & WELSH, A.H. (2002). Distribution-function-based bivariate quantiles. *J. Multivariate Anal.* **83**, 208–231.

GNANADESIKAN, R. & KETTENRING, J.R. (1972). Robust estimates, residuals, and outlier detection with multi-response data. *Biometrics* **28**, 81–124.

HUBER, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.

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

KOLTCHINSKII, V. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25**, 435–477.

LIU, R.Y. (1988). On a notion of simplicial depth. *Proc. Natl. Acad. Sci. USA* **18**, 1732–1734.

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.

RUPPERT, D. & CARROLL, R.J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75**, 828–838.

SHIAU, J-J.H. & CHEN, L-A. (2002). *The Multivariate Parallelogram and its Application to Multivariate Trimmed Means.* Technical Report. Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

SMALL, C.G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.* **58**, 273–277.