Proceedings of the World Congress on Engineering 2008 Vol II
WCE 2008, July 2 - 4, 2008, London, U.K.

Multivariate Analysis of Variance Using a Kotz Type Distribution

Kusaya Plungpongpun, and Dayanand N. Naik *

Abstract—**Most standard inferential statistical methods for multivariate data are developed under the fundamental assumption that the data are from a multivariate normal distribution. Unfortunately, one can never be sure whether a set of data is really from a multivariate normal distribution. There are numerous methods for checking (testing) multivariate normality, but these tests are generally not very powerful, especially for smaller sample sizes. Hence it is always beneficial to have alternative multivariate distributions and the methodology for using them.**

In this article, we consider a Kotz type multivariate distribution which has fatter tail regions than that of multivariate normal distribution and show how multivariate analysis of variance can be performed using this distribution as model.

Keywords: Generalized spatial median, Kotz type distribution, simultaneous confidence intervals, testing the equality of mean vectors.

1 Introduction

Multivariate normal distribution is fundamental for multivariate analysis of variance. Elegant results are obtained under this model. However, in practice, the assumption of this distribution may not be valid. Numerous classes of multivariate distributions have been used in practice in place of multivariate normal distribution. See [3]-[5] and [11]. In this article, we consider a Kotz type multivariate distribution (of a $p-$ variate random vector $\mathbf{X}$) which has fatter tail regions than that of multivariate normal distribution and its probability density function (*pdf*) is given by:

$$f(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma}) = c_p \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{\frac{1}{2}}\}, \quad (1)$$

where $\boldsymbol{\mu} \in \Re^p$, $\boldsymbol{\Sigma}$ is a positive definite matrix (p.d.) and $c_p = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$.

This *pdf* has appeared in the literature in different forms. For example, the pdf is a special case of the following families of distributions:

Multivariate distributions proposed by Simoni (see [18]): These have the *pdf* proportional to

$$\exp\{-\frac{1}{r}[(\mathbf{x}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})]^{\frac{r}{2}}\},$$

where $\mathbf{A}$ is p. d. and $r \geq 1$. For $r = 1$ one obtains our multivariate distribution.

Elliptically symmetric distributions (see [8]): Let $\mathbf{X}$ be a $p \times 1$ random vector, $\boldsymbol{\mu}$ be a $p \times 1$ vector in $\Re^p$, and $\boldsymbol{\Sigma}$ be a $p \times p$ non-negative definite matrix. Then $\mathbf{X}$ has an elliptically contoured distribution, denoted by $EC_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\psi)$ if the characteristic function $\boldsymbol{\phi}_{\mathbf{X}-\boldsymbol{\mu}}(t) = E[\exp(i\mathbf{t}'(\mathbf{X}-\boldsymbol{\mu}))]$ of $\mathbf{X}-\boldsymbol{\mu}$ is a function of the quadratic form $\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}$ as $\boldsymbol{\phi}_{\mathbf{X}-\boldsymbol{\mu}}(t) = \psi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$ for some function ψ.

Therefore, the elliptically symmetric distributions denoted by $EC_p(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$, have the *pdf* (here $\boldsymbol{\Sigma}$ is p.d.) in the form

$$f(\mathbf{x}) = k_p|\boldsymbol{\Sigma}|^{-\frac{1}{2}}g[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})],$$

where g is a one-dimensional real-valued function independent of p and

$$k_p = \frac{p\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}}.$$

For our distribution $g(t) = \exp\{-t^{\frac{1}{2}}\}$.

Power exponential distributions (see [5]): A random vector $\mathbf{X}$ has a p-dimensional power exponential distribution, denoted by $PE_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\beta)$, with $\boldsymbol{\mu},\boldsymbol{\Sigma}$, and β, where $\boldsymbol{\mu} \in \Re^p$, $\boldsymbol{\Sigma}$ is a $p \times p$ p. d. matrix, and $\beta \in (0,\infty)$. Its density function is

$$f(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma},\beta) = k|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{\beta}\},$$

where $k = \frac{p\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}}$.
For $\beta = \frac{1}{2}$ one obtains our distribution. This function is actually the *pdf* of an elliptically contoured random vector $EC_p(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$.

Kotz type distributions (see [4]): If $\mathbf{X} \sim EC_p(\boldsymbol{\mu},\boldsymbol{\Sigma},g)$ and the density generator g is of the form $g(u) = c_p u^{N-1}\exp(-ru^s)$, $\ r,s > 0,\ 2N+p > 2\ $ then we say

*Manuscript submitted March 20, 2008. K. Plungpongpun is with Department of Statistics, Silpakorn University, Bangkok, Thailand (e-mail: kusaya@su.ac.th), and D. N. Naik is with Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia 2329, USA (e-mail: dnaik@odu.edu)

that $\mathbf{X}$ possesses a symmetric Kotz distribution. The *pdf* of $\mathbf{X}$ is given by

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_p \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} [(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{N-1} \cdot \exp\ \{-r[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^s\},$$

where $c_p = \frac{s\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(\frac{2N+p-2}{2s})}\ r^{\frac{2N+p-2}{2s}}$.

When $N = 1, s = \frac{1}{2}$, and $r = 1$ the distribution reduces to our distribution.

The pdf (1) can also be written as a multivariate normal mixtures as in [9] and [10]. The Kotz type distribution with the *pdf* given in (1) has heavier tail regions than those covered by the multivariate normal distribution and hence can be useful in providing robustness against "outliers" (see [13]). For $p = 1$, the *pdf* (1) reduces to that of a double exponential (or Laplace) distribution. Hence we may treat this distribution as a multivariate generalization of double exponential distribution. However, this is not a multivariate double exponential distribution because, its marginal distributions are not double exponential distributions. We note that double exponential distribution is symmetric around a location parameter μ, and the maximum likelihood estimate of μ is the median. It is well known that a median is more robust estimator of a location parameter than the mean. For this reason, many times in practice double exponential (Laplace) distribution is used for data analysis instead of normal distribution.

In our earlier paper, [15], we have discussed various characteristics of the distribution (1), including its marginal and conditional distributions and moments. We note that $E(\mathbf{X}) = \boldsymbol{\mu}$, and $Var(\mathbf{X}) = (p+1)\boldsymbol{\Sigma}$. Also, we provided an algorithm for simulating samples from this distribution. The maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, were also derived and the asymptotic distribution of the maximum likelihood estimate of $\boldsymbol{\mu}$ was given. Further, using Mardia's multivariate measures of skewness and kurtosis, we provided a goodness-of-fit test for Kotz type distribution. Inference for parameter vector $\boldsymbol{\mu}$ was also discussed in [15]. In this article we discuss how this distribution can be used to perform multivariate analysis of variance.

In the next section, for ease of reading, we will provide the maximum likelihood estimate of $\boldsymbol{\mu}$ and its asymptotic distribution and also provide some details on how to construct simultaneous confidence intervals using Bonferroni probability inequality. It is worth noting that the most interesting property of the distribution in hand is that the maximum likelihood estimators under this distribution are the generalized spatial median (GSM) estimators as defined in [16]. Sections 3 and 4 discuss one way multivariate analysis of variance. An example to illustrate the methods is considered in Section 5 and concluding remarks are provided in Section 6.

2 Estimation of Parameters

Many researchers have discussed statistical inference using elliptical distributions. For example, see [3] and the references therein. However, the maximum likelihood theory developed in [3] assumes that the joint distribution of the random sample, $\mathbf{X}_1, ..., \mathbf{X}_n$, is elliptically symmetric. In fact, in this case the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are essentially same as those in the multivariate normal case (see [3]).

Several authors have performed statistical inference based on certain elliptical distributions. For example, [12] used multivariate t-distribution and maximum likelihood method to analyze certain regression and repeated measurements, and [13] used multivariate power exponential distribution to analyze a certain repeated measurements. In each case numerical algorithms are used to find the estimates of the parameters. In the following we discuss estimation of parameters using maximum likelihood methods when an independent identically distributed sample from (1) is available.

Suppose $\mathbf{X}_1, ..., \mathbf{X}_n$ is a random sample from Kotz type distribution (1). Then the log-likelihood function is given by

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = n \ln c - \frac{n}{2} \ln |\boldsymbol{\Sigma}| - \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}.$$

The MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by minimizing

$$\frac{n}{2} \ln \mid \boldsymbol{\Sigma} \mid + \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})} \qquad (2)$$

simultaneously w.r.t. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

When $\boldsymbol{\Sigma} = \mathbf{I}$, the solution to the above problem or the MLE of $\boldsymbol{\mu}$ is the spatial median introduced in [7] and for general $\boldsymbol{\Sigma}$ it is generalized spatial median introduced in [16] and studied in [14].

J. B. S. Haldane defined (see [7]) the spatial median of multivariate data vectors $\mathbf{x}_1, ..., \mathbf{x}_n$, as a point (vector) $\hat{\boldsymbol{\mu}} \in \Re^p$ which minimizes

$$\sum_{i=1}^{n} \|\mathbf{x}_i-\boldsymbol{\mu}\| = \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'(\mathbf{x}_i-\boldsymbol{\mu})}$$

with respect to $\boldsymbol{\mu}$. For $p > 1$, the vector $\hat{\boldsymbol{\mu}}$ is unique except when all the mass of the distribution is concentrated on a line and is invariant under orthogonal transformation, but not under affine transformation (see [1], [2]). C. R. Rao (see [16]) defined two generalized spatial medians which are invariant under affine transformation as:

(i) a vector $\hat{\boldsymbol{\mu}}$ which minimizes

$$\sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}$$

with respect to $\boldsymbol{\mu}$, where $\mathbf{S}$ is the usual sample variance covariance matrix, and

(ii) a vector $\hat{\boldsymbol{\mu}}$ which minimizes

$$\frac{n}{2}\ \ln|\boldsymbol{\Sigma}| + \sum_{i=1}^{n}\sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}$$

simultaneously with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Thus, we note that the MLE of $\boldsymbol{\mu}$ under the assumption of Kotz type distribution (1) for $\mathbf{X}_1, ..., \mathbf{X}_n$ is same as the generalized spatial median defined in [16].

Computation of GSM and $\widehat{\boldsymbol{\Sigma}}$: Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be a random sample from (1). Then the GSM of $\boldsymbol{\mu}$ which minimizes (2) can be computed in two stages as follows (see [14]).

Suppose $\boldsymbol{\Sigma}$ is known or set to an initial value and $\boldsymbol{\Sigma} = \mathbf{G}\mathbf{G}'$, for a nonsingular $\mathbf{G}$. Then the generalized spatial median $\hat{\boldsymbol{\mu}}$ which minimizes

$$\sum_{i=1}^{n}\sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}$$

w.r.t. $\boldsymbol{\mu}$ can be obtained as $\hat{\boldsymbol{\mu}} = \mathbf{G}\hat{\boldsymbol{\nu}}$, where $\hat{\boldsymbol{\nu}}$ is the spatial median which minimizes $\sum_{i=1}^{n}\sqrt{(\mathbf{y}_i-\boldsymbol{\nu})'(\mathbf{y}_i-\boldsymbol{\nu})}$ w.r.t. $\boldsymbol{\nu}$. Here $\mathbf{y}_i = \mathbf{G}^{-1}\mathbf{x}_i$ and $\boldsymbol{\nu} = \mathrm{G}^{-1}\boldsymbol{\mu}$. Spatial median can be computed using an algorithm given in [6]. Next using $\hat{\boldsymbol{\mu}}$ the maximum likelihood estimate of $\boldsymbol{\Sigma}$ is obtained as the matrix $\widehat{\boldsymbol{\Sigma}}$ which minimizes (2) with respect to $\boldsymbol{\Sigma}$ as a solution to the non-linear equations given by

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\frac{(\mathbf{x}_i-\hat{\boldsymbol{\mu}})(\mathbf{x}_i-\hat{\boldsymbol{\mu}})'}{\sqrt{(\mathbf{x}_i-\hat{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\hat{\boldsymbol{\mu}})}}.$$

We use nonlinear optimization methods to obtain maximum likelihood estimates of all the parameters. We have adopted SAS' IML procedure for writing the computer programs. Using the $Newton-Raphson$ method the optimization yields unique estimates in the feasible regions under most covariance structures.

Theorem (The asymptotic distribution of GSM): Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be a random sample from $p-$variate ($p > 1$) Kotz type distribution (1) with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\mu}}$ be the maximum likelihood estimate of $\boldsymbol{\mu}$. Then

$$\sqrt{n}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\mathbf{0},\ \boldsymbol{\Sigma}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\Sigma}),$$

where $\boldsymbol{B} = E\left[\frac{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'}{(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}\right]$ and

$$\boldsymbol{A} = E[\frac{1}{\sqrt{(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}}(\boldsymbol{\Sigma} - \frac{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'}{(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})})].$$

Further $\mathbf{B}$ and $\mathbf{A}$ can be estimated by

$$\begin{aligned}
\widehat{\boldsymbol{B}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'}{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})},\\
\widehat{\boldsymbol{A}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})}}\cdot\\
&\quad [\widehat{\boldsymbol{\Sigma}} - \frac{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'}{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})}],
\end{aligned}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$.

Many times in practice, we may be interested in performing simultaneous inference on a set of k parameters, for example, on components of vector $\boldsymbol{\mu}$. One convenient and easy way to build simultaneous confidence intervals on these parameters is using the Bonferroni method. It is a simple method that allows the construction of many confidence intervals maintaining an overall confidence coefficient. The method is based on Bonferroni's probability inequality, $P(\cap_{i=1}^{k} A_i) \geq 1 - \sum_{i=1}^{k} P(A_i^c)$, where A_i is the event that the i^{th} confidence interval contains the corresponding parameter and A_i^c is the complement of that event. Hence the left hand side of Bonferroni's inequality is the probability that all the confidence intervals simultaneously contain their corresponding true parameter values and the right hand side is one minus the sum of the probabilities that the intervals do not contain the corresponding true values. Thus if we want the overall confidence coefficient to be $1-\alpha$ then we should construct the individual confidence interval with a confidence level of $1 - \alpha/k$.

Proposition 1 (Simultaneous Confidence Intervals): The $100(1-\alpha)\%$ Bonferroni simultaneous confidence intervals for m linear functions of $\mu_i's$, are given by

$$\left(\mathbf{a}_i'\hat{\boldsymbol{\mu}} - z_{\alpha/2m}\sqrt{\frac{\mathbf{a}_i'\widehat{\boldsymbol{\tau}}\mathbf{a}_i}{n}},\ \mathbf{a}_i'\hat{\boldsymbol{\mu}} + z_{\alpha/2m}\sqrt{\frac{\mathbf{a}_i'\widehat{\boldsymbol{\tau}}\mathbf{a}_i}{n}}\right),\ i = 1, ..., m,$$

where $\mathbf{a}_i's$ are vectors of known constants and $z_{\alpha/2m}$ is the upper $100(1-\alpha/2m)$ th percentile of a standard normal distribution, and $\widehat{\boldsymbol{\tau}} = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}\widehat{\boldsymbol{\Sigma}}$.

In [15], we have used these results to perform inference on the components of vector $\boldsymbol{\mu}$.

3 Testing Equality of Mean Vectors

Suppose $\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{in_i}$ is a random sample of size n_i from Kotz type population with the parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, ..., g$. The random samples from different g populations are assumed to be independent. Let $\boldsymbol{\mu}_i = (\mu_{i1},\ \mu_{i2}, ...,\ \mu_{ip})'$ and $\boldsymbol{\sigma}_i = (\sigma_{i,11}, ...,\ \sigma_{i,1p}, ...,\ \sigma_{i,p-1,p},\ \sigma_{i,pp})'$, $i = 1, ..., g$. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_g', \boldsymbol{\sigma}_1', ..., \boldsymbol{\sigma}_g')'$ be the vector of all unknown parameters.

ISBN:978-988-17012-3-7

Consider the problem of testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$ when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$, that is, when $\boldsymbol{\sigma}_1 = \cdots = \boldsymbol{\sigma}_g = \boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = (\sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1,p}, \sigma_{pp})'$.
Under H_0, let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\mu}}', \widetilde{\boldsymbol{\sigma}}')'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum of the likelihood function under H_0 is given by

$$L(\widetilde{\boldsymbol{\theta}}) = (c_p)^n |\widetilde{\boldsymbol{\Sigma}}|^{-n/2} \ e^{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})}},$$

where $n = (\sum_{i=1}^{g} n_i)$.

Next, let $\widehat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}_1', ..., \hat{\boldsymbol{\mu}}_g', \widehat{\boldsymbol{\sigma}}')'$ be the MLE of $\boldsymbol{\theta}$ under no restrictions. Then the maximum of the likelihood function is given by

$$L(\widehat{\boldsymbol{\theta}}) = (c_p)^n |\widehat{\boldsymbol{\Sigma}}|^{-n/2} \ e^{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}}.$$

Then the likelihood ratio test for testing H_0 rejects H_0 if

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

where c is the critical value to be obtained appropriately.

If $\boldsymbol{\sigma}_i's$ are different and we wanted to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$, the likelihood ratio test for testing H_0 is given as follows. Let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\mu}}', \widetilde{\boldsymbol{\sigma}}_1', ..., \widetilde{\boldsymbol{\sigma}}_g')'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum of the likelihood function is given by

$$L(\widetilde{\boldsymbol{\theta}}) = (c_p)^n \prod_{i=1}^{g} |\widetilde{\boldsymbol{\Sigma}}_i|^{-n_i/2} e^{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})}}.$$

Under no restrictions, let $\widehat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}_1', ..., \hat{\boldsymbol{\mu}}_g', \widehat{\boldsymbol{\sigma}}_1', ..., \widehat{\boldsymbol{\sigma}}_g')'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum likelihood function is given by

$$L(\widehat{\boldsymbol{\theta}}) = (c_p)^n \prod_{i=1}^{g} |\widehat{\boldsymbol{\Sigma}}_i|^{-n_i/2} e^{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}}.$$

Then the likelihood ratio test rejects H_0 if

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

where c is a suitably chosen constant. When the sample size n is large,

$$-2\ln\Lambda = -2\ln\Big(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\Big) \text{ is approximately distributed as } \chi_r^2$$

random variable, where the degrees of freedom, r = (dimension of $\boldsymbol{\theta}$ under no restrictions) - (dimension of $\boldsymbol{\theta}$ under H_0).

4 Simultaneous Confidence Intervals

Let $\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{in_i}, \ i = 1, .., g,$ be g independent random samples of size n_i each from Kotz type distributions with parameters $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i, \ i = 1, ..., g$. Suppose the tests have revealed that a significant difference exists between the population means. In order to pinpoint the differences we construct simultaneous confidence intervals on various contrasts of difference between any two mean vectors. The following results provide distributional results that enable constructing simultaneous confidence intervals for linear combinations of $\mu_{ij}'s$.

Proposition 2 (Simultaneous Confidence Intervals): Let $\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{in_i}$ be a random sample of size n_i from Kotz type distribution with parameters $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i, \ i = 1, ..., g$ and suppose the samples from different g populations are independent. Suppose $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$. Using the Theorem and Proposition 1, the $100(1-\alpha)\%$ Bonferroni simultaneous confidence intervals for m linear combinations of $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'}, \ l < l' = 1, ..., g$ are given by

$$\mathbf{a}_k'(\hat{\boldsymbol{\mu}}_l - \hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m}\ \sqrt{\mathbf{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l + \frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}_k}, \quad k = 1, ..., m \quad (3)$$

where $\mathbf{a}_k's$ are vectors of known constants, $z_{\alpha/2m}$ is the upper $100(1-\alpha/2m)th$ percentile of the standard normal distribution, $\widehat{\boldsymbol{\tau}}_i = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{A}}_i^{-1}\widehat{\mathbf{B}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\boldsymbol{\Sigma}}, \ \ i = l, l' = 1, ..., g,$ and

$$\begin{aligned}\widehat{\boldsymbol{B}}_i &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},\\ \widehat{\boldsymbol{A}}_i &= \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{\sqrt{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}}\cdot\\ &\quad[\widehat{\boldsymbol{\Sigma}} - \frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}].\end{aligned}$$

If the variance covariance matrices from different populations are different then we have the following result.

Proposition 3 (Simultaneous Confidence Intervals): Like before, let $\mathbf{X}_{i1}, \mathbf{X}_{i2}, ..., \mathbf{X}_{in_i}$ be a random sample of size n_i from Kotz type distribution with $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i, \ i = 1, ..., g$, and the samples from different g populations are independent. Then $100(1-\alpha)\%$ Bonferroni simultaneous confidence intervals for m linear combinations of $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'}, \ l < l' = 1, ..., g$ are given by

$$\mathbf{a}_k'(\hat{\boldsymbol{\mu}}_l - \hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m}\ \sqrt{\mathbf{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l + \frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}_k}, \quad k = 1, ..., m \quad (4)$$

where $\mathbf{a}_k's$ are vectors of known constants, $z_{\alpha/2m}$ is the upper $100(1-\alpha/2m)th$ percentile of the standard normal distribution, $\widehat{\boldsymbol{\tau}}_i = \widehat{\boldsymbol{\Sigma}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\mathbf{B}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i, \ \ i = l, l' = 1, ..., g,$ and

$$\widehat{\boldsymbol{B}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},$$

ISBN:978-988-17012-3-7

$$\widehat{\boldsymbol{A}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{\sqrt{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}} \cdot [\widehat{\boldsymbol{\Sigma}}_i - \frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}].$$

Using the following example, we illustrate the computation of the maximum likelihood estimates and perform some statistical inference under Kotz type distribution (1). All the computations are done using programs written in SAS/IML software.

5 An Example

In the following, we illustrate the procedure for testing the equality of several population means using the Football helmet data given in the example below. Before testing the equality of the means, we first test the equality of the variance covariance matrices using the likelihood ratio test. Data on three variables, $\mathbf{x}_1$ = eye-to-top-of-head measurement, $\mathbf{x}_2$ = ear-to-top-of-head measurement, and $\mathbf{x}_3$ = jaw width are given for three groups of players, namely, high school football players, college football players, and non-football players. There are 30 observations in each group. The helmet data collected as part of a preliminary study of a possible link between football helmet design and neck injuries are provided in [17]. The hypotheses and the results from testing are as follows:

(i) Test $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}$.

The test statistic $= -2\ ln\Lambda = 13.054878$. Using $-2\ ln\Lambda \sim \chi^2_{12}$, the P-value = 0.3650646. Hence, we do not reject H_0 and conclude that the variance covariance matrices are the same for the three groups.

(ii) Next, we test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}$, given $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}$.

The test statistic $= -2\ ln\Lambda = 91.70311$. The P-value $= P[\chi^2_6 > 91.70311] < 0.0001$. Hence we reject H_0 and conclude that at least two $\boldsymbol{\mu}_i^{,}s$ are different.

(iii) Since we rejected the hypothesis of equality of means we want to find simultaneous confidence intervals for linear functions of $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'},\ l < l' = 1,2,3$. Let $\boldsymbol{\mu}_i = (\mu_{i1},..,\mu_{ip})',\ i = 1,2,3$. Then with the choices, $\mathbf{a}_1' = (1,0,0)$, $\mathbf{a}_2' = (0,1,0)$, and $\mathbf{a}_3' = (0,0,1)$ using (3), the 95% Bonferroni simultaneous confidence intervals for $\mu_{1j} - \mu_{2j},\ j = 1,2,3$ are:

$$\begin{aligned}
\mu_{11} - \mu_{21} &\in (2.2932274,\ 3.9341077),\\
\mu_{12} - \mu_{22} &\in (0.4294995,\ 2.0009086),\\
\mu_{13} - \mu_{23} &\in (-0.212608,\ 0.923095).
\end{aligned}$$

The 95% Bonferroni simultaneous confidence intervals for $\mu_{1j} - \mu_{3j},\ j = 1,2,3$ are:

$$\begin{aligned}
\mu_{11} - \mu_{31} &\in (1.3875155,\ 3.1273591),\\
\mu_{12} - \mu_{32} &\in (0.3040446,\ 1.6128665),\\
\mu_{13} - \mu_{33} &\in (-0.000264,\ 1.0659512).
\end{aligned}$$

The 95% Bonferroni simultaneous confidence intervals for $\mu_{2j} - \mu_{3j},\ j = 1,2,3$ are:

$$\begin{aligned}
\mu_{21} - \mu_{31} &\in (-1.709631,\ -0.002829),\\
\mu_{22} - \mu_{32} &\in (-0.835963,\ 0.3224655),\\
\mu_{23} - \mu_{33} &\in (-0.374984,\ 0.7301842).
\end{aligned}$$

6 Concluding Remarks

In our earlier paper [15], we proposed the Kotz type distribution given in (1) as an alternative to the multivariate normal distribution for performing multivariate inference. We introduced various properties of the distribution there and discussed the maximum likelihood estimation of the parameters. Further we have provided a goodness-of-fit test and a simulation algorithm. Application of this distribution to multivariate analysis of variance and simultaneous confidence interval construction is provided here in this article. Using an example, we have illustrated the computations. However, one need to perform an in-depth study, perhaps using an extensive simulation, to compare the performance of this distribution under the presence of outliers and other scenario against the multivariate normal (the gold standard) distribution. We intend to undertake such a study in the near future.

References

[1] Brown, B. M., “Statistical Use of the Spatial Median,” *Journal of the Royal Statistical Society, Series B*, vol. 45, 1983, pp. 25-30.

[2] Ducharme, G. R., and Milasevic, P., “Spatial Median and Directional Data,” *Biometrika*, vol. 74, 1987, pp. 212-215.

[3] Fang, K. T., and Anderson, T. W., *Statistical Inference in Elliptically Contoured and Related Distributions*, Allerton Press, New York, 1990.

[4] Fang, K. T., Kotz, S., and Ng, K. W., *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London, 1990.

[5] Gómez, E., Gómez-Villegas, M. A., and Marín, J. M., “A Multivariate Generalization of the Power Exponential Family of Distributions,” *Communications in Statistics, Theory and Methods*, vol. 27, 1998, pp. 589-600.

[6] Gower, J. S., "The Mediancentre," *Applied Statistics*, vol. 23, 1974, pp. 466-470.

[7] Haldane, J. B. S., "Note on the Median of a Multivariate Distribution," *Biometrika*, vol. 35, 1948, pp. 414-415.

[8] Johnson, M. E., *Multivairiate Statistical Simulation*, J. Wiley, New York, 1987.

[9] Kano, Y., "Consistency Property of Elliptical Probability Density Functions," *Journal of Multivariate Analysis*, vol. 51, 1994, pp. 139-147.

[10] Kariya, T. K., and Sinha, B. K., *Robustness of Statistical Tests*, Academic Press, Boston, 1989.

[11] Kotz, S., "Multivariate Distributions at a Cross-Road," in *Statistical Distributions in Scientific Work*, Patil, G. P., Kotz, S., and Ord, J. K., Eds., D. Reidel, The Netherlands, 1975, pp. 247-270.

[12] Lange, K. L., Little, R. J. A., and Taylor, J. M. G., "Robust Statistical Modeling Using the t Distribution," *Journal of the American Statistical Association*, vol. 84, 1989, pp. 881-896.

[13] Lindsey, J. K., "Multivariate Elliptically Contoured Distributions for Repeated Measurements," *Biometrics,* vol. 55, 1999, pp. 1277-1280.

[14] Naik, D. N., "Multivariate Medians: A Review," in *Probability and Statistics*, Basu, S. K., and Sinha, B. K., Eds., Narosa Publishing House, New Delhi, 1993, pp. 80-90.

[15] Naik, D. N., and Plungpongpun, K., "A Kotz Type Distribution for Multivariate Statistical Inference," in *Advances in Distribution Theory, Order Statistics, and Inference*, Balakrishnan, N., Castillo, E., and Sarabia, J. M., Eds., Birkhäuser, Boston, 2006, pp. 111-124.

[16] Rao, C. R., "Methodology Based on the L_1-Norm in Statistical Inference," *Sankhyā, Series A*, vol. 50, 1988, pp. 289-313.

[17] Rencher, A. C., *Methods of Multivariate Analysis*, 2nd ed., J. Wiley, New York, 2002, pp. 280-281.

[18] Simoni, S. de, "Su una Estensione dello Schema delle Curve Normalize di Ordina r alle Variabili Doppie," *Statistica (Bologna)*, vol. 28, 1968, pp. 151-170.