Performance of Some Spatial Median Tests under Elliptical Symmetry

J. Somorčík

Comenius University, Bratislava, Slovakia
E-mail: somorcik@fmph.uniba.sk

Abstract. *Finite sample performance of two non-affine invariant multi-sample location parameter tests based on spatial medians is studied by simulations in case of spherically symmetric distributions of the samples. We demonstrate that the decrease of the powers of the tests in case of extreme ellipticity depends also on the distances of the true location parameters. It is also shown how can the ellipticity of the samples have also a positive effect.*

Keywords: multi-sample location problem, spatial median, spherical and elliptical symmetry, affine invariance, power of test

1. Introduction

In practice, the experimenter has often q random samples from multivarite d-dimensional distributions and he has to decide if the location parameters of the underlying distributions are equal or not. To explain the term *location parameter*, suppose that the underlying densities are of the same type, i.e. the density of the i-th sample has the form $f(. - \mu_i)$. In this notation μ_i's are called the location parameters. The task of the experimenter is to test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_q.$$

This is known as *multivariate multi-sample location problem*. We note that in practice location parameters μ_i's are often expected values of the distributions but they do not have to be! When the expected value does not exist the interpretation of the location parameter is the centre of symmetry of the density function or, more generally, the spatial median of the distribution etc.

There exist many tests for testing the above hypothesis, a good overview can be found in [1]. Some of these tests enjoy the so called *affine invariance* which means that the value of the test statistic does not change after an affine transformation (i.e. multiplication by a regular matrix) of the data before testing. As a practical result, for example, the value of the test statistic (and hence also the result of the testing) does not change after linear change of the units of measurements or after a linear transformation of the coordinate system. This property is natural but there is also another benefit of the affine invariance. By *spherical* or *elliptical symmetry* of a distribution one understands that that the contours of the density are concentric circles or ellipses respectively. It was mentioned in many papers (see e.g. [2], [3], [4] and especially [5]) that the performance of the tests which are not affine invariant tends to be poor when the data come from an elliptically symmetric distribution and not from a spherically symmetric one (of the same type).

This was the motivation of our simulation study, which is aimed at investigation of the behaviour of tests based on spatial median. As usual, by the *spatial median* of the random sample $X_1, \ldots, X_n$ of

d-dimensional data we understand the vector $\hat{\mu} \in \mathbf{R}^d$ such that

$$\sum_{i=1}^{n} \|X_i - \hat{\mu}\| = \min_{M \in \mathbf{R^d}} \sum_{i=1}^{n} \|X_i - M\|,$$

where $\|.\|$ denotes the usual Euclidean norm in $\mathbf{R}^d$. Details about spatial median can be found in [6]. In [7] we have introduced two test statistics based on spatial median:

$$M_1 := \sum_{a=1}^{q} n_a (\hat{\mu}_a - \bar{\mu})^T \hat{V}^{-1} (\hat{\mu}_a - \bar{\mu})$$

and

$$M_2 := \sum_{a=1}^{q} n_a (\hat{\mu}_a - \hat{\mu})^T \hat{V}^{-1} (\hat{\mu}_a - \hat{\mu}),$$

where n_a is the size of the sample from the a-th population, $\hat{\mu}_a$ is the spatial median of the a-th sample, $\bar{\mu} := \frac{1}{n} \sum_{a=1}^{q} n_a \hat{\mu}_a$, is the weighted mean of the sample spatial medians and $\hat{\mu}$ is the spatial median of the joint sample. $\hat{V}$ is a consistent estimate of the asymptotic covariance matrix of the sample spatial median under H_0 (see [8] for details).

The asymptotic distribution under H_0 of both test statistics is $\chi^2_{(q-1)d}$ (which occurs very frequently in multi-sample situations, see e.g. [9]). Therefore for $i = 1, 2$ the test based on the statistic M_i is carried out in such a way that H_0 is rejected if M_i is greater than the $\alpha\%$ critical value of the chi-square distribution with $(q-1)d$ degrees of freedom. This rule is a test having asymptotic significance level $\alpha\%$.

We presented some other properties of these test statistics and their relations to other well-known tests in [7]. There we also studied by simulations their finite sample performance in case of spherically symmetric distributions. It turned out that tests based on M_1 and M_2 are preferable especially in case of heavy-tailed distributions and they seem to be robust against outliers. However, M_1 and M_2 are not affine invariant so an affine transformation of the data could affect their performance. Hence the aim of our simulation study will be to investigate their finite sample performance in case of elliptically symmetric sample distributions.

2. Method

We have performed computer simulations to compare the performance of the proposed spatial median test statistics M_1 and M_2 with the performance of the well-known affine invariant Lawley-Hotelling T^2 (see e.g. [1] for definition). We compared $q = 3$ samples of $n_1 = n_2 = n_3 = 30$ data points from $\mathbf{R}^3$ (i.e. the dimension is $d = 3$). In two cases we also compared unbalanced samples of $n_1 = 20$, $n_2 = 30$ and $n_3 = 40$ data points but we obtained virtually the same behaviour as in the balanced case. Hence the rest was performed with $n_1 = n_2 = n_3 = 30$. In each of the 5000 simulation all the data points were generated from spherically symmetric multivariate distributions of the same type centered around the location parameters μ_1, μ_2, μ_3, i.e the underlying densities had the form $g((x - \mu_i)^T (x - \mu_i))$; $i = 1, 2, 3$. We sampled only from a heavy-tailed multivariate Cauchy distribution since it was shown in [7] that in case of light-tailed distributions M_1 and M_2 are outperformed by T^2 already in spherical situations.

Then the three data clouds were elliptically "deformed" by multiplying with the square root of the positive-definite matrix

$$\Sigma := \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \; ; \rho \in (-1, 1).$$

It is equivalent to sampling from the distributions given by the densities $|\det(\Sigma)|^{-\frac{1}{2}} g((x - \Sigma^{\frac{1}{2}}\mu_i)^T \Sigma^{-1}(x - \Sigma^{\frac{1}{2}}\mu_1))$; $i = 1, 2, 3$. Contours of these densities are ellipsoids (around the location parameters $\Sigma^{\frac{1}{2}}\mu_1$, $\Sigma^{\frac{1}{2}}\mu_2$, $\Sigma^{\frac{1}{2}}\mu_3$) given by the equations

$$(x - \Sigma^{\frac{1}{2}}\mu_i)^T \Sigma^{-1}(x - \Sigma^{\frac{1}{2}}\mu_i) = const; \quad i = 1, 2, 3.$$

As we can see, the matrix Σ is the matrix of the symmetry of the ellipsoidal contours. The values of ρ we used were 0 (=no deformation; spherical symmetry), 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 (=extreme deformation; eccenric elliptical contours) because we wanted to focus on stronger deformations rather than on the "nearly spherical" cases (ρ close to zero).

For affine invariant tests (e.g. the Lawley-Hotelling test) the "deformed" situation is the same as spherical symmetry around the location parameters μ_1, μ_2, μ_3. But for our median-based test statistics M_1, M_2 it is not the case, the performance of the tests could depend on the choice of the matrix Σ.

3. Performance under H_0

First we simulated the case when H_0 is true, i.e. the μ_i's were set up to vectors of zeroes. The tests were carried out in such a way that the null hypothesis was rejected if the employed test statistic exceeded the 5% critical value of the $\chi^2_{(q-1)d} = \chi^2_{(3-1)\cdot 3}$ distribution, so we were testing on the nominal level of significance 5%. Table 1 shows how the simulated probabilities of type I error differ from the "ideal" 0.05.

Table 1: Simulated probabilities of type I error - balanced and unbalanced samples

	$n_1 = n_2 = n_3 = 30$			$n_1 = 20, n_2 = 30, n_3 = 40$		
ρ	M_1	M_2	T^2	M_1	M_2	T^2
0	0.073	0.079	0.023	0.076	0.084	0.032
0.5	0.080	0.086	0.023	0.088	0.096	0.032
0.6	0.081	0.089	0.023	0.091	0.100	0.032
0.7	0.085	0.094	0.023	0.095	0.107	0.032
0.8	0.097	0.107	0.023	0.103	0.117	0.032
0.9	0.114	0.132	0.023	0.128	0.141	0.032
0.95	0.151	0.174	0.023	0.162	0.186	0.032

It is obvious from the table that the stability of the spatial median tests gets worse as ρ increases, i.e. the data clouds are more ellipsoidal. We note that asymptotically the true values of probabilities of type I error must attain 0.05 for larger sample sizes because the asymptotic distribution of M_1 and M_2 is $\chi^2_{(q-1)d}$ also under elliptical symmetry (see [7]). The impact of elliptical symmetry is that for smaller sample sizes the use of $\chi^2_{(q-1)d}$ critical values is less accurate than in case of spherical symmetry. Finally, note that the results for unbalanced samples are very similar to the balanced case: simulated probability of type I error is increasing with ρ increasing.

4. Different location parameter distances

Now we present the results of simulations when H_0 did not hold. We set up the location parameters to

$$\mu_1 := \begin{pmatrix} 0 \\ 0.4 \\ 0 \end{pmatrix}, \quad \mu_2 := \begin{pmatrix} -0.4 \\ 0.4 \\ 0.4 \end{pmatrix}, \quad \mu_3 := \begin{pmatrix} 0.4 \\ 0 \\ 0 \end{pmatrix}$$

and then to $1.5\mu_1$, $1.5\mu_2$, $1.5\mu_3$ and $2\mu_1$, $2\mu_2$, $2\mu_3$. Table 2 shows the simulated powers of the considered tests.

Table 2: Power under various violations of H_0

	Location parameters μ_1, μ_2, μ_3			Location parameters $1.5\mu_1, 1.5\mu_2, 1.5\mu_3$			Location parameters $2\mu_1, 2\mu_2, 2\mu_3$		
ρ	M_1	M_2	T^2	M_1	M_2	T^2	M_1	M_2	T^2
0	0.430	0.449	0.049	0.769	0.784	0.087	0.937	0.942	0.146
0.5	0.438	0.454	0.049	0.769	0.782	0.087	0.944	0.948	0.146
0.6	0.436	0.454	0.049	0.766	0.777	0.087	0.943	0.947	0.146
0.7	0.433	0.451	0.049	0.759	0.773	0.087	0.942	0.947	0.146
0.8	0.425	0.446	0.049	0.744	0.760	0.087	0.937	0.942	0.146
0.9	0.414	0.440	0.049	0.708	0.727	0.087	0.911	0.924	0.146
0.95	0.404	0.441	0.049	0.670	0.696	0.087	0.874	0.888	0.146

Similarly to the situation when H_0 was true the performance of the spatial median tests gets poorer because their simulated powers are decreasing when ρ is increasing. Table 2 also shows that the decrease of simulated powers depends on "how far" from H_0 we are: when the distances between location parameters are longer (cases $1.5\mu_1$, $1.5\mu_2$, $1.5\mu_3$ and $2\mu_1$, $2\mu_2$, $2\mu_3$) the decrease of powers after elliptical deformation could be sometimes more dramatical (nearly 10%) when comparing with a moderate violation of H_0 (the case μ_1, μ_2, μ_3). However, from the simulation results it is not clear if the fall-off of the power increases monotonically with an increas of distances between location parameters.

In the case of location parameters $1.5\mu_1$, $1.5\mu_2$, $1.5\mu_3$ we simulated also the case of unbalanced samples: $n_1 = 20$, $n_2 = 30$ and $n_3 = 40$ (see Table 3). As in the "H_0-situation" the results are similar to the balanced case (Table 2): simulated power is decreasing with ρ increasing.

5. Changed orientation of location parameter vectors

Since spatial median and arithmetic mean are rotationally equivariant, the test statistics M_1, M_2 and T^2 are rotationally invariant (i.e. the value of the test statistic does not change after a rotation of the data).

Table 3: Location parameters $1.5\mu_1$, $1.5\mu_2$, $1.5\mu_3$ and unlabanced sample sizes $n_1 = 20$, $n_2 = 30$ and $n_3 = 40$

ρ	M_1	M_2	T^2
0	0.798	0.811	0.104
0.5	0.807	0.821	0.104
0.6	0.806	0.819	0.104
0.7	0.801	0.817	0.104
0.8	0.789	0.803	0.104
0.9	0.754	0.769	0.104
0.95	0.714	0.738	0.104

This can be seen in the first row of Table 4: first we generated data points around the location parameters

$$\mu_1 := \begin{pmatrix} 0.4 \\ -0.4 \\ 0 \end{pmatrix}, \quad \mu_2 := \begin{pmatrix} -0.4 \\ 0.4 \\ 0 \end{pmatrix}, \quad \mu_3 := \begin{pmatrix} 0.4 \\ 0 \\ 0 \end{pmatrix}$$

and then we rotated the data 90 degrees around the z-axis in $\mathbf{R}^3$ and obtained the same simulation results. Note that rotation of the data corresponds (i.e. the distribution of the test statistics is the same) to sampling around rotated location parameters

$$\mu_1' := \begin{pmatrix} 0.4 \\ 0.4 \\ 0 \end{pmatrix}, \quad \mu_2' := \begin{pmatrix} -0.4 \\ -0.4 \\ 0 \end{pmatrix}, \quad \mu_3' := \begin{pmatrix} 0 \\ 0.4 \\ 0 \end{pmatrix}$$

because the distributions of the samples are spherically symmetric and the test statistics are rotationally invariant.

Table 4: Powers of the tests for "deformed" and "rotated & deformed" data

	Location parameters μ_1, μ_2, μ_3			Location parameters μ_1', μ_2', μ_3'		
ρ	M_1	M_2	T^2	M_1	M_2	T^2
0	0.558	0.574	0.062	0.558	0.574	0.062
0.5	0.553	0.570	0.062	0.536	0.558	0.062
0.6	0.547	0.567	0.062	0.531	0.555	0.062
0.7	0.539	0.558	0.062	0.522	0.550	0.062
0.8	0.525	0.547	0.062	0.516	0.552	0.062
0.9	0.502	0.528	0.062	0.527	0.564	0.062
0.95	0.487	0.520	0.062	0.564	0.603	0.062

If we now elliptically deform the original data (samples around μ_1, μ_2, μ_3) by the matrix $\Sigma^{\frac{1}{2}}$ and do the same with the rotated data (samples around μ_1', μ_2', μ_3') we again obtain the same powers for the

Lawley-Hotelling test but not for the median tests based on M_1 and M_2 (compare the corresponding columns in Table 4).

Now we try to explain the different behaviour of the tests for "deformed" and "rotated & deformed" data. Note that spherical symmetry of the distributions of the samples and rotational invariance of our test statistics ensure that the powers of all three tests depend on the location parameters μ_1, μ_2, μ_3 only through the mutual distances between them and not on the spatial orientation of the vectors μ_1, μ_2, μ_3. But when we deform the data by the matrix $\Sigma^{\frac{1}{2}}$ the distances between the location parameters μ_1, μ_2, μ_3 will change and this change depends on the spatial orientation of μ_1, μ_2, μ_3 with respect to the elliptical contours given by the matrix Σ. For the affine invariant Lawley-Hotelling T^2 it does not cause problems because the Lawley-Hotelling test can be seen as retransformation of the data to the spherically symmetric case with original location parameters μ_1, μ_2, μ_3 and then computing of the test statistic T^2. Hence we obtained the same power. But M_1 and M_2 are not able to "return back" to the undeformed data so they have to deal with a completely new situation: changed distances between the new location parameters and elliptical distribution of the samples.

Looking at Table 4 one can notice that for location parameters μ_1, μ_2, μ_3 there is an decrease of powers of the median tests when ρ is increasing (similar behaviour as in Table 2) but for the rotated location parameters μ_1', μ_2', μ_3' there is an increase of powers for ρ close to 1! The reason is that whereas before the deformation of the data there was no difference (in the distances between the location parameters) between the original (location parameters μ_1, μ_2, μ_3) and rotated situation (location parameters μ_1', μ_2', μ_3'):

$$\|\mu_1 - \mu_2\| = \|\mu_1' - \mu_2'\| = 1.13,$$

$$\|\mu_2 - \mu_3\| = \|\mu_2' - \mu_3'\| = 0.89,$$

$$\|\mu_2 - \mu_3\| = \|\mu_2' - \mu_3'\| = 0.4,$$

after the elliptical deformation (we take $\rho = 0.95$ as an example) the distances got shorter for the "deformed" data:

$$\|\Sigma^{\frac{1}{2}}\mu_1 - \Sigma^{\frac{1}{2}}\mu_2\| = 0.25,$$

$$\|\Sigma^{\frac{1}{2}}\mu_2 - \Sigma^{\frac{1}{2}}\mu_3\| = 0.44,$$

$$\|\Sigma^{\frac{1}{2}}\mu_2 - \Sigma^{\frac{1}{2}}\mu_3\| = 0.4,$$

and longer for the "rotated & deformed" data:

$$\|\Sigma^{\frac{1}{2}}\mu_1' - \Sigma^{\frac{1}{2}}\mu_2'\| = 1.58,$$

$$\|\Sigma^{\frac{1}{2}}\mu_2' - \Sigma^{\frac{1}{2}}\mu_3'\| = 1.19,$$

$$\|\Sigma^{\frac{1}{2}}\mu_2' - \Sigma^{\frac{1}{2}}\mu_3'\| = 0.4.$$

Hence it should be easier for the median tests to reveal the violation of H_0 in case of the "rotated & deformed" data but the situation is also influenced by the fact that the distributions in the samples are no more spherically but elliptically symmetric. So one can see that an elliptical deformation of the data does not necessary imply a decrease of the powers of the test. It depends on the spatial orientation of the location parameter vectors with respect to the elliptical contours given by the matrix Σ.

6. Conclusions

The results of the simulation study show that the elliptical symmetry does not affect the performance of the spatial median tests M_1 and M_2 that much as the performance of some non-affine invariant test studied by simulations in [1] and [5]. And M_1 and M_2 were still better than the affine invariant Lawley-Hotelling test in case of the heavy-tailed multivariate Cauchy distributon. Moreover, it turned out that the performance of M_1 and M_2 under elliptical symmetry strongly depends on the orientation of the true location parameter vectors and in some situations the elliptical deformation of the data could even improve the powers of M_1 and M_2!

A “negative” result for the spatial median tests is that the elliptical deformation increases their finite sample probabilities of type I error. This phenomenon was not observed in case of non-affine invariant tests studied in [1] and [5]. Further we found out that the impact of elliptical deformation is stronger when violation of H_0 is more serious (observed also in [1] and [5]).

Despite of some promising properties of the spatial median test statistics M_1 and M_2 it is still a challenging problem for us to adjust them so that they become affine invariant.

Acknowledgements

The author is very grateful to his supervisor František Rublík for several helpful suggestions and stimulating discussions which have improved the quality of this paper. The author also would like to express his appreciation to two anonymous referees for their careful reading and suggested improvements. The work has been supported by the VEGA grant No. 1/3016/06 of the Science Grant Agency of the Slovak Republic.

References

[1] Um, Y. & Randles, R. H. (1998). *Nonparametric tests for the multivariate multi-sample location problem.* Statistica Sinica, 8, 801-812.

[2] Brown, B. M. (1983). Statistical use of the spatial median. *Journal of the Royal Statistical Society Series B*, **49**, 25-30.

[3] Chakraborty, B., Chaudhuri, P. (1996). On a transformation and re-transformation technique for constructing affine equivariant multivariate median. *Proceedings of American Mathematical Society*, **124**, 2539-2547.

[4] Hettmansperger, T. P., Möttönen, J., Oja H. (1998). Affine invariant multivariate rank tests for several samples. *Statistica Sinica* **8**, 785-800.

[5] Chakraborty, B., Chaudhuri, P., Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica* **8**, 767-784.

[6] Chaudhuri, P. (1992). *Multivariate location estimation using extension of R-estimates through U-statistics type approach.* Annals of Statistics, 20, 897-916.

[7] Somorčík, J. (2006). *Tests Using Spatial Median.* Austrian Journal of Statistics, 35, 331-338.

[8] Bai, Z. D., Chen, X. R., Miao, B. Q., Rao, C. R. (1990). Asymptotic theory of least distances estimate in multivariate linear models. *Statistics*, **21**, 503-519.

[9] Rublík, F. (2001). *Test of some hypotheses on charakteristic roots of covariance matrices not requiring normality assumptions.* Kybernetika, 37, 61-78.