A note on Left-Spherically Distributed Test with covariates

Livio Finos
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this note we extend the Left-Spherically Distributed linear scores test (LSD) of Läuter, Glimm and Kropf test (1998, The Annals of Statistics). The LSD test is a method for multivariate testing also applicable to the $p >> n$ (much more variables than observations). As a key feature, the score coefficients are chosen such that a left-spherical distribution of the scores is reached under the null hypothesis. Here the test is extended to account for nuisance parameters, particularly for covariates that are assumed to explain (part of) the response variables but are not under test. Moreover it is shown how to deal with the random matrix D which is a Borel function of the residuals under the tested null hypothesis instead of a Borel function of total residuals. This provides a D matrix - and a test as well - which is more focused on the effects of the predictors under test. A R code is available on the `someMTP` package in `CRAN`.

Keywords: Multiple Testing, Hight dimensional data

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

First Draft (2010-07-12)

Contents

Department of Statistical Sciences   **Corresponding author:**
Via Cesare Battisti, 241   Livio Finos
35121 Padova   tel: +39 049 827 4113
Italy   `livio@stat.unipd.it`
                                     `http://www.stat.unipd.it/~livio`

tel: +39 049 8274168

fax: +39 049 8274170

`http://www.stat.unipd.it`

A note on Left-Spherically Distributed Test with covariates

Livio Finos
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this note we extend the Left-Spherically Distributed linear scores test (LSD) of Läuter, Glimm and Kropf test (1998, The Annals of Statistics). The LSD test is a method for multivariate testing also applicable to the $p >> n$ (much more variables than observations). As a key feature, the score coefficients are chosen such that a left-spherical distribution of the scores is reached under the null hypothesis. Here the test is extended to account for nuisance parameters, particularly for covariates that are assumed to explain (part of) the response variables but are not under test. Moreover it is shown how to deal with the random matrix D which is a Borel function of the residuals under the tested null hypothesis instead of a Borel function of total residuals. This provides a D matrix - and a test as well - which is more focused on the effects of the predictors under test. A `R` code is available on the `someMTP` package in `CRAN`.

Keywords: Multiple Testing, Hight dimensional data

1   Introduction

In this brief manuscript we make the Left-Spherically Distributed linear scores test (LSD) of Läuter *et al.* (1998) easier to deal with covariates not under test. The LSD test is a method for multivariate testing also applicable to the $p >> n$ (much more variables than observations). It is very general and works with any multivariate spherical error term and any set of predictors. It is also possible to deal with covariates known to be associated with the response and not of interest for the test. Despite this result, it is hard to figure out a general framework to deal with them.

The following is the theoretical framework used in Läuter *et al.* (1998) (with slight notational changes). The presented theory refers to $n \times p$ left-spherically distributed matrices $\mathbf{Y}$ (n refers to observations and p to variables). Such random matrices are characterized by the fact that

$$\mathbf{Y} =^d C\mathbf{Y} \quad \text{for every fixed } n \times n \text{ orthogonal matrix } C$$

and by a characteristic function of the form $\phi(T'T)$. The symbol $=^d$ denotes the equality of two distributions. This class of distributions is too wide for the construction of optimal tests according to the likelihood ratio criterion. Therefore, other authors (Fang and Zhang (1990), Anderson (1993), Gupta and Varga

(1993)) prefer a restricted class of distributions $\mathbf{Y}$ with a characteristic function $\psi(tr(T'TA)) = \psi(tr(TAT'))$, where A is a $p \times p$ positive definite symmetric matrix. The matrices $\mathbf{Y}$ of this class have the representation $\mathbf{Y} =^d \mathbf{U}A^{1/2}$ with a vector-spherically distributed matrix $\mathbf{U}$ (Fang and Zhang (1990), page 96). The notation $A^{1/2}$ indicates the positive definite symmetric matrix that satisfies $(A^{1/2})^2 = A$. Under these assumptions Läuter *et al.* (1998) prove the following theorem:

THEOREM 1. *Assume* $1 \le f_H$, $1 \le f_G$ *and* $1 \le q \le f_H + f_G$. *Assume a test statistic* $F = F(\mathbf{H}_Z, \mathbf{G}_Z)$ *as a Borel function defined for all* $q \times q$ *positive semidefinite symmetric matrices* $\mathbf{H}_Z$ *and* $\mathbf{G}_Z$ *and satisfying the invariance condition*

$$(9) \qquad\qquad F(\mathbf{A}\mathbf{H}_Z\mathbf{A}, \mathbf{A}\mathbf{G}_Z\mathbf{A}) = F(\mathbf{H}_Z, \mathbf{G}_Z)$$

for every $q \times q$ *positive definite symmetric matrix* $\mathbf{A}$.

Now, suppose a dimension p *with* $p \ge q$. *Let* $\mathbf{H}, \mathbf{G}$ *and* $\mathbf{L}$ *be three* $p \times p$ *random positive semidefinite symmetric matrices that are mutually stochastically independent, where* $\mathbf{H}$ *and* $\mathbf{G}$ *have the Wishart distributions*

$$(10) \qquad\qquad \mathbf{H} \sim W_p(\mathbf{\Sigma}, f_H), \qquad \mathbf{G} \sim W_p(\mathbf{\Sigma}, f_G)$$

for a positive definite $\mathbf{\Sigma}$. *Let* $\mathbf{D}$ *be a* $p \times q$ *random matrix defined as a Borel function of* $\mathbf{H} + \mathbf{G} + \mathbf{L}$ *and having rank* q *with probability* 1.

Then the distribution of $F = F(\mathbf{D}'\mathbf{H}\mathbf{D}, \mathbf{D}'\mathbf{G}\mathbf{D})$ *is the same for each* p, *each* $\mathbf{\Sigma}$, *each* $\mathbf{D}$ *function and each suitable* $\mathbf{L}$ *distribution.*

D is defined as a function of $\mathbf{H} + \mathbf{G} + \mathbf{L}$, where $\mathbf{L}$ is a matrix of "neutral" information. For example, this is the case of multivariate model having predictors which are not under test (see application 4 Läuter *et al.* (1998), page 1979). The authors present the case of multivariate two-way classification with orthogonal design. *In the classical MANOVA approach, tests of hypotheses are performed by means of various* $p \times p$ *sums of products matrices* $\mathbf{H}_A$, $\mathbf{H}_B$, *and* $\mathbf{H}_{A \times B}$, *say, associated with main effects and interactions. Using Theorem 1, a coefficient vector* d *for weighting the* p *variables may be determined as a function of* $\mathbf{H}_A + \mathbf{H}_B + \mathbf{H}_{A \times B} + \mathbf{G}$, *where* $\mathbf{G}$ *is the residual sum of products matrix. The main effect* A *may be tested by*

$$F = \frac{d'\mathbf{H}_A d / f_A}{d'\mathbf{G}d / f_G}$$

According to Theorem 1, this statistic is F *distributed with* f_A *and* f_G *degrees of freedom under* H_A *(the null hypothesis of null effect of factor A), because* $\mathbf{L} = \mathbf{H}_B + \mathbf{H}_{A \times B}$ *is stochastically independent of* $\mathbf{H}_A$ *and* $\mathbf{G}$, *even if the hypotheses* H_B *and* $H_{A?B}$ *are not true.* [...]. *Of course, in any practical application, one would also have to consider the power of the resulting tests.*

In this example the authors highlight the fact that a test is possible when the matrix $\mathbf{Y}'\mathbf{Y}$ can be decomposed in three orthogonal matrix: the explained part (e.g. $\mathbf{H}_A$), the "nuisance" part (e.g. $\mathbf{H}_B + \mathbf{H}_{A \times B}$) and a residual part (eg $\mathbf{G}$). However the proposed solution is possible only for the very special case of orthogonal variables/factor. In this work we discuss a solution for the general case of testing part of the predictors when they are not reciprocally orthogonal.

A second remarkable point is the following. The reason of the last comment in the quoted text is due to the fact that the weights (i.e. the elements of D) given to any single variables will be affected by the size of the effect B and $A \times B$. A second result of this work is to provide a transformation of $\mathbf{Y}$ that allow to define a nuisance-free D matrix for any kind of covariates (evene if non orthogonal to the tested predictors).

2   The LSD test

Let us define a general framework where the $n \times p$ matrix $\mathbf{Y}$ is

$$\mathbf{Y} = \mathbf{X}\mathbf{B_X} + \mathbf{Z}\mathbf{B_Z} + \mathbf{E}, \tag{1}$$

$\mathbf{X}$ and $\mathbf{Z}$ being a $n \times k$ and $n \times h$ matrix respectively and $k + h < n$. $\mathbf{E} = (\epsilon'_1, \ldots, \epsilon'_n)$ is a $n \times p$ spherical matrix and has n independent multivariate errors ($\epsilon_i \sim (\mathbf{0}, \Sigma)$ $\forall i = 1, \ldots, n$) $\mathbf{B_X}$ and $\mathbf{B_Z}$ are $k \times p$ and $h \times p$ matrix respectively.

The null hypothesis to be tested is

$$H_0 : \mathbf{B_X} = \mathbf{0_{k \times q}}.$$

In this case we can not easily reach the decomposition $\mathbf{Y}'\mathbf{Y} = \mathbf{H_X} + \mathbf{H_Z} + \mathbf{G}$ as have been done in the motivating example quoted in the section since $\mathbf{H_X} = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ will be not independent from $\mathbf{H_Z} = \mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ except for the special case of orthogonal designs.

Define $\mathbf{I}_{n \times n} - \mathbf{P_Z} = \mathbf{I}_{n \times n} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and premultiply it in left and right member of equation (1)f:

$$(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{Y} = (\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{X}\mathbf{B_X} + (\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{E} \tag{2}$$

The equation is now free from the nuisance parameters $\mathbf{B_Z}$ but has n observations which are not independent. Kherad-Pajouh and Renaud (2010) cope with a similar problem but in the case of permutation tests with nuisance parameters. They suggest to reduce the $n \times q$ matrix $(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{Y}$ in a $n - h \times q$ matrix of independent rows by using the eigenvalues decomposition of $\mathbf{I}_{n \times n} - \mathbf{P_Z}$. Since it is a symmetric and idempotent matrix, it possesses only two eigenvalues: 0 and 1. Let now $\mathbf{I}_{n \times n} - \mathbf{P_Z} = UDU'$ be the eigen-decomposition, where D is a diagonal matrix containing the eigenvalues, and U is a unitary matrix, whose columns are the eigenvectors. Since $\mathbf{I}_{n \times n} - \mathbf{P_Z}$ has rank $n - h$, there are exactly $n - h$ ones and h zeros in the diagonal of D.

Therefore pre-multiplying by $(DU)'$ both sides of 2, we get the modified model as follows:

$$\mathbf{Y_0} = \mathbf{X_0}\mathbf{B_X} + \mathbf{E_0} \tag{3}$$

where $\mathbf{V_0} = U'D'(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{V}$ ($\mathbf{V} = \{\mathbf{Y}, \mathbf{X}, \mathbf{E}\}$). Hence under the null hypothesis and through the lemma 4 in Dawid (1977) $\mathbf{Y_0} = \mathbf{E_0}$ is left-spherical.

Now we can apply again the theorem 1 on (3) defining $D = f(G + H)$ (and not on L). Recalling the quoted application 4 inLäuter *et al.* (1998) we see that the D matrix can be chosen as a function of $\mathbf{H}_A + \mathbf{G}$ which rise up to be a big gain

when effects of factors B and $A \times B$ are big. Let's define $\mathbf{H_0}$ and $+\mathbf{G_0}$ as $\mathbf{Y_0}'\mathbf{Y_0} = \mathbf{Y_0}'\mathbf{X_0}(\mathbf{X_0}'\mathbf{X_0})^{-1}\mathbf{X_0}'\mathbf{Y_0} + \mathbf{Y_0}(\mathbf{I}_{n \times n} - \mathbf{X_0}(\mathbf{X_0}'\mathbf{X_0})^{-1}\mathbf{X_0}')\mathbf{Y_0} =: \mathbf{Y}'(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{Y} = \mathbf{H_0} + \mathbf{G_0} =$, then a valid test has the form

$$F = \frac{trace(D'\mathbf{H_0}D)/(k - 1 + q)}{trace(D'\mathbf{G_0}D)/(n - h - k + 1 - q)}$$

and $D = f(\mathbf{H_0} + \mathbf{G_0}) = f(\mathbf{Y_0}'\mathbf{Y_0}) = f(\mathbf{Y}'(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{Y})$.

Remark 1. The two samples case described in application 2 of Läuter *et al.* (1998) fall to be a special case of this generalization.

Remark 2. The permutation-based version of the test is also possible of course. It is based on permutations of the rows of $\mathbf{X_0}$. The solution is exact when $\mathbf{Z}$ is a constant vector (the test is permutationally equivalent to the not mean-centered multivariate test) or when $\mathbf{E}$ is spherical (Kherad-Pajouh and Renaud (2010), theorem 4). When $\mathbf{E}$ has (row-wise) exchangeable terms, then $\mathbf{Y_0}$ has weakly exchangeable (rows) elements (Kherad-Pajouh and Renaud (2010), theorem 3) and the test is exact up to the second moment.

2.1   Software and simulation

The calculations needed to perform the LSD test are quite simple, however we have produced a R code for the users. This is the function `lsd.test()` in some MTP package available on CRAN. To get the test for the one-sample case of application 1 in Läuter *et al.* (1998) we set $\mathbf{X} = \mathbf{1}$ and no covariates. The R call is as follow:

```
> library(someMTP)
> set.seed(1)
> X = matrix(rnorm(50), 5, 10) + 1
> lsd.test(resp = X, alternative = rep(1, 5))


F = 201.1744; p =  0.0001434661
```

The $C \leq 2$ sample case can be handled setting $\mathbf{Z} = \mathbf{1}$ and $\mathbf{X}$ as a matrix of dummy variables.

```
> X2 = X + matrix(c(0, 0, 1, 1, 1), 5, 10) * 1
> lsd.test(resp = X2, null = rep(1, 5),
+              alternative = c(0, 0, 1,1, 1))


F = 43.1053943; p =  0.007186903
```

The general setting is handle in a similar way

```
> lsd.test(resp = X2, null = cbind(rep(1, 5), c(0, 0, 1, 1, 1)),
+              alternative = 1:5)


F = 0.3277843; p =  0.6247483
```

There are many studies evaluating the power of LSD. The performance strongly depends on the D matrix. We just run three quick simulations to see the effect of covariates in the test. We fix $n = 15$ (sample size) and $p = 10$ (variables). The **Y** matrix are generated from a 10-variate normal distribution with identity matrix σ and mean determined by the subject ID (ranging between 1 and $n = 15$). The predictor **X** is a two sample contrast comparing the first 5 observations vs the remaining 10. This make **X** and **Z** non orthogonal. In Figure 1 we report the results. In Sim1 the LSD test does not consider the covariate **Z** and the type I error is out of control as expected. The Sim2 performs the LSD taking in account **Z** and the type I error overlap the principal bisector. In Sim3 the data are generate as in Sim1 and Sim2 but 0.5 is added on the last ten observations. The test is under the alternative hypothesis.

```
> library(someMTP)
> X <- matrix(c(rep(0, 5), rep(2, 10)), 15, 1)
> Z <- matrix(1:15, 15, 1)
> Sim1 = replicate(10000, {
+     Y = matrix(rnorm(450), 15, 30) + matrix(1:15, 15, 30)
+     lsd.test(resp = Y, null = cbind(rep(1, 15)), alternative = X)
+ })
> Sim2 = replicate(10000, {
+     Y = matrix(rnorm(450), 15, 30) + matrix(1:15, 15, 30)
+     lsd.test(resp = Y, null = cbind(1, Z), alternative = X)
+ })
> Sim3 = replicate(10000, {
+     Y = matrix(rnorm(450), 15, 30) + matrix(1:15, 15, 30) +
+                         matrix(c(rep(0, 5), rep(0.5, 10)), 15, 30)
+     lsd.test(resp = Y, null = cbind(1, Z), alternative = X)
+ })
```
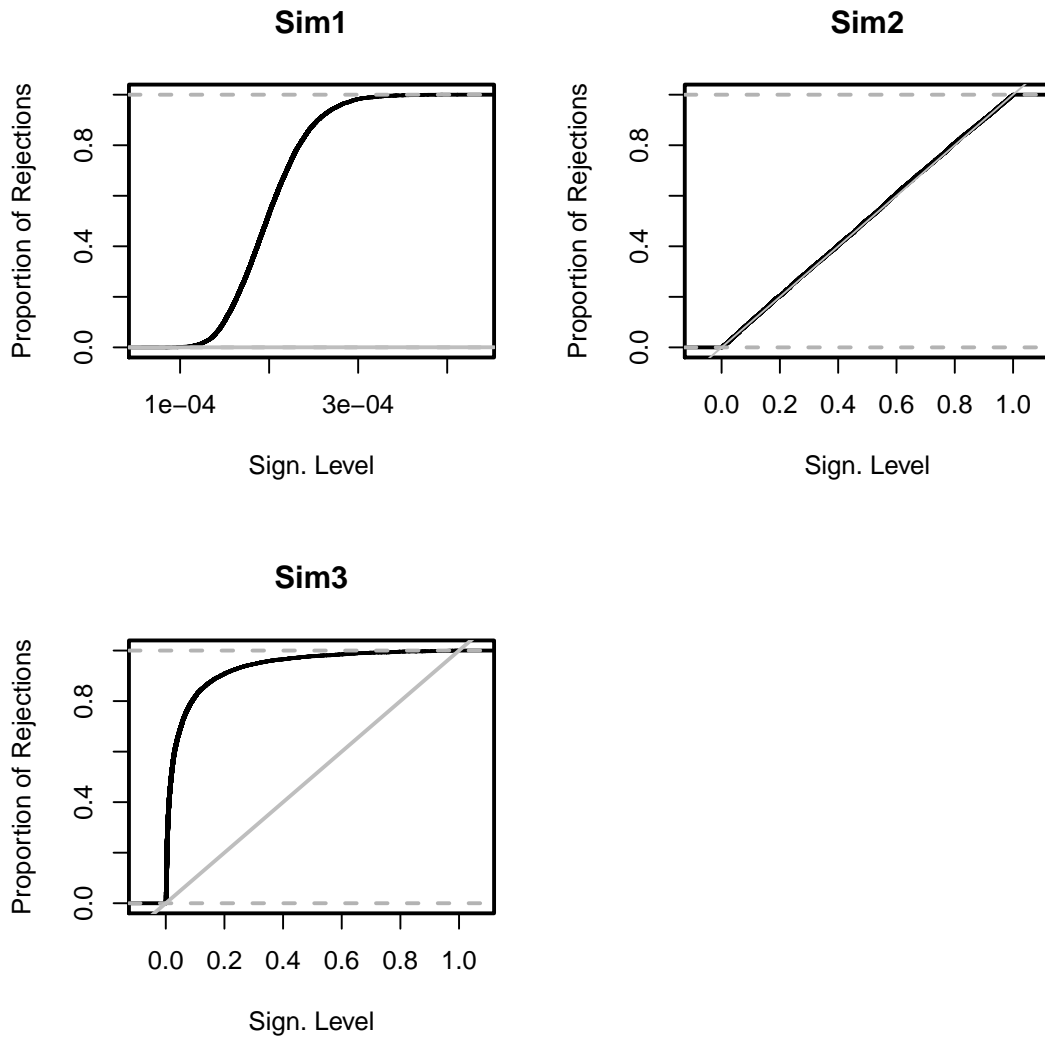
Figure 1: Data are generated as described in the text. In Sim1 the LSD test does not consider the covariate **Z**. The Sim2 performs the LSD taking in account **Z**. In Sim3 the data are generate as in Sim1 and Sim2 but 0.5 is added on the last ten observations.

3   Conclusion

In this brief note we define the LSD test in a slight more general framework and make it easier to deal with covariates (i.e. predictors to be putted in the model but note tested). We also show how to define $D = f(\mathbf{H_0} + \mathbf{G_0}) = f(\mathbf{Y}'(\mathbf{I}_{n \times n} - \mathbf{P_Z})\mathbf{Y})$ instead of $D = f(\mathbf{H} + \mathbf{G} + \mathbf{L}) = f(\mathbf{Y}'\mathbf{Y})$.

The related R code is available with the package `someMTP` in CRAN.

References

T. W. ANDERSON (1993). Nonnormal multivariate distributions: inference based on elliptically contoured distributions. *Multivariate Analysis: Future Directions (C. R. Rao, ed.)1 - 24. North-Holland, Amsterdam*, 1 – 24.

A. P. DAWID (1977). Spherical matrix distributions and a multivariate model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39, 2**, 254–261.

K.T. FANG AND Y.T. ZHANG (1990). *Generalized Multivariate Analysis*. Science Press, Beijing.

A. K. GUPTA AND T. VARGA (1993). *Elliptically Contoured Models in Statistics*. Kluwer, Dordrecht.

SARA KHERAD-PAJOUH AND OLIVIER RENAUD (2010). An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova. *Computational Statistics & Data Analysis*, **54, 7**, 1881 – 1893.

J. LÄUTER, E. GLIMM, AND S. KROPF (1998). Multivariate tests based on left-spherically distributed linear scores. *Annals of Statistics*, **26**, 1972–1988.

Acknowledgements

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy