Exact Multivariate Tests for Brain Imaging Data

Rita Almeida^{1,2*} and Anders Ledberg¹

¹Division of Human Brain Research, Department of Neuroscience, Karolinska Institute, Sweden

²Institute of Biophysics and Biomedical Engineering, University of Lisbon, Portugal

Abstract: In positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) data sets, the number of variables is larger than the number of observations. This fact makes application of multivariate linear model analysis difficult, except if a reduction of the data matrix dimension is performed prior to the analysis. The reduced data set, however, will in general not be normally distributed and therefore, the usual multivariate tests will not be necessarily applicable. This problem has not been adequately discussed in the literature concerning multivariate linear analysis of brain imaging data. No theoretical foundation has been given to support that the null distributions of the tests are as claimed. Our study addresses this issue by introducing a method of constructing test statistics that follow the same distributions as when the data matrix is normally distributed. The method is based on the invariance of certain tests over a large class of distributions of the data matrix. This implies that the method is very general and can be applied for different reductions of the data matrix. As an illustration we apply a test statistic constructed by the method now presented to test a multivariate hypothesis on a PET data set. The test rejects the null hypothesis of no significant differences in measured brain activity between two conditions. The effect responsible for the rejection of the hypothesis is characterized using canonical variate analysis (CVA) and compared with the result obtained by using univariate regression analysis for each voxel and statistical inference based on size of activations. The results obtained from CVA and the univariate method are similar. *Hum. Brain Mapping* 16:24–35, 2002. © 2002 Wiley-Liss, Inc.

Key words: multivariate analysis; PET; JMRI; linear models

INTRODUCTION

The most widely used approach to the analysis of positron emission tomography (PET) and magnetic resonance imaging (fMRI) data is to apply a univariate linear model (ULM) for each voxel, followed by a correction for multiple comparisons. Although this

univariate approach is the most common, brain imaging data has clear multivariate features, since the signal measured at different voxels covaries. Concomitantly, several multivariate approaches have been proposed, ranging from descriptive characterizations of patterns of activity to multivariate linear models [for reviews of both univariate and multivariate strategies see Lange et al., 1999, Petersson et al., 1999a, b].

Our study is about the application of multivariate linear model (MLM) analysis. MLM analysis has previously been used for brain imaging data [Friston et al., 1995, 1996]. In these reports the original data matrix was reduced before the analysis. The reduced data matrix was assumed to be normally distributed. That is not the case. This could imply that the test statistics used would not follow the distributions claimed. In

Grant sponsor: Ministério da Ciência e Tecnologia, Portugal; Grant number: PRAXIS XXI/BD/13979/97; Grant sponsor: European Union; Grant number: QLG3-CT-1999-00677.

*Correspondence to: Rita Almeida, Division of Human Brain Research, Department of Neuroscience A3:3, Karolinska Institutet, Retzius väg 8, 17177 Stockholm, Sweden.
E-mail: Rita.Almeida@neuro.ki.se.

Received for publication 17 April 2001; accepted 9 November 2001.

the referred reports no derivation of the distributions of the test statistics under the null hypothesis is given. The aim of this study is to explain how to test hypothesis on the reduced data matrix using test statistics that have known distributions under the null hypothesis.¹

Our study is organized as follows. In the first section MLM analysis and its applications are explained in comparison to ULM analysis. Subsequently a problem of applying MLM analysis to brain imaging data is discussed. Finally, the concept of invariant tests is introduced and their applicability to solve the mentioned problem is discussed. In the theory section the left-spherical distributions and some of its properties that are relevant for this study are introduced. Tests for MLM analysis when the number of variables exceeds the number of observations are constructed. Some technical steps concerning these tests are presented in appendices A, B, and C. We describe the application of one multivariate test to a PET data set. To characterize the effect responsible for the rejection of the null hypothesis canonical variate analysis is used. The results are then compared with the ones obtained by using univariate linear model analysis and statistical inference based on the size of activations.

Multivariate linear model analysis for brain imaging data explained in comparison to univariate linear model analysis

One important difference between using a MLM or a ULM is the way the data is considered. In both cases, the data consists of a set of measurements reflecting brain activity at different times, scans, for a set of units of volume, voxels. With a ULM the signals from different voxels are considered separately, even if the same model (often expressed in a design matrix) is usually fitted to all voxels. For each variable (measure reflecting brain activity in a voxel) there is one observation for every time point. This approach allows one to make spatially restricted statistical inferences and therefore, to conclude that a certain volume within the brain has significantly increased activity during the performance of a task relatively to the activity during the performance of another task [Friston et al., 1991; Worsley et al., 1992]. In a MLM, a model is also fitted to the measured signal from all the voxels. In this case, each observation comprises all the variables. This im-

plies that inference cannot be made on just a subset of voxels. The tests applied in MLM analysis assess effects in all voxels at the same time and account for dependencies between them. Inference is usually made for the whole volume. This allows the conclusion that there is a difference in the activation over the whole volume when two different tasks are compared. ULM and MLM analyses allow testing of different hypotheses and so the situations in which it is more interesting to use one or the other might differ. MLM analysis might be a sensitive way of testing the hypothesis that a given effect (for example an experimental condition) has no influence on the whole brain activity. Using this idea, MLM analysis has been suggested as a means for comparing models [Friston et al., 1996]. For this purpose, one can test a succession of models including more explanatory effects. In the same article, MLM analysis has also been suggested as a possible way of testing the validity of using subtraction between tasks, by testing that the interaction between factors is not significant. In this case one would test that the interaction between two factors has no effect in the activity of the whole brain. MLM analysis in combination with other multivariate techniques have also been suggested as a method to test for significant activity in the brain with unpredicted spatiotemporal patterns [Worsley et al., 1997], supplying information that ULM analysis and related tests do not provide.

Multivariate linear model analyses when the number of variables exceeds the number of observations

MLM analysis cannot be applied directly to brain imaging data. Let the data be organized in a $n \times p$ matrix where the n rows correspond to the scans and the p columns to the voxels. To apply a MLM and test hypothesis on brain imaging data the following two characteristics are important: 1) the number of degrees of freedom (dependent of the number of observations) has to be larger than the number of variables; and 2) the data is assumed to be normally distributed. The first characteristic is needed to assure that the estimated covariance matrix is invertible, in order to use the test statistics. The normality assumption, (2), is needed for the derivation of the distributions of the tests used. The problem is that (1) and (2) cannot be fulfilled at the same time. The data matrix, in the format described above, can be assumed normally distributed. In this case, (2) is verified but not (1), because for PET and fMRI data the number of voxels is always larger than the number of scans. To over-

¹Note that there are other methods that allow testing of multivariate hypothesis on imaging data. In particular, Partial Least Squares analysis [McIntosh et al., 1996, 1998] has been used for this effect.

come the situation, the data matrix can be reduced to conform to (1). In this case, (1) is verified but not (2), because the reduced data matrix is in general no longer normally distributed. Applications to brain imaging data have usually followed the approach of reducing the data matrix [Friston et al., 1995, 1996]. The aim of the reduction is to summarize what is thought to be important information contained in the data in a number of new variables smaller than the number of degrees of freedom. Principal component analysis has been used to this effect. As stated before, however, this reduction implies a violation of normality. In this situation the multivariate test statistics used for normally distributed data (see below) might not keep their levels of significance. This problem has not been addressed in the literature concerning applications of MLM analysis to functional brain data. In the previous work no foundation is given to support the claimed null distributions of the test statistics. The aim of this study is to show how to assure that the tests for reduced data follow the same distributions of the tests derived for normal data.

Invariant tests for multivariate brain imaging data

The concept of invariant tests over a class of distributions is introduced and illustrated with an example. This concept is important for the construction of multivariate tests for brain imaging data. Many tests concerning multivariate analysis have been derived for normally distributed data [for examples see Christensen, 1991; Mardia et al., 1979]. The multivariate normal distribution is characterized by two parameters, the mean vector and the covariance matrix. The tests derived for multivariate normally distributed data are valid for independent and identically distributed samples from any normal distribution, independently of its mean and covariance matrix, as long as it is non singular. Some of these tests are valid over a wider class of distributions. The following example illustrates the concept of invariance of a test over a class of distributions.² The example, for the sake of simplicity, deals with the t -test for univariate data.

Example 1

Let $u_1, \dots, u_n$ be an independent identically distributed sample from a normal distribution with mean μ

and variance σ^2 . To test the hypothesis $\mu = 0$ the t -test is usually employed, given its statistical properties. The test statistic is

$$t_u = \sqrt{n} \frac{\bar{u}}{s_u}, \quad \text{where} \quad \bar{u} = \frac{1}{n} \sum_{k=1}^n u_k \quad \text{and} \quad s_u^2 = \frac{1}{n-1} \sum_{k=1}^n (u_k - \bar{u})^2.$$

The test follows, under the null hypothesis, a Student's t -distribution with $n - 1$ degrees of freedom.

Now let z_k be defined as $z_k = ru_k$ with $k = 1, \dots, n$, where r is a positive random variable independent of u . Note that the z_k 's are not independent and are in general not normally distributed. In analogy with the previous test statistics we can define:

$$t_z = \sqrt{n} \frac{\bar{z}}{s_z}, \quad \text{where} \quad \bar{z} = \frac{1}{n} \sum_{k=1}^n z_k \quad \text{and} \quad s_z^2 = \frac{1}{n-1} \sum_{k=1}^n (z_k - \bar{z})^2.$$

It is easy to see that $\bar{z} = r\bar{u}$ and $S_z = rS_u$ and so that $t_u = t_z$. That is, the distribution of the test statistic is invariant over the class of distributions of the type $z_k = ru_k$.

The example above shows that the test statistics (t_u or t_z) follow the same distribution although one is obtained through a transformation of the original variables, making them non-normal. That is, the test is invariant under a certain type of transformations. We will now describe how to use the invariance property to test multivariate hypothesis on functional imaging data.

Several common tests used to test multivariate hypotheses are invariant over a class of distributions called the class of left spherical distributions. Henceforth an invariant test refers to an invariant test over the class of left spherical distributions. The class of left spherical distributions contains the normal distribution. This implies that invariant test statistics constructed from a left spherically distributed data matrix, will follow the same distribution as if the tests were constructed from a normally distributed data matrix, in analogy with Example 1.

As stated above, to test hypotheses on brain imaging data the data matrix has to be reduced. If the

²The example is modified from an example given by K.T. Fang on a lecture held at the University of Uppsala, Sweden, in 1999 and on the related course notes [Fang, 1999].

reduction is such that the reduced data matrix belongs to the class of left spherical distributions any invariant test can be applied and the distribution of the test statistic under the null hypothesis will be known. The present study is based on the results of Fang and coauthors [Fang and Chen, 1989; Fang and Zhang, 1990] and of Lauter et al. [Lauter, 1996; Lauter et al., 1996, 1997].

THEORY

Left spherical matrix distributions

Capital bold face letters denote matrices, small bold face letters denote vectors, $\mathbf{A}_{n \times q}$ is a matrix with n rows and q columns and $\mathbf{A}_n$ is a square matrix with dimension n .

Let $\mathbf{Z}$ be a $n \times q$ matrix. In terms of columns and rows $\mathbf{Z}$ can be written as:

$$\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_q) = (\mathbf{z}_{(1)}, \dots, \mathbf{z}_{(n)})'.$$

Let the n rows of $\mathbf{Z}$ be an independent and identically distributed sample of size n from a population following a multivariate distribution with q dimensions. $\mathbf{Z}$ is said to be a random matrix. This matrix follows a matrix variate normal distribution with mean equal to the null matrix $\mathbf{0}_{n \times q}$ and covariance matrix $\mathbf{I}_n \otimes \Sigma_q$, $N(\mathbf{0}_{n \times q}, \mathbf{I}_n \otimes \Sigma_q)$ if:

$$\text{vec}(\mathbf{Z}') \sim N(\mathbf{0}_{nq}, \mathbf{I}_n \otimes \Sigma_q),$$

where vec is the operator defined by $\text{vec}(\mathbf{Z}) =$

$$\begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_q \end{bmatrix},$$

$\otimes$ is the Kronecker product and $N(\mathbf{0}_{nq}, \mathbf{I}_n \otimes \Sigma_q)$ is the multinormal distribution with mean equal to the null vector $\mathbf{0}_{nq}$ and covariance matrix $\mathbf{I}_n \otimes \Sigma_q$, $\mathbf{I}_n$ being the n dimensional identity matrix.

To test multivariate hypotheses on data where the number of variables is larger than the number of degrees of freedom a reduction of the dimensions of the data matrix is performed. The reduction can be written as a multiplication of the original data matrix $\mathbf{Y}$ with another matrix, say $\mathbf{W}$. $\mathbf{W}$ is in general not a constant matrix but a random matrix, therefore in

general the reduced matrix does not follow a normal distribution.³

We now introduce a class of matrix distributions that allows the generalization of some standard multivariate tests developed for normal distributions. This class of distributions, the class of left spherical distributions, includes the multivariate normal distribution above. The distributions in this class are characterized by their invariance over orthogonal transformations. For the definition the concept of orthogonal matrices is needed. A square matrix $\Gamma_{n \times n}$ is orthogonal if $\Gamma' \Gamma = \Gamma \Gamma' = \mathbf{I}$.

Definition 1

Let $\mathbf{Z}$ be a $n \times q$ random matrix. $\mathbf{Z}$ follows a left spherical distribution if $\mathbf{Z} \stackrel{d}{=} \Gamma \mathbf{Z}$, for every Γ belonging to the class of orthogonal matrices with dimension n , where $\stackrel{d}{=}$ denotes equally distributed.

The reduced matrix, obtained from the original normally distributed data matrix, must have less variables than degrees of freedom (df) and must follow a left spherical distribution. The problem is to determine $\mathbf{W}$ such that the reduced data matrix $\mathbf{Z} = \mathbf{Y}\mathbf{W}$ is left spherical. It turns out that it suffices that $\mathbf{W}$ is a function dependent on $\mathbf{Y}$ only through $\mathbf{Y}'\mathbf{Y}$, $\mathbf{W} = f(\mathbf{Y}'\mathbf{Y})$, given that $\mathbf{Y}$ follows a left spherical distribution. This follows from Definition 1, because for every matrix Γ belonging to the class of orthogonal matrices:

$$\Gamma \mathbf{Z} = \Gamma \mathbf{Y} f(\mathbf{Y}'\mathbf{Y}) = \Gamma \mathbf{Y} f(\mathbf{Y}' \Gamma' \Gamma \mathbf{Y}) \stackrel{d}{=} \mathbf{Y} f(\mathbf{Y}'\mathbf{Y}) = \mathbf{Z}.$$

The result is stated formally in Theorem 1 (proved in the proof of Theorem 2 of Lauter, 1996⁴).

Theorem 1

Let $\mathbf{Y}_{n \times p} \sim N(\mathbf{0}, \mathbf{L}_n \otimes \Sigma)$ and $\mathbf{W}_{p \times q}$ be a function uniquely determined by $\mathbf{Y}'\mathbf{Y}$ and $q < df$ ⁵. Then

$$\mathbf{Z}_{n \times q} = \mathbf{Y}_{n \times p} \mathbf{W}_{p \times q}$$

³This is the point overlooked in previous works [Friston et al., 1995, 1996].

⁴In fact, in the paper the proof has been made for right spherical distributions. The proof of the theorem now stated is easily obtained from the one of Lauter [1996], knowing that if $\mathbf{Z}'$ is left spherically distributed, $\mathbf{Z}$ follows a right spherical distribution.

⁵The condition $q < df$ assures that the reduced matrix has less variables than degrees of freedom, what is necessary to test multivariate hypothesis.

follows a left spherical distribution.

To test multivariate hypothesis on the reduced data matrix we want to use a test following a known distribution under the null hypothesis and that is invariant over the class of left spherical distributions. The following theorem (Theorem 9 of Fang and Chen [1989]) is an important result concerning the invariance of a test over this class of distributions.

Theorem 2

Let $\mathbf{Z}$ be a $n \times q$ random matrix following a left spherical distribution. A statistic $t(\mathbf{Z})$ is invariant over the class of left spherical distributions if and only if $t(\mathbf{Z}) \stackrel{d}{=} t(\mathbf{Z}\mathbf{A})$ where $\mathbf{A}$ is any $q \times q$ nonsingular matrix.

Lawley-Hotelling trace test, Wilk's Λ test, Roy's maximum root test and Pillai-Bartlett test, among other test statistics, have been proposed to test multivariate hypotheses on normally distributed data [Mardia et al., 1979; Christensen, 1991]. In the next subsection Theorem 2 is used to demonstrate the invariance of these test statistics.

Application to multivariate linear models

In this section, we describe the construction of multivariate tests for brain imaging data, according to the theory presented above.

The multivariate linear model

Let us introduce the multivariate linear model:

$$\mathbf{U}_{n \times p} = \mathbf{X}_{n \times t} \mathbf{B}_{t \times p} + \mathbf{E}_{n \times p}$$

where $\mathbf{U}$ is a matrix containing the n observations on each p dependent variable ($p > n$), $\mathbf{X}$ is the matrix of independent variables (design matrix), $\mathbf{B}$ is a matrix of unknown regression parameters and $\mathbf{E} = [E_{ij}]$ is the error matrix. We assume that $\mathbf{E}_{n \times p} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$. The assumption of an error matrix with such a structure implies the assumption of independence between observations. This assumption is reasonable for PET data sets and for fMRI data sets where each observation consists of one session [Ledberg et al., 2001]. For fMRI data sets where each observation consists of one scan, the assumption of independence between observations is not valid, precluding a straight forward application of linear model analysis [Zarahn et al., 1997].

Parameter estimation and statistics for hypothesis testing

The regression parameters can be estimated by using least squares estimates (these estimates are optimal in the sense that they are the linear unbiased estimates with minimum variance). The estimated parameter matrix $\hat{\mathbf{B}}$ satisfies:

$$\mathbf{X}\hat{\mathbf{B}} = \mathbf{M}\mathbf{U}; \quad \mathbf{M}_{n \times n} = \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'$$

where $\mathbf{X}'$ denotes the transpose of $\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^-$ denotes a generalized inverse of $(\mathbf{X}'\mathbf{X})$ and $\mathbf{M}$ is the perpendicular projection operator onto the column space of $\mathbf{X}$.

Let us consider hypothesis of the form:

$$H_0 : \mathbf{c}'\mathbf{B} = 0 \quad \text{against} \quad K : \mathbf{c}'\mathbf{B} \neq 0$$

where

$$\mathbf{c}' = \mathbf{l}'\mathbf{X}.$$

The vector $\mathbf{l}$ is introduced to define the contrast vector $\mathbf{c}'$ in such a way that its estimability is assured. For reasons explained below, we will consider the test of only one contrast at a time, so $\mathbf{l}$ and $\mathbf{c}$ are vectors.

To test hypotheses on the model, test statistics based on the hypothesis and error statistics, $\mathbf{H}$ and $\mathbf{G}$, have been proposed. The hypothesis statistics is the sum of squares and products matrix under the alternative hypothesis. The error statistics is the sum of squares and products matrix due to error. They can be defined as $\mathbf{H} = \mathbf{U}'\mathbf{P}\mathbf{U}$ and $\mathbf{G} = \mathbf{U}'\mathbf{R}\mathbf{U}$ where $\mathbf{P}$ is the perpendicular projection operator onto the column space of $\mathbf{M}$

$$\mathbf{P} = \mathbf{M}(\mathbf{l}'\mathbf{M}\mathbf{l})^{-1} \mathbf{l}'\mathbf{M}$$

and $\mathbf{R}$ is the perpendicular projector operator onto the space complementary to the column space of $\mathbf{X}$

$$\mathbf{R} = \mathbf{I}_n - \mathbf{M}.$$

For the following we need to write $\mathbf{P} = \mathbf{l}\mathbf{l}'$, valid when $\mathbf{l}'$ is a vector with dimension n , is in the column space of $\mathbf{X}$ and $\mathbf{l}'\mathbf{l} = 1$. The vector $\mathbf{l}$ can be obtained from the contrast $\mathbf{c}$ and from the design matrix $\mathbf{X}$, as explained in Appendix A. The construction of $\mathbf{l}$ assures that its transpose lies in the column space of $\mathbf{X}$.

Reduction of the dimensions of the data matrix

The matrix $\mathbf{U}$, following the distribution $N(\mathbf{XB}, \mathbf{I} \otimes \mathbf{\Sigma})$, must be transformed into a reduced data matrix that follows a left spherical distribution. This step is divided in two parts. First $\mathbf{U}$ is transformed into a new matrix $\mathbf{Y}$:

$$\mathbf{Y}_{n' \times p} = \begin{bmatrix} \mathbf{I}'_{1 \times n} \\ \mathbf{K}'_{s \times n} \end{bmatrix} \mathbf{U}_{n \times p},$$

where $\mathbf{K}\mathbf{K}'$ and $\mathbf{I}\mathbf{I}'$ respectively define the projectors $\mathbf{R}$ and $\mathbf{P}$ and $n' = 1 + s$. $\mathbf{Y}$ follows the distribution $N(\mathbf{0}, \mathbf{I} \otimes \mathbf{\Sigma})$ under the null hypothesis. The detailed construction of the matrix $\mathbf{Y}$ and the derivation of its distribution are presented in Appendix B.

From $\mathbf{Y}$ we can construct a reduced data matrix $\mathbf{Z}$, following a left spherical distribution. The reduced data matrix is obtained by the expression

$$\mathbf{Z}_{n' \times q} = \mathbf{Y}_{n' \times p} \mathbf{W}_{p \times q},$$

where $\mathbf{W}$ is an adequate random weight matrix and $q < df$. According to Theorem 1, for $\mathbf{Z}$ to follow a left spherical distribution it suffices that $\mathbf{W}$ is a function uniquely determined by $\mathbf{Y}'\mathbf{Y}$.

An example of a suitable weight matrix is presented further on. Other weight matrices can be found in the work of Lauter and coauthors [Lauter, 1996; Lauter et al., 1996, 1997].

Statistics for hypothesis testing on the reduced data matrix

Statistics for the reduced data matrix $\mathbf{Z}$ are constructed and shown to correspond to the usual hypothesis and error statistics.

Let $\mathbf{H}_Z$ and $\mathbf{G}_Z$ be defined as:

$$\mathbf{H}_Z = \mathbf{Z}'_1 \mathbf{Z}_1; \quad \mathbf{G}_Z = \mathbf{Z}'_2 \mathbf{Z}_2,$$

where $\mathbf{Z}_1$ is a vector containing the first row of $\mathbf{Z}$

$$\mathbf{Z}_1 = \mathbf{I}'_{1 \times n} \mathbf{U}_{n \times p} \mathbf{W}_{p \times q}$$

and $\mathbf{Z}_2$ a matrix containing all other rows of $\mathbf{Z}$

$$\mathbf{Z}_2 = \mathbf{K}'_{s \times n} \mathbf{U}_{n \times p} \mathbf{W}_{p \times q}.$$

It can be shown that, because $\mathbf{P} = \mathbf{I}\mathbf{I}'$ (as a consequence of the restrictions imposed on $\mathbf{I}$), $\mathbf{H}_Z = \mathbf{W}'\mathbf{H}\mathbf{W}$. It can also be easily demonstrated that $\mathbf{G}_Z = \mathbf{W}'\mathbf{G}\mathbf{W}$.

To see that $\mathbf{H}_Z$ and $\mathbf{G}_Z$ correspond to the usual hypothesis and error statistics let us assume a data set with such a dimensionality that it does not need to be reduced. Using the formulation introduced above we define:

$$\mathbf{H}_Y = \mathbf{Y}'_1 \mathbf{Y}_1 \quad \mathbf{G}_Y = \mathbf{Y}'_2 \mathbf{Y}_2,$$

where

$$\mathbf{Y}_1 = \mathbf{I}'_{1 \times n} \mathbf{U}_{n \times p} \quad \text{and} \quad \mathbf{Y}_2 = \mathbf{K}'_{s \times n} \mathbf{U}_{n \times p}.$$

The restrictions imposed on $\mathbf{I}$ assure that $\mathbf{H}_Y$ and $\mathbf{G}_Y$ are equal to the usual hypothesis and error statistics. Hence, the distributions of test statistics based on $\mathbf{H}_Y$ and $\mathbf{G}_Y$ have been derived and are known under the null hypothesis. It suffices that these tests are also invariant to define test statistics based on $\mathbf{H}_Z$ and $\mathbf{G}_Z$ with known distributions under the null hypothesis.

Invariant tests for the reduced data matrix

The Lawley-Hotelling trace test, the Wilk's Λ test, the Roy's maximum root test and the Pillai-Bartlett test are invariant over the class of left spherical distributions. These tests are all functions of the eigenvalues of $\mathbf{H}_Z \mathbf{G}_Z^{-1}$. Their invariance is a consequence of the following result.

Theorem 3

A test function of the eigenvalues of $\mathbf{H}_Z \mathbf{G}_Z^{-1}$ is invariant over the class of left spherical distributions.

Theorem 3 can be shown using Theorem 2 above. The demonstration is done in Appendix C. The invariance of these tests has also been shown in Lauter [1996] and in Lauter et al. [1996, 1997], in a more general framework.

The principal component test

We present a possible multivariate test for brain imaging data. In this test the new variables of the reduced data matrix are the principal components (PC) of $\mathbf{Y}'\mathbf{Y}$. The weights' matrix $\mathbf{W}$ has dimensions $p \times q$ where q is the number of PC we want to use in the reduced data matrix. q must obey $q < n - \text{rank}(\mathbf{X})$, where $\text{rank}(\mathbf{X})$ is the rank of the matrix $\mathbf{X}$. $\mathbf{W}$ is a solution of:

$$\mathbf{Y}'\mathbf{Y}\mathbf{W} = \mathbf{W}\mathbf{\Lambda},$$

where $\mathbf{\Lambda}$ is a diagonal matrix

$$\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_q \end{pmatrix},$$

$\lambda_1, \dots, \lambda_q$ are the first q eigenvalues of $\mathbf{Y}'\mathbf{Y}$ and $\mathbf{W}$ contains the corresponding eigenvectors.⁶

Any test statistic invariant over the class of left spherical distributions can be used, such as the Lawley-Hotelling trace test, the Wilk's Λ test, the Roy's maximum root test and the Pillai-Bartlett test already mentioned. There are tables and approximations for the distributions of these test statistics under the null hypothesis. We choose to use the Lawley-Hotelling trace test statistic because its exact distribution under the null hypothesis is known in the case when only one contrast is tested. The statistic is given by:

$$T^2 = \text{tr}[\mathbf{H}_Z \mathbf{G}_Z^{-1}](n - \text{rank}(\mathbf{X})),$$

where tr is the trace operator. The null hypothesis is rejected if T^2 is too large. The statistics follows a F distribution [Christensen, 1991]:

$$aT^2 \sim F(q, d)$$

where

$$d = 4 + \frac{q+2}{b-1}$$

$$b = \frac{(n-r-q)(n-r-1)}{(n-r-q-3)(n-r-q)}$$

$$a = \frac{1}{q} \left[\frac{d}{d-2} \right] \left[\frac{n-r-q-1}{n-r} \right]$$

and

$$r = \text{rank}(\mathbf{X}).$$

APPLICATION TO PET DATA

We use a PET data set to illustrate the method described for testing multivariate hypothesis. This data set has been used and described previously in Ledberg [2000]. We briefly explain the data acquisition and the experimental paradigm. We apply the princi-

pal component test to test the null hypothesis that the two different experimental conditions do not induce significant changes in the measured brain activity. The construction of the test assures that its distribution is the classical one, derived for multivariate normal data. Finally we use CVA to characterize the differences between the two conditions.

Experimental data

Six subjects participated in this study. Each subject was scanned in a PET scanner (SIEMENS-CTI ECAT EXACT HR) during 12 injections of 15-O-labeled butanol. The experimental paradigm consisted of two conditions each repeated six times. One scan for each condition had to be discarded for one of the subjects. The conditions are referred to as TEST and CONTROL. The TEST consisted of a two alternative forced choice tactile discrimination paradigm. The subjects discriminated which of two aluminum parallelepiped [Roland and Mortensen, 1987] was the most oblong. They made the discrimination using active palpation with their right hand. In the CONTROL the subjects were lying in the PET scanner with their eyes closed and were instructed not to do or think of anything in particular. All data was acquired with the PET scanner operating in 3D mode. The images were reconstructed with a ramp filter having a cutoff frequency of 0.5 to a voxel size of $2 \times 2 \times 3.125$ mm. The images were then normalized for global difference in uptake by division with the whole brain average value and filtered with a Gaussian filter of 12 mm FWHM in all dimensions. Information from individual MR scans was used to reformat all images into a standard anatomical space [Roland et al., 1994] with cubic voxels of $2 \times 2 \times 2$ mm.

To analyze the data the following model was used for each voxel:

$$U_{ijk} = \mu + t_i + s_j + e_{ijk},$$

where μ accounts for a mean effect, t_i accounts for the effect of the two tasks ($i = 1, 2$), s_j accounts for the effect of the six different subjects ($j = 1, \dots, 6$) and e_{ijk} is an error term ($k = 1, \dots, 6$ for all subjects i except one where $k = 1, \dots, 5$). The corresponding design matrix $\mathbf{X}$ is shown in Figure 1.

Testing hypothesis

We want to test the null hypothesis that the two different experimental conditions do not induce sig-

⁶ $\mathbf{W}$ is usually obtained using singular value decomposition, given the high dimensionality of the matrix $\mathbf{Y}'\mathbf{Y}$.

Figure 1.

The design matrix. Black represents one and white zero. The first column accounts for a mean effect. The second and third columns represent the CONTROL and TEST conditions. The six remaining columns correspond to the six subjects.

nificant differences in the measured brain activity. For this purpose we use the contrast:

$$\mathbf{c}' = \gamma[0 \quad 1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0],$$

where γ is a constant introduced due to the normalization of the vector $\mathbf{l}$.

The test is constructed according to the steps indicated in the theory section. First we transform the data matrix $\mathbf{U}$ into the matrix $\mathbf{Y}$ as explained in Appendix B. The vector $\mathbf{l}'$ can be computed from $\mathbf{c}'$ according to Appendix A. The vector $\mathbf{l}$ follows the assumptions specified in the theory section. From $\mathbf{Y}$ we construct a reduced data matrix with the largest principal components. The number of principal components used was the largest integer less than half of the total number of components, 32. To test our hypothesis we use the principal components test. The value of aT^2 is 48.9562 and the degrees of freedom are $q = 32$ and $d = 38$. Therefore, the null hypothesis of no significant differences between the two tasks is rejected ($p \approx 0$).

Characterization of the main effect and comparison

Having rejected the null hypothesis, we want to characterize the differences between CONTROL and TEST. We aim at determining which combinations of variables (PCs) are responsible for the rejection of the null hypothesis and for the differences between the two conditions. To achieve this goal we can use a

technique known as canonical variate analysis [Chatfield and Collins, 1980]. This technique, in conjunction with multivariate linear model analysis, has previously been applied to brain imaging data [Friston et al., 1995, 1996]. The linear combination of variables that accounts for most inconsistency between the null hypothesis and the data is the vector ($\mathbf{a}_1$) that maximizes the ratio:

$$\frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{G}\mathbf{a}}.$$

The vector $\mathbf{a}_1$ is called the first canonical vector and $\mathbf{Z}^1 = \mathbf{Z}\mathbf{a}_1$ is the first canonical variate. Higher order canonical variates can be defined by maximizing again the same ratio with the restriction that the covariance between the new canonical variate and the previously defined is zero. In Figure 2 on the right hand columns of both sets of images we show the first canonical variate of our example. The image is thresholded at zero for visualization.

We analyze, for comparison, the same data set using univariate regression analysis. The method in Ledberg [2000] is used to make statistical inference based on the size of activations. The results are shown in Figure 2 on the left hand side columns of both sets of images. The clusters shown have size above 1,105 voxels. This corresponds to an omnibus p value of $p < 0.05$ and a t -threshold of 2.388.

Figure 2 shows a great concordance between the patterns of activation obtained using univariate analysis and the first canonical variate thresholded at zero. This means that the regions that are mainly responsible for the rejection of the multivariate null hypothesis are in agreement with the regions found to be significantly active using a univariate approach. A detailed quantitative comparison between the two patterns of activation is not meaningful because the first canonical variate is arbitrarily thresholded at zero.

CONCLUSION

This study demonstrates a general approach to test multivariate hypothesis on PET data using multivariate linear model analysis. Because the number of variables is larger than the number of observations, the tests are performed on a reduced data set obtained by linear combinations of the original variables. The potentialities of applying multivariate linear model analysis to brain imaging data have been described in previous works [Friston et al., 1995, 1996; Worsley, 1997; Worsley et al., 1997]. The aim of this study is to

Figure 2.

The functional results are superimposed on MRI anatomical images. The first and third columns show the results of the univariate analysis. The clusters considered have size above 1,105, corresponding to an omnibus p value of $p < 0.05$ and a t -threshold of 2.388. The second and fourth columns show the first canonical

variate thresholded at zero. The arbitrary color scale is shown at the lowest right hand side of the figure. For both cases horizontal sections separated by 10 mm are shown. The first sections on the left have $z = 65$ mm and the last on the right have $z = -45$ mm.

address a question concerning the usage of multivariate linear model analysis that has not been considered in the literature before, namely the non-normality of the reduced data matrix. In previous works [Friston et al., 1995, 1996], no foundation justifying the claimed distributions of the test statistics under the null hypothesis is given. In this study, we present a way of constructing multivariate tests for brain imaging data, such that their exact distributions under the null hypothesis are known.

As an illustration, an application of the theory exposed to a PET data set is presented. The reduced data matrix is constructed using PCA. The test rejects the null hypothesis of two different experimental tasks not inducing differences in brain activity. The combination of variables of the reduced data set that contributes more to the rejection of the null hypothesis is computed by means of CVA. The obtained pattern of brain activity is compared with the results obtained using a univariate linear model for all voxels. The results from the two different tests are very similar. This is not surprising because we investigate similar questions with both methods. PCA and CVA have been used previously to characterize brain imaging data in the same context as in this study [Friston et al., 1995, 1996]. The novelty of this study is the theoretical-founded construction of a test with known exact levels of significance, as explained in the paragraph above.

The difficulty of applying this method to test multivariate hypothesis on fMRI data is the temporal dependency between the observations. If we view the data as four dimensional (three spatial dimensions plus one temporal) [Ledberg et al., 2001] the temporal correlations are accounted for and this technique can directly be applied. In this case, one observation comprises all the voxels and all the scans within a session. Different sessions correspond to different observations. We can assume that consecutive observations are independent provided that the time interval between sessions is large enough. This approach allows in principle to use the same general framework to test multivariate hypothesis on fMRI data.

APPENDIX A

In this section it is explained how to construct the vector $\mathbf{l}$. The vector $\mathbf{l}$ is used in the construction of the matrix $\mathbf{Y}$ from the original data matrix $\mathbf{U}$ (see Appendix B). Let the matrix $\mathbf{V}$ be defined by $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$. Multiplying the equation defining the contrast with $\mathbf{V}$ the following equation is obtained:

$$\mathbf{l}'\mathbf{XV} = \mathbf{c}'\mathbf{V}.$$

Because $\mathbf{XV} = \mathbf{X}$ [Searle, 1997] then

$$\mathbf{l}'\mathbf{X} = \mathbf{c}'\mathbf{V}.$$

Replacing $\mathbf{V}$ by its definition:

$$\mathbf{l}'\mathbf{X} = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}.$$

Dividing by a constant for normalization, $\mathbf{l}$ is given by the expression:

$$\mathbf{l}' = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' / \sqrt{(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'}.$$

The role of the normalizing constant is made clear in the main text and in Appendix B.

To see that $\mathbf{l}'$ lies in the column space of $\mathbf{X}$ it suffices to show that:

$$\mathbf{M}\mathbf{l} = \mathbf{l}$$

or equivalently that

$$\mathbf{l}'\mathbf{M} = \mathbf{l}'.$$

The equality is easily proven by using the definitions of $\mathbf{l}'$ and $\mathbf{M}$. In the demonstration below the following property of the generalized inverse is used [Mardia et al., 1979] (Appendix A):

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}.$$

It finally follows that:

$$\mathbf{l}'\mathbf{M} = \xi \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \xi \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{l}',$$

where ξ denotes the normalization constant.

APPENDIX B

In this section it is shown how to construct a matrix $\mathbf{Y}$ with distribution $N(\mathbf{0}, \mathbf{I} \otimes \Sigma)$ under the null hypothesis, from the original data matrix $\mathbf{U}$ distributed as $N(\mathbf{XB}, \mathbf{I} \otimes \Sigma)$. Note that the matrix $\mathbf{R}$

$$\mathbf{R} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

defines a perpendicular projection onto the space complementary to the one spanned by the columns of $\mathbf{X}$. This implies that there is a matrix $\mathbf{K}_{n \times s}$ such that

$$\mathbf{K}\mathbf{K}' = \mathbf{R}_{n \times n}$$

and

$$\mathbf{K}'\mathbf{K} = \mathbf{I}_s,$$

where $s = n - \text{rank}(\mathbf{X})$. By the definition of $\mathbf{R}$: $\mathbf{R}\mathbf{X} = \mathbf{0}$.
From the equations

$$\mathbf{R}\mathbf{X} = \mathbf{0}$$

and

$$\mathbf{K}'\mathbf{K} = \mathbf{I}_s,$$

it is easy to see that:

$$\mathbf{K}\mathbf{K}'\mathbf{X} = \mathbf{0}$$

and so

$$\mathbf{K}'\mathbf{X} = \mathbf{0}.$$

Now consider a matrix defined as:

$$\mathbf{F}_{(1+s) \times n} = \begin{bmatrix} \mathbf{1}'_{1 \times n} \\ \mathbf{K}'_{s \times n} \end{bmatrix}.$$

The mean of $\mathbf{Y} = \mathbf{F}\mathbf{U}$ is given by:

$$E[\mathbf{Y}] = \mathbf{F}\mathbf{X}\mathbf{B} = \begin{bmatrix} \mathbf{c}'\mathbf{B} \\ \mathbf{0} \end{bmatrix},$$

equal to $\mathbf{0}$ under the null hypothesis. Its variance-covariance structure under the null hypothesis is:

$$\text{vec}[(\mathbf{F}\mathbf{U})']\text{vec}[(\mathbf{F}\mathbf{U})']'.$$

$\text{vec}(\mathbf{U}'\mathbf{F}')$ can be written as $\text{vec}(\mathbf{I}\mathbf{U}'\mathbf{F}')$. From a known property of the Kronecker product [Mardia et al., 1979] (Appendix A):

$$\text{vec}(\mathbf{I}\mathbf{U}'\mathbf{F}') = (\mathbf{F} \otimes \mathbf{I})\text{vec}(\mathbf{U}').$$

Using this expression the variance-covariance structure is:

$$(\mathbf{F} \otimes \mathbf{I})\text{vec}(\mathbf{U}')\text{vec}(\mathbf{U}')'(\mathbf{F}' \otimes \mathbf{I}).$$

Because the variance-covariance structure of $\mathbf{U}$ is equal to $\mathbf{I} \otimes \mathbf{\Sigma}$ the expression above can be written as:

$$(\mathbf{F} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{\Sigma})(\mathbf{F}' \otimes \mathbf{I})$$

Using the properties of the Kronecker product [Mardia et al., 1979] (Appendix A) the equation above is equal to:

$$\mathbf{F}\mathbf{F}' \otimes \mathbf{\Sigma} = \begin{bmatrix} \mathbf{I}' \\ \mathbf{K}' \end{bmatrix} [\mathbf{I} \quad \mathbf{K}] \otimes \mathbf{\Sigma} = \mathbf{I} \otimes \mathbf{\Sigma},$$

once $\mathbf{K}$ and $\mathbf{I}$ are orthogonal, and both $\mathbf{K}\mathbf{K}'$ and $\mathbf{I}\mathbf{I}'$ define projectors. So, under the null hypothesis $\mathbf{Y}$ has the desired distribution:

$$\mathbf{Y}_{n' \times p} = (\mathbf{F}\mathbf{U})_{n' \times p} \sim N(\mathbf{0}, \mathbf{I}_{n'} \otimes \mathbf{\Sigma}),$$

with $n' = 1 + s$.

APPENDIX C

In this section we demonstrate Theorem 3. Theorem 3 states that a test $t(\mathbf{Z})$ function of the eigenvalues of $\mathbf{H}_Z\mathbf{G}_Z^{-1}$, written $t(\mathbf{Z}) = t(\text{eig}(\mathbf{H}_Z\mathbf{G}_Z^{-1})) = t(\text{eig}((\mathbf{Z}'_1\mathbf{Z}_1)(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}))$, is invariant over the class of left spherical distributions. To see this let $\mathbf{A}$ be any $q \times q$ nonsingular matrix. By a known property of the eigenvalues [Mardia et al., 1979] (Appendix A):

$$\text{eig}((\mathbf{Z}'_1\mathbf{Z}_1)(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}) = \text{eig}(\mathbf{A}'(\mathbf{Z}'_1\mathbf{Z}_1)(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{A}^{-1}).$$

Then we have:

$$\begin{aligned} t(\mathbf{Z}) &= t(\text{eig}(\mathbf{A}'(\mathbf{Z}'_1\mathbf{Z}_1)(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{A}^{-1})) \\ &= t(\text{eig}(\mathbf{A}'(\mathbf{Z}'_1\mathbf{Z}_1)\mathbf{A}\mathbf{A}^{-1}(\mathbf{Z}'_2\mathbf{Z}_2)^{-1}\mathbf{A}^{-1})). \end{aligned}$$

The right hand side of the equality can be written as

$$t(\text{eig}((\mathbf{A}'\mathbf{Z}'_1\mathbf{Z}_1\mathbf{A})(\mathbf{A}'\mathbf{Z}'_2\mathbf{Z}_2\mathbf{A})^{-1})),$$

which is equal to $t(\mathbf{Z}\mathbf{A})$. It follows that $t(\mathbf{Z}\mathbf{A}) = t(\mathbf{Z})$ and hence that the tests are invariant, according to Theorem 2.

REFERENCES

- Chatfield C, Collins A (1980): Introduction to multivariate analysis. London: Chapman and Hall. p 153–159.
- Christensen R (1991): Linear models for multivariate, time series and spatial data. New York: Springer-Verlag. p 1–68.
- Fang KT (1999): Multivariate analysis in large number of variables. Course Notes.
- Fang KT, Chen HF. (1989): Relationships among classes of spherical distributions. In: Fang KT, Anderson TW, editors. Statistical inference in elliptically contoured and related distributions. New York: Allerton Press. p 25–36.

- Fang KT, Zhang YT. (1990): Generalized multivariate analysis. New York: Springer-Verlag.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RS. (1991): Comparing functional (PET) images: the assessment of significant change. *Cereb Blood Flow Metab* 11:690–699.
- Friston KJ, Frith CD, Frackowiak RS, Turner R. (1995): Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage* 2:166–172.
- Friston KJ, Poline JB, Holmes AP, Frith CD, Frackowiak RSJ. (1996): A multivariate analysis of PET activation studies. *Hum Brain Mapp* 4:140–151.
- Lange N, Strother SC, Anderson JR, Nielsen FA, Holmes AP, Koldenda T, Savoy R, Hansen LK. (1999): Plurality and resemblance in fMRI data analysis. *Neuroimage* 10:228–303.
- Läuter J. (1996): Exact t - and F -tests for analysing studies with multiple endpoints. *Biometrics* 52:964–970.
- Läuter J, Glimm E, Kropf S. (1996): New multivariate tests for data with an inherent structure. *Biom Z* 38:5–23.
- Läuter J, Glimm E, Kropf S. (1997): Multivariate tests based on left-spherically distributed linear scores. *Proceedings of the conference, "Contemporary multivariate analysis and its applications,"* May 19–22, 1997, Hong Kong. pC34–C44.
- Ledberg A. (2000): Robust estimation of the probabilities of 3D clusters in functional brain images: application to PET data. *Hum Brain Mapp* 9:143–155.
- Ledberg A, Fransson P, Larsson J, Petersson KM. (2001): A 4D approach to the analysis of functional brain images: application to fMRI data. *Hum Brain Mapp* 13:185–198.
- Mardia KV, Kent JTB, Bibby JM. (1979): *Multivariate analysis*. London: Academic Press. p 59–184, appendix A.
- McIntosh AR, Bookstein FL, Haxby JV, Grady CL. (1996): Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* 3:143–157.
- McIntosh AR, Lobaugh NJ, Cabeza FLB, Houle S. (1998): Convergence of neural systems processing stimulus associations and coordinating motor responses. *Cereb Cortex* 8:648–659.
- Petersson KM, Nichols TE, Poline JB, Holmes AP. (1999a): Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Philos Trans R Soc Lond B Biol Sci* 354:1239–1260.
- Petersson KM, Nichols TE, Poline JB, Holmes AP. (1999b): Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference. *Philos Trans R Soc Lond B Biol Sci* 354:1261–1281.
- Roland PE, Mortensen E. (1987): Somatosensory detection of microgeometry, macrogeometry and kinesthesia in man. *Brain Res* 434:1–42.
- Roland PE, Graufelds CJ, Wählin J, Ingelman L, Andersson M, Ledberg A, Pedersen J, Åkerman S, Dabringhaus A, Zilles K. (1994): Human brain atlas: for high-resolution functional and anatomical mapping. *Hum Brain Mapp* 1:173–184.
- Searle SR. (1997): *Linear models*. New York: John Wiley & Sons. p 1–30.
- Worsley KJ. (1997). An overview and some new developments in the statistical analysis of PET and fMRI data. *Hum Brain Mapp* 5:254–258.
- Worsley KJ, Evans AC, Marrett S, Neelin P. (1992): A 3D statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab* 12:900–918.
- Worsley KJ, Poline JB, Friston KJ, Evans AC. (1997): Characterizing the response of PET and fMRI data using multivariate linear models. *Neuroimage* 6:305–319.
- Zarahn E, Aguirre GK, D'Esposito M. (1997): Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage* 5:179–197.