fMRI Analysis 101 - Univariate Analysis[1]
Douglas N. Greve

`$Id: univar-analysis.tex,v 1.3 2005/05/04 17:00:49 greve Exp $`

1   Introduction

This document gives a terse description of how to perform linear univariate analysis of fMRI data (though the techniques generalize to other modalities). The name *univariate* is somewhat of a misnomer that has been adopted in fMRI. In this context, univariate refers to all the processing done on the time course from a *single* voxel, and *multivariate* refers to analysis done on multiple voxels (eg, Gaussian Random Fields). In this discussion, the *observable* is the raw data after any preprocessing (eg, slice-timing correction, motion correction, spatial smoothing). The techniques described are all *linear*[2] in that the observable is modeled as a linear combination of regressors. The *design matrix* is the matrix whose columns are the regressors and contains information about the stimulus *schedule*, i.e., which stimulus was presented when. It may also have *nuisance* regressors (e.g., linear drift).

2   Forward Model

$$y = X\beta + n, n : N(0, \sigma_n^2 \Sigma) \tag{1}$$

y is the $N_t \times 1$ observable (1 = "univariate")
X is the $N_t \times N_\beta$ design matrix (stimulus convolution matrix)
β is the $N_\beta \times 1$ vector of regression coefficients (1 = "univariate")
σ_n^2 is the noise variance
Σ is $N_t \times N_t$ noise covariance matrix; identity for white noise.
N_t is the number of time points
N_β is the number of regression coefficients

The forward model represents a set of N_t linear equations with N_β unknowns (the βs).

3   Ordinary Least Squares (OLS) Solution

Assume white noise ($\Sigma = I$).

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{2}$$

$(X^T X)^{-1} X^T$ is the pseudo-inverse of X (aka X^+)

Signal Estimate:

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y \tag{3}$$

Residual Error:

$$r = y - \hat{y} = y - X\hat{\beta} = (I - X(X^T X)^{-1} X^T)y = Ry \tag{4}$$

R is the *residual error forming matrix*. Idempotent: $RR = R$.

Degrees of Freedom:

$$\nu = DOF = trace(R) = N_t - N_\beta \tag{5}$$

Residual Error Variance:

$$\sigma_r^2 = \frac{r^t r}{DOF} \approx \sigma_n^2 \tag{6}$$

---

[1] This document can be obtained from http://surfer.nmr.mgh.harvard.edu/fsfast/univar-analysis.

[2] This is sometimes referred to as a *General Linear Model*, or GLM. This is a slight exaggeration because the GLM allows for several types of noise distributions, whereas, in the fMRI literature, almost exclusively assumes Gaussian.

Covariance of $\hat{\beta}$ under Null Hypothesis ($\beta = 0$):

$$Cov(\hat{\beta}) = E(\hat{\beta}\hat{\beta}^T) = \sigma_r^2(X^TX)^{-1}X^T\Sigma X(X^TX)^{-1} \tag{7}$$

Under the assumption of white noise ($\Sigma = I$)

$$Cov(\hat{\beta}) = \sigma_r^2(X^TX)^{-1} \tag{8}$$

Contrast Effect Size:

$$\gamma = C\hat{\beta} \tag{9}$$

C is the $J \times N_\beta$ *Contrast Matrix*.

t-Test ($J = 1$):

$$t_{DOF} = \frac{C\hat{\beta}}{Cov(C\hat{\beta})} \tag{10}$$

F-Test:

$$F_{J,DOF} = \frac{(C\hat{\beta})^T(Cov(C\hat{\beta}))^{-1}(C\hat{\beta})}{J} \tag{11}$$

where

$$Cov(C\hat{\beta}) = CCov(\hat{\beta})C^T \tag{12}$$

Under the assumption of white noise ($\Sigma = I$), this becomes

$$Cov(C\hat{\beta}) = \sigma_r^2 C(X^TX)^{-1}C^T \tag{13}$$

Note: if a white noise model is assumed but the noise is not really white, then the statistics are said to be "invalid", meaning that the false positive rate (FPR) is not ideal. In fMRI, the FPR usually becomes "liberal" (ie, there are too many false positives).

4   Generalized Least Squares (GLS) Solution

Non-white noise and temporal filtering.

Forward Model with temporal filtering of the residuals:

$$Fy = FX\beta + Fn, n : N(0, \sigma_n^2\Sigma) \tag{14}$$

F is the $N_t \times N_t$ filter matrix. Note: the equations below will simplify a lot when $F = \Sigma^{-\frac{1}{2}}$, a special case which is discussed below.

New solution:

$$\hat{\beta} = (X^TF^TFX)^{-1}X^TF^TFy \tag{15}$$

if $\tilde{X} = FX$, $\tilde{y} = Fy$, then

$$\hat{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{y} \tag{16}$$

Note: $\hat{\beta}$ is still unbiased. Proof: solving the above equation with $n = 0$ shows that $\hat{\beta} = \beta$.

Residual Error forming Matrix:

$$R = I - FX(X^TF^TFX)^{-1}X^TF^TF \tag{17}$$

where the residual error is still $r = Ry$.

Degrees of Freedom (Statistics):

$$\nu = DOF = \frac{(trace(RF\Sigma F^T))^2}{trace((RF\Sigma F^T)^2)} \tag{18}$$

Degrees of Freedom (Variance):

$$\eta = trace(RF\Sigma F^T) \tag{19}$$

2

Note: this parameter has no name in the literature.

Residual Error Variance:

$$\sigma_r^2 = \frac{r^t r}{\eta} \tag{20}$$

Covariance of $\hat{\beta}$ under Null Hypothesis ($\beta = 0$):

$$Cov(\hat{\beta}) = \sigma_r^2 (X^T F^T F X)^{-1} X^T F^T \Sigma F X (X^T F^T F X)^{-1} \tag{21}$$

Problem: Σ is unknown.

5 How to set the Filter Matrix F? What about Σ?

5.1 SPM

Make $F >> \Sigma^{\frac{1}{2}}$, make the smoothing induced by the temporal filter much more than the inherent smoothness of the noise, in which case:

$$F\Sigma F^T \approx FF^T \tag{22}$$

One only needs some vague knowledge about Σ, enough to be able to produce an F that will overwhelm it. This approximation allows $Cov(\hat{\beta})$ to be computed in a way that sort of takes the inherent noise covariance into account and so makes the false-positive rate closer to the ideal. This is inefficient, but not bad for blocked designs. For event-related designs, this can cause a massive loss in efficiency. It will also cause the DOF (Equation **??**) to be less than $N_t - N_\beta$.

5.2 Whitening

Setting $F = \Sigma^{-\frac{1}{2}}$ is referred to as *whitening the residuals*. Under this condition, $F\Sigma F^T = I$. This results in a *fully efficient estimator* (ie, $DOF = N_t - N_\beta$). Note that there are an infinite number of F's that meet the above condition. Typically F is obtained from the Cholesky or singular value decompositions.

5.3 Obtaining Σ

In order to whiten, we need an estimate of Σ. To do this, we exploit the fact that the covariance matrix of a time-invariant Gaussian process is a Toeplitz matrix of the *autocorrelation function*, or ACF. This reduces estimating Σ to estimating the ACF of the noise. This is done by estimating the ACF of the residuals. The ACF itself is often parameterized (eg, AR(1)). The ACF is often very noisy, so it's usually a good idea to apply some spatial regularization. Also, the residual ACF will be biased with respect to the true ACF, though there are ways to compensate for this (Worsley, et al, 2002).

6 Efficiency and Design Optimization

Goal: choose an event schedule (ie, an X) that will minimize the expected sum-square error (SSE) of the estimates ($\hat{\beta}$).

The error in $\hat{\beta}$ is given by:

$$\hat{\beta}_e = \beta - \hat{\beta} = (X^T X)^{-1} X^T n \tag{23}$$

under the white noise assumption. Note: this is different than the residual error.

The expected SSE is then

$$E\{SSE\} = E\{\hat{\beta}_e^T \hat{\beta}_e\} = E\{trace(\hat{\beta}_e \hat{\beta}_e^T)\} = \sigma_n^2 trace((X^T X)^{-1}) \tag{24}$$

again under the white noise assumption. The SSE can be minimized by minimizing $trace((X^T X)^{-1})$. The *efficiency* is defined as:

$$eff = \frac{1}{trace((X^T X)^{-1})} \tag{25}$$

Algorithm: randomly select a stimulus schedule, compute X, compute the efficiency, do this a million times, keep the schedules that give the largest efficiencies.

7 Exploring Signal Models

7.1 The Forward Model Revisited

The forward equation is

$$y = X\beta + n, n : N(0, \sigma_n^2 \Sigma) \tag{26}$$

Each column of X is a regressor, and each component of β is a regression coefficient. The interpretation of each β depends upon how its corresponding regressor was constructed. All the assumptions about the signal are embedded in X. X may actually be made up of many different explanatory variables. For example, several event-related stimuli, periodic stimuli, polynomial drift components, physiological regressors, etc.

7.2 Event-Related Models (includes Block too)

7.2.1 Finite-Impulse Response (FIR)

The event-related model (ERM) is best understood by first considering a Finite-Impulse Response (FIR) model. The FIR makes no assumptions about the shape of the hemodynamic response. In this model, a *time window* is hypothesized to exist around and event. The response to the event is zero outside the window but can be anything in the window (discretized to some sampling time). In the X for the FIR, the number of rows equals N_t (this is always the case) and the number of columns equals the number of samples in the window (N_w). This causes n^{th} β to be interpreted as the value of the response at the nth sample in the window. Example: for a short event, the response may last for 20 sec, so one could start the time window at $t = 0$ (ie, stimulus onset) and end it at $t = 20$ sampling at the TR (eg, TR=2). In this case, $N_w = 10$. Note that if the duration of the event is non-trivial, just extend the time window to be long enough to capture the entire response. A blocked design is just an event-related design with long events.

The FIR X is built in the following way. The components are either 0 or 1. The fist column is built by putting a 1 everywhere a stimulus appears and a 0 everywhere else. The next column is the same as the first but shifted down by one, filling the first row with a 0. Each additional column constructed in the same way.

7.2.2 Modeling an Assumed Response Shape

Often one wants to make assumptions about the shape of the hemodynamic response. Such assumptions can greatly increase the statistical power, though there are some risks. In general, the X matrix can be computed as:

$$X = X_{fir} A \tag{27}$$

where X_{fir} is the FIR design matrix as described above, and A is the matrix that holds the assumptions about the hemodynamic response sampled at the points in the time window. For example, A could be a single column representing a gamma function. In this case there would be only one β and it would represent the amplitude of the best-fit gamma function. One could make a second column representing the derivative of the gamma function (this is what SPM usually does).

7.2.3 Benefits to Assuming a Shape

Assuming a shape will usually increase the statistical power, sometimes dramatically so. To see why, consider assuming the shape using a simple gamma function. This model only has one free parameter: the amplitude. The equivalent FIR model may have 10 free parameters. Thus, for the gamma model, each event gives us an opportunity to observe that one parameter 10 times (approximately) whereas in the FIR, each parameter is only observed once. One roughly expects the variance of an estimated parameter to be reduced by a factor equal to the number of observations, so, all other things being equal, one expects the gamma parameter to

have one tenth the variance of the FIR parameters. Note: this is why there is a penalty in the FIR model for extending the time window.

7.2.4 Risks to Assuming a Shape

There are three main risks to assuming a shape to the hemodynamic response; all have to due with mis-characterizing the shape. First, if the true response is in the null space of A, then β will be zero, even in the case where there is no noise. In this case, all statistical tests will report that there is no effect even though the effect can be huge. The closer to the null space of A the true response is, the lower the final significance will be. Second, mis-specifying the shape will cause the residual error to be larger; but this is usually a small effect. The third risk is more difficult to explain. If the assumed shape is wrong, then it cannot explain all the signal associated with an event type. If the event schedules are not thoroughly randomized, then the regressors from a second event type can soak up some of the signal left by the first. This can make the effect of the second event type appear to be larger than it is. This can cause a false positive when compared to baseline or a false negative when compared to the first stimulus. It is not known how large this effect is.

7.2.5 Multiple Event-Types

If there are multiple event types, then build the $X = X_{fir}A$ matrix for each one and horizontally concatenate them together.

8 Nuisance Variables

A *nuisance* variable is generally considered to be any systematic effect that is not part of the experimental paradigm. For example, fMRI time courses always have a mean offset. Others include motion artifacts and drift due to scanner heating. These effects can be modeled by adding regressors to the design matrix. Ideally, the nuisance regressors span the noise, thereby reducing the residual variance and improving the statistical power. However, it is possible (and likely) that the nuisance regressor will not be orthogonal to the task-related components. When this happens, the task effect size may drop significantly when the nuisance regressor is added. It is not possible to say which task effect size is "correct".

8.1 Polynomial Drift Models

The BOLD signal always has a mean offset that must be taken into account when estimating the task-related effects. There can also be slow increases/decreases due to amplifiers heating up, etc. These effects can be accounted for in several ways. Here, we show how to include polynomial regressors. The X matrix will have N_t rows (always) and a number of columns equal to the $order + 1$ of the polynomial, where $order = 0$ is a constant (models the mean offset). The first column will be all ones, the next will be linear from 1 to N_t, the next will be quadratic, etc. Note that this will be different than simply subtracting the mean from the raw time course unless you also subtract the mean from the (non-polynomial) design matrix. Once you have the X for the polynomial regressors, horizontally concatenate it with the design matrices from other explanatory components.

8.2 Motion Correction Regressors

When a functional volume is motion corrected, the program performing the correction will output the six "motion correction parameters" for each volume in the series. These parameters are the translation (3) and rotation at each time point. Together, they make an $N_t \times 6$ matrix. This matrix can be included as nuisance variables (though it is best to do data reduction by taking the three largest singular vectors). Given that the volume has been motion corrected, why would you want to include motion correction regressors? The motion correction algorithm cannot remove all the effects of motion. For example, it cannot remove the *spin history* effect, which is when tissue not in the original slice plane gets moved into the slice plane. This tissue was not excited by the RF pulse and so will have no signal. Note: using motion correction regressors can be tricky because it is often the canse that the subject moves in time to the stimulus, which causes the motion correction regressors to span the space of the task-related signal. This can cause a drop in the task-related effect size. On the other hand, if there is stimulus-locked motion, one does not know

whether an effect is due to motion or to stimulus. This is much more likely to happen during a blocked design.

References:

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SCR, Frackowiak RSJ, Turner R. Analysis of fMRI Time Series Revisited. Neuroimage 2:45-53 (1995).

Worsley KJ, Friston KJ. Analysis of fMRI Time-Series Revisited - Again Neuroimage 2:173-181 (1995) .

Dale, A.M. Optimal experimental design for event-related fMRI. Human Brain Mapping 8:109-114 (1999).

Burock, M.A. and Dale, A.M. Estimation and Detection of Event-Related fMRI Signals with Temporally Correlated Noise: A Statistically Efficient Approach. Human Brain Mapping, 11:249-260 (2000).

Worsley, et al. A general statistical analysis for fMRI Data. NeuroImage 15, 1-15 (2002).