The general linear model and Statistical Parametric Mapping II: GLM for fMRI

Alexa Morcom
Edinburgh SPM course, April 2010

Centre for Cognitive & Neural Systems/
Department of Psychology
University of Edinburgh
Overview

• Introduction
  – Two GLMs in 2-stage procedure

• General linear model(s) for fMRI
  – Low frequency noise
  – Haemodynamic response
  – Temporal basis functions
  – Time series
Modelling fMRI data

Why?
Make inferences about effects of interest

How?
1. Decompose data into effects and error
2. Form statistic using estimates of effects and error

Model?
Use any available knowledge

data -> model -> effects estimate -> error estimate -> statistic

- "Modelling fMRI data"
Overview of SPM

- Preprocessing
  - Image time-series
  - Preprocessing

- Design matrix
  - General linear model
  - Variance components

- Contrasts
  - SPMs
  - Thresholding

- Template
  - Kernel
Modelling in SPM

Smoothed normalised data – i.e., single voxel timeseries

Design matrix

General linear model

Parameter estimates

Variance components
2-stage GLM

Single subject
Each has an independently acquired set of data
These are modelled separately
Models account for \textit{within subjects variability}
Parameter estimates apply to individual subjects

Group/s of subjects
To make population inferences, 2\textsuperscript{nd} level models
account for \textit{between subjects variability}
Parameter estimates apply to group effect/s

Single subject \textbf{contrasts of parameter estimates} taken
forward to 2\textsuperscript{nd} level as (spm\_con*.img) \textit{‘con images’}

Statistics compare \textbf{contrasts of 2\textsuperscript{nd} level parameter estimates} to 2\textsuperscript{nd} level error

‘Summary statistic’ random effects method
2-stage GLM

Each has an independently acquired set of data. These are modelled separately. Models account for *within subjects variability*. Parameter estimates apply to individual subjects.
Mass-univariate analysis: voxel-wise GLM

The design matrix embodies all available knowledge about experimentally controlled factors and potential confounds.

\[ y = X\beta + e \]

\[ e \sim N(0, \sigma^2 I) \]

Model is specified by
1. Design matrix \( X \)
2. Assumptions about \( e \)

\( N \): number of scans
\( p \): number of regressors
fMRI example

One session
Passive word listening versus rest
7 cycles of rest and listening
Each epoch 6 scans with 7 sec TR

Question: Is there a change in the BOLD response between listening and rest?
The GLM applied to fMRI

What are the problems?

1. The BOLD signal includes substantial amounts of low-frequency noise.

2. BOLD responses have a delayed and dispersed form (‘sluggish’)

3. The data are a timeseries, so are serially correlated (temporally autocorrelated; for TR < ~8s)

→ Therefore they are not independent observations - violates the assumptions of the GLM’s noise model
The GLM applied to fMRI

What are the solutions?

1. The data can be filtered to remove low-frequency (1/f) noise

2. Effects are convolved with haemodynamic (BOLD) response function (HRF), to capture the sluggish nature of the response

3. The data are modelled as a timeseries, taking account of temporal autocorrelation
The GLM applied to fMRI

What are the solutions?

1. **The data can be filtered to remove low-frequency (1/f) noise**

2. Effects are convolved with haemodynamic (BOLD) response function (HRF), to capture the sluggish nature of the response

3. The data are modelled as a timeseries, taking account of temporal autocorrelation
1. Low frequency noise

**Physical** (scanner drifts)

**Physiological** (aliased)
- cardiac (~1 Hz)
- respiratory (~0.25 Hz)
Regression model

\[ Y = X_1 \hat{\beta} + \hat{\epsilon} \]

Number of scans

No. of effects in model

Single subject

[Diagram showing the relationship between the number of scans, the number of effects in the model, and the regression equation.]
Add high pass filter

This means ‘taking out’ fluctuations below the specified frequency
SPM implements by fitting low frequency fluctuations as effects of no interest

Single subject

Frequency domain
128 second High-pass filter

relative spectral density

Discrete Cosine Transform (DCT) set
High pass filtering: example

blue = data
black = mean + low-frequency drift
green = predicted response, taking into account low-frequency drift
red = predicted response, NOT taking into account low-frequency drift
Fitted & adjusted data

Raw fMRI timeseries
Fitted & adjusted data

Raw fMRI timeseries

High-pass filtered (and scaled)

Fitted high-pass filter
Fitted & adjusted data

Raw fMRI timeseries

Adjusted data

high-pass filtered (and scaled)

fitted high-pass filter

fitted box-car
Fitted & adjusted data

Raw fMRI timeseries

Adjusted data

high-pass filtered (and scaled)

fitted high-pass filter

Residuals

fitted box-car
Regression model

\[ Y = X_1 \ast \hat{\beta}_1 + \hat{\epsilon}_1 \]
Regression model

\[ Y = X_1 \ast \hat{\beta} + \hat{e} \]

What's wrong with this model?

1. Stimulus function is not expected BOLD response
2. Data are serially correlated
The GLM applied to fMRI

What are the solutions?

1. The data can be filtered to remove low-frequency (1/f) noise

2. Effects are convolved with haemodynamic (BOLD) response function (HRF), to capture the sluggish nature of the response

3. The data are modelled as a timeseries, taking account of temporal autocorrelation
2. The BOLD Haemodynamic Response

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s post stimulus; baseline after 20-30s
- May observe an initial undershoot
- Similar across V1, A1, S1
- May differ across other regions
- Differs across individuals (Aguirre et al, 1998)
2. The BOLD Haemodynamic Response

- Particularly important for event-related fMRI
- Early studies used long Stimulus Onset Asynchrony (SOA) to allow BOLD to return to baseline
- But can accommodate overlap of successive responses at short SOAs if the BOLD response is explicitly modeled (esp. if responses assumed to combine linearly)
GLM for a single voxel:

\[ y(t) = u(t) \otimes h(\tau) + \varepsilon(t) \]

\[ u(t) = \text{neural causes} \]

(stimulus train)

\[ u(t) = \sum \delta(t - nT) \]

\[ h(\tau) = \text{hemodynamic (BOLD) response} \]

\[ h(\tau) = \sum \beta_i f_i(\tau) \]

\[ f_i(\tau) = \text{temporal basis functions} \]

\[ y(t) = \sum \sum \beta_i f_i(t - nT) + \varepsilon(t) \]

\[ y = X \beta + \varepsilon \]
Convolution with the HRF

Unconvolved fit = hæmodynamic response ⊗ Residuals

Boxcar function \times \text{haemodynamic response} = \text{convolved with HRF}
Convolution with the HRF

Boxcar function \( \otimes \) haemodynamic response = convolved with HRF

Unconvolved fit

Convolved fit

Residuals

Residuals (less structure)
The BOLD Haemodynamic Response

- May differ across regions
- Differs across individuals
- This matters in event-related fMRI
- Different possible shapes are modelled using multiple temporal basis functions
- Each is entered separately in the GLM and each is fitted separately (separate parameter estimates)

Impulse Response Function

- Peak
- Brief Stimulus
- Initial Undershoot
- Undershoot

0 5 10 15 20 PST (s)
Temporal basis functions
Temporal basis functions

- Fourier Set
  - Windowed sines & cosines
  - Any shape (up to frequency limit)
  - Inference via F-test
Temporal basis functions

- Finite Impulse Response (FIR)
  - Mini “timebins” (selective averaging)
  - Any shape (up to bin-width)
  - Inference via F-test
Temporal basis functions

- **Fourier Set/ FIR**
  Any shape (up to frequency limit / bin width)
  Inference via F-test

- **Gamma Functions**
  Bounded, asymmetrical (like BOLD)
  Set of different lags
  Inference via F-test
Temporal basis functions

- **Fourier Set / FIR**
  Any shape (up to frequency limit / bin width)
  Inference via F-test

- **Gamma Functions**
  Bounded, asymmetrical (like BOLD)
  Set of different lags
  Inference via F-test

- **‘Informed Basis Set’**
  Best guess of canonical BOLD response
  Variability captured by Taylor expansion
  ‘Magnitude’ inference via t-test…?
Temporal basis functions

‘Informed’ Basis Set (Friston et al. 1998)
• Canonical HRF (2 gamma functions)
Temporal basis functions

‘Informed’ Basis Set
(Friston et al. 1998)
• Canonical HRF (2 gamma functions)
  plus Multivariate Taylor expansion in:
    time (Temporal Derivative)
Temporal basis functions

CanIn canical
Temporal
Dispersion

‘Informed’ Basis Set
(Friston et al. 1998)

• Canonical HRF (2 gamma functions)

plus Multivariate Taylor expansion in:

  time (Temporal Derivative)
  width (Dispersion Derivative)
Temporal basis functions

Canonical Temporal Dispersion

‘Informed’ Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions)
- Multivariate Taylor expansion in:
  - time (Temporal Derivative)
  - width (Dispersion Derivative)
- Magnitude’ inference via t-test on canonical parameters (provided canonical is a reasonable fit)
Temporal basis functions

‘Informed’ Basis Set
(Friston et al. 1998)
• Canonical HRF (2 gamma functions)

plus Multivariate Taylor expansion in:
  time (Temporal Derivative)
  width (Dispersion Derivative)

• Magnitude’ inference via t-test on canonical parameters (provided canonical is a reasonable fit)

• ‘Latency’ inference via tests on ratio of derivative : canonical parameters
The GLM applied to fMRI

What are the solutions?

1. The data can be filtered to remove low-frequency (1/f) noise

2. Effects are convolved with haemodynamic (BOLD) response function (HRF), to capture the sluggish nature of the response

3. The data are modelled as a timeseries, taking account of temporal autocorrelation
3. Serial correlations

Each observation is related to adjacent observations.

1\textsuperscript{st} order autoregressive process: AR(1)

\[
e_t = ae_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma^2)
\]

\[eCov\] (autocovariance function)

\[Cov(e)\]
3. Serial correlations

- Why are they a problem?
- If we have fewer independent observations than we think, this affects degrees of freedom and statistical inference (see RFX lecture)
- But this is a single subject model and our inference will likely be at group level
- …we still want to optimise parameter estimation in 2-stage procedure (see also RFX, Covariance Components lectures)
Parameter estimation (OLS)

Objective:
estimate parameters
to minimize

$$\sum_{t=1}^{N} e_t^2$$

Ordinary least squares estimation (OLS):

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

$$y = X\beta + e$$
Assumes Gaussian ‘spherical’ (i.i.d.) errors

Sphericity = i.i.d. error covariance is a multiple of the identity matrix:
\[
\text{Cov}(e) = \sigma^2 I
\]

Examples of non-sphericity:

\[
\text{Cov}(e) = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}
\]

Non-identity

\[
\text{Cov}(e) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}
\]

Non-independence
Dealing with serial correlations

Pre-whitening

• Use an enhanced noise model with multiple error covariance components
  i.e. \( e \sim N(0, \sigma^2 V) \) instead of \( e \sim N(0, \sigma^2 I) \)
  
  \( V \) is modelled using an AR (1) + white noise model estimated across all active voxels

• Use the estimated \( V \) to specify a filter matrix \( W \) for whitening the data – ‘undoing’ the serial correlations

\[
\begin{align*}
We & \sim N(0, \sigma^2 W^2 V) \\
\Rightarrow W^2 V & = I \\
\Rightarrow W & = V^{-1/2}
\end{align*}
\]

\[
Wy = WX\beta + We
\]
Dealing with serial correlations

- Once data are ‘pre-whitened’, estimation can proceed using Ordinary Least Squares
- The parameter estimates are again maximum likelihood
- This is Weighted Least Squares (WLS)
- (see Covariance Components lecture)
- The parameter estimates are ready to be used for statistical inference
Summary: GLM for fMRI

- Mass-univariate approach: same GLM for each voxel
- 2-stage procedure: single subject then group level
- GLM includes all known experimental effects and confounds
- High-pass filter to account for low-frequency noise
- Convolution with an HRF
  - both have design implications: later lectures
- Prewhiten the data to account for serial correlations
  - Covariance component estimation: later lecture

Many thanks to R Henson, A Holmes, S Kiebel, C Ruff, JB Poline and K Stephan for slides
Which temporal basis set?

In this example (rapid motor response to faces; Henson et al, 2001)

Canonical + 2 derivatives appear sufficient to capture most activity
For more complex trials (e.g. stimulus-prolonged delay (>~2 s)-response) better modelled with separate neural components (i.e., activity no longer delta function) + constrained HRF (Zarahn, 1999)