Signal Processing for Functional Brain Imaging: General Linear Model

Melissa Saenz
Auditory Neuroscience Group, CHUV/EPFL
melissa.saenz@epfl.ch
Check the website (by this evening)

- Exercise Set 1 (run GLM in MATLAB)
  - reinforce lectures, not graded but good practice for final exam
  - solutions posted the Tuesday night before the next class
  - question period at beginning of the next class

- Class Slides

- Review - matrix algebra

- Journal club dates: 28.03, 11.04, 18.04, 02.05, 16.05, 30.05
  - assignments given next week
Overview

- Basics: the fMRI BOLD signal
- GLM method (part 1, today 28.02.13)
  - intuitive explanation
  - matrix algebra explanation
    - model generation
    - parameter estimation
    - hypothesis testing
- GLM method (part 2, next week 07.03.13)
  - hypothesis testing continued
    - t-test, f-test, contrasts
  - enriching the model
    - accounting for imaging artifacts, physiological noise
  - from single-subject to group-level analysis
MRI versus functional MRI (fMRI)

The idea behind fMRI is simple: we measure a series of MRI images over time (i.e. a movie) and we test for small changes in signal intensity over time that are related to brain activity.

- **Single 3D volume**
  - 1x1x1 mm
  - takes couple of minutes

- **Series of 3D volumes**
  - 3x3x3 mm
  - every 2-4 sec
  - during 5-10 minutes
The idea behind fMRI is simple: we measure a series of MRI images over time (i.e. a movie) and we test for small changes in signal intensity over time that are related to neural activity.

- Series of 3D volumes
  - 3x3x3 mm
  - every 2-4 sec
  - during 5-10 minutes

voxel (in this case, 3x3x3 mm)

voxel timecourse or timeseries
What is the BOLD signal?

- **Blood Oxygenation Level Dependent (BOLD) signal**
  - measure of the ratio of oxygenated to deoxygenated hemoglobin

- Hemoglobin carries oxygen in the blood (6 binding sites per molecule)

- Oxygenation vs. deoxygenated hemoglobin have different magnetic properties. Higher ratio of oxy vs. deoxy yields higher local MRI signal

- Following neural activity, metabolic demand rises, blood flow locally increases bringing oxygenated hemoglobin --> the BOLD signal rises
What is the BOLD signal?

- **Blood Oxygenation Level Dependent (BOLD) signal**
  - measure of the ratio of oxygenated to deoxygenated hemoglobin

- Hemoglobin carries oxygen in the blood (6 binding sites per molecule)

- Oxygenation vs. deoxygenated hemoglobin have different magnetic properties. Higher ratio of oxy vs. deoxy yields higher local MRI signal

- Following neural activity, metabolic demand rises, blood flow locally increases bringing oxygenated hemoglobin --> the BOLD signal rises
What is the BOLD signal?

- Blood Oxygenation Level Dependent (BOLD) signal - measure of the ratio of oxygenated to deoxygenated hemoglobin

- Hemoglobin carries oxygen in the blood (6 binding sites per molecule)

- Oxygenation vs. deoxygenated hemoglobin have different magnetic properties. Higher ratio of oxy vs. deoxy yields higher local MRI signal

- Following neural activity, metabolic demand rises, blood flow locally increases bringing oxygenated hemoglobin --> the BOLD signal rises

*BOLD is an indirect measure of neural activity. Not perfect, but allows non-invasive, repeatable measures of human brain activity at a reasonable spatial resolution.*
Hemodynamic Response Function (HRF)

- the BOLD impulse response function

Empirical hemodynamic responses to brief events

Canonical hrf model (gamma function)

- Most analyses assume the HRF constant across brain areas and individuals,
- for now let’s assume that’s a valid assumption (although some differences do exist!)
- often modeled by a gamma function (or difference between two gamma functions)
Aim of the game
FMRI brain map
  - Task positive and task negative voxels
  - Contrasting conditions
High-dimensional data
  - 100’000 intracranial voxels
  - 100-1’000 timepoints
  - 10-100 trials
  - 10-100 subjects
General linear model

- The most commonly used tool in fMRI data analysis
  - hypothesize the BOLD response for a given task (*predictors or regressors*)
  - look for evidence in the measure data (confirmatory analysis)

- Parametric model-based hypothesis testing
  - **model-based** vs. blind
  - **univariate** vs. multivariate
    - Applies to data a single voxel at a time. Thus if we measure the whole brain, the analysis is repeated ~ 100,000 times (mass-univariate).
  - **parametric** vs. non-parametric
    - assume null distribution, estimate its parameters

*The GLM equation*

\[ y = X\beta + e \]
Tutorial fMRI experiment

- Three conditions

Time

Is there a face-selective brain region?

- faces
- objects
- baseline

[inspired from SPM course on GLM]
Tutorial fMRI experiment

Is there a face-selective brain region?
Model: a set of hypothetical time-series

For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series.

\[ \text{Model: a set of hypothetical time-series} \]

- For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series.

\[ \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot \]

![Diagram](image)

faces objects baseline

measured

“known” regressors

unknown parameters
Estimation: find “best” parameters

- For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series

\[
\beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot
\]
Estimation: find “best” parameters

- For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series.

\[
\approx \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot
\]

Not brilliant!
Estimation: find “best” parameters

For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series

Neither this faces objects baseline

\[ \approx \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot \]
For a given voxel (time-series), we try to figure out just what type that is by “modelling” it as a linear combination of the hypothetical time-series.

\[ \approx \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot \]

Nice Fit!

0.83 0.16 2.98
And the same model can be fitted to each time-series, with different parameter values of course.

Estimation: find “best” parameters

\[ \beta_1 \cdot + \beta_2 \cdot + \beta_3 \cdot \]

Another voxel time course with more weight given to the second predictor.

0.68  0.82  2.17
Estimation: find “best” parameters

- And the same model can be fitted to each time-series, with different parameter values of course.

\[ y \approx \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 \]

Doesn’t care 0.03 0.06 2.04
Estimation: putting things together

- Same model for all voxels, different parameter values

\[ \hat{\beta} = \begin{bmatrix} 0.83 \\ 0.16 \\ 2.98 \end{bmatrix} \]

\[ \hat{\beta} = \begin{bmatrix} 0.03 \\ 0.06 \\ 2.04 \end{bmatrix} \]

\[ \hat{\beta} = \begin{bmatrix} 0.68 \\ 0.82 \\ 2.17 \end{bmatrix} \]
Parameter estimation

- What is a “best” fit?

\[ \hat{\beta} = \begin{bmatrix} 0 \\ 0 \\ 3.31 \end{bmatrix} \]

\[ e_i = 0 \]
\[ \sum_i e_i = 0 \]
\[ \sum_i e_i^2 = 17.16 \]

\[ \hat{\beta} = \begin{bmatrix} 0.83 \\ 0.16 \\ 2.98 \end{bmatrix} \]

\[ e_i = 0 \]
\[ \sum_i e_i = 0 \]
\[ \sum_i e_i^2 = 9.47 \]

Parameter estimation aims to reduce the sum of the squared errors (ordinary least squares method)
Model revisited

- General linear model (GLM): the “design matrix” fashion

\[
y \approx X\beta
\]

or, in matrix notation:
FMRI statistical analysis

- Voxel-wise General Linear Model (GLM)

\[ y = X\beta + e \]

- \( y \): \( N \times 1 \) — measured timecourse
- \( X \): \( N \times L \) — columns contain \( L \) regressors
- \( \beta \): \( L \times 1 \) — unknown parameters
- \( e \): \( N \times 1 \) — estimation error

- Model specification: setup “design matrix” \( X \)
- Parameter estimation: find “best” \( \hat{\beta} \)
- Hypothesis testing and inference
Model specification

- From stimuli to modeled BOLD response
- Convolve stimulus time course with HRF

- blocks (epochs)

- events
Parameter Estimation

- General Linear Model (GLM)

\[(\text{eq. 1}) \quad y = X\beta + e\]

- Ordinary least-squares (OLS) estimator

  - Geometric interpretation: \(\hat{\beta} = \arg \min_{\beta} ||y - X\beta||^2 = \arg \min_{\beta} \sum_k e_k^2\)

Equivalent to solving for \(\hat{\beta}\) using simple matrix algebra:

- Projection matrix: \(P = X X^+\)
- Residual forming matrix: \(R = I - P\)

MATLAB code to implement eq. 3 is simply:

```matlab
beta = inv(X' * X) * X' * y
```

Assumes that \(X^T X\) exists!

- For \(X\) to be invertible, the columns of \(X\) must have full rank; no column is a linear combination of any other column
- If \(X\) is rank deficient, there is not a unique solution for the parameter estimates.
Parameter Estimation

- Let’s make some assumptions about the noise ($\epsilon$)
  - $\epsilon$ is normally distributed, such that mean = 0, variance = $\sigma^2$.
  - elements of the vector are uncorrelated
  - typically written as $\mathcal{N}(0, \sigma^2 I)$

- Then we meet the conditions of the Gauss-Markov theorem and we are guaranteed that the OLS estimator

$$\hat{\beta} = X^+ y = \beta + X^+ \epsilon$$

produces a minimum variance linear unbiased estimate (MVUE) of $\beta$

- unbiased: $E[\hat{\beta}] = E[X^+ y] = \beta + E[X^+ \epsilon] = 0$

- variance: (reaches Cramèr-Rao bounds)

$$\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = E[\hat{\beta}\hat{\beta}^T] - \beta\beta^T$$

$$= E[X^+ \epsilon\epsilon^T X^+] = \sigma^2 (X^TX)^{-1}$$

Nice! MVUE is the gold standard in the parameter estimation world!

will use for hypothesis testing

[Worsley et al., 1996]
Contrasts

- Questioning the fitted model—extracting “contrast” $c^T \widehat{\beta}$

- response to “faces”: $c^T = [1 \ 0 \ 0]$  
  - contrast to extract the first parameter for hypothesis testing  
  \[ c^T \widehat{\beta} = 0.83 \]

- response to “faces vs objects”: $c^T = [1 \ -1 \ 0]$  
  - contrast to extract the difference between the first two parameters for hypothesis testing  
  \[ c^T \widehat{\beta} = 0.67 \]
Hypothesis test of the contrast

Consider contrast $c^T \beta = 0$ (null hypothesis)

- we know $\hat{\beta}$ is asymptotically normal $\mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$
- then $c^T \hat{\beta}$ is asymptotically normal $\mathcal{N}(c^T \beta, \sigma^2 c^T (X^T X)^{-1} c)$

- test statistic (effect size / uncertainty of effect size)

$$t = \frac{c^T \hat{\beta}}{\hat{\sigma} \sqrt{c^T (X^T X)^{-1} c}},$$

where we estimate (reminder: $R^T R = R^2 = R$ and $\hat{e} = R y$)

$$\hat{\sigma}^2 = \frac{\hat{e}^T \hat{e}}{\text{tr}(R)}, \text{since } E[\hat{e}^T \hat{e}] = E[e^T R^T R e] = \sigma^2 \text{tr}(R)$$

and $t$ follows Student t-distribution with $\text{tr}(R) = \text{rank}(R) = N - L$ degrees of freedom
The full picture

get data

model

get effect size

\[ \hat{\beta} = \begin{bmatrix} 0.83 \\ 0.16 \\ 2.98 \end{bmatrix} \]

contrast

\[ \mathbf{c}^T = [1 \ 0 \ 0] \]

t

\[ t = \frac{6.42}{2} = 6.42 \]
Overview

- Basics: the fMRI BOLD response
- GLM method (part 1, today 28.02.13)
  - intuitive explanation
  - matrix algebra explanation
    - model generation
    - parameter estimation
    - hypothesis testing
- GLM method (part 2, next week 07.03.13)
  - hypothesis testing continued
    - t-test, f-test, contrasts
  - enriching the model
    - accounting for imaging artifacts, physiological noise
  - from single-subject to group-level analysis
Generality of GLM

Introduction to linear regression

From Wiki page

Given a data set \( \{y_i, x_{i1}, \ldots, x_{ip} \}_{i=1}^n \) of \( n \) statistical units, a linear regression model assumes that the relationship between the dependent variable \( y_i \) and the \( p \)-vector of regressors \( x_i \) is linear. This relationship is modelled through a disturbance term or error variable \( \varepsilon_i \) — an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

\[
y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \ldots, n,
\]

where \( ^T \) denotes the transpose, so that \( x_i^T \beta \) is the inner product between vectors \( x_i \) and \( \beta \).

Often these \( n \) equations are stacked together and written in vector form as

\[
y = X \beta + \varepsilon,
\]

where

\[
y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.
\]

- **Ordinary least squares (OLS)** is the simplest and thus most common estimator. It is conceptually simple and computationally straightforward. OLS estimates are commonly used to analyze both experimental and observational data.

The OLS method minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of the unknown parameter \( \beta \):

\[
\hat{\beta} = (X^T X)^{-1} X^T y = \left( \frac{1}{n} \sum x_i x_i^T \right)^{-1} \left( \frac{1}{n} \sum x_i y_i \right).
\]