The General Linear Model, Part I

Based on slides from: JB. Poline

Contents

• Introduction
• The General Linear Model
  – Data & model
  – Design matrix
  – Parameter estimates & interpretation
  – Simple contrast
• « Take home » message

Realignment & motion correction

Normalisation

Anatomical reference

General Linear Model
  → model fitting
  → statistic image

Statistical Parametric Map

Corrections for multiple comparisons

Random effect analysis

Dynamic causal modelling, Functional & effective connectivity, PPI, ...

Parameter estimates

Statistical Parametric Map

Realigned, normalised, smoothed image data

Design matrix
General Linear Model

What does it mean?

- **General**: The model can be used to answer a wide variety of questions.
- **Linear**: The model uses simple linear relationships between the variables.

Model

GLM is the basic model or general framework underlying the analysis of variance and multiple regression.

SPM key concepts...

...a voxel by voxel hypothesis testing approach

→ reliably identify regions showing a significant experimental effect of interest

Key concepts

- Type I error
  - Significance test at each voxel
- Parametric statistics
  - Parametric model for voxel data, test model parameters
- No exact prior anatomical hypothesis
  - Multiple comparisons

Functional neuroimaging signal

- Neuronal activity
- Glucose and oxygen metabolism
- Oxyhemoglobin changes
- Blood oxygenation

Blood oxygen-level dependent (BOLD) signal

- MRI (T2*-weighted image)
- Decay time (T2*)

Source: Doug Noll's primer
Hemodynamic Response Function

% signal change = (point – baseline)/baseline usually 0.5-3%

initial dip -more focal and potentially a better measure -somewhat elusive so far, not everyone can find it

time to rise signal begins to rise soon after stimulus start

time to peak signal peaks 4-6 sec after stimulus begins

post stimulus undershoot signal suppressed after stimulation ends

A simple experiment

Stimuli: passive word listening versus rest

BOLD response in the primary auditory cortex
Looking at 2 individual scans

ON-OFF, just one scan per condition

Simple fMRI example dataset: can we do better?

- One session, one subject
- 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?

Stimulus function

Time series of BOLD responses in one voxel

Voxel by voxel statistics...

Test your hypothesis visually

do it voxel by voxel

statistic image or SPM
Why do we need stats?

- Clearly voxel surfing isn’t a viable option. We’d have to do it 1,000 of times and it would require a lot of subjective decisions about whether activation was real.

- Statistics:
  - tell us where to look for activation that is related to our paradigm
  - help us decide how likely it is that activation is “real”

Voxel by voxel statistics...

- Statistics: tell us where to look for activation that is related to our paradigm.
- Help us decide how likely it is that activation is “real”.
- Voxel by voxel statistics: Time, Intensity, single voxel time series, statistic image or SPM.

Classical statistics...

- **parametric**
  - one sample t-test
  - two sample t-test
  - paired t-test
  - Anova
  - AnCova
  - correlation
  - linear regression
  - multiple regression
  - F-tests
  - etc...

- **all cases of the General Linear Model**
  - to account for serial correlations:
  - Generalised Linear Model

- **non-parametric?** → SnPM

Statistics

- Formal statistics are just doing what your eyeball test of significance did.
  - Estimate how likely it is that the signal is real given how noisy the data is.

- “p value” = probability value of the “null hypothesis”:
  - Null hypothesis = “No activation”!
  - confidence: how likely is it that the results could occur purely due to chance?
  - If “p = .03”, that means there is a 3% chance that the results are bogus.

- By convention, if the probability that a result could be due to chance is less than 5% (p < .05), we say that result is statistically significant.

- Significance depends on:
  - signal (differences between conditions)
  - noise (other variability)
  - sample size (more time points are more convincing)
...e.g. two-sample t-test?  

**Question:** Is there a change in the BOLD response between listening and rest?  

The standard t-test assumes independence, which ignores temporal autocorrelation!  

The two-sample t-test compares size of effect to its error standard deviation.  

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Hypothesis test:  
$$\beta_1 = 0?$$  
(using t-statistic)

---

**t distribution**  

Depends on 1 set of degrees of freedom!  

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>163</td>
<td>128</td>
<td>116</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>153</td>
<td>119</td>
<td>106</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>143</td>
<td>113</td>
<td>100</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>131</td>
<td>105</td>
<td>91</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>118</td>
<td>92</td>
<td>75</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>106</td>
<td>82</td>
<td>64</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>94</td>
<td>72</td>
<td>53</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>88</td>
<td>69</td>
<td>50</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**Regression example...**

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

(\(\epsilon \sim N(0, \sigma^2 I)\))

**Question:** Is there a change in the BOLD response between listening and rest?  

(\(\epsilon \) is normal and independently and identically distributed)

Hypothesis test:  
$$\beta_1 = 0?$$  
(using t-statistic)

---

**Regression example...**

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

(\(\epsilon \sim N(0, \sigma^2 I)\))

**Correlation:**  
$$H_0: \rho = 0$$  
equivalent  
to test  
$$H_0: \beta_1 = 0$$

**Two-sample t-test:**  
$$H_0: \mu_0 = \mu_1$$  
equivalent  
to test  
$$H_0: \beta_1 = 0$$
Regression example...

Model as basis functions

Design Matrix

General Linear Model

What's wrong with this model?
1. Stimulus function is not expecting BOLD response
2. Data is serially correlated

Model can be extended to account for temporal autocorrelation!

Data vector
Design Matrix
Parameter vector
Error vector

Model is specified by
1. Design matrix $X$
2. Assumptions about $\beta$

$Y = X\beta + \epsilon$

This is for a SINGLE voxel! Design matrix $X$ is the same for ALL voxels!
General Linear Model

- **fMRI time series:** $Y_1, \ldots, Y_s, \ldots, Y_N$ acquired at times $t_1, \ldots, t_s, \ldots, t_N$
- **Model:** Linear combination of basis functions
  
  $$Y_s = \beta_1 f_1(t_s) + \ldots + \beta_l f_l(t_s) + \ldots + \beta_L f_L(t_s) + \varepsilon_s$$

  - $f(\cdot)$: basis functions – “reference waveforms” – dummy variables
  - $\beta_l$: parameters (fixed effects) – amplitudes of basis functions (regression slopes)
  - $\varepsilon_s$: residual errors: $\varepsilon_s \sim N(0, \sigma^2)$ – identically distributed – independent, or serially correlated

Parameter estimation

$$Y = X\beta + \varepsilon$$

Estimate parameters

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

such that $\sum_{i=1}^{N} \hat{\varepsilon}_i^2$ minimal

Mass univariate approach

$$Y = X\beta + \varepsilon$$

Design Matrix and contrast

$$Y = X\beta + \varepsilon$$

**Question:** Is there a change in the BOLD response between listening and rest? (using t-statistic)

**Hypothesis test:** $\beta_1 = 0$?
**Model — contrasts — SPM\{t\}**

**Contrast**: specifies linear combination of parameter vector: $c \hat{\beta}$

\[
c' = +1 \quad 0
\]

box-car amplitude > 0 ?

\[
\hat{\beta}_1 > 0 \quad (\hat{\beta}_1: \text{estimation of } \beta_1)
\]

\[
1 \times \hat{\beta}_1 + 0 \times \hat{\beta}_2 > 0 ?
\]

test $H_0: c' \times \hat{\beta} > 0 ?$

\[
T = \frac{c' \hat{\beta}}{s' c (X'X)^{-1} c}
\]

**How is this computed? (t-test)**

Estimation $[Y, X; \hat{b}, s]$

\[
Y = X \hat{\beta} + \epsilon \quad \epsilon \sim \sigma^2 N(0,1) \quad (Y: \text{at one position})
\]

\[
\hat{\beta} = (X'X)^{-1} X'Y \quad (\hat{\beta} = \text{estimation of } \beta) \rightarrow \text{compute images}
\]

\[
e = Y - X \hat{\beta} \quad (e = \text{estimation of } \epsilon)
\]

\[
s^2 = (e'e/(n-p)) \quad (s = \text{estimation of } \sigma, n: \text{scans, } p: \text{parameters})
\]

\[
\text{Test } \{h,c,t\} [c': h, t]
\]

\[
\text{Var}(c' \hat{\beta}) = s' c (X'X)^{-1} c \quad \text{compute for each contrast } c
\]

\[
t = c' \hat{\beta} / \sqrt{s'c' (X'X)^{-1} c} \quad \text{compute the } i \text{ images} \rightarrow \text{compute } t \text{ images}
\]

under the null hypothesis $H_0: t \sim \text{Student( } df \text{) } \quad df = n-p$

**t distribution**

Depends on 1 set of degrees of freedom!

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.5%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.574</td>
<td>2.776</td>
<td>2.365</td>
<td>2.132</td>
<td>1.886</td>
<td>1.645</td>
<td>1.282</td>
<td>1.000</td>
<td>0.606</td>
<td>0.430</td>
<td>0.319</td>
<td>0.276</td>
</tr>
<tr>
<td>2</td>
<td>2.920</td>
<td>2.353</td>
<td>2.132</td>
<td>2.015</td>
<td>1.859</td>
<td>1.746</td>
<td>1.533</td>
<td>1.282</td>
<td>0.990</td>
<td>0.765</td>
<td>0.606</td>
<td>0.554</td>
</tr>
<tr>
<td>3</td>
<td>2.998</td>
<td>2.552</td>
<td>2.306</td>
<td>2.160</td>
<td>2.015</td>
<td>1.859</td>
<td>1.637</td>
<td>1.341</td>
<td>1.031</td>
<td>0.811</td>
<td>0.680</td>
<td>0.642</td>
</tr>
<tr>
<td>4</td>
<td>2.966</td>
<td>2.447</td>
<td>2.230</td>
<td>2.086</td>
<td>1.968</td>
<td>1.833</td>
<td>1.645</td>
<td>1.397</td>
<td>1.124</td>
<td>0.900</td>
<td>0.717</td>
<td>0.671</td>
</tr>
<tr>
<td>5</td>
<td>2.920</td>
<td>2.403</td>
<td>2.201</td>
<td>2.060</td>
<td>1.932</td>
<td>1.812</td>
<td>1.630</td>
<td>1.383</td>
<td>1.119</td>
<td>0.920</td>
<td>0.740</td>
<td>0.690</td>
</tr>
<tr>
<td>6</td>
<td>2.878</td>
<td>2.364</td>
<td>2.179</td>
<td>2.035</td>
<td>1.914</td>
<td>1.796</td>
<td>1.613</td>
<td>1.356</td>
<td>1.093</td>
<td>0.899</td>
<td>0.718</td>
<td>0.668</td>
</tr>
<tr>
<td>7</td>
<td>2.831</td>
<td>2.328</td>
<td>2.153</td>
<td>1.995</td>
<td>1.871</td>
<td>1.753</td>
<td>1.571</td>
<td>1.317</td>
<td>1.052</td>
<td>0.871</td>
<td>0.684</td>
<td>0.645</td>
</tr>
</tbody>
</table>

**Example: a line through 3 points**

\[
Y_i = \beta_1 x_i + \beta_2 x_i x_i + e_i \quad i = 1,2,3
\]

simple linear regression

\[
Y_i = \beta_1 x_i + \beta_2 x_i x_i + e_i
\]

\[
\hat{Y}_i = \hat{\beta}_1 x_i + \hat{\beta}_2 x_i x_i + e_i
\]

\[
\hat{\beta}_1, \hat{\beta}_2 \quad \text{parameter estimates}
\]

\[
\hat{Y}_1, \hat{Y}_2, \hat{Y}_3 \quad \text{fitted values}
\]

\[
e_1, e_2, e_3 \quad \text{residuals}
\]
Geometrical perspective...

\[
Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon
\]

Estimation, geometrically...

Question: Does \( X_1 \) explain anything?

Model: \( Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \)

test the null hypothesis:

e.g. \( H_0: \beta_1 = 0 \) (zero slope...)

i.e. does \( X_{i1} \) explain anything? (after \( X_{i2} \))
Contents

• Introduction
• The General Linear Model
  – Data & model
  – Design matrix
  – Parameter estimates & interpretation
  – Simple contrast
• « Take home » message

Why modelling?

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

Way to proceed

Prepare your questions.
ALL the questions!

Find a model which
• allows contrasts that translates these questions.
• takes into account ALL the effects (interaction, sessions, etc)

Devise task & stimulus presentation.

Acquire the data & analyse.

Not the other way round!!!