The General Linear Model, Part II

Contents

- Introduction & recap
  - The General Linear Model
  - T test
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message

Realigned, normalised, smoothed image data

Parameter estimates

Statistical Parametric Map

Realigned, normalised, smoothed image data

Parameter estimates

Statistical Parametric Map
**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**
A set of equations are used to describe the data. Questions about the data can then be stated as mathematical expressions.

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

---

**Functional neuroimaging signals**
- Measurable electromagnetic fields on/outside the head
- Nitrogen (N2)
- Glucose and oxygen metabolism
- Blood oxygenation
- Cerebral blood volume (CBV)
- Decay time (T2*)
- Metabolic change
- Physiological effects
- Physical effects
- Magnetic field uniformity (microscopic)
- Generation of microscopic current sources
- Combination in space and time of the current sources
- Blood oxygen-level dependent (BOLD) signal
- FDG PET
- H215O PET
- EEG/MEG

---

**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

**Time series of BOLD responses in one voxel**

**Stimulus function**

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

---

**Voxel by voxel statistics...**

- Model specification
- Parameter estimation
- Hypothesis test
- Statistic

**Statistic image or SPM**

**Source:** Doug Noll's primer
General Linear Model

\[ Y = X\beta + \varepsilon \]

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

- This is for a SINGLE voxel!
- Design matrix \( X \) is the same for ALL voxels!

MRI 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
\[ Y_i = \beta_1 f_1(t_i) + \ldots + \beta_p f_p(t_i) + \ldots + \beta_L f_L(t_i) + \varepsilon \]
- basis functions
- “reference waveforms”
- dummy variables
- \( \beta_p \): parameters (fixed effects)
- amplitudes of basis functions (regression slopes)
- \( \varepsilon \): residual errors
- \( \varepsilon_s \sim N(0, \sigma^2) \)
- identically distributed
- independent, or serially correlated

Mass univariate approach

\[ Y = X\beta + \varepsilon \]

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

Contrast:
- e.g. \([1 -1] \)
Estimate parameters:
\[
\hat{\beta} = (X^t X)^{-1} X^t Y
\]

Let's assume iid error:
\[
\hat{\varepsilon} = Y - X\hat{\beta}
\]

The residuals are estimated such that:
\[
\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \text{ minimal}
\]

Parameter estimation:
\[
Y = X\beta + \varepsilon
\]

Assume iid error:
\[
\hat{\beta} = (X^t X)^{-1} X^t Y
\]

Contrast: specifies linear combination of parameter vector:
\[
c'\hat{\beta}
\]

box-car amplitude > 0 ?
\[
\hat{\beta}_1 > 0 ? \quad (\hat{\beta}_i : \text{estimation of } \beta_i)
\]

Test \( H_0 : c'\hat{\beta} > 0 ? \)

The contrast of estimated parameters is computed for each contrast:
\[
T = \frac{c'\hat{\beta}}{\sqrt{s c'(X(X^t X)^{-1})c}}
\]

How is this computed? (t-test)

Estimation \([Y, X] [b, s]\)

\[
Y = X\beta + \varepsilon
\]

Error \(\varepsilon \sim \sigma^2 N(0, I)\) (\(Y\) at one position)

\[
b = (X^t X)^{-1} X^t Y
\]

\(b\) = estimation of \(\beta\)

\[
e = Y - Xb
\]

\(e\) = estimation of \(\varepsilon\)

\[
s^2 = (e'e/(n-p))
\]

\(s^2\) = estimation of \(\sigma^2\), \(n\): scans, \(p\): parameters

Test \([b, e, c] [c'b, t]\)

\[
\text{Var}(c'b) = s^2 c'(X(X^t X)^{-1})c
\]

Compute for each contrast \(c\).

\[
t = c'b / \sqrt{s^2 c'(X(X^t X)^{-1})c}
\]

Compute the \(t\) images under the null hypothesis \(H_0 : t \sim \text{Student(} df \text{)}\)

\(df = n-p\)
Tests multiple linear hypotheses:

Does $X_1$ model anything?

$H_0$: True model is $X_0$

$F = \frac{S_0^2 - S^2}{S^2}$

Additional variance accounted for by tested effects

Error variance estimate

F test (SPM{F}): a reduced model or multi-dimensional contrasts?

Estimation $[Y, X]$ in $[b, \beta]$

$Y = X \beta + \varepsilon$

$\varepsilon \sim N(0, \sigma^2 I)$

$Y_0 = X_0 \beta_0 + \varepsilon_0$

$\varepsilon_0 \sim N(0, \sigma_0^2 I) \quad X_0 : X$ Reduced

Estimation $[Y, X]$ in $[b, \beta]$ (not really like that)

$b_0 = (X_0' X_0)^{-1} X_0 Y$

$e_0 = Y - X_0 b_0$

$s_0^2 = \frac{e_0' e_0}{n - p_0}$

$c_0 = \frac{1}{s_0^2}$

Test $[b, \beta, c]$ [ess, F]

$F = \frac{(e_0' e_0 - e' c_0^{-1} e_0)}{(p - p_0) / s^2}$

Under the null hypothesis: $F \sim F(df1, df2)$

This model? Or this one?
F distribution

Depends on 2 sets of degrees of freedom!

T and F test: take home ...

- T tests are simple combinations of the betas; they are either positive or negative ($b_1 - b_2$ is different from $b_2 - b_1$)
- F tests can be viewed as testing for the additional variance explained by a larger model wrt. a simpler model, or
- F test the sum of the squares of one or several combinations of the betas
- In testing "single contrast" with an F test, for ex. $b_1 - b_2$, the result will be the same as testing $b_2 - b_1$. It will be exactly the square of the t-test, testing for both positive and negative effects, and the p-value will be twice as big.

« Additional variance » : Again

Testing for the green
correlated regressors, for example:
green: subject age
yellow: subject score

Independent contrasts
« Additional variance » : Again

Testing for the red

correlated contrasts

Testing for the green

Entirely correlated contrasts?
Non estimable!

Testing for the green and yellow

If significant?
Could be G or Y!

Contents

- Introduction & recap
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message
True signal (---) and observed signal

Model (green, peak at 6 sec) and TRUE signal (blue, peak at 3 sec)

Fitting: $b_1 = 0.2$, mean = .11

Noise (still contains some signal)

⇒ Test for the green regressor not significant

True signal + observed signal

Model (green and red) and true signal (blue ---)
Red regressor: temporal derivative of the green regressor
Global fit (blue) and partial fit (green & red)
Adjusted and fitted signal
Noise (a smaller variance)

⇒ Test of the red regressor very significant

Residual Variance = 0.3
We rather test flexible models if there is little a priori information, and precise ones with a lot a priori information.

In general, use the F-tests to look for an overall effect, then look at the betas or the adjusted signal to characterise the origin of the signal.

Interpreting the test on a single parameter (one function) can be very confusing: cf the delay or magnitude situation.

Summary ...

- The residuals should be looked at ...(non random structure ?)
- We rather test flexible models if there is little a priori information, and precise ones with a lot a priori information.
- In general, use the F-tests to look for an overall effect, then look at the betas or the adjusted signal to characterise the origin of the signal.
- Interpreting the test on a single parameter (one function) can be very confusing: cf the delay or magnitude situation.

Correlation between regressors

- True signal
- Model (green and red)
- Fitting (blue : global fit)
- Noise

Correlation between regressors

- Residual var. = 0.2
- $P( b_1 = 0 ) = 0.08$ (t test $b_1>0$)
- $P( b_2 = 0 ) = 0.07$ (t test $b_2>0$)
- $P( [b_1 b_2] = 0 ) = 0.002$ (F test $[b_1 b_2] ≠ 0$)

Correlation between regressors - 2

- True signal
- Model : red regressor orthogonalised with respect to the green one = remove every thing that can correlate with the green regressor
- Fit
- Noise
Y = \( Xb + e \)

\( X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \)

\( C1 \) and \( C2 \) are not independent!

- \( C1 + C2 \) is not estimable
- \( C1 - C2 \) is estimable

Summary ...

- We are implicitly testing additional effect only, so we may miss the signal if there is some correlation in the model using t tests.
- Orthogonalisation is not generally needed - parameters and test on the changed regressor don’t change.
- It is always simpler (when possible !) to have orthogonal (uncorrelated) regressors.
- In case of correlation, use F-tests to see the overall significance. There is generally no way to decide where the ‘common’ part shared by two regressors should be attributed to.
- In case of correlation and you need to orthogonalise a part of the design matrix, there is no need to re-fit a new model - the contrast only should change.
Contents

- Introduction & Recap
- F-test and added variance
- Good & bad models
  - Improved model
    - Haemodynamic response function
    - High pass filter
    - Serial correlation
    - Global effect
- HRF and ER fMRI
- « Take home » message

Haemodynamic response function

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s post-stimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1...
- ...but differences across: other regions (Schacter et al, 1997) individuals (Aguirre et al, 1998)

Improved model

Convolve stimulus function with model of BOLD response

Fitted data

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
Low frequency nuisance effects

- Drifts
  - physical
  - physiological

- Aliased high frequency effects
  - cardiac (~1 Hz)
  - respiratory (~0.25 Hz)

⇒ Power in the low frequencies

Physiological “noise”

\[ \beta = X Y \]

Discrete cosine transform set

High pass filter

\[ f(t) = \cos(\pi t / T) \]

Discrete cosine transform set

GLM fitted

Raw MRI time series

Adjusted for global & low Hz effects

Scaled for global changes

Fitted “high-pass filter”

Respiratory cycle every 4-10 s (0.3 Hz)
Cardiac cycle every ~1 s (0.9 Hz)
Serial correlation (fMRI)

**General Linear Model**

⇒ **Generalised Linear Model**

fMRI time series are auto-correlated:
- adaptation of general linear model necessary for valid test
- estimation of autocorrelation
- Optimal low-pass filter

Serial correlation

\[ Y_i = aY_{i-1} + e_i \quad \text{with} \quad e_i \sim N(0, \sigma^2) \]

autoregressive process of order 1 (AR(1))

autocovariance function

\[ \text{Cov}(Y) \]

Error covariance matrix

\[ Y = X\beta + \varepsilon \]

\[ \text{Cov}(\varepsilon) \]

i.i.d.

 sampled error covariance matrices (10^3 voxels)
Serial correlations

\[ Y = X\beta + \varepsilon \]
\[ \varepsilon \sim N(0, \sigma^2 V) \text{ – intrinsic autocorrelation} \]

**Problem:**
Estimate \( \sigma^2 \) at each voxel and make inference about \( c^T \beta \)

**Model:**
Model \( V \) as linear combination of \( m \) variance components
\[ V = \lambda_1 Q_1 + \lambda_2 Q_2 + \ldots + \lambda_m Q_m \]

**Assumptions:**
- \( V \) is the same at each voxel
- \( \sigma^2 \) is different at each voxel

**Example:**
For one fMRI session, use 2 variance components. Choice of \( Q_1 \) and \( Q_2 \) motivated by autoregressive model of order 1 plus white noise (AR(1)+wn)

**Estimation in SPM**

1. **Ordinary least-squares**
   \[ y_j = X\theta_j + \varepsilon_j \]
2. **ReML (pooled estimate)**
   \[ \hat{\theta}_j,\text{OLS} = X^+ y_j \]
   \[ \hat{\theta}_j,\text{ML} = (X^T V^{-1} X)^{-1} X^T V^{-1} y_j \]

**Restricted Maximum Likelihood**

\[ Y = X\beta + \varepsilon \]
\[ \text{Cov}(\varepsilon)? \]

**Computation of sample covariance matrix of data at all activated voxels:**
\[ \hat{C}_Y = \frac{1}{K} \sum k Y_k Y_k^T \]

**Important:** Data \( Y_k \) must be high-pass filtered.

**Model**
\[ C_Y = X\beta + \sum \lambda_i Q_i \]

**Estimate** \( \lambda_i \) using “Restricted Maximum Likelihood” (ReML)

**Estimate** \( \sigma^2 \) at each voxel in the usual way by
\[ \sigma^2 - (R^T \hat{V}^{-1} R) / \text{trace}(R V) \text{ – unbiased} \]
where \( R = 1 - X X^T / X \)

**Assume, at voxel \( j \):**
\[ \hat{C}_{\varepsilon,j} = \sigma^2 V \]
Serial correlations...inference

Inference:
To test null hypothesis $c^T \beta = 0$, compute t-value by dividing size of effect by its standard deviation: $t = \frac{c^T \beta}{\text{std}[c^T \beta]}$

where $\text{std}[c^T \beta] = \sqrt{\sigma^2 c^T (X^T X)^{-1} X^T V V X (X^T X)^{-1} c}$

... but ... $\text{std}[c^T \beta]$ is not a $\chi^2$ variable because of $V$

Approximating $\chi^2$ distribution using Satterthwaite approximation:

$\text{Var}[\sigma^2] = \frac{2 \sigma^4 \text{trace}(R V V R)}{\text{trace}(R V V)^2}$

$\nu = \frac{2 \text{E}[\sigma^2]^2 / \text{Var}[\sigma^2]}{\text{trace}(R V V) / \text{trace}(R V V R)}$ – effective degrees of freedom

Use t-distribution with $\nu$ degrees of freedom to compute $p$-value for $t$

Absolute value of BOLD signal is meaningless

fMRI signal of an individual voxel across scans and sessions

⇒ Scale the scans by the session global mean

fMRI Global scaling artefact

Scale each scan by its own global mean?

No scaling.

Stimulus
Global Voxel 1
Voxel 2

With scaling:
voxel/global

Voxel 1, no effect
Voxel 2, pos. effect

Voxel 1, neg. effect
Voxel 2, "no" effect

PET Global effects: AnCova...

• AnCova – classic way to include a nuisance covariate into a comparison
• Assumptions
  - linear / parallel
  - Constant across conditions

⇒ Scale the scans by the session global mean
Single subject activation (AnCova)

\[ Y_{ij} = \alpha_i + \mu + \xi (e_{ij} - \bar{e}) + \epsilon_{ij} \]

- Null hypothesis (at this voxel)
  - \( H_0: \alpha_i = 0 \)
- Parameter vector
  - \( \beta = (\alpha_1, \alpha_2, \alpha_3, \mu, \xi) \)
- Contrast weights \( g \)
  - Activation: \( H_1: \alpha_1 > 0 \)
    - \( (-1, 1, 0) \)
  - Deactivation: \( H_1: \alpha_1 < 0 \)
    - \( (+1, -1, 0) \)
- F-test
  - \( H_0: \alpha_i = \alpha_i - \alpha_i = 0 \)

Proportional scaling by gCBF...

- Scale gCBF to 50ml/min/dl
  \[ Y_{ij} = Y_{ij} / (r_{CBF} / 50) \]

Confounded covariates...

- E.g. global effects
  - Frequently correct for global changes
  - Nuisance effect?
  - Global mean affected by response?
- Motion effects in fMRI

Contents

- Introduction & Recap
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message
BOLD Impulse Response

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s poststimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1...
- ... but differences across: other regions (Schacter et al 1997) individuals (Aguirre et al, 1998)

Epoch vs Event-related fMRI

"PET Blocked conception" (scans assigned to conditions)

A
B
A
B...

"fMRI Epoch conception" (scans treated as timeseries)

Boxcar function

Condition A
Condition B

Convolved with HRF

Overview

1. Advantages of eMfRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

Advantages of Event-related fMRI

1. Randomised trial order
   c.f. confounds of blocked designs
Advantages of Event-related fMRI

1. Randomised trial order
c.f. confounds of blocked designs

2. Post hoc / subjective classification of trials
e.g., according to subsequent memory

3. Some events can only be indicated by subject (in time)
e.g., spontaneous perceptual changes
Advantages of Event-related fMRI

1. Randomised trial order
c.f. confounds of blocked designs
2. Post hoc / subjective classification of trials
e.g., according to subsequent memory
3. Some events can only be indicated by subject (in time)
e.g., spontaneous perceptual changes
4. Some trials cannot be blocked
e.g., “oddball” designs
5. More accurate models even for blocked designs?
e.g., “state-item” interactions
### Overview

1. Advantages of efMRI
2. **BOLD impulse response**
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

### BOLD Impulse Response

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s poststimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1...
- ... but differences across: other regions (Schacter et al 1997) individuals (Aguirre et al, 1998)

---

### Disadvantage of Randomised Designs

1. Less efficient for detecting effects than are blocked designs *(see later...)*
2. Some psychological processes may be better blocked *(eg task-switching, attentional instructions)*

---

### Blocked Design

- **“Epoch” model**
- **“Event” model**
Early event-related fMRI studies used a long Stimulus Onset Asynchrony (SOA) to allow BOLD response to return to baseline. However, if the BOLD response is explicitly modelled, overlap between successive responses at short SOAs can be accommodated, particularly if responses are assumed to superpose linearly. Short SOAs are more sensitive.

BOLD Impulse Response

Overview

1. Advantages of eFMI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

General Linear (Convolution) Model

GLM for a single voxel:

\[ Y(t) = x(t) \otimes h(t) + \epsilon \]

- \( x(t) \) = stimulus train (delta functions)
- \( x(t) = \sum \delta(t - nT) \)
- \( h(t) \) = hemodynamic (BOLD) response
- \( h(t) = \sum \beta_i f_i(t) \)
- \( f_i(t) \) = temporal basis functions
- \( Y(t) = \sum \sum \beta_i f_i(t - nT) + \epsilon \)

General Linear Model (in SPM)

Auditory words every 20s

(Orthogonalised) Gamma functions \( f_i(u) \) of peristimulus time \( u \)

Sampled every TR = 1.7s

Design matrix, \( X \)

\[ f_i(u) \otimes x(t) \quad f_i(u) \otimes x(t) \quad ... \]

\( SPM(F) \)
Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

Temporal Basis Functions

- Canonical HRF (2 gamma functions)
- Multivariate Taylor expansion in:
  - Temporal Derivative
  - Dispersion Derivative

Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions) plus Multivariate Taylor expansion in:
  - Temporal Derivative
  - Dispersion Derivative
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" inferences via t-test on canonical parameters (providing canonical is a good fit...more later)

Informed Basis Set (Friston et al. 1998)
- Canonical HRF (2 gamma functions) plus Multivariate Taylor expansion in:
  - time (Temporal Derivative)
  - width (Dispersion Derivative)
- "Magnitude" inferences via t-test on canonical parameters (providing canonical is a good fit...more later)
- "Latency" inferences via tests on ratio of derivative : canonical parameters (more later...)

Temporal Basis Functions

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

Finite Impulse Response (FIR)
- Mini timebins (selective averaging)
- Any shape (up to bin-width)
- Inference via F-test
Temporal Basis Functions

- **Fourier Set**
  - Windowed sines & cosines
  - Any shape (up to frequency limit)
  - 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" inferences via t-test…?

Temporal Basis Sets: Which One?

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

Canonical + Temporal + Dispersion + FIR

...canonical + temporal + dispersion derivatives appear sufficient
...may not be for more complex trials (eg stimulus-delay-response)
...but then such trials better modelled with separate neural components (ie activity no longer delta function) + constrained HRF (Zarahn, 1999)
Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

Timing Issues

- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0, 4, 8, 12, ...] post-stimulus may miss peak signal
- Higher effective sampling by: 1. Asynchrony, eg. SOA=1.5TR
Timing Issues

- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0,4,8,12...] post-stimulus may miss peak signal
- Higher effective sampling by:
  1. Asynchrony, eg. SOA=1.5TR
  2. Random Jitter, eg. SOA=(2±0.5)TR

...but "Slice-timing Problem" (Henson et al, 1999)
Slices acquired at different times, yet model is the same for all slices

Better response characterisation (Miezin et al, 2000)
Timing Issues

• but "Slice-timing Problem" (Henson et al., 1999)
  Slices acquired at different times, yet model is the same for all slices
  => different results (using canonical HRF) for different reference slices

• Solutions:
  1. Temporal interpolation of data
     ... but less good for longer TRs

BOLD Response Latency (Linear)

- Assume the real response, \( r(t) \), is a scaled (by \( \alpha \)) version of the canonical, \( f(t) \), but delayed by a small amount \( dt \):
  \[
  r(t) = \alpha f(t + dt) \sim \alpha f(t) + \alpha f'(t) dt
  \]
  1st-order Taylor

- If the fitted response, \( R(t) \), is modelled by the canonical + temporal derivative:
  \[
  R(t) = \beta_1 f(t) + \beta_2 f'(t)
  \]
  GLM fit

- Then canonical and derivative parameter estimates, \( \beta_1 \) and \( \beta_2 \), are such that:
  \[
  \alpha = \beta_1, \quad dt = \beta_2 / \beta_1
  \]

BOLD Response Latency: example

- If the fitted response, \( R(t) \), is modelled by the canonical + temporal derivative:
  \[
  R(t) = \beta_1 f(t) + \beta_2 f'(t)
  \]
  GLM fit

- Then canonical and derivative parameter estimates, \( \beta_1 \) and \( \beta_2 \), are such that:
  \[
  \alpha = \beta_1, \quad dt = \beta_2 / \beta_1
  \]

ie, Latency can be approximated by the ratio of derivative-to-canonical parameter estimates (within limits of first-order approximation, +/- 1s)
BOLD Response Latency (Linear)

- Delayed Responses {green/yellow}
- Canonical Basis Functions
- Actual latency, $dt$, vs. $\beta_2/\beta_1$

BOLD Response Latency (Iterative)

- Numerical fitting of explicitly parameterised canonical HRF (Henson et al, 2001)
- Distinguishes between Onset and Peak latency...
  ...unlike temporal derivative...
  ...and which may be important for interpreting neural changes (see previous slide)
- Distribution of parameters tested nonparametrically (Wilcoxon’s T over subjects)

Neural Response Latency?

- Neural
  A. Decreased
  B. Advanced
  C. Shortened (same integrated)
  D. Shortened (same maximum)
- BOLD
  A. Smaller Peak
  B. Earlier Onset
  C. Earlier Peak
  D. Smaller Peak and earlier Peak

Neural Response Latency?

- Neural
  D. Shortened (same maximum)
- BOLD
  D. Smaller Peak and earlier Peak

Most parsimonious account is that repetition reduces duration of neural activity.
Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

Fixed SOA = 16s

Stimulus (“Neural”)  HRF  Predicted Data

Fixed SOA = 4s

Stimulus (“Neural”)  HRF  Predicted Data

Not particularly efficient...

Randomised, SOA_{min} = 4s

Stimulus (“Neural”)  HRF  Predicted Data

More Efficient...

Very Inefficient...
Blocked, $\text{SOA}_{\text{min}} = 4\text{s}$

Even more Efficient...

Blocked, epoch = 20s

Blocked-epoch (with small SOA) and Time-Freq equivalences

Sinusoidal modulation, $f = 1/33\text{s}$

The most efficient design of all!

Blocked (80s), $\text{SOA}_{\text{min}} = 4\text{s}$, highpass filter = 1/120s

“Effective HRF” (after highpass filtering) (Josephs & Henson, 1999)

Don’t have long (>60s) blocks!
Randomised, SOA\textsubscript{min} = 4s, highpass filter = 1/120s

(Randomised design spreads power over frequencies)

Stimulus ("Neural") ⊗ HRF = Predicted Data

Design Efficiency

\[ T = \beta^T \sigma^{2 \cdot \text{std}(\beta)} \]

\[ \text{std}(\beta) = \sqrt{\sigma^2 c^T (X^T X)^{-1} c} \quad (\text{i.i.d}) \]

- For max. \( T \), want min. contrast variability (Friston et al, 1999)
- If assume that noise variance (\( \sigma^2 \)) is unaffected by changes in \( X \), then want maximal efficiency, \( e \):\[ e(c,X) = \frac{c^T (X^T X)^{-1} c}{\text{maximal bandpassed signal energy}} \] (Josephs & Henson, 1999)

Efficiency - Multiple Event-types

- Design parametrised by: SOA\textsubscript{min}, Minimum SOA
- \( p_i(h) \): Probability of event-type \( i \) given history \( h \) of last \( m \) events
- With \( n \) event-types \( p_i(h) \) is a \( n^m \times n \) Transition Matrix
- Example: Randomised AB

\[
\begin{array}{cc}
A & B \\
0.5 & 0.5 \\
0.5 & 0.5 \\
\end{array}
\]

\[ \Rightarrow \text{ABBABAABABAAA...} \]

4s smoothing; 1/60s highpass filtering

Efficiency - Multiple Event-types

- Example: Alternating AB

\[
\begin{array}{cc}
A & B \\
0 & 1 \\
1 & 0 \\
\end{array}
\]

\[ \Rightarrow \text{ABABABABABABAB...} \]

- Example: Permuted AB

\[
\begin{array}{cc}
A & B \\
AA & 0 \\
AB & 0.5 \\
BA & 0.5 \\
BB & 1 \\
\end{array}
\]

\[ \Rightarrow \text{ABBAABABABBA...} \]

4s smoothing; 1/60s highpass filtering
Efficiency - Multiple Event-types

- Example: Null events
  
  \[
  \begin{array}{c|cc}
  & A & B \\
  A & 0.33 & 0.33 \\
  B & 0.33 & 0.33 \\
  \end{array}
  \]

  \[\Rightarrow \text{AB-BA-A-B} \ldots\]

- Efficient for differential and main effects at short SOA
- Equivalent to stochastic SOA (Null Event like third unmodelled event-type)
- Selective averaging of data (Dale & Buckner 1997)

Efficiency - Conclusions

- Optimal design for one contrast may not be optimal for another
- Blocked designs generally most efficient with short SOAs (but earlier restrictions and problems of interpretation...)
- With randomised designs, optimal SOA for differential effect (A-B) is minimal SOA (assuming no saturation), whereas optimal SOA for main effect (A+B) is 16-20s
- Inclusion of null events improves efficiency for main effect at short SOAs (at cost of efficiency for differential effects)
- If order constrained, intermediate SOAs (5-20s) can be optimal; if SOA constrained, pseudo-randomised designs can be optimal (but may introduce context-sensitivity)

Contents

- Introduction & Recap
- Good & bad models
- Improved model
- HRF and ER fMRI
- "Take home" message

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!!!
Three Stages of an Experiment

1. Sledgehammer Approach
   • brute force experiment: powerful stimulus &
     don’t try to control for everything
   • look at what was done before or by others
   • run a couple of subjects -- see if it looks
     promising
   • if it doesn’t look great, tweak the stimulus or
     task
   • try to be a subject yourself so you can notice
     any problems with stimuli or subject strategies

2. Real Experiment
   • at some point, you have to stop changing things
     and collect enough subjects run with the same
     conditions to publish it
   • how many subjects do you need
     - some psychophysical studies test two or three subjects,
     - many studies test 6-10 subjects, random effects analysis
       requires at least 15 subjects, ...
     - some subjects WILL be rejected, so acquire more than
       the minimum !
   • can run all subjects in one or two days
     - pro: minimize setup and variability
     - con: “bad magnet day” means a lot of wasted time
     - make sure all the data are treated the “same way”. (script)

3. “Whipped Cream” experiment
   • after the real experiment works, then think
     about a “whipped cream” version
   • going straight to whipped cream is a huge
     endeavor, especially if you’re new to imaging
   • and it gives you a second paper!