First-level fMRI modeling

Monday, Lecture 3
Jeanette Mumford
University of Wisconsin - Madison

What do we need to remember from the last lecture?
• What is the general structure of a t-statistic?
  – How about the specific structure for the GLM?
• What is the residual?
• What are the assumptions we make with the linear model (Gauss Markov)?

Goal in fMRI analysis

Find voxels with BOLD time series that look like this
Goal in fMRI analysis

Task on

Find voxels with BOLD time series that look like this

btw, does this remind you of anything??

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
\vdots \\
Y_{197} \\
Y_{198} \\
Y_{199} \\
Y_{200}
\end{pmatrix} = \begin{pmatrix} 1&0 \\ 1&0 \\ 1&0 \\ \vdots \\ 1&0 \\ 0&1 \\ 0&1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon
\]
BOLD issues

- BOLD response is delayed
  - Convolution
  - FIR modeling
- BOLD time series suffer from low frequency noise
  - Highpass filtering
  - Prewhitening
  - Precoloring
- Scaling the data
  - Grand mean scaling
  - Intensity normalization
Recall the GLM

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

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{pmatrix} = 
\begin{pmatrix}
1 & x_{11} & x_{21} & \cdots & x_{n1} \\
1 & x_{12} & x_{22} & \cdots & x_{n2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{1n} & x_{2n} & \cdots & x_{nn}
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_n
\end{pmatrix} + 
\begin{pmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_n
\end{pmatrix}
\]

Single voxel time series

Recall the basic structure of a t-statistic

\[ t = \frac{c e}{\sqrt{c(X'X)^{-1}c'c}} \]

How to make a good model

• Explain as much variability in the data as possible
  – If you miss something it will go into the residual error, \( e \)
  \[
  \hat{\sigma}^2 = \frac{c e}{N - p}
  \]
  \[ t = \frac{c(X'X)^{-1}X'Y}{\hat{\sigma}\sqrt{c(X'X)^{-1}c'}} \]

Big residuals \( \Rightarrow \) Big variance \( \Rightarrow \) Small t stat
Understanding the data

- Time series drifts down in beginning
- BOLD response is delayed

Simplest Model

\[ Y = X\beta \]
Simplest Model

Modeling the delay

• Hemodynamic response function
  – Real data was used to find good models for the hemodynamic response

Stimulus

HRF (double gamma)
Convolution

- Combine HRF and expected neural response

Typically model derivative of convolved HRF to adjust for small differences in onset (<1s)

Different HRF’s

Too symmetric

Basic shape okay, but no post stimulus undershoot
Different HRF’s

Too symmetric

Basic shape okay, but no post stimulus undershoot

Includes post stimulus undershoot

How convolution is actually done

• Since timing of trials will not occur on TR and to use a better resolution HRF, we convolve in a finer time resolution
• Otherwise any trial occurring at any point during the period of the TR will look exactly the same in the model

Usampling/downsampling
Assumptions of canonical HRF

• BOLD increases linearly

Dale & Buckner, 1997

Assumptions of canonical HRF

• The width, height and delay are correct
• Lindquist & Wager (2007)
  – Fit H/W/T separately
  – Works okay-ish

Finite impulse response model

• FIR
  – Make no assumption about the shape of the HRF
Constrained basis set

- Lower the number of regressors in the model by using a basis set
- Constrained to shapes that are reasonable for HRF shapes

FLOBS

- fMRIB Linear Optimal Basis Sets
  - Generates a set of basis sets to model signal
  - Specify ranges for different portions of the hrf
Comparison

More thoughts about canonical HRF

• Advantages:
  – Simpler analysis
  – Easily interpretable outcome
  – Simplifies group analysis

• Disadvantages
  – Biased if canonical HRF is incorrect

Unbiased basis sets

• Advantages
  – Not biased towards a particular shape
  – Allows testing of hypotheses about specific HRF parameters

• Disadvantages
  – Less powerful
  – Makes group analysis more difficult
  – Tend to overfit the data (i.e., fit noise)
We can make a design matrix!

- Start with task blocks or delta functions
- Convolve
- Estimate the GLM and carry out hypothesis!

Convolved Boxcar

\[ t = \frac{0.66}{1.22 \times 0.06} = 9.02 \]
The Noise

- White noise
  - All frequencies have similar power
  - Not a problem for OLS

More Noise

- Colored noise
  - Has structure
  - OLS needs help!

What about the drift?

- Sources
  - Head motion
  - Cardiac noise
  - Respiratory noise
  - Scanner noise
What the noise looks like

Power spectra of noise data (Zarahn, Aguirre, D'Esposito, Nl, 1997)

1/f structure

More noise

Average spectrum of principal components (Mitra & Pesaran, Biophysical Journal, 1999)

Low frequency noise

Average spectrum of principal components (Mitra & Pesaran, Biophysical Journal, 1999)
More noise

Average spectrum of principal components
(Mitra & Pesaran, Biophysical Journal, 1999)

More noise

Average spectrum of principal components
(Mitra & Pesaran, Biophysical Journal, 1999)

The 1-2 punch

- Punch 1: Highpass filtering
  - FSL uses gaussian weighted running line smoother
  - SPM fits a DCT basis set
- Punch 2: Prewhitening
  - We’ll get to that later
- There’s also a thing called lowpass filtering (precoloring), but generally it isn’t so great and nobody uses it
Highpass filtering

• Simply hack off the low frequency noise
  – SPM: Adds a discrete cosine transform basis set to design matrix

Highpass filtering

• FSL: Gaussian-weighted running line smoother
  – Step 1: Fit a Gaussian weighted running line

Fit at time t is a weighted average of data around t
Highpass filtering

- Step 2: Subtract Gaussian weighted running line fit
- IMPORTANT: Must apply filter to both the data and the design.
  - FSL has ‘apply temporal filter’ box in design setup. Leave it checked!

Highpass filtered design (FSL)

If it wasn’t filtered, this trend wouldn’t be here.

Highpass filtering

Filter below .01 Hz
Filter cutoff

- High, but not higher than paradigm frequency
  - Look at power spectrum of your design and base cutoff on that
  - Block design: Longer than 1 task cycle... usually twice the task cycle
  - Event related design: Larger than 66 s (based on the power spectrum of a canonical HRF of a single response)

High-pass Filtering

- Removes the worst of the low frequency trends

Highpass filtering

- What does it do to the signal??

[Images and graphs]

Woolrich et al, NI 2001
Highpass filtering

• What does it do to the signal??

<table>
<thead>
<tr>
<th>Signal</th>
<th>Power spectrum</th>
<th>Lowpass filter</th>
<th>Highpass filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Filtering conclusions

• Lowpass filtering
  - Idea is to swamp out high frequency noise
  - Easily removes important signal in ER designs
  - Choose cutoff to remove noise, but avoid your signal

• Highpass filtering
  - Removes low frequency drift
  - We typically avoid designs with low frequencies, so highpass filtering is always used
Bandpass filtering

- High and lowpass filtering
  - Common in functional connectivity analysis since it allows you to focus on a specific frequency
  - Not typically used in standard fMRI analyses

Model with HP filter

\[
Y = X\beta
\]

Parameter of interest

Use contrast

\[
c = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}
\]

\[
H_0 : c\beta = 0
\]
Convolution & HP filter

\[ t = \frac{0.64}{1.04 \times 0.06} = 10.26 \]

Punch 2: Prewhitening

- Highpass filtering
  - Analogous to using a roller to paint a wall… you can’t get the edges very neatly
- Prewhitening
  - More precise estimate of correlation…like using a brush for the edges

Prewhitening

- Remember Gauss Markov?
  - If our errors are distributed with mean 0, constant variance and not temporally autocorrelated then our estimates are unbiased and have the smallest variance of all unbiased estimators.
  - Uh oh,
    \[ \epsilon \sim N(0, \sigma^2 V) \]
    (even after highpass filtering)
Prewhitening

- Find $K$ such that
  
  $$KV'K = I$$

- Premultiply GLM by $K$
  
  $$KY = KX\beta + K\epsilon$$

  $$\text{Var}(KY) = \text{Var}(K\epsilon) = \sigma^2K'VK = \sigma^2I$$

Awesome! G-M holds for our "new" model

Whitening

- OLS can be used on whitened model
  
  $$\hat{\beta} = (X'X)^{-1}X'y$$

  $$\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$
Prewhitening

• Step 1: Fit the linear regression ignoring temporal autocorrelation
• Step 2: Use residuals from first regression to estimate temporal autocorrelation to obtain K
• Step 3: Create prewhitened model and estimate in usual way

Estimating V

• We don’t know V, so we estimate it
• There’s a bias problem….
  – SPM uses a global covariance estimate to help with this
  – FSL uses a local estimate, but smooths it

fMRI noise

• Tends to follow 1/f trend
• Autoregressive (AR) models fit it well
  – Cor(\(e_i, e_j\)) = \(\rho^{|i-j|}\)
Whitening SPM

- Globally estimates correlation
  - Correlation of time series averaged over voxels
- Structured correlation estimate
  - Scaled AR(1) with correlation 0.2 plus white noise

Only 2 parameters are estimated

Convolution, HP filter, Whitening

\[
t = \frac{0.66}{0.954 \times 0.08} = 8.65
\]
Scaling

- **Grand Mean Scaling**
  - Removes intersession variance in global signal due to changes in gain of scanner amplification
  - Allows us to combine data across subjects
  - Whole 4D data set is scaled by a single number
  - Automatically done in software packages
  - Doesn’t change variability between time points

- **Proportional scaling (Intensity normalization)**
  - Forces each volume of 4D dataset to have the same mean
  - Also done by modeling the global signal
  - Idea is to remove background activity
  - Problems can occur if true activation is wide spread
  - Negative activations may result
Intensity Normalization

- without
- with

Signal is lost and negative activation artifacts

Other modeling considerations

- Adding the derivative of the HRF
- Adding motion parameters to the model

Model HRF & Derivative
**Temporal derivative**

- We model the derivative, but don’t study inferences of it
  - Lingquist, et al (NI, 2008) suggest this is a bad idea…may lead to bias
- If there’s time later, I’ll show you how to incorporate the derivative parameter estimate

**Collinearity**

- When designing your study, you want your tasks to be uncorrelated
- Correlation between regressors lowers the efficiency of the parameter estimation
- Parameter estimates are highly variable
  - Can even flip signs
Why is it a problem?

There are an infinite # of solutions for \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \):

\[ \hat{\beta}_1 = 10, \hat{\beta}_2 = 2 \quad \hat{\beta}_1 = 100, \hat{\beta}_2 = -88 \quad \text{...etc} \]

**Collinearity illustration**

\[ Y = \beta_0 + X_1 \beta_1 + X_2 \beta_2 + \epsilon \quad \beta_0 = 1, \beta_1 = 2, \beta_2 = 4 \]
Correlated Regressors

Intercept

\( X_1 \)

\( X_2 \)

Highly variable over experiments

Inflated for correlated case (green)

\( \beta \)

\( \text{Cov}(\beta) \)

T statistic

stats are smaller (due to inflated variance)
Residuals don’t change

- The designs explain the same amount of variability

Collinearity

- You can’t fix it after the data have been collected
- You can’t tell from the t statistic if you had collinearity
  - FSL has some diagnostics

Variance Inflation Factor

- That matrix is okay, but only points out one type of collinearity
- What if $X_1 = X_2 + X_3$?
  - Cor($X_1, X_2$) or Cor($X_1, X_3$) won’t catch this well
Variance Inflation Factor

- That matrix is okay, but only points out one type of collinearity
- What if $X_1 = X_2 + X_3$?
  - $\text{Cor}(X_1, X_2)$ or $\text{Cor}(X_1, X_3)$ won't catch this well
- Oh wait, that thing in the previous bullet point looks like a regression equation!

Variance Inflation Factor

- If one regressor is a linear combination of others, we could see that in a linear regression
  - Assume 3 regressors
  - To check for any collinearity with $X_1$, run the following regression
    \[
    X_1 = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \epsilon
    \]

Variance Inflation Factor

- If the regression fits well, then we have a problem
- Assess fit of regression by correlating $X_1$ with the estimate of $X_1$ using the model
  \[
  \text{Cor}(X_1, \hat{\beta}_0 + \hat{\beta}_1 X_2 + \hat{\beta}_2 X_3) = \text{Cor}(X_1, \hat{X}_1)
  \]
Example

\[ \begin{bmatrix} X_1 & X_2 & X_3 \\ 1 & 1 & 0 \\ 2 & 2 & 0 \\ 3 & 3 & 0 \\ 4 & 4 & 0 \\ 5 & 5 & 0 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \\ 3 & 0 & 3 \\ 4 & 0 & 4 \\ 5 & 0 & 5 \end{bmatrix} \]

\[ \text{cor}(X_1, X_2) = \text{cor}(X_1, X_3) = .39 \]

What's the solution for the betas?

\[ X_1 = \beta_1 X_2 + \beta_2 X_3 \]
Example

\[
\begin{array}{ccc}
X_1 & X_2 & X_3 \\
1 & 1 & 0 \\
2 & 2 & 0 \\
3 & 3 & 0 \\
4 & 4 & 0 \\
5 & 5 & 0 \\
1 & 0 & 1 \\
2 & 0 & 2 \\
3 & 0 & 3 \\
4 & 0 & 4 \\
5 & 0 & 5 \\
\end{array}
\]

\[\text{cor}(X_1, X_2) = \text{cor}(X_1, X_3) = .39\]

\[X_1 = \beta_1 X_2 + \beta_2 X_3\]

What's the solution for the betas?

\[X_1 = \overline{X}_1 \implies \text{cor}(X_1, \overline{X}_1) = 1\]

What is this correlation?

- Squaring this gives the model's R\(^2\) value

\[R^2 = [\text{cor}(X_1, \overline{X}_1)]^2\]

Finish VIF

- VIF uses the model's R\(^2\)

\[VIF = \frac{1}{1 - R^2} = \frac{1}{1 - [\text{cor}(X_1, \overline{X}_1)]^2}\]

- Goal: VIF < 5
  - As long as it isn't >10, you're probably okay
Summary of collinearity

- Parameter estimates are highly variable when collinearity is present
- Technically, inferences are still valid!
  - Outliers could possibly cause problems
- Power takes a hit due to inflated variance
  - Collinearity is directly related to efficiency

You are now first level modeling experts!

- What canonical HRF will you use in SPM?
- What does adding the derivative do?
- How can you incorporate the derivative's parameter estimate in the activation magnitude

You are now first level modeling experts!

- How many parameters does SPM use in the whitening matrix?
- What two things remove low frequency noise in fMRI data?
- Can you trust p-values from models with collinearity?
You are now first level modeling experts!

- Should you ever model the global mean in task fMRI data?

Break!