First-level fMRI modeling

UCLA Advanced NeuroImaging Summer School, 2008
Recall the GLM

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

\[
\begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n \\
\end{pmatrix} = 
\begin{pmatrix}
1 & X_{11} & X_{21} & X_{31} \\
1 & X_{12} & X_{22} & X_{32} \\
\vdots & \vdots & \vdots & \vdots \\
1 & X_{1n} & X_{2n} & X_{3n} \\
\end{pmatrix} 
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\end{pmatrix} + 
\begin{pmatrix}
\epsilon_1 \\
\epsilon_1 \\
\vdots \\
\epsilon_n \\
\end{pmatrix}
\]

Single voxel time series
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, $e$

\[
\hat{\sigma}^2 = \frac{e'e}{N - p}
\]

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

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

\[ Y = X\beta \]
Simplest Model
Simplest Model

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

\[ t = \frac{0.3}{1.41 \times 0.06} = 3.55 \]
Modeling the delay

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

Stimulus

HRF
(double gamma)
Using the HRF

- Convolve the HRF with stimulus function
Using the HRF

- Convolve the HRF with stimulus function

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

• The good of the canonical hrf
  – Easy to model
  – Easy to interpret
  – Easy to do group analysis

• The bad
  – If it is wrong your results can be biased

• Stay tuned for Martin Lindquist’s talk on Monday
Convolved Boxcar

\[ t = \frac{0.66}{1.22 \times 0.06} = 9.02 \]
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!
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
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 to smooth

– IMPORTANT: Must apply filter to both the data and the design.

  • FSL has ‘apply temporal filter’ box in design setup. Leave it checked!
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)
Model with HP filter

\[ Y = X\beta \]

Parameter of interest
Model with HP filter

\[ Y = X\beta \]

Parameter of interest

\[ H_0 : c\beta = 0 \]

Use contrast

\[ c = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) \]
Convolution & HP filter

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

- $Y = X\beta + \epsilon$
  \[ \text{Cov}(\epsilon) = \sigma^2 V \]

- Gauss Markov: If the error has mean 0, with constant variance and uncorrelated, the least squares estimators are unbiased and have minimum variance among all unbiased linear estimates
Whitening

• If $V$ is known
  – Due to the structure of correlation matrix there exists $K$ such that $V = KK'$

\[
V = \begin{pmatrix}
1 & 0.2 & 0.04 & 0.008 & 0.0016 & 0.0003 \\
0.2 & 1 & 0.2 & 0.04 & 0.008 & 0.0016 \\
0.04 & 0.2 & 1 & 0.2 & 0.04 & 0.008 \\
0.008 & 0.04 & 0.2 & 1 & 0.02 & 0.04 \\
0.0016 & 0.008 & 0.04 & 0.2 & 1 & 0.2 \\
0.0003 & 0.0016 & 0.008 & 0.04 & 0.2 & 1
\end{pmatrix}
\]
Whitening

- If V is known
  - Due to the structure of correlation matrix there exists $K$ such that $V = KK'$
  - Whitened model
    - $K^{-1}Y = K^{-1}X\beta + K^{-1}\epsilon$
    - $Y^* = X^*\beta + \epsilon^*$
    - $\text{Cov}(K^{-1}\epsilon) = \sigma^2 K^{-1}V(K')^{-1}$
      $= \sigma^2 K^{-1} KK'(K')^{-1}$
      $= \sigma^2 I$
Whitening

• OLS can be used on whitened model
  \[ \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \]

\[ \hat{\text{Cov}}(\hat{\beta}) = \hat{\sigma}^2(X'V^{-1}X)^{-1} \]

• But we don’t know V, we have to estimate it
fMRI noise

• Tends to follow 1/f trend

Zarahn et al, 1997, NI
fMRI noise

• Tends to follow $1/f$ trend
• Autoregressive (AR) models fit it well
  \[
  \text{Cor}(\epsilon_i, \epsilon_j) = \rho^{|i-j|}
  \]

\[
\begin{pmatrix}
  1 & 0.2 & 0.04 & 0.008 & 0.0016 & 0.0003 \\
  0.2 & 1 & 0.2 & 0.04 & 0.008 & 0.0016 \\
  0.04 & 0.2 & 1 & 0.2 & 0.04 & 0.008 \\
  0.008 & 0.04 & 0.2 & 1 & 0.02 & 0.04 \\
  0.0016 & 0.008 & 0.04 & 0.2 & 1 & 0.2 \\
  0.0003 & 0.0016 & 0.008 & 0.04 & 0.2 & 1
\end{pmatrix}
\]
Whitening FSL

- Estimates $V$ locally
- Step 1: Estimate raw autocorrelations
Whitening FSL

- Estimates V *locally*
- Step 1: Estimate raw autocorrelations

$$r(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} e(t)e(t + \tau)/(N - \tau)$$

Lag 1

$$[e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \ e_7 \ e_8 \ e_9 \ e_{10} \ e_{11} \ e_{12}]$$

Take products and average
Whitening FSL

- Estimates $V$ locally
- Step 1: Estimate raw autocorrelations

$$r(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} e(t)e(t+\tau)/(N-\tau)$$

Lag 2

$$\begin{bmatrix} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \\
\end{bmatrix}$$

Take products and average
Whitening FSL

• Estimates V locally
• Step 1: Estimate raw autocorrelations

\[ r(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} e(t)e(t+\tau)/(N - \tau) \]

Lag 7

\[
\begin{bmatrix}
e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \\
e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & & & & \\
\end{bmatrix}
\]

Take products and average
Whitening FSL

- Estimates $V$ locally
- Step 1: Estimate raw autocorrelations
- Step 2: Smooth spatially
- Step 3: Within voxel, smooth correlation estimates (Tukey taper)
  - Correlation estimates at high lags aren’t estimated well, so they are down-weighted
Whitening FSL

Woolrich et al., 2001, NI
Whitening FSL

Time Domain

Spectral Domain

Smoothed estimate

Woolrich et al., 2001, NI
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
Whitening SPM

Only 2 parameters are estimated
Convolution, HP filter, Whitening

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

• Add your own correlation
  – Lowpass filter
  – Not good for designs with high frequency
Coloring

• Bias/Variance tradeoff
  – Coloring has less bias but higher variance than whitening (less sensitive)
  – Bias corrected whitening is almost always used now

• Bandpass filtering
  – High and lowpass filtering
Scaling

• **Grand mean scaling**
  – All time series are scaled by the same factor to have approximately the same mean (FSL uses $100^2$)
  – Necessary when combining subjects

• **Intensity normalization**
  – Forces each volume to have the same mean intensity
  – Discouraged since you can lose signal
  – Turned off by default in FSL
Intensity Normalization

without

with

Signal is lost and negative activation artifacts

Junghofer et al, 2004, NI
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?

\[
\begin{pmatrix}
12 \\
12 \\
12 \\
3 \\
3 \\
3 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\end{pmatrix}
= 
\begin{pmatrix}
\beta_1 + \beta_2 \\
\beta_1 + \beta_2 \\
\beta_1 + \beta_2 \\
\beta_3 \\
\beta_3 \\
\beta_3 \\
\end{pmatrix}
\]
Why is it a problem?

\[
\begin{pmatrix}
12 \\
12 \\
12 \\
3 \\
3 \\
3
\end{pmatrix} = \begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3
\end{pmatrix} = \begin{pmatrix}
\beta_1 + \beta_2 \\
\beta_1 + \beta_2 \\
\beta_1 + \beta_2 \\
\beta_3 \\
\beta_3 \\
\beta_3
\end{pmatrix}
\]

- There are an infinite # of solutions for \( \beta_1 \) and \( \beta_2 \)

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

$\hat{\beta}$

Highly variable over experiments

Inflated for correlated case (green)

T statistic

Bias can go in either direction
Correlated Regressors

Intercept

\[ \hat{\beta} \]

Highly variable over experiments

\[ \text{Cov}(\hat{\beta}) \]

Inflated for correlated case (green)

T statistic
Bias can go in either direction
Correlated Regressors

Highly variable over experiments

\[ \hat{\beta} \]

\[ \text{Cov}(\hat{\beta}) \]

Inflated for correlated case (green)

T statistic

Bias can go in either direction
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

Absolute value of correlation
Bad=white off diagonal

Eigenvalues from SVD
Bad=near 0
Have fun in lab!