Last Week

- Paradigm Design
  - Block Design
  - Event-related Design
  - Jittered Design
  - Mixed Design

This week

- The General Linear Model
- fMRI Timeseries
- GLM Demo
- Design Efficiency (if time)

Some Terminology

- SPM ("Statistical Parametric Mapping") is a massively univariate approach - meaning that a statistic (e.g., T-value) is calculated for every voxel - using the "General Linear Model"
- Experimental manipulations are specified in a model ("design matrix") which is fit to each voxel to estimate the size of the experimental effects ("parameter estimates") in that voxel…
- … on which one or more hypotheses ("contrasts") are tested to make statistical inferences ("p-values"), correcting for multiple comparisons across voxels (using "Random Field Theory")
- The parametric statistics assume continuous-valued data and additive noise that conforms to a "Gaussian" distribution ("nonparametric" version SNPM eschews such assumptions)

Some Terminology

- SPM usually focused on "functional specialization" - i.e. localizing different functions to different regions in the brain
- One might also be interested in "functional integration" - how different regions (voxels) interact
- Multivariate approaches work on whole images and can identify spatial/temporal patterns over voxels, without necessarily specifying a design matrix (PCA, ICA)...
- … or with an experimental design matrix (PLS, CVA), or with an explicit anatomical model of connectivity between regions - "effective connectivity" - eg using Dynamic Causal Modeling

Overview

1. General Linear Model

   Design Matrix
   Estimation/Contrasts
   Covariates (eg global)
   Estimability/Correlation

2. fMRI timeseries

   Highpass filtering
   HRF convolution
   Autocorrelation (nonsphericity)

Overview

1. General Linear Model

   Design Matrix
   Estimation/Contrasts
   Covariates (eg global)
   Estimability/Correlation

2. fMRI timeseries

   Highpass filtering
   HRF convolution
   Autocorrelation (nonsphericity)

General Linear Model…

- **Parametric statistics**
  - 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

General Linear Model

- **Equation for single (and all) voxels:**

$$y_j = x_{j1}\beta_1 + \dots x_{jL}\beta_L + \varepsilon_j \qquad \varepsilon_j \sim N(0,\sigma^2)$$

y_j : data for scan, $j = 1 \dots J$
x_{jl} : explanatory variables / covariates / regressors, $l = 1 \dots L$
β_l : parameters / regression slopes / fixed effects
ε_j : residual errors, *independent & identically distributed ("iid")*
(Gaussian, mean of zero and standard deviation of σ)

- **Equivalent matrix form:**

$$y = X\beta + \varepsilon$$

X : "design matrix" / model

Matrix Formulation

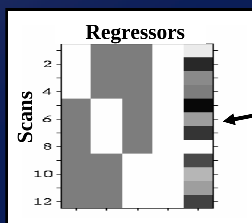
$$Y_j = x_{j1}\beta_1 + \dots + x_{jl}\beta_l + \dots + x_{jL}\beta_L + \epsilon_j$$

Equation for scan j

Simultaneous equations for scans *1.. J*

$$\begin{aligned} Y_1 &= x_{11}\beta_1 + \dots + x_{1l}\beta_l + \dots + x_{1L}\beta_L + \epsilon_1 \\ \vdots &= \vdots \\ Y_j &= x_{j1}\beta_1 + \dots + x_{jl}\beta_l + \dots + x_{jL}\beta_L + \epsilon_j \\ \vdots &= \vdots \\ Y_J &= x_{J1}\beta_1 + \dots + x_{Jl}\beta_l + \dots + x_{JL}\beta_L + \epsilon_J \end{aligned}$$

…that can be solved for parameters $\beta_{1..L}$



$$\underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_j \\ \vdots \\ Y_J \end{pmatrix}}_{J\times 1} = \underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1l} & \cdots & x_{1L} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & \cdots & x_{jl} & \cdots & x_{jL} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{J1} & \cdots & x_{Jl} & \cdots & x_{JL} \end{pmatrix}}_{J\times L} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_l \\ \vdots \\ \beta_L \end{pmatrix}}_{L\times 1} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_l \\ \vdots \\ \epsilon_L \end{pmatrix}}_{J\times 1}$$

Overview

1. General Linear Model
   - Design Matrix
   - Estimation/Contrasts
   - Covariates (eg global)
   - Estimability/Correlation
2. fMRI timeseries
   - Highpass filtering
   - HRF convolution
   - Autocorrelation (nonsphericity)

General Linear Model (Estimation)

- Estimate parameters from least squares fit to data, y:

$$\hat{\beta} = (X^T X)^{-1} X^T y = X^+ y \qquad \textit{(OLS estimates)}$$

- Fitted response is:

$$Y = X\hat{\beta}$$

- Residual errors and estimated error variance are:

$$\hat{\varepsilon} = y - Y \qquad \hat{\sigma}^2 = \hat{\varepsilon}^T \hat{\varepsilon} / df$$

where df are the degrees of freedom (assuming iid):

$$df = J - rank(X) \qquad (=J\text{-}L \text{ if } X \text{ full rank})$$

$$(\; R = I - XX^+ \qquad \varepsilon = Ry \qquad df = trace(R) \;)$$

Simple LS Derivation

$$y_i = \beta_1 x_i + e_i$$

$$e_i = y_i - \beta_1 x_i$$

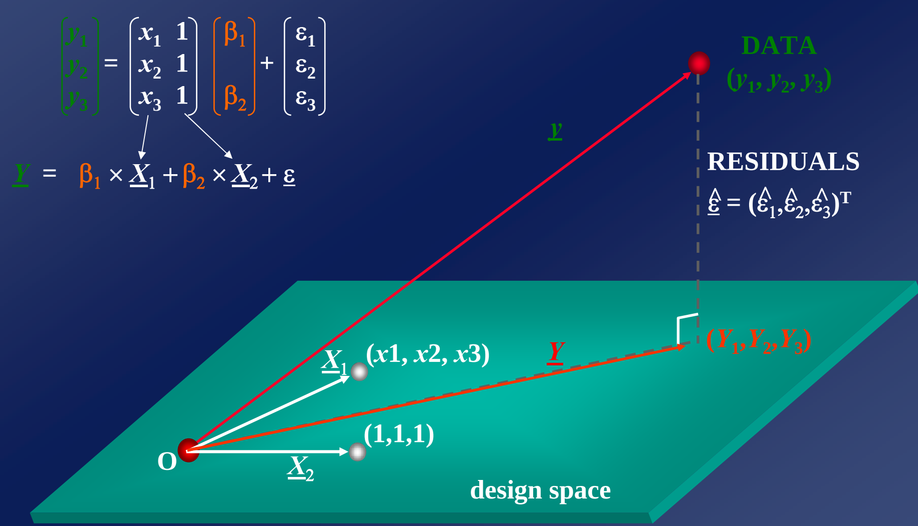
$$\min \Sigma (e_i)^2$$

$$E = \Sigma (y_i - \beta_1 x_i)^2 = \Sigma (y_i^2 - 2\beta_1 x_i y_i + \beta_1^2 x_i^2)$$

$$\frac{dE}{d\beta_1} = \Sigma(-2x_i y_i + 2\beta_1 x_i^2) = 0$$

$$\beta_1 = \frac{\Sigma (x_i y_i)}{\Sigma x_i^2} = \frac{X^T y}{X^T X} = (X^T X)^{-1} X^T y$$

GLM Estimation – Geometric Perspective



General Linear Model (Inference)

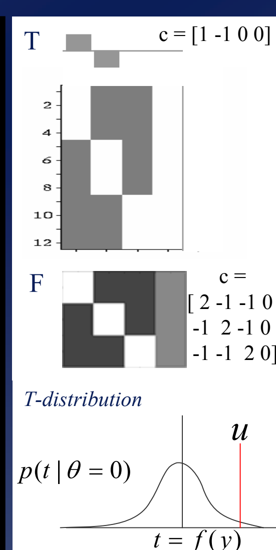
- Specify contrast (hypothesis), c, a linear combination of parameter estimates, $c^T\hat{\beta}$
- Calculate T-statistic for that contrast:

  $T = c^T\hat{\beta} / std(c^T\hat{\beta}) = c^T\hat{\beta} / sqrt(\hat{\sigma}^2 c^T(X^TX)^{-1}c)$

  (c is a vector), or an F-statistic:

  $F = [(\varepsilon_0^T\varepsilon_0 - \varepsilon^T\varepsilon) / (L-L_0)] / [\varepsilon^T\varepsilon / (J-L)]$

  where ε_0 and L_0 are residuals and rank resp. from the reduced model specified by c (which is a matrix)
- Prob. of falsely rejecting Null hypothesis, H_0: $c^T\beta=0$ ("p-value")



Z-score and T-statistic

$$Z = \frac{x - \mu}{\sigma}$$

$$T = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

- *The z-score describes the relative location of a particular score (x) when the mean (μ) and standard deviation (σ) are known*
- *The t-score describes the relative location of the sample mean (x) when the population mean is known and the population standard deviation is estimated with the sample standard deviation (s)*

Simple "ANOVA-like" Example

- 12 scans, 3 conditions (1-way ANOVA)

  $$y_j = x_{1j}\,\beta_1 + x_{2j}\,\beta_2 + x_{3j}\,\beta_3 + x_{4j}\,\beta_4 + \varepsilon_j$$

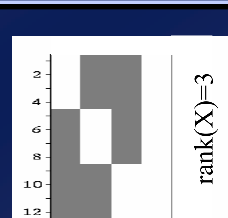
  where (dummy) variables:

  x_{1j} = *[0,1] = condition A (first 4 scans)*
  x_{2j} = *[0,1] = condition B (second 4 scans)*
  x_{3j} = *[0,1] = condition C (third 4 scans)*
  x_{4j} = *[1] = grand mean*

- T-contrast : [1 -1 0 0] tests whether A>B

  [-1 1 0 0] tests whether B>A

- F-contrast: [2 -1 -1 0
  -1 2 -1 0
  -1 -1 2 0] tests main effect of A,B,C



$$\begin{bmatrix} 11 \\ 9 \\ 12 \\ 8 \\ 21 \\ 19 \\ 22 \\ 18 \\ 31 \\ 29 \\ 32 \\ 28 \end{bmatrix} = \begin{bmatrix} 1&0&0&1 \\ 1&0&0&1 \\ 1&0&0&1 \\ 1&0&0&1 \\ 0&1&0&1 \\ 0&1&0&1 \\ 0&1&0&1 \\ 0&1&0&1 \\ 0&0&1&1 \\ 0&0&1&1 \\ 0&0&1&1 \\ 0&0&1&1 \end{bmatrix} \begin{bmatrix} -10 \\ 0 \\ 10 \\ 20 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 2 \\ -2 \\ 1 \\ -1 \\ 2 \\ -2 \\ 1 \\ -1 \\ 2 \\ -2 \end{bmatrix}$$

c=[-1 1 0 0], T=10/sqrt(3.3*8)
df=12-3=9, T(9)=1.94, $p<.05$

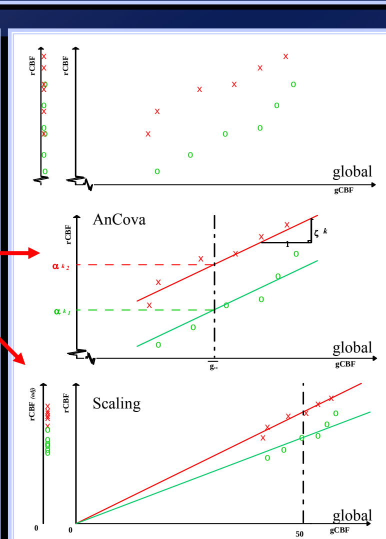
Overview

1. General Linear Model
   - Design Matrix
   - Estimation/Contrasts
   - Covariates (eg global)
   - Estimability/Correlation

2. fMRI timeseries
   - Highpass filtering
   - HRF convolution
   - Autocorrelation (nonsphericity)

Global Effects

- May be variation in overall image intensity from scan to scan
- Such “global” changes may confound local / regional induced by experiment
- Adjust for global effects by:
  - AnCova (Additive Model) - *PET?*
  - Proportional Scaling - *fMRI?*
- Can improve statistics when orthogonal to effects of interest (as here)…
- …but can also worsen when effects of interest correlated with global (as next)



Simple ANCOVA Example

- 12 scans, 3 conditions, 1 confounding covariate

  $y_j = x_{1j}\,\beta_1 + x_{2j}\,\beta_2 + x_{3j}\,\beta_3 + x_{4j}\,\beta_4 + x_{5j}\,\beta_5 + \varepsilon_j$

  where (dummy) variables:

  x_{1j} *= [0,1] = condition A (first 4 scans)*

  x_{2j} *= [0,1] = condition B (second 4 scans)*

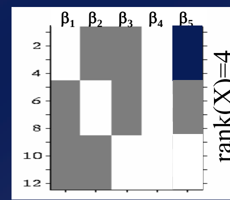
  x_{3j} *= [0,1] = condition C (third 4 scans)*

  x_{4j} *= grand mean*

  x_{5j} *= global signal (mean over all voxels)*

  (further mean-corrected over all scans)

- Global correlated here with conditions (and time)



$$\begin{bmatrix} 11 \\ 9 \\ 12 \\ 8 \\ 21 \\ 19 \\ 22 \\ 18 \\ 31 \\ 29 \\ 32 \\ 28 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.7 \\ 5.0 \\ 8.3 \\ 15 \\ 6.7 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 2 \\ -2 \\ 1 \\ -1 \\ 2 \\ -2 \\ 1 \\ -1 \\ 2 \\ -2 \end{bmatrix}$$

c=[-1 1 0 0], T=3.3/sqrt(3.8*8)
df=12-4=8, T(8)=0.61, p>.05

Global Effects (fMRI)

- Two types of scaling: Grand Mean scaling and Global scaling
- Grand Mean scaling is automatic, global scaling is optional
- Grand Mean scales by 100/mean over all voxels and ALL scans
  (i.e, single number per session)
- Global scaling scales by 100/mean over all voxels for EACH scan
  (i.e, a different scaling factor every scan)

- Problem with global scaling is that TRUE global is not (normally) known…
- …we only estimate it by the mean over voxels
- So if there is a large signal change over many voxels, the global estimate will be confounded by local changes
- This can produce artifactual deactivations in other regions after global scaling
- Since most sources of global variability in fMRI are low frequency (drift), high-pass filtering may be sufficient, and many people to not use global scaling

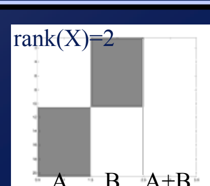
Overview

1. General Linear Model
   - Design Matrix
   - Estimation/Contrasts
   - Covariates (eg global)
   - Estimability/Correlation

2. fMRI timeseries
   - Highpass filtering
   - HRF convolution
   - Autocorrelation (nonsphericity)

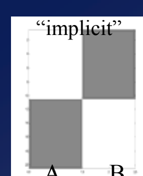
A word on correlation/estimability

- If any column of X is a linear combination of any others (X is rank deficient), some parameters cannot be estimated uniquely (inestimable)
- … which means some contrasts cannot be tested (eg, only if sum to zero)
- This has implications for whether "baseline" (constant term) is explicitly or implicitly modelled
- Rank deficiency can be thought of as perfect correlation…



$c_m = [1\ \ 0\ \ 0]$ ✗

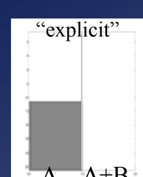
$c_d = [1\ \text{-}1\ \ 0]$ √



$\beta_1 = 1.6$ $\quad c_m = [1\ \ 0]$ √

$\beta_2 = 0.7$ $\quad c_d = [1\ \text{-}1]$ √

$c_d{*}\boldsymbol{\beta} = [1\ \text{-}1]{*}\boldsymbol{\beta} = 0.9$



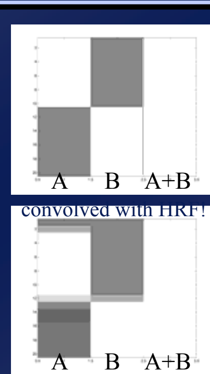
$\beta_1 = 0.9$ $\quad c_d = [1\ \ 0]$ √

$\beta_2 = 0.7$

$c_d{*}\boldsymbol{\beta} = [1\ \ 0]{*}\boldsymbol{\beta} = 0.9$

A word on correlation/estimability

- When there is high (but not perfect) correlation between regressors, parameters can be estimated…
- …but the estimates will be inefficiently estimated (ie highly variable)
- …meaning some contrasts will not lead to very powerful tests
- SPM shows pairwise correlation between regressors…
- …but this will NOT tell you that, eg, X_1+X_2 is highly correlated with X_3…
- … so some contrasts can still be inefficient/efficient, even though *pairwise* correlations are low/high

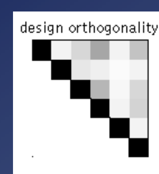


$c_m = [1\ \ 0\ \ 0]$ ✗

$c_d = [1\ \text{-}1\ \ 0]$ √

$c_m = [1\ \ 0\ \ 0]$ (√)

$c_d = [1\ \text{-}1\ \ 0]$ √

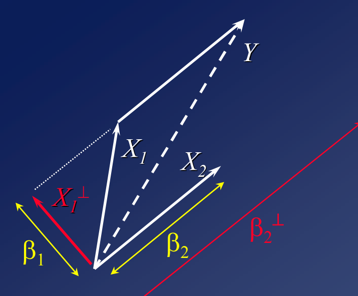


A word on orthogonalization

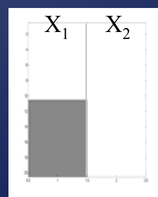
- To remove correlation between two regressors, you can explicitly orthogonalize one (X_1) with respect to the other (X_2):

$$X_1^{\perp} = X_1 - (X_2X_2^{+})X_1 \qquad \text{(Gram-Schmidt)}$$

- Paradoxically, this will NOT change the parameter estimate for X_1, but will for X_2
- In other words, the parameter estimate for the orthogonalized regressor is unchanged!
- This reflects fact that parameter estimates automatically reflect orthogonal component of each regressor…
- …so no need to orthogonalize, UNLESS you have a priori reason for assigning common variance to the other regressor



A word on orthogonalization

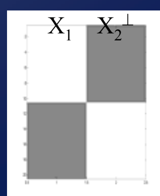


$\beta_1 = 0.9$

$\beta_2 = 0.7$

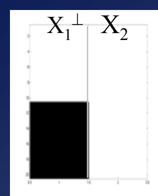
Orthogonalize X_2
(Model M1)

Orthogonalize X_1
(Model M2)



$\beta_1^{(M1)} = 1.6$

$\beta_2^{(M1)} = 0.7$



$\beta_1^{(M2)} = 0.9 \quad = \beta_1^{(M1)} - \beta_2^{(M1)}$

$\beta_2^{(M2)} = 1.1 \quad = (\beta_1^{(M1)} + \beta_2^{(M1)})/2$

Overview

1. General Linear Model
   - Design Matrix
   - Estimation/Contrasts
   - Covariates (eg global)
   - Estimability/Correlation

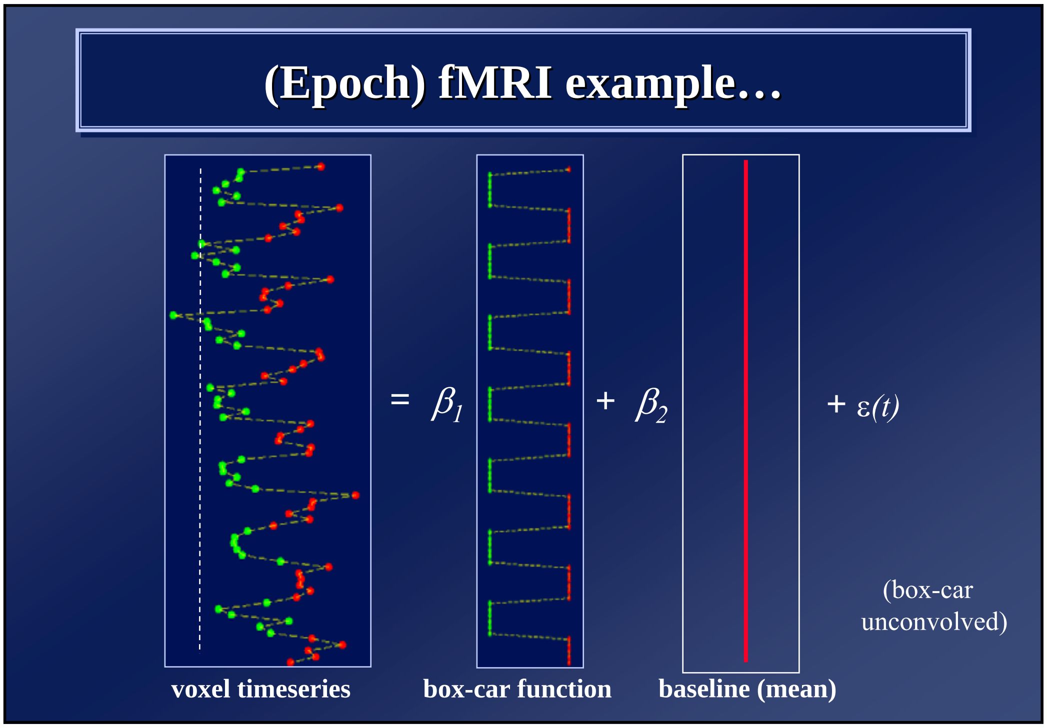
2. fMRI timeseries
   - Highpass filtering
   - HRF convolution
   - Autocorrelation (nonsphericity)

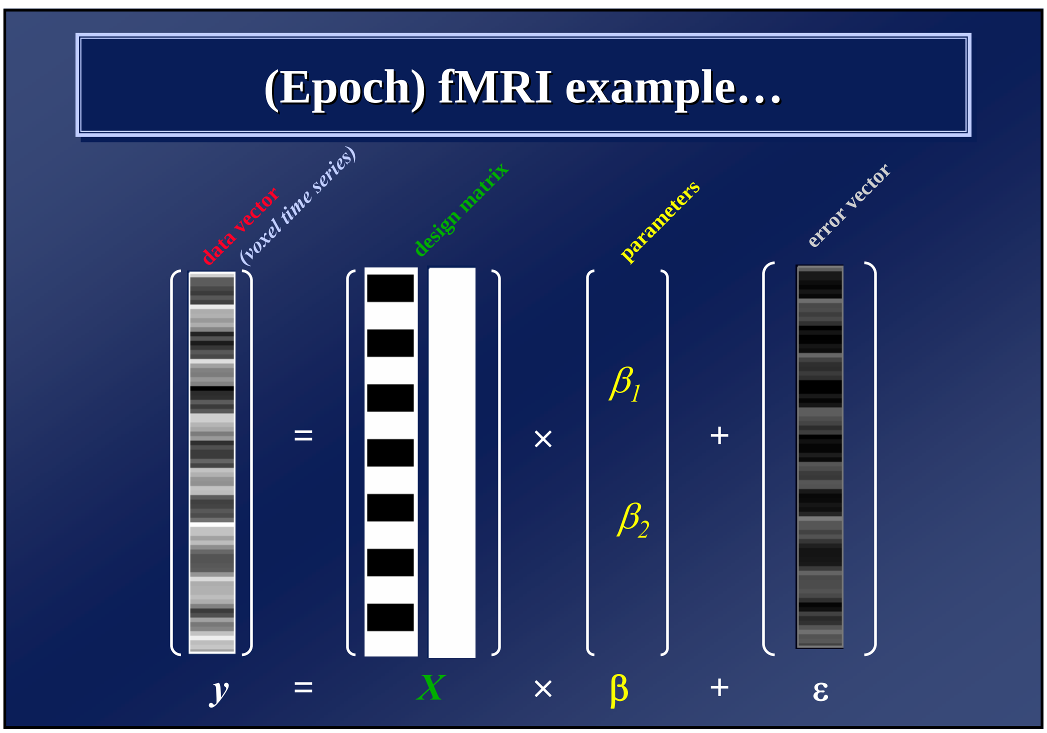
fMRI Analysis

- Scans are treated as a timeseries…
  … and can be filtered to remove low-frequency (1/f) noise
- Effects of interest are convolved with haemodynamic response function (HRF), to capture sluggish nature of (BOLD) response
- Scans can no longer be treated as independent observations…
  … they are typically temporally autocorrelated (for TRs<8s)

fMRI Analysis

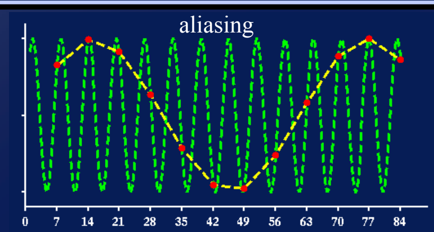
- Scans are treated as a timeseries…
  … and can be filtered to remove low-frequency (1/f) noise
- Effects of interest are convolved with haemodynamic response function (HRF), to capture sluggish nature of (BOLD) response
- Scans can no longer be treated as independent observations…
  … they are typically temporally autocorrelated (for TRs<8s)

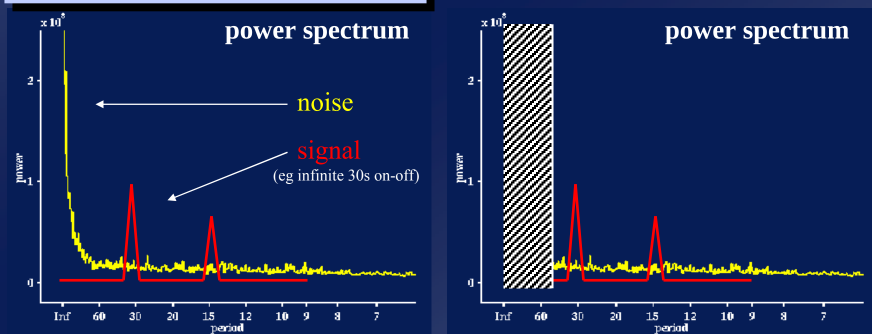

(Epoch) fMRI example…
= β1
+ β2
+ ε(t)
(box-car
unconvolved)
voxel timeseries
box-car function
baseline (mean)


(Epoch) fMRI example…
data vector
(voxel time series)
design matrix
parameters
error vector
=
×
β1
β2
+
y = X × β + ε

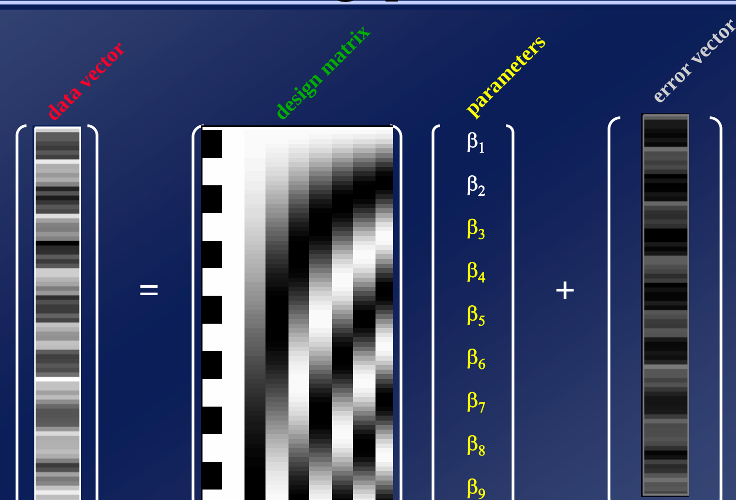
Low frequency noise

- Low frequency noise:
  - Physical (scanner drifts)
  - Physiological (aliased)
    - cardiac (~1 Hz)
    - respiratory (~0.25 Hz)



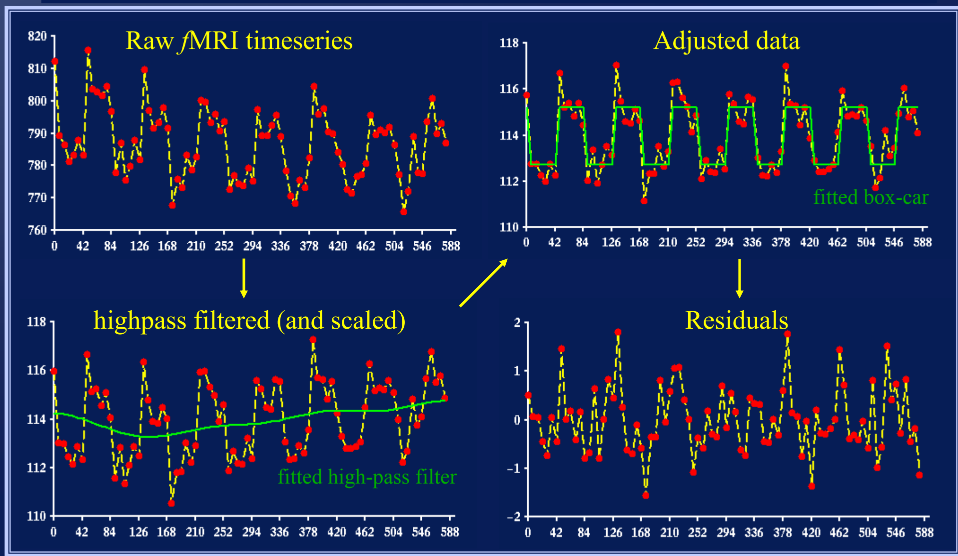


(Epoch) fMRI example… …with highpass filter



$$y = X \times \beta + \varepsilon$$

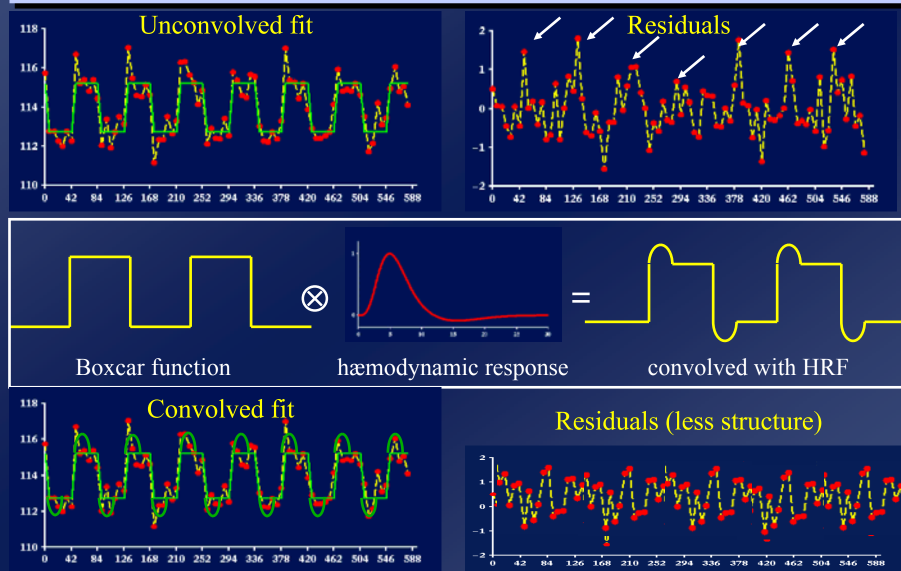
(Epoch) fMRI example…
…fitted and adjusted data



fMRI Analysis

- Scans are treated as a timeseries…
  … and can be filtered to remove low-frequency (1/f) noise
- Effects of interest are convolved with haemodynamic response function (HRF), to capture sluggish nature of (BOLD) response
- Scans can no longer be treated as independent observations…
  … they are typically temporally autocorrelated (for TRs<8s)

Convolution with HRF



fMRI Analysis

- Scans are treated as a timeseries…
  … and can be filtered to remove low-frequency (1/f) noise
- Effects of interest are convolved with haemodynamic response function (HRF), to capture sluggish nature of (BOLD) response
- Scans can no longer be treated as independent observations…
  … they are typically temporally autocorrelated (for TRs<8s)

Temporal autocorrelation…

- Because the data are typically correlated from one scan to the next, one cannot assume the degrees of freedom (dfs) are simply the number of scans minus the dfs used in the model – need "effective degrees of freedom"
- In other words, the residual errors are not independent:

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N(0,\sigma^2 V) \qquad V \neq I,\ V = AA'$$

where A is the intrinsic autocorrelation

- Generalized least squares:

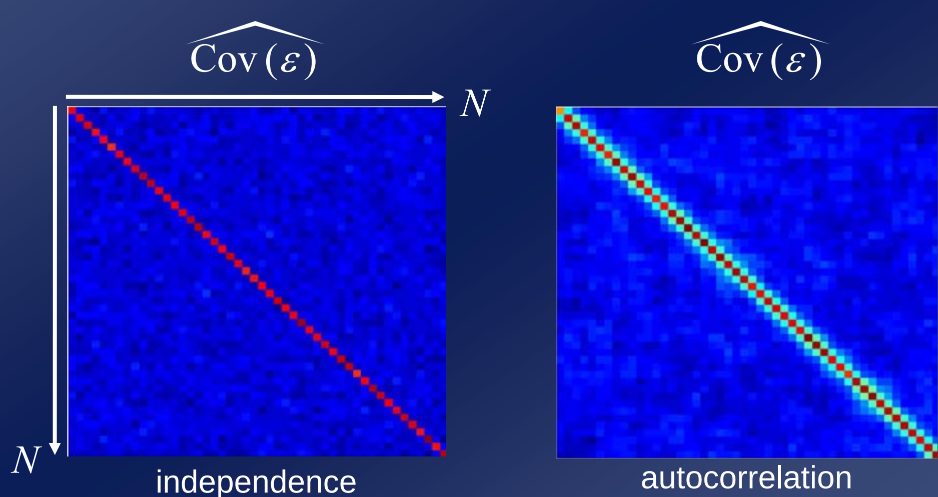
$$KY = KX\beta + K\varepsilon \qquad K\varepsilon \sim N(0,\ \sigma^2 V) \qquad V = KAA'K'$$

(autocorrelation is a special case of "nonsphericity"…)

Linear Model Errors

$Y = X\beta + \varepsilon$

Two sample covariance matrices...



Conclusion

- GLM
  - Encompasses almost any statistical model you need
  - Important to account for fMRI autocorrelation
- Contrasts
  - Way of “interrogating” GLM
  - T contrasts (is this effect positive [negative]?)
  - F contrasts
    (are one or more of these effects different from zero?)