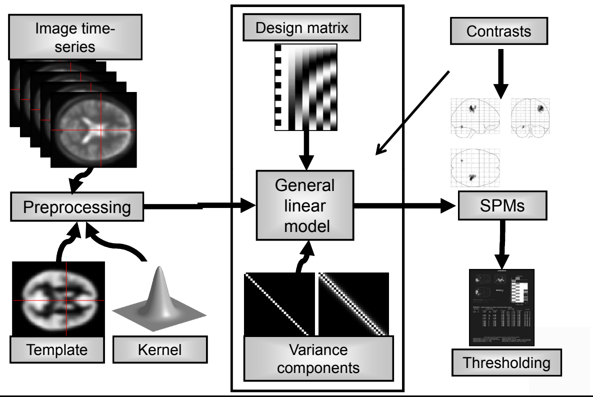
Group Inference, Non-sphericity & Covariance Components in SPM

Alexa Morcom
Edinburgh SPM course, April 2011

Centre for Cognitive & Neural Systems/
Department of Psychology
University of Edinburgh

SPM

Overview of SPM



Overview

- Making the group inferences we want
  - Two stage GLM revisited
- Non-sphericity
- Beyond Ordinary Least Squares
  - Non-sphericity at the first level
  - Multiple Covariance Components
- Model estimation
- A word on power

2-stage GLM

'Summary statistic' random effects method

Single subject

Each has an independently acquired set of data
These are modelled separately
Models account for **within subjects variability**
Parameter estimates apply to individual subjects

1st level

Single subject **contrasts of parameter estimates** taken forward to 2nd level as (spm_con*.img) **'con images'**

Group/s of subjects

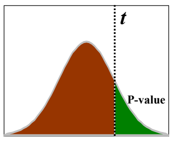
To make population inferences, 2nd level models account for **between subjects variability**
Parameter estimates apply to group effect/s

2nd level

Statistics compare **contrasts of 2nd level parameter estimates to 2nd level error**

Models for fMRI

1. Non-sphericity & why it matters
2. Hierarchical models
   - Why they are needed
   - Issues and SPM solutions
3. We need to estimate
   - Effect magnitude
   - Effect variability
   - p values



Null Distribution of T

$$p(T > t \mid H_0)$$

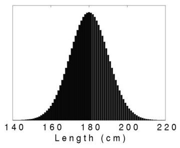
$$t = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

Covariance and non-sphericity

- Classical inference is about what is surprising
- Compare observed (estimated) parameters with their expected behaviour under the null hypothesis
- A statistic is formed from estimates of effects and their variability, but how surprising is this?
- Degrees of freedom must reflect how related (correlated) different observations are
- If observations are not independent (i.e. covary), then there are fewer observations than we think, and the significance of statistics is overrated

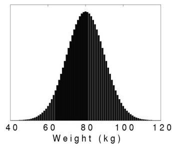
Variance

Length of men



μ=180cm, σ=14cm (σ^2=200)

Weight of men



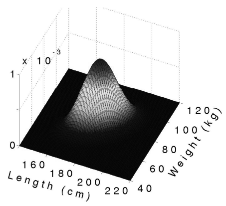
μ=80kg, σ=14kg (σ^2=200)

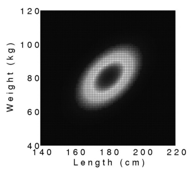
Each 1-dimensional variable is completely characterised by μ (mean) and σ^2 (variance)

i.e. can calculate $p(l|\mu,\sigma^2)$ for any *l* and $p(w|\mu,\sigma^2)$ for any *w*

Variance-covariance matrix

- Can also view length and weight as a 2-dimensional stochastic variable (p(*l*,*w*)).





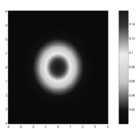
$p(l,w|\boldsymbol{\mu},\boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} 180 \\ 80 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 200 & 100 \\ 100 & 200 \end{bmatrix}$$

Length and weight are related – i.e., covary

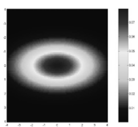
What is (and isn't) sphericity?

sphericity => i.i.d.
error covariance is a multiple of the identity matrix:
Cov(e) = σ^2I



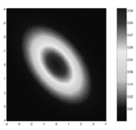
$$Cov(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Examples of non-sphericity:



$$Cov(e) = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

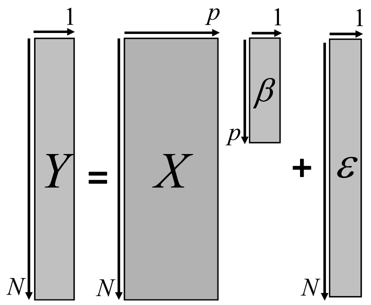
non-identity



$$Cov(e) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

non-independence

The voxel-wise GLM revisited



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

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

Estimate with **Ordinary Least Squares** (OLS)

N: number of scans
p: number of regressors

Model is specified by
1. Design matrix X
2. **Assumptions about ε**

Ordinary Least Squares estimation revisited

Find $\hat{\beta}$ that minimises $\|y - X\beta\|^2 = \varepsilon^T \varepsilon$

The Ordinary Least Squares parameter estimates are:

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

Under i.i.d. assumptions i.e. sphericity, these estimates are unbiased, and have maximum precision (minimum variance)

$$\varepsilon \sim N(0, \sigma^2 I) \longrightarrow Y \sim N(X\beta, \sigma^2 I)$$

$$\hat{\beta} \sim N(\beta, \boxed{\sigma^2 (X^T X)^{-1}})$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - p}$$

Estimate of error variance

Covariance of parameter estimates

Ordinary Least Squares conditions

- Estimated covariance of parameter estimates

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

  $$\hat{C}_\varepsilon = \sigma^2 \, \mathrm{I} \qquad \text{i.i.d.}$$

- Estimation is direct – find the (pseudo) inverse of the design matrix X & multiply data by it
- This works because there is a single covariance component, the variance σ^2
- But only valid if errors are i.i.d. because covariance affects the statistics…

Covariance and statistics

$$T = \frac{c^T \hat{\beta}}{\sqrt{\mathrm{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \sim t_{N-p}$$

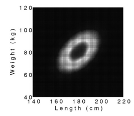
$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

- How good an estimator (precise) is β?
- How much do we think betas covary? – a minimum C_β maximises T
- df are also a function of C_ε & design matrix X…

The traditional solution (e.g. SPSS)

- A measure of departure from sphericity: ε
- Using ε, distribution of SS ratios is approximated by F with Greenhouse-Geisser df – i.e. fewer

Heights & weights



$$\Sigma = \begin{bmatrix} 200 & 100 \\ 100 & 200 \end{bmatrix} \quad \varepsilon = 0.8$$

= Satterthwaite correction

(in theory sl. liberal – but see Mumford & Nichols, 2009)

Sphericity, df and surprise

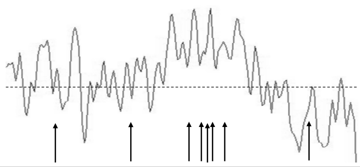
How much do the following observations tell us?

Rain on 4 consecutive days in June

Rain on the same day in May, June, July and August

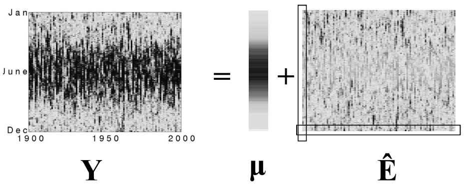
…which is more likely to indicate a wet summer?

Can we determine the patterns of correlation?



The rain in Bergen

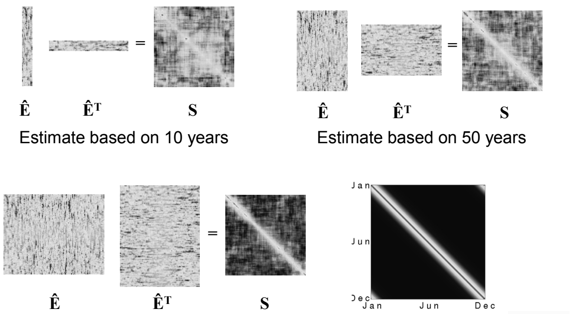
12 months for 100 years



A simple GLM: model monthly rainfall using mean

Data from whole 20th century

Estimating nonsphericity



Estimate based on 10 years

Estimate based on 50 years

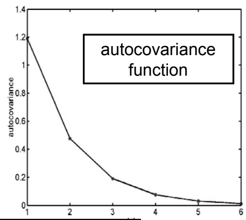
Estimate based on 100 years

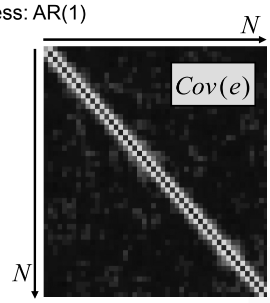
True **Σ** – as if there were not 100*365=36500 data points, but 2516!

Serial correlations in fMRI

$e_t = ae_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,\sigma^2)$

1st order autoregressive process: AR(1)





Also: high-pass filtering

Dealing with serial correlations

Pre-whitening

- Use an enhanced noise model with multiple error covariance components

  i.e. $e \sim N(0, \sigma^2 V)$ instead of $e \sim N(0, \sigma^2 I)$

  V is modelled using AR (1) + white noise model estimated across all active voxels
- Use the estimated V to specify a filter matrix *W* for whitening the data – *'undoing'* the serial correlations

$$We \sim N(0, \sigma^2 W^2 V)$$
$$\Rightarrow W^2 V = I$$
$$\Rightarrow W = V^{-1/2}$$

$$Wy = WX\beta + We$$

Dealing with serial correlations

- Once data are 'pre-whitened', estimation can proceed using Ordinary Least Squares
- The parameter estimates are again optimal – unbiased and minimum variance
- This is Generalised Least Squares (GLS)
- However
  - How do we estimate V?
  - How robust is this method?

Prewhitening in SPM

- Model using $e_t = ae_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$
- 1st order autoregressive process: AR(1)
  - Cannot be estimated precisely at each voxel
  - But precision is key, or estimates are worse than OLS – biased and imprecise
  - Use spatial regularisation
  - Pool estimation over active voxels, defined using 1st pass OLS estimate ($P < .001$)
- PLUS White noise – voxel-specific variance σ^2
- AND – this introduces another issue...

Discovering the 'colour'

- In order to prewhiten we want to know the error covariance
  - Estimate it using C_ε - BUT now not multiple of I
  - $C_\varepsilon = \hat{e}\hat{e}^T + X C_\beta X^T$
  - C_β is a function of C_ε!
- So to prewhiten we need to know
  - Covariance of residuals
  - Covariance of parameter estimates that produced the residuals

•…Use EM/ ReML

$$C_{\hat{\beta}} = (X^T C_\varepsilon^{-1} X)^{-1}$$
$$\hat{C}_\varepsilon \neq \sigma^2 \mathrm{I}$$

Why bother with 2 stages?

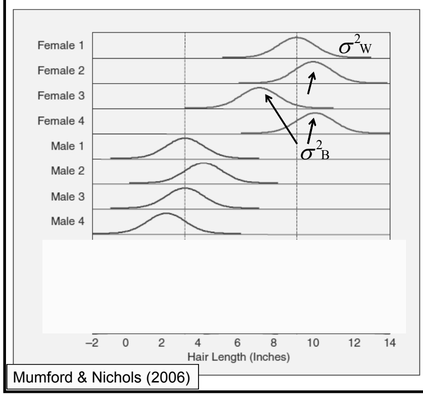
- We want to make an inference to the population, not a single subject, so why do we care?
- Why can't we just do group stats on data for each voxel, as in SPSS?

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

Use if data Y are simple values per voxel – precisely **known**

Estimate with **Ordinary Least Squares** (OLS)

Hierarchical models



Mumford & Nichols (2006)

Does hair length differ by gender?

2 sources of variability

Within-subject: σ^2_W

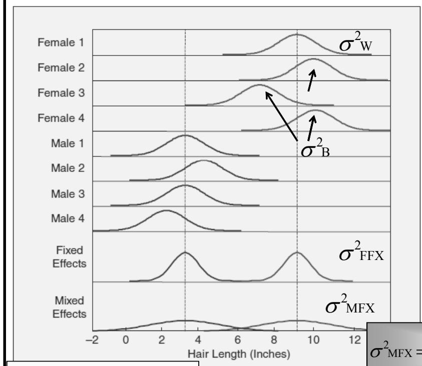
Between-subjects: σ^2_B

To generalise across this sample, combine data from hairs measured in all subjects, get σ^2_{FFX}

To generalise to population, use estimates of hair length for each subject, get σ^2_{MFX}

MIX of between/ within variability

Hierarchical models



Mumford & Nichols (2006)

Does hair length differ by gender?

2 sources of variability

Within-subject (1)

Between-subjects (49)

To generalise across this sample if p = 25 hairs per subject

$$\sigma^2_{FFX} = \frac{1}{4} * \frac{\sigma^2_W}{25} = 0.01$$

To generalise to population, given N = 4 subjects per group

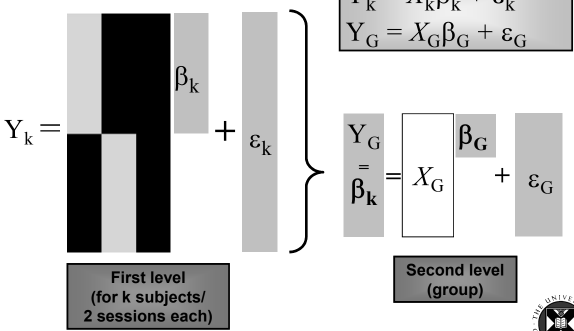
$$\sigma^2_{MFX} = \frac{1}{4} * \frac{\sigma^2_W}{25} + \boxed{\frac{1}{4}} \sigma^2_B = 12.26$$

Why bother with 2 stages?

- We want to make an inference to the population, not a single subject, so why do we care?
- Why can't we just do group stats on data for each voxel, as in SPSS?
- ...that could be valid but would not be optimal
- Hierarchical models deal with mixed sources of variance, not just between-subject variance
- Model both scan-to-scan and subject-to-subject variability

A hierarchical model for fMRI

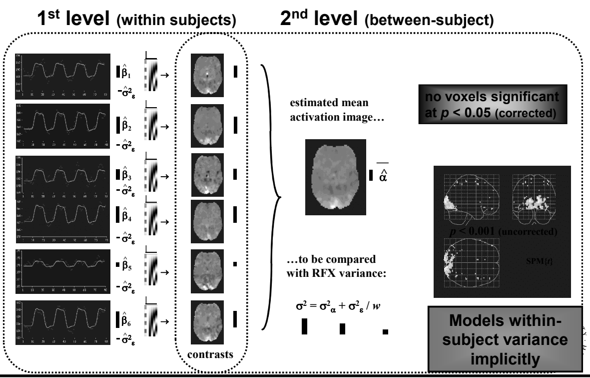
$$Y_k = X_k\beta_k + \varepsilon_k$$
$$Y_G = X_G\beta_G + \varepsilon_G$$



Hierarchical modelling in SPM

- Two approaches
  1. Simple summary statistic – Holmes & Friston
  2. Non-sphericity modelling at group level
- Pros and cons – assumptions vs. flexibility
  - Subject variances equivalent
  - Subject design matrices equivalent
  - (2) enables a wide range of 2nd level models

Summary statistic ‘HF’ approach



Simple HF approach - assumptions

- Distribution
  - Normality, independent subjects
- Homogeneous variance
  - Subjects’ residual errors same
  - Subjects’ design matrices same

  - 2 covariance components
  - Collapse into 1 if the elements of Cov(Y_G) are homogenous over subjects

$$Y_G = X_G\beta_G + \varepsilon_G^m$$
$$\mathrm{Cov}(\varepsilon_G^m) = \mathrm{Cov}(Y_G) + \sigma_G^2 \mathrm{I}_N$$
$$\mathrm{Cov}(Y_G) = \sum_i \sigma_i^2 c (X_i' V_i^{-1} X_i)^{-1} c'$$

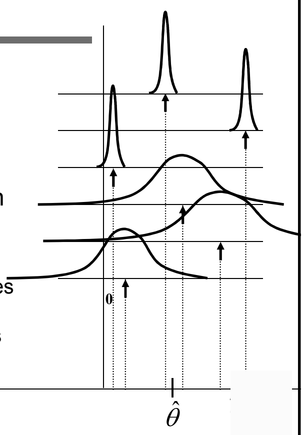
Simple HF approach

- Only single image per subject
- Limits analysis to 1- or 2-sample t-tests at the 2nd level
- Balanced designs

- Limitation = strength
  - No 2nd level sphericity assumption
  - 'Partitioned' error term @ 2nd level

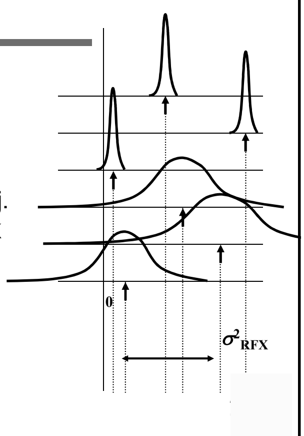
HF – efficiency

- If assumptions true
  - Optimal, fully efficient
- If σ^2_{FFX} differs between subjects
  - Reduced efficiency
  - Here, optimal $\hat{\theta}$ requires down-weighting the 3 highly variable subjects



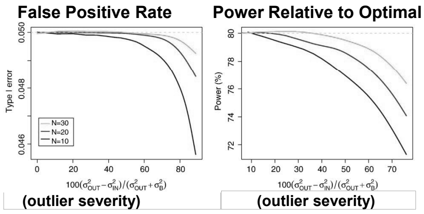
HF – validity

- If assumptions true
  - Exact P-values
- If σ^2_{FFX} differs btw subj.
  - Standard errors not OK
    - Est. of σ^2_{RFX} may be biased
  - df not OK
    - Here, 3 Ss dominate
    - DF < 5 = 6-1



HF – robustness

- In practice, Validity & Efficiency are excellent
  - For one sample case, HF very robust



  - Potential concern with 2-sample or correlation if outliers/ large imbalance

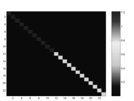
A more flexible approach

- Can model non-sphericity at the 2nd level
- Model within-level just as at 1st level
- Represent different sources of covariance using linear combination of basis functions
- Multiple covariance components
  - Need to estimate using ReML as at 1st level
  - Prewhitening approach, cross-voxel 'pooling'

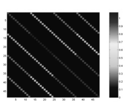
Modelling 2nd level covariance

Error Covariance

- Errors are independent but not identical



- Errors are not independent and not identical

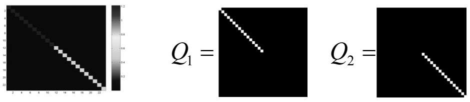


Non-identical data

Errors can be Independent but Non-Identical when…

1) One parameter but from different groups – 2-sample t-test

e.g. patients and control groups



Non-independent data

Error can be Non-Independent and Non-Identical when…

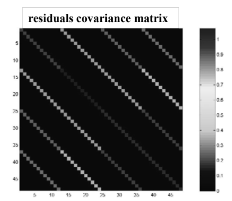
Several contrasts per subject are taken to 2nd level

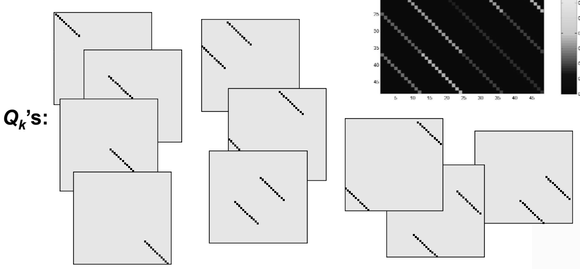
e.g. Repeated Measures ANOVA

Omnibus test is needed across several basis functions characterising the hemodynamic response
e.g. F-test combining HRF, temporal derivative and dispersion regressors

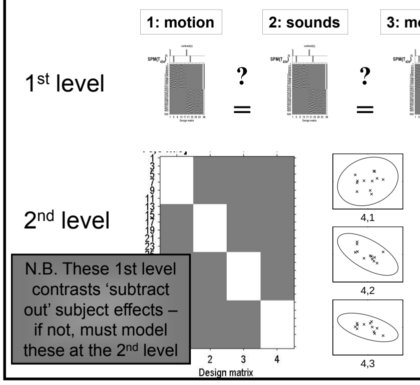
Non-independent data

Errors are not independent and not identical





Example



N.B. These 1st level contrasts 'subtract out' subject effects – if not, must model these at the 2nd level

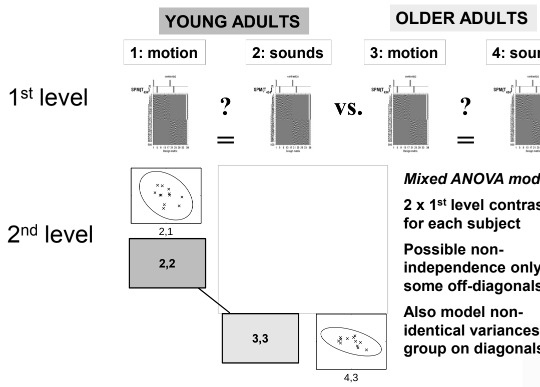
Block design study

Repeated measures ANOVA model

Which regions are sensitive to semantic content of words across 4 conditions?

Noppeney et al.

Example



Mixed ANOVA model

2 x 1st level contrasts for each subject

Possible non-independence only on some off-diagonals

Also model non-identical variances by group on diagonals

A more flexible approach

- Assumptions
  - Fewer than HF but may be more at risk of violations
  - of cross-voxel pooling, homogenous across 'active' voxels
  - Within subject covariance still homogenous
- Advantages
  - Fast relative to 'full' mixed-effects procedures
  - Flexibility of 2nd level models e.g. Multiple basis functions

Summary

- fMRI models need to take account of
  - Multiple sources of variability at 1st level
  - Hierarchical nature of data
  - Multiple sources of variability at 2nd level
- If estimate correctly, get maximum precision, unbiased estimates of parameters & errors
  - Iterative methods are used (EM/ ReML)
  - Spatial regularisation by cross-voxel pooling
  - SPM8 enables very flexible 2nd level models

Bibliography

- *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2007.
- *Generalisability, Random Effects & Population Inference*. Holmes & Friston, NeuroImage,1999.
- *Classical and Bayesian inference in neuroimaging: theory*. Friston et al., NeuroImage, 2002.
- *Classical and Bayesian inference in neuroimaging: variance component estimation in fMRI*. Friston et al., NeuroImage, 2002.
- Simple group fMRI modeling and inference.
  Mumford & Nichols, *Neuroimage*, 2009
- *Flexible factorial tutorial* by Glascher and Gitelman www.sbirc.ed.ac.uk/cyril/cp_fmri.html

Many thanks to J Andersson, J Daunizeau, R Henson, A Holmes, S Kiebel, T Nichols for slides