Statistical Inference

Guillaume Flandin
Wellcome Trust Centre for Neuroimaging
University College London
Image time-series

- Realignment
  - Normalisation
  - Anatomical reference

- Spatial filter

- Smoothing

- Design matrix

- General Linear Model

- Parameter estimates

- Statistical Parametric Map

- Statistical Inference

- RFT

- p < 0.05
A mass-univariate approach
Estimation of the parameters

i.i.d. assumptions: \[ \varepsilon \sim N(0, \sigma^2 I) \]

OLS estimates: \[ \hat{\beta} = (X^T X)^{-1} X^T y \]

\[ \hat{\beta}_1 = 3.9831 \]

\[ \hat{\beta}_{2-7} = \{0.6871, 1.9598, 1.3902, 166.1007, 76.4770, -64.8189\} \]

\[ \hat{\beta}_8 = 131.0040 \]

\[ \hat{\varepsilon} = \]

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

\[ \hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N-p} \]
Contrasts

- A contrast selects a specific effect of interest.
  - A contrast \(c\) is a vector of length \(p\).
  - \(c^T \beta\) is a linear combination of regression coefficients \(\beta\).

\[
c = [1 \ 0 \ 0 \ 0 \ ...]^T
\]

\[
c^T \beta = 1 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3 + 0 \times \beta_4 + \ldots
= \beta_1
\]

\[
c = [0 \ 1 \ -1 \ 0 \ ...]^T
\]

\[
c^T \beta = 0 \times \beta_1 + 1 \times \beta_2 + -1 \times \beta_3 + 0 \times \beta_4 + \ldots
= \beta_2 - \beta_3
\]

\[c^T \hat{\beta} \sim N(c^T \beta, \sigma^2 c^T (X^TX)^{-1} c)\]
Hypothesis Testing

To test an hypothesis, we construct “test statistics”.

- **Null Hypothesis $H_0$**
  Typically what we want to disprove (no effect).
  ⇒ The Alternative Hypothesis $H_A$ expresses outcome of interest.

- **Test Statistic $T$**
  The test statistic summarises evidence about $H_0$.
  Typically, test statistic is small in magnitude when the hypothesis $H_0$ is true and large when false.
  ⇒ We need to know the distribution of $T$ under the null hypothesis.
Hypothesis Testing

- **Significance level $\alpha$:**
  Acceptable *false positive rate* $\alpha$.
  $\Rightarrow$ threshold $u_\alpha$
  Threshold $u_\alpha$ controls the false positive rate
  $$\alpha = p(T > u_\alpha \mid H_0)$$

- **Conclusion about the hypothesis:**
  We reject the null hypothesis in favour of the alternative hypothesis if $t > u_\alpha$

- **$p$-value:**
  A *$p$-value* summarises evidence against $H_0$.
  This is the chance of observing value more extreme than $t$ under the null hypothesis.
  $$p(T > t \mid H_0)$$
**T-test** - one dimensional contrasts – SPM\{t\}

\[ c^T = 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \]

Null hypothesis:
\[ H_0: c^T \beta = 0 \]

Question:
box-car amplitude > 0 ?
\[ \beta_1 = c^T \beta > 0 ? \]

Test statistic:
\[ T = \frac{c^T \hat{\beta}}{\sqrt{\text{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} \sim t_{N-p} \]
\( T \)-contrast in SPM

For a given contrast \( c \):

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

beta_???? images

\[
\sigma^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{N - p}
\]

ResMS image

con_???? image

\[
c^T \hat{\beta}
\]

spmT_???? image

SPM\{\}
**T-test**: a simple example

- Passive word listening versus rest

\[
c^T = [1 \, 0 \, 0 \, 0 \, 0 \, 0 \, 0 \, 0]
\]

Q: activation during listening?

Null hypothesis: \( \beta_1 = 0 \)

\[
t = \frac{c^T \hat{\beta}}{\sqrt{\text{var}(c^T \hat{\beta})}}
\]

**SPM results:**
Height threshold \( T = 3.2057 \) \( \{p<0.001\} \)

<table>
<thead>
<tr>
<th>T</th>
<th>Z</th>
<th>( p_{\text{uncorrected}} )</th>
<th>mm</th>
<th>mm</th>
<th>mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.94</td>
<td>Inf</td>
<td>0.000</td>
<td>-63</td>
<td>-27</td>
<td>15</td>
</tr>
<tr>
<td>12.04</td>
<td>Inf</td>
<td>0.000</td>
<td>-48</td>
<td>-33</td>
<td>12</td>
</tr>
<tr>
<td>11.82</td>
<td>Inf</td>
<td>0.000</td>
<td>-66</td>
<td>-21</td>
<td>6</td>
</tr>
<tr>
<td>13.72</td>
<td>Inf</td>
<td>0.000</td>
<td>57</td>
<td>-21</td>
<td>12</td>
</tr>
<tr>
<td>12.29</td>
<td>Inf</td>
<td>0.000</td>
<td>63</td>
<td>-12</td>
<td>-3</td>
</tr>
<tr>
<td>9.89</td>
<td>7.83</td>
<td>0.000</td>
<td>57</td>
<td>-39</td>
<td>6</td>
</tr>
<tr>
<td>7.39</td>
<td>6.36</td>
<td>0.000</td>
<td>36</td>
<td>-30</td>
<td>-15</td>
</tr>
<tr>
<td>6.84</td>
<td>5.99</td>
<td>0.000</td>
<td>51</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>6.36</td>
<td>5.65</td>
<td>0.000</td>
<td>-63</td>
<td>-54</td>
<td>-3</td>
</tr>
<tr>
<td>6.19</td>
<td>5.53</td>
<td>0.000</td>
<td>-30</td>
<td>-33</td>
<td>-18</td>
</tr>
<tr>
<td>5.96</td>
<td>5.36</td>
<td>0.000</td>
<td>36</td>
<td>-27</td>
<td>9</td>
</tr>
<tr>
<td>5.84</td>
<td>5.27</td>
<td>0.000</td>
<td>-45</td>
<td>42</td>
<td>9</td>
</tr>
<tr>
<td>5.44</td>
<td>4.97</td>
<td>0.000</td>
<td>48</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>5.32</td>
<td>4.87</td>
<td>0.000</td>
<td>36</td>
<td>-27</td>
<td>42</td>
</tr>
</tbody>
</table>
**T-test: summary**

- T-test is a *signal-to-noise* measure (ratio of estimate to standard deviation of estimate).

- Alternative hypothesis:
  
  \[
  H_0: \quad c^T \beta = 0 \quad \text{vs} \quad H_A: \quad c^T \beta > 0
  \]

- T-contrasts are simple combinations of the betas; the T-statistic does not depend on the scaling of the regressors or the scaling of the contrast.
The $T$-statistic does not depend on the scaling of the regressors.

The $T$-statistic does not depend on the scaling of the contrast.

Contrast $c^T \hat{\beta}$ depends on scaling.

Be careful of the interpretation of the contrasts $c^T \hat{\beta}$ themselves (eg, for a second level analysis):

$T = \frac{c^T \hat{\beta}}{\sqrt{\text{var}(c^T \hat{\beta})}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}$
**F-test** - the extra-sum-of-squares principle

- Model comparison:

  **Null Hypothesis** $H_0$: True model is $X_0$ (reduced model)

  **Test statistic**: ratio of explained variability and unexplained variability (error)

  $$ F \propto \frac{RSS_0 - RSS}{RSS} $$

  $$ F \propto \frac{ESS}{RSS} \sim F_{v_1,v_2} $$

  $v_1 = \text{rank}(X) - \text{rank}(X_0)$

  $v_2 = N - \text{rank}(X)$
**F-test** - multidimensional contrasts – SPM$\{F\}$

- Tests multiple linear hypotheses:

  $H_0$: True model is $X_0$

  $H_0$: $\beta_4 = \beta_5 = \ldots = \beta_9 = 0$

  Test $H_0$: $c^T\beta = 0$?

$X_0 \mid X_1 (\beta_{4-9}) \mid X_0$

$X_0$

$X_0$

$X_0$

$c^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
F-contrast in SPM

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

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

\[ \text{ess}_???? \text{ images} \]

\[ (\text{RSS}_0 - \text{RSS}) \]

\[ \text{spmF}_???? \text{ images} \]

\[ \text{SPM}\{F\} \]
$F$-test example: movement related effects
**F-test: summary**

- F-tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler (nested) model ⇒ *model comparison*.

- F tests a weighted **sum of squares** of one or several combinations of the regression coefficients $\beta$.

- In practice, we don’t have to explicitly separate $X$ into $[X_1 X_2]$ thanks to **multidimensional contrasts**.

- Hypotheses:
  
  $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

  Null Hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

  $\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$

  Alternative Hypothesis $H_A : \text{at least one } \beta_k \neq 0$

- In testing uni-dimensional contrast with an F-test, for example $\beta_1 - \beta_2$, the result will be the same as testing $\beta_2 - \beta_1$. It will be exactly the square of the $t$-test, testing for both positive and negative effects.
Orthogonal regressors

Variability described by $X_1$

Testing for $X_1$

Variability described by $X_2$

Testing for $X_2$

Variability in $Y$
Correlated regressors

Variability described by $X_1$

Variability described by $X_2$

Shared variance

Variability in $Y$
Correlated regressors

Testing for $X_1$

Variability described by $X_1$

Variability described by $X_2$

Variability in $Y$
Correlated regressors

Variability described by $X_1$

Variability described by $X_2$

Testing for $X_2$

Variability in $Y$
Correlated regressors

Variability described by $X_1$

Variability described by $X_2$

Variability in $Y$
Correlated regressors

Testing for $X_1$

Variability described by $X_1$

Variability described by $X_2$

Variability in $Y$
Correlated regressors

Testing for $X_2$

Variability described by $X_1$

Variability described by $X_2$

Variability in $Y$
Correlated regressors

Testing for $X_1$ and/or $X_2$

Variability described by $X_1$

Variability described by $X_2$

Variability in $Y$
For each pair of columns of the design matrix, the orthogonality matrix depicts the magnitude of the cosine of the angle between them, with the range 0 to 1 mapped from white to black.

If both vectors have zero mean then the cosine of the angle between the vectors is the same as the correlation between the two variates.
Correlated regressors: summary

- We implicitly test for an additional effect only. When testing for the first regressor, we are effectively removing the part of the signal that can be accounted for by the second regressor:
  ⇒ implicit orthogonalisation.

\[ x_{2}^\perp = x_{2} - x_{1} \cdot x_{2} \cdot x_{1} \]

- Orthogonalisation = decorrelation. Parameters and test on the non modified regressor change.
  Rarely solves the problem as it requires assumptions about which regressor to uniquely attribute the common variance.
  ⇒ change regressors (i.e. design) instead, e.g. factorial designs.
  ⇒ use F-tests to assess overall significance.

- Original regressors may not matter: it’s the contrast you are testing which should be as decorrelated as possible from the rest of the design matrix.
Design efficiency

- The aim is to minimize the standard error of a $t$-contrast (i.e. the denominator of a $t$-statistic).

$$\text{var}(c^T \hat{\beta}) = \hat{\sigma}^2 c^T (X^T X)^{-1} c$$

- This is equivalent to maximizing the efficiency $e$:

$$e(\hat{\sigma}^2, c, X) = (\hat{\sigma}^2 c^T (X^T X)^{-1} c)^{-1}$$

  - Noise variance
  - Design variance

- If we assume that the noise variance is independent of the specific design:

$$e(c, X) = (c^T (X^T X)^{-1} c)^{-1}$$

- This is a relative measure: all we can really say is that one design is more efficient than another (for a given contrast).
High correlation between regressors leads to low sensitivity to each regressor alone.

We can still estimate efficiently the difference between them.
Bibliography:


- **Ambiguous results in functional neuroimaging data analysis due to covariate correlation.** A. Andrade et al., NeuroImage, 1999.

Estimability of a contrast

- If $X$ is not of **full rank** then we can have $X\beta_1 = X\beta_2$ with $\beta_1 \neq \beta_2$ (different parameters).

- The parameters are **not** therefore ‘unique’, ‘identifiable’ or ‘estimable’.

- For such models, $X^TX$ is not invertible so we must resort to generalised inverses (SPM uses the pseudo-inverse).

- Example:

  $[1 \ 0 \ 0], [0 \ 1 \ 0], [0 \ 0 \ 1]$ are not estimable.

  $[1 \ 0 \ 1], [0 \ 1 \ 1], [1 \ -1 \ 0], [0.5 \ 0.5 \ 1]$ are estimable.
Three models for the two-samples t-test

- Model 1:
  \[ \beta_1 = y_1 \]
  \[ \beta_2 = y_2 \]
  \[ [1 0] \beta = y_1 \]
  \[ [0 1] \beta = y_2 \]

- Model 2:
  \[ \beta_1 = y_1 \]
  \[ \beta_2 = y_2 \]
  \[ [0 -1] \beta = y_1 - y_2 \]
  \[ [.5 .5] \beta = \text{mean}(y_1, y_2) \]

- Model 3:
  \[ \beta_1 + \beta_3 = y_1 \]
  \[ \beta_2 + \beta_3 = y_2 \]
  \[ [1 0 1] \beta = y_1 \]
  \[ [0 1 1] \beta = y_2 \]
  \[ [1 -1 0] \beta = y_1 - y_2 \]
  \[ [.5 0.5 1] \beta = \text{mean}(y_1, y_2) \]
Multidimensional contrasts

Think of it as constructing 3 regressors from the 3 differences and complement this new design matrix such that data can be fitted in the same exact way (same error, same fitted data).
Example: working memory

- **A**: Time (s) with corresponding stimulus and response.
  - Correlation = -0.65
  - Efficiency ([1 0]) = 29

- **B**: Time (s) with corresponding stimulus and response.
  - Correlation = +0.33
  - Efficiency ([1 0]) = 40

- **C**: Time (s) with corresponding stimulus and response.
  - Correlation = -0.24
  - Efficiency ([1 0]) = 47

- **B**: Jittering time between stimuli and response.
- **C**: Requiring a response on a randomly half of trials.