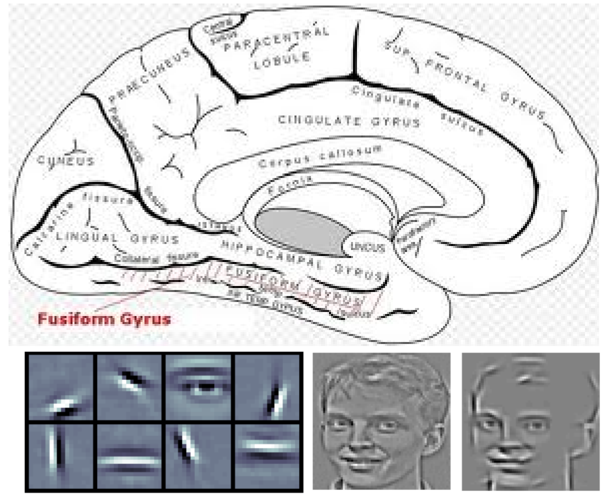
FMRI Data modeling

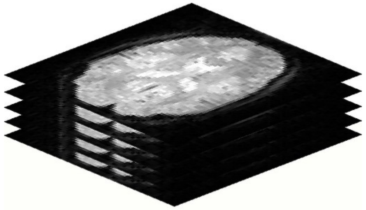
Master course

Bertrand Thirion
bertrand.thirion@inria.fr

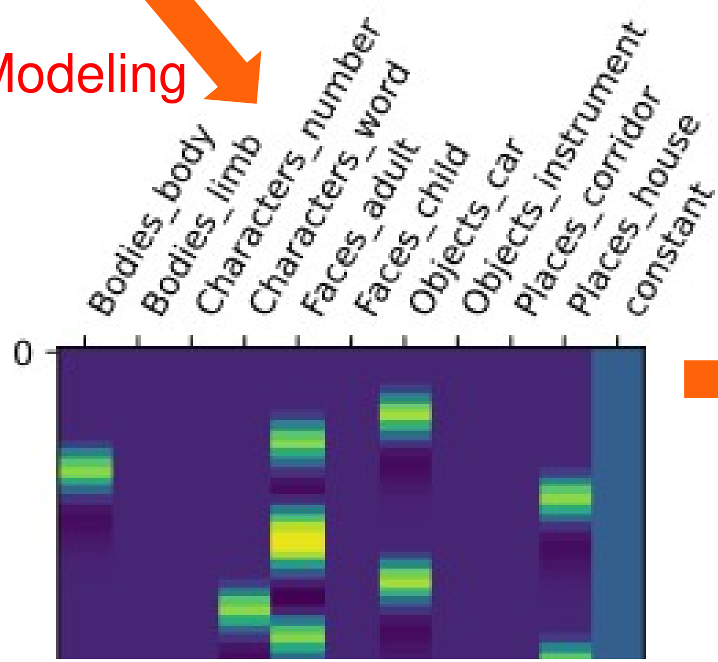
fMRI data analysis



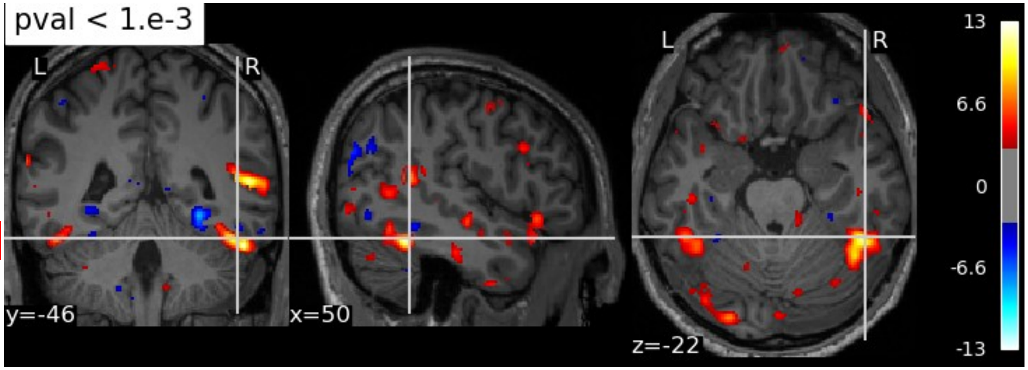
Complex metabolic pathway



Modeling



Statistical analysis



On localizing face-specific activity

What was most exciting to me about our work was that it seemed to address directly a major and long-standing theoretical question in cognitive psychology: the degree to which mental architecture is “domain specific,” that is, specialized for particular kinds of information such as faces or places or language.

This question had been debated heatedly in our field for nearly 200 years (Finger, 2001) and now here was a little piece of the brain that seemed to do just one thing: perceive faces.

This finding fit the broader idea that the mind is not a general purpose device, but is instead composed of a set of distinct components, some of them highly specialized for solving a very specific problem (Fodor, 1983).

Kanwisher, the quest for the FFA, 2017

Part 1: FMRI data Modeling

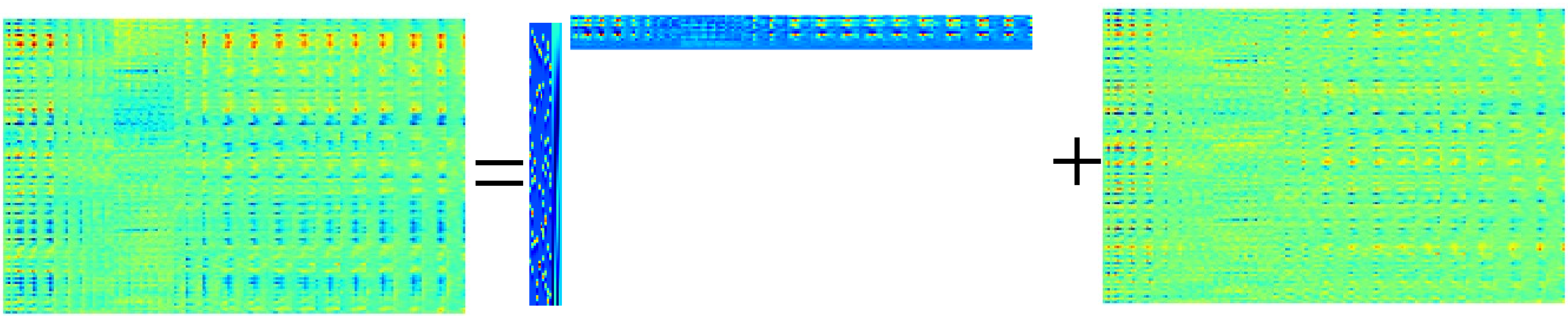
Y: fMRI data,
(time points, voxels)

E: noise,
(time points, voxels)
i.e. unmodeled signal

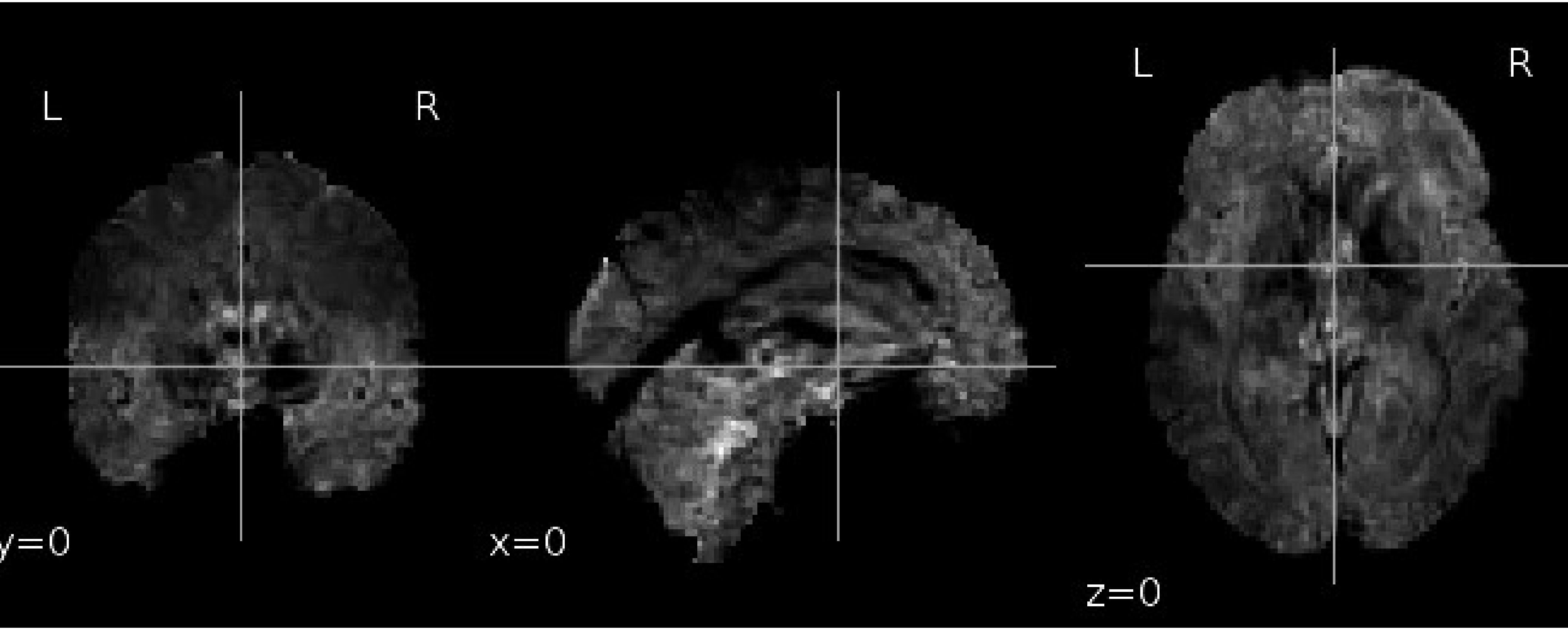
$$Y = XB + E$$

X: design matrix
(time points, regressors)
= various effects that occur during the experiments and/or nuisance effects

B: parameters
(regressors, voxels)
parameters of the model, to be estimated



fMRI data



Linear model fitting

Normal equation

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

Covariance estimator

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n}$$

Decision statistic

$$t = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{ii}}}$$

Linear model fitting

Normal equation $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Covariance estimator $$\hat{\mathrm{Cov}}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\frac{\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n}$$

Decision statistic $$t = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{ii}}}$$

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]\ ;\ \|\mathbf{X}_1\| = \|\mathbf{X}_2\| = 1, \mathbf{X}_1^T\mathbf{X}_2 = \rho \in [0,1]\ ;\ \|\mathbf{y}\| = 1$$

$$\hat{\boldsymbol{\beta}}_1 = \frac{1}{1-\rho^2}\left(\mathbf{X}_1^T\mathbf{y} - \rho\mathbf{X}_2^T\mathbf{y}\right)$$

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{1,1} = \frac{1}{1-\rho^2}\frac{1}{n}\|\mathbf{y} - P_{\mathbf{X}}(\mathbf{y})\|^2$$

Linear model fitting

Normal equation

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

Covariance estimator

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\frac{\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n}$$

Decision statistic

$$t = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{ii}}}$$

- **Correlated effects → less power**

$$\hat{\boldsymbol{\beta}}_1 = \frac{1}{1-\rho^2}\left(\mathbf{X}_1^T\mathbf{y} - \rho\mathbf{X}_2^T\mathbf{y}\right)$$

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{1,1} = \frac{1}{1-\rho^2}\frac{1}{n}\|\mathbf{y}-P_{\mathbf{X}}(\mathbf{y})\|^2$$

Linear model fitting

Normal equation

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

Covariance estimator

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\frac{\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n}$$

Decision statistic

$$t = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{ii}}}$$

Include nuisance effects → more power

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2];\ \|\mathbf{X}_1\| = \|\mathbf{X}_2\| = 1;\ \mathbf{X}_1^T\mathbf{X}_2 = \rho \in [0,1];\ \|\mathbf{y}\| = 1$$

$$\hat{\boldsymbol{\beta}}_1 = \frac{1}{1-\rho^2}\left(\mathbf{X}_1^T\mathbf{y} - \rho\mathbf{X}_2^T\mathbf{y}\right)$$

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{1,1} = \frac{1}{1-\rho^2}\frac{1}{n}\|\mathbf{y} - P_{\mathbf{X}}(\mathbf{y})\|^2$$

Linear model fitting

Normal equation
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Covariance estimator
$$\hat{\mathrm{Cov}}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\frac{\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n}$$

Decision statistic
$$t = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{ii}}}$$

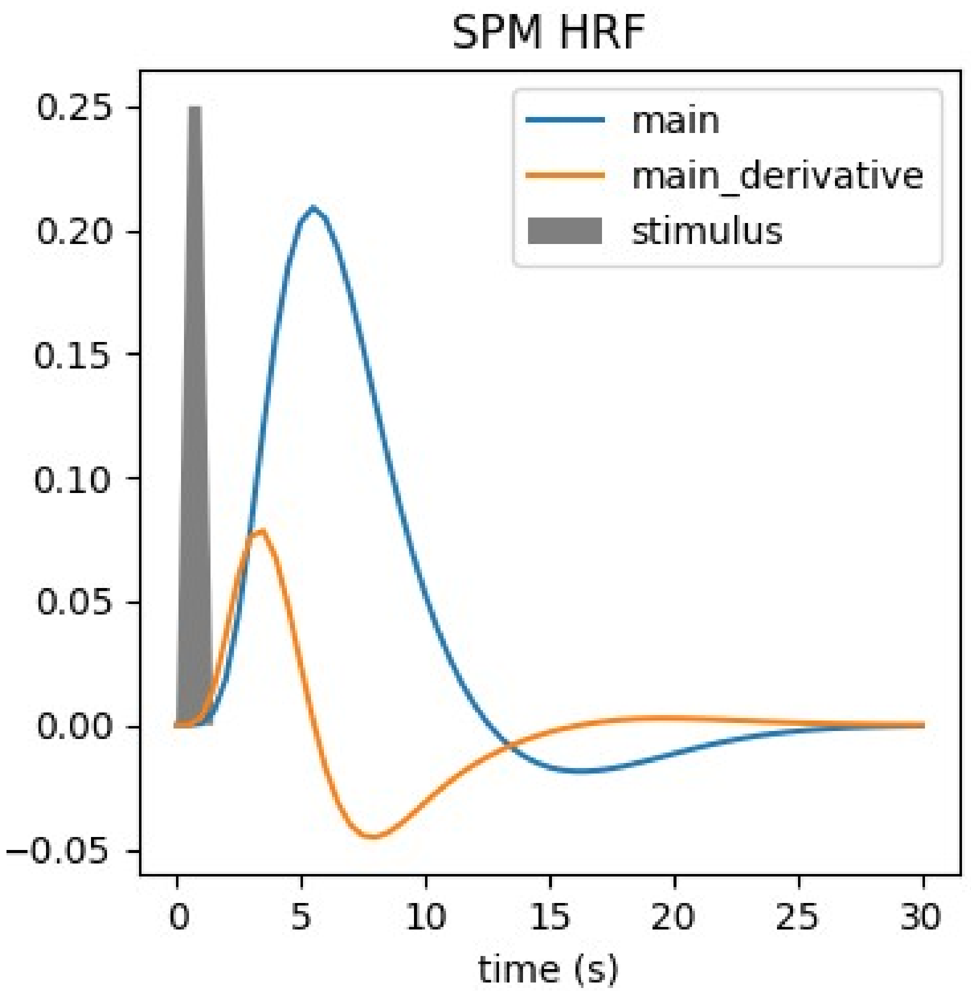
More samples → more power

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2] \; ; \; \|\mathbf{X}_1\| = \|\mathbf{X}_2\| = 1 \; ; \; \mathbf{X}_1^T\mathbf{X}_2 = \rho \in [0,1] \; ; \; \|\mathbf{y}\| = 1$$

$$\hat{\boldsymbol{\beta}}_1 = \frac{1}{1-\rho^2}\left(\mathbf{X}_1^T\mathbf{y} - \rho\mathbf{X}_2^T\mathbf{y}\right)$$

$$\hat{\mathrm{Cov}}(\boldsymbol{\beta})_{1,1} = \frac{1}{1-\rho^2}\frac{1}{n}\left\|\mathbf{y} - P_{\mathbf{X}}(\mathbf{y})\right\|^2$$

The hemodynamic response function



- Phenomenological model
- ~ low-pass filter of neural signal
- Time to peak: ~ 5s
- Response width: ~ 5s
- Subsequent undershoot
- Return to baseline after ~ 20s

https://nilearn.github.io/stable/auto_examples/04_glm_first_level/plot_hrf.html

FMRI Model: the design matrix

Defines the temporal effects observed in the fMRI data:

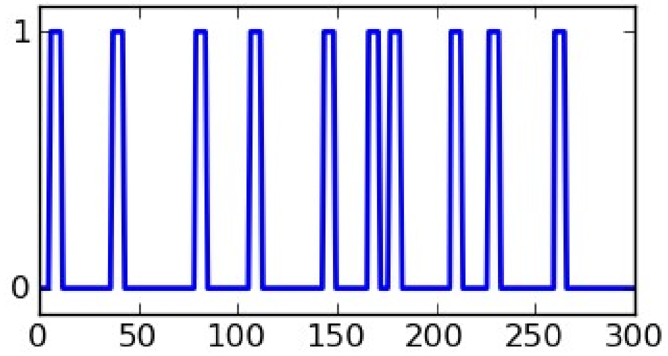
- Task-related activity
- Nuisance events (motion, respiratory cycle, cardiac rhythm fluctuations)
- Low-frequency signals

FMRI model: the design matrix

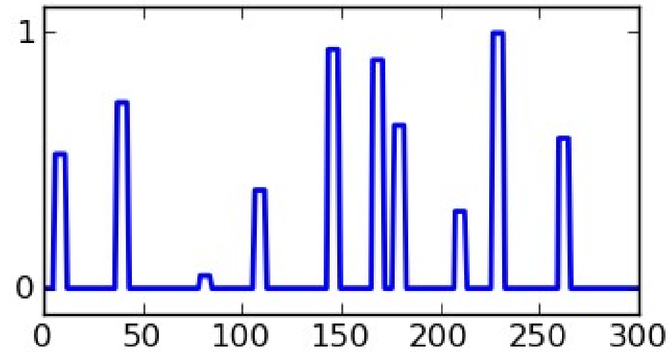
- Modeling of task-related activity during an experiment

Different kinds of stimulation paradigms

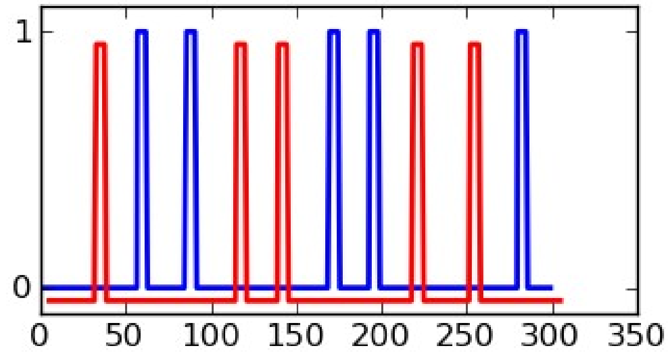
Event-related design



Parametric modulation

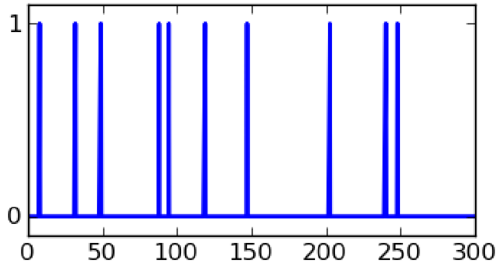


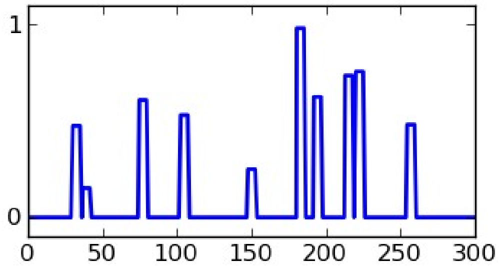
Multiple experimental conditions



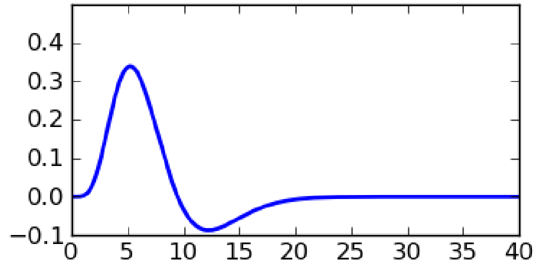
FMRI model: the design matrix

- fMRI resolution, neural activity = stimulus model
- But BOLD signal slower: modeled as a linear filter

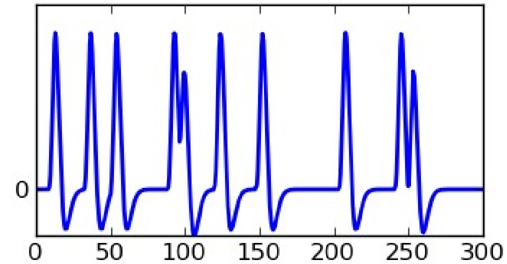


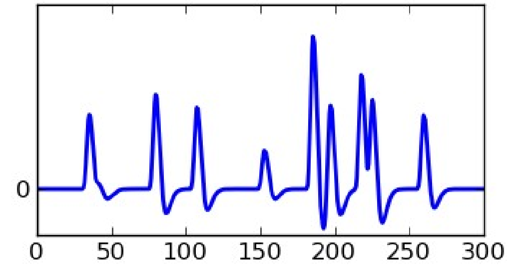


*



=



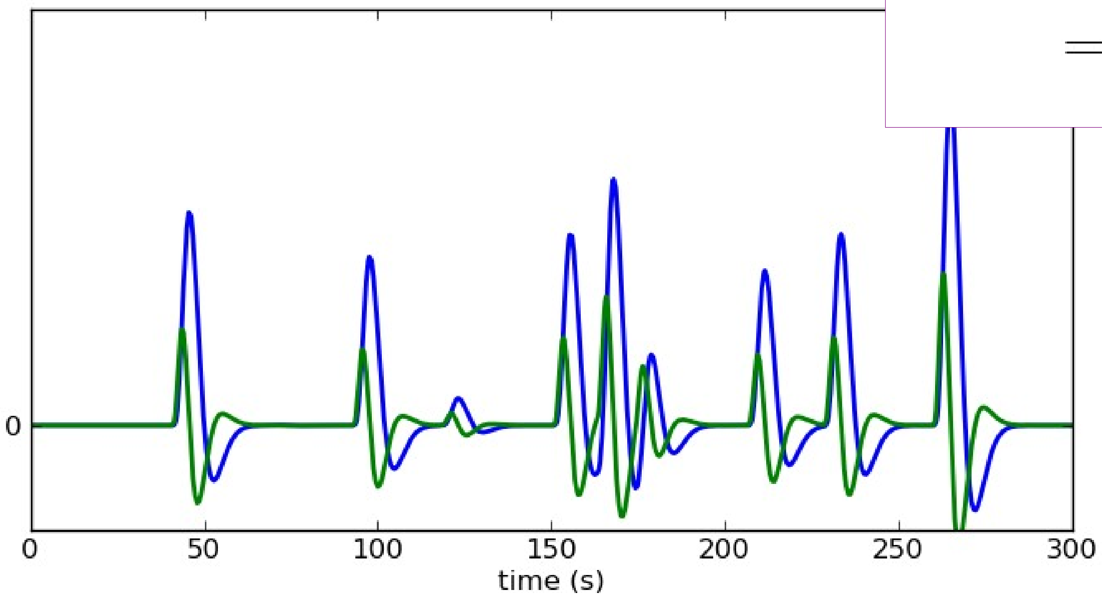


2 core hypotheses on BOLD:
- additivity/superposition
- time invariance

FMRI model: the design matrix

- Linear model is sufficient in many cases
  - Corrected for inaccurate delay: use regressors time derivative

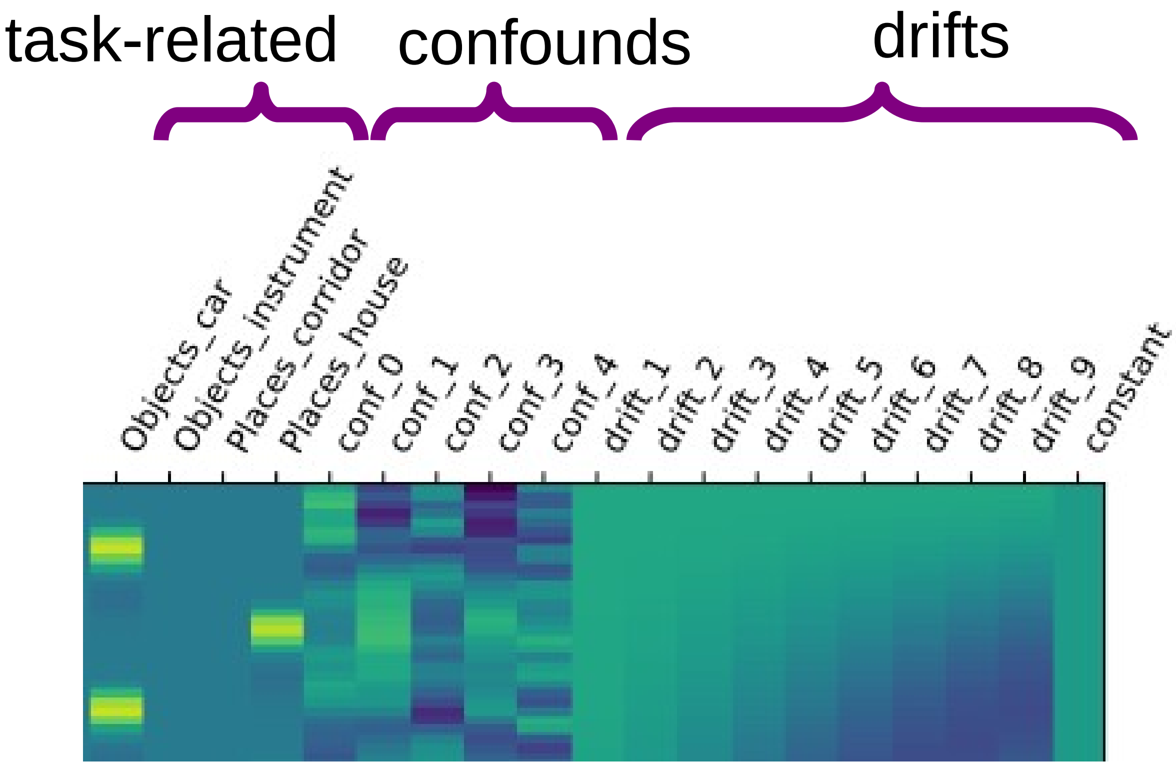
$$
\begin{aligned}
y(t) &= \beta(p*h)(t+\tau)+\varepsilon(t)\\
&\simeq \beta(p*h)(t)+\beta\tau\frac{d}{dt}(p*h)(t)+\varepsilon(t)\\
&= \beta(p*h)(t)+\beta\tau(p*\frac{dh}{dt})(t)+\varepsilon(t)
\end{aligned}
$$



FMRI model: the design matrix

Effects of no interest

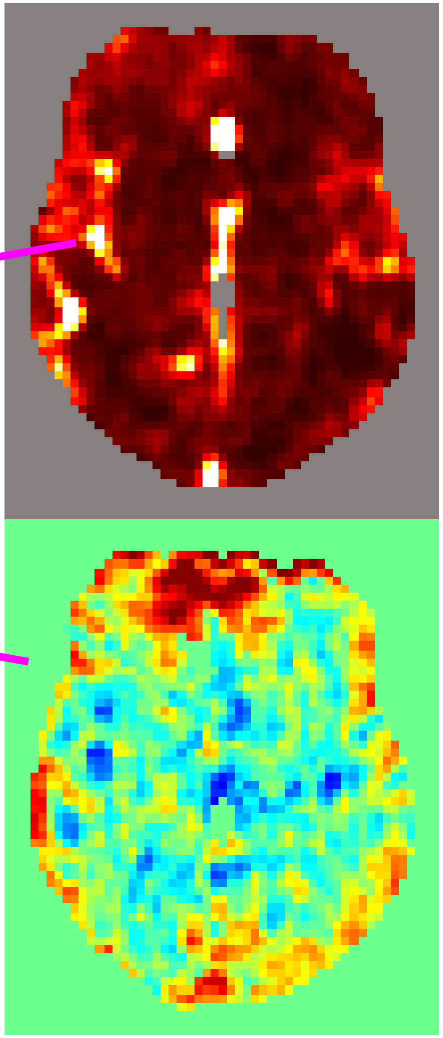
- Low frequency drifts [0- 0.01 Hz]
- Confounds = Physiological data (respiratory signal, cardiac rhythm), motion



Dealing with autocorrelated noise

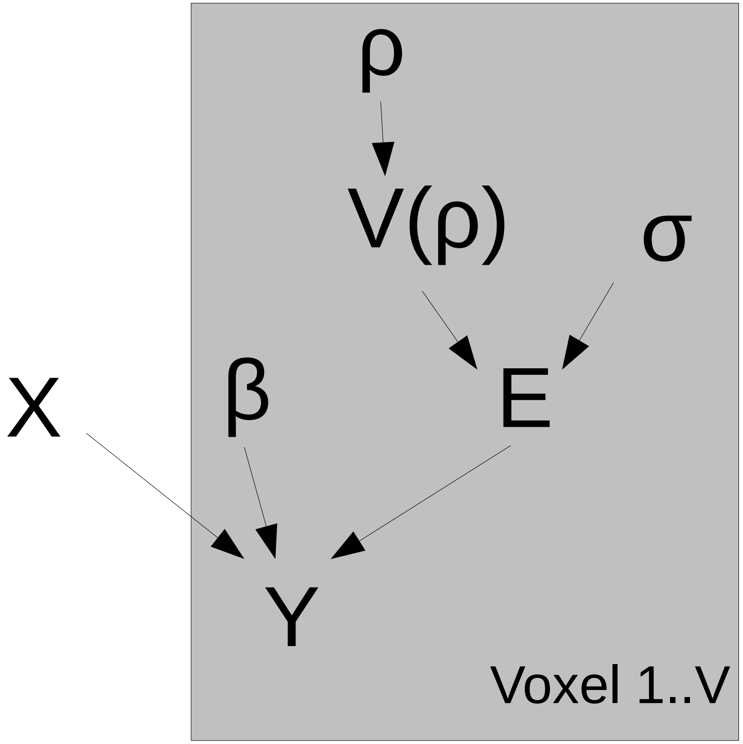
spatial/temporal covariance

- Noise varies spatially in amplitude
- Positive correlation in space (neglected)
- Positive temporal correlation
  - Important for unbiased variance-covariance estimation
  - Modeled as AR(k) process
  - 2 parameters: variance, autocorrelation



$$p(E) = \prod_{v=1}^{n_{voxels}} \mathcal{N}(E_v; 0, \sigma_v^2 V(\rho_v))$$

Dealing with autocorrelated noise



- Estimation procedure

  given (X, Y) estimate the remaining quantities at each voxel

Dealing with autocorrelated noise

- Mass univariate framework: per-voxel MLE

$$\min_{\boldsymbol{\beta},\sigma^2,\rho} \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_\rho^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{1}{2}\log|2\pi\mathbf{V}_\rho\sigma^2|$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$
$$\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\hat{\rho} = \frac{\langle \hat{\mathbf{r}}(t), \hat{\mathbf{r}}(t-1)\rangle}{\langle \hat{\mathbf{r}}, \hat{\mathbf{r}}\rangle} \longrightarrow \mathbf{V}_{\hat{\rho}}$$

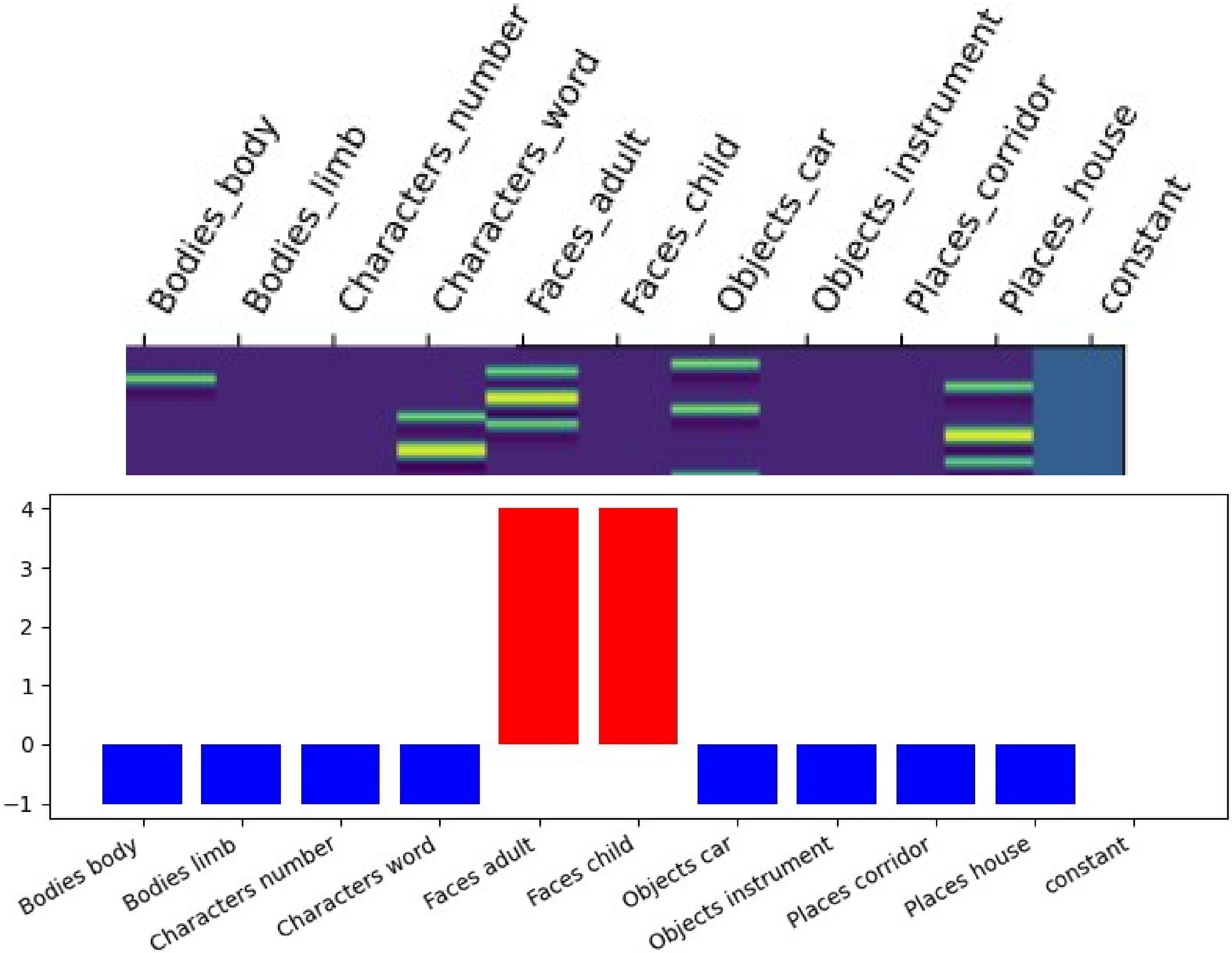
$$\hat{\boldsymbol{\beta}}_{\mathrm{AR1}} = (\mathbf{X}^T\mathbf{V}_{\hat{\rho}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_{\hat{\rho}}^{-1}\mathbf{y}$$
$$\hat{\sigma}^2 = \frac{1}{n}\left(\mathbf{y}^T\mathbf{V}_{\hat{\rho}}^{-1}\mathbf{y} - (\mathbf{y}-\mathbf{X}^T\hat{\boldsymbol{\beta}})^T\mathbf{V}_{\hat{\rho}}^{-1}(\mathbf{y}-\mathbf{X}^T\hat{\boldsymbol{\beta}})\right)$$
$$\mathrm{dof} = n - \mathrm{rank}(\mathbf{X})$$

Part II Statistical analysis

- First, formulate a question “Which regions show more activity for condition A than for condition B ?” = functional contrast
- Decide if the response is '*yes*' in a certain region
- And if we consider all the brain together ?

Statistical analysis: contrasts

Contrast = linear combination of experimental conditions



More specific effects

Statistical analysis: contrast

- Two main alternatives:
- Testing mono-dimensional, signed effects “A>B” $\mathbf{c}^T\boldsymbol{\beta} > 0$
- Testing the importance of an effect: is “are the effects A et B greater than chance” $\mathbf{c}^T\boldsymbol{\beta} \neq 0$
  - Multi-dimensional, unsigned contrasts

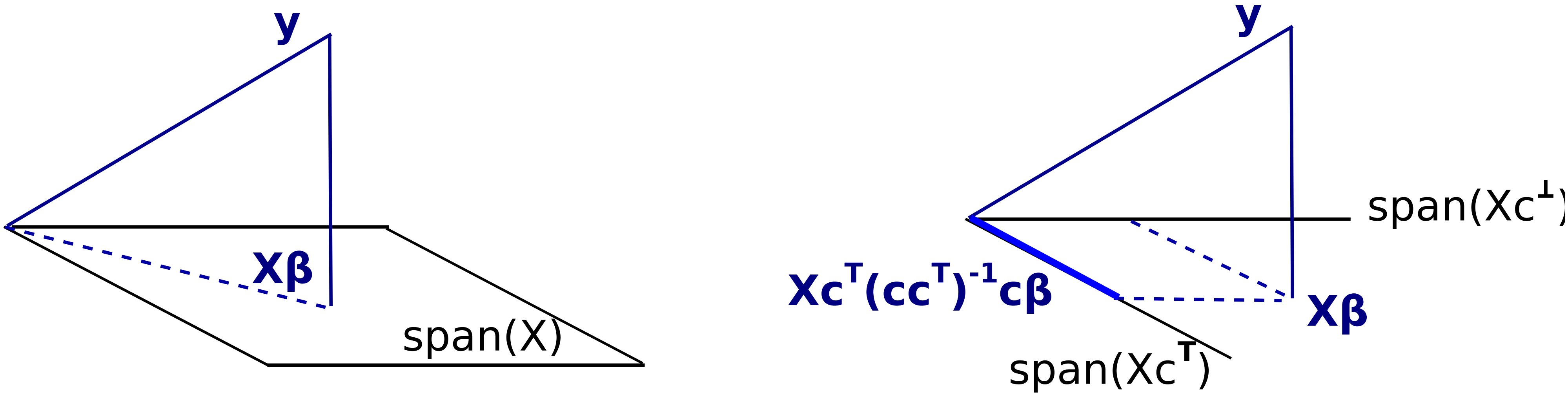
$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \end{pmatrix}^T$$

$$\|\mathbf{c}^T\boldsymbol{\beta}\| \neq 0$$

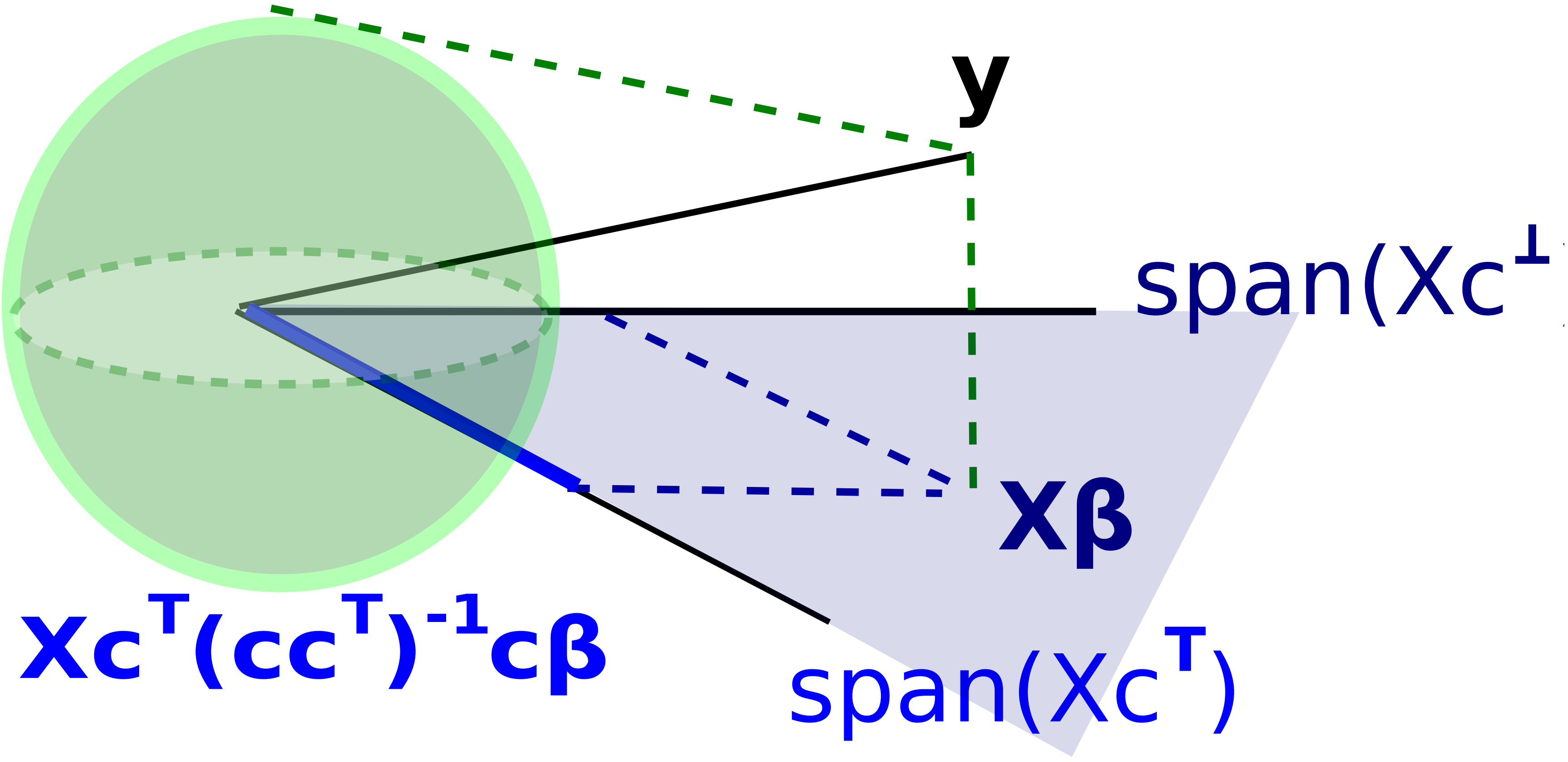
$$\begin{pmatrix} 1 & -1 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots \end{pmatrix}^T$$

Statistical analysis: contrast

- These "linear contrasts" measure the projection of the data into particular subspaces



Contrasts and test



Is the effect of interest larger than the noise level ?

Statistical analysis: classical inference

- Decide between two hypotheses:

$$H_0 : \mathbf{c}^T \boldsymbol{\beta} = 0$$
$$H_1 : \mathbf{c}^T \boldsymbol{\beta} \neq 0$$

$$\mathbf{y} = \mathbf{X}\mathbf{c}(\mathbf{c}^T\mathbf{c})^{-1}\mathbf{c}^T\boldsymbol{\beta} + \mathbf{X}(\mathbf{I} - \mathbf{c}(\mathbf{c}^T\mathbf{c})^{-1}\mathbf{c}^T)\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{y} = \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{X}_\perp\boldsymbol{\beta}_\perp + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$$

Statistical analysis: classical inference

- Decide between two hypotheses:

$$H_0 : \mathbf{c}^T \boldsymbol{\beta} = 0$$
$$H_1 : \mathbf{c}^T \boldsymbol{\beta} \neq 0$$

$$\mathbf{y} = \mathbf{X}\mathbf{c}(\mathbf{c}^T\mathbf{c})^{-1}\mathbf{c}^T\boldsymbol{\beta} + \mathbf{X}(\mathbf{I} - \mathbf{c}(\mathbf{c}^T\mathbf{c})^{-1}\mathbf{c}^T)\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{y} = \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{X}_\perp\boldsymbol{\beta}_\perp + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$$

- *Optimal decision* provided by the likelihood ratio test

$$\Lambda = \frac{\sup_{\boldsymbol{\beta}_c, \boldsymbol{\beta}_\perp, \sigma^2} \mathcal{N}(\mathbf{y}; \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{X}_\perp\boldsymbol{\beta}_\perp, \sigma^2\mathbf{I})}{\sup_{\boldsymbol{\beta}_\perp, \sigma^2} \mathcal{N}(\mathbf{y}; \mathbf{X}_\perp\boldsymbol{\beta}_\perp, \sigma^2\mathbf{I})}$$

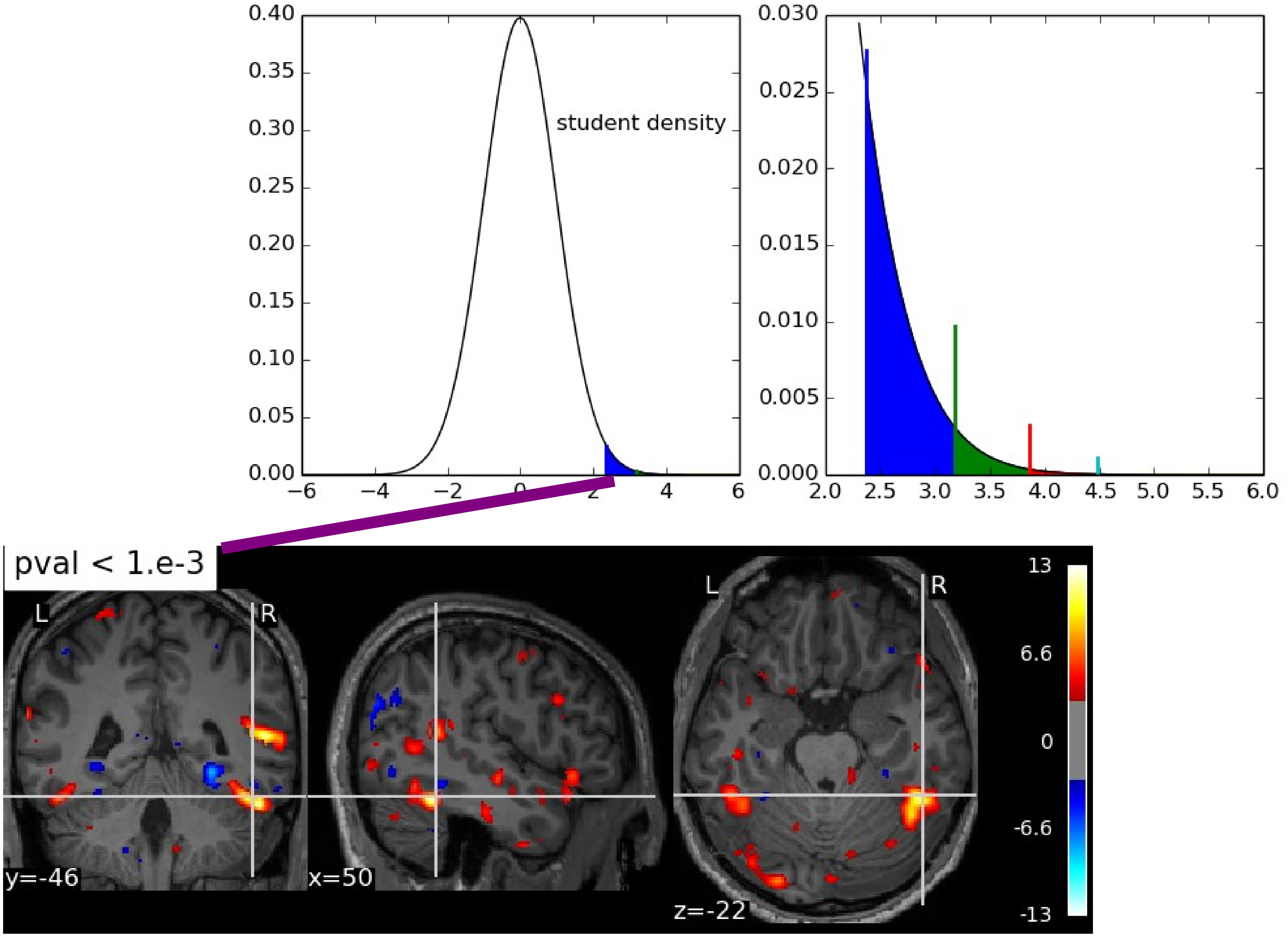
- Statistical decision: threshold the LRT with false positive control

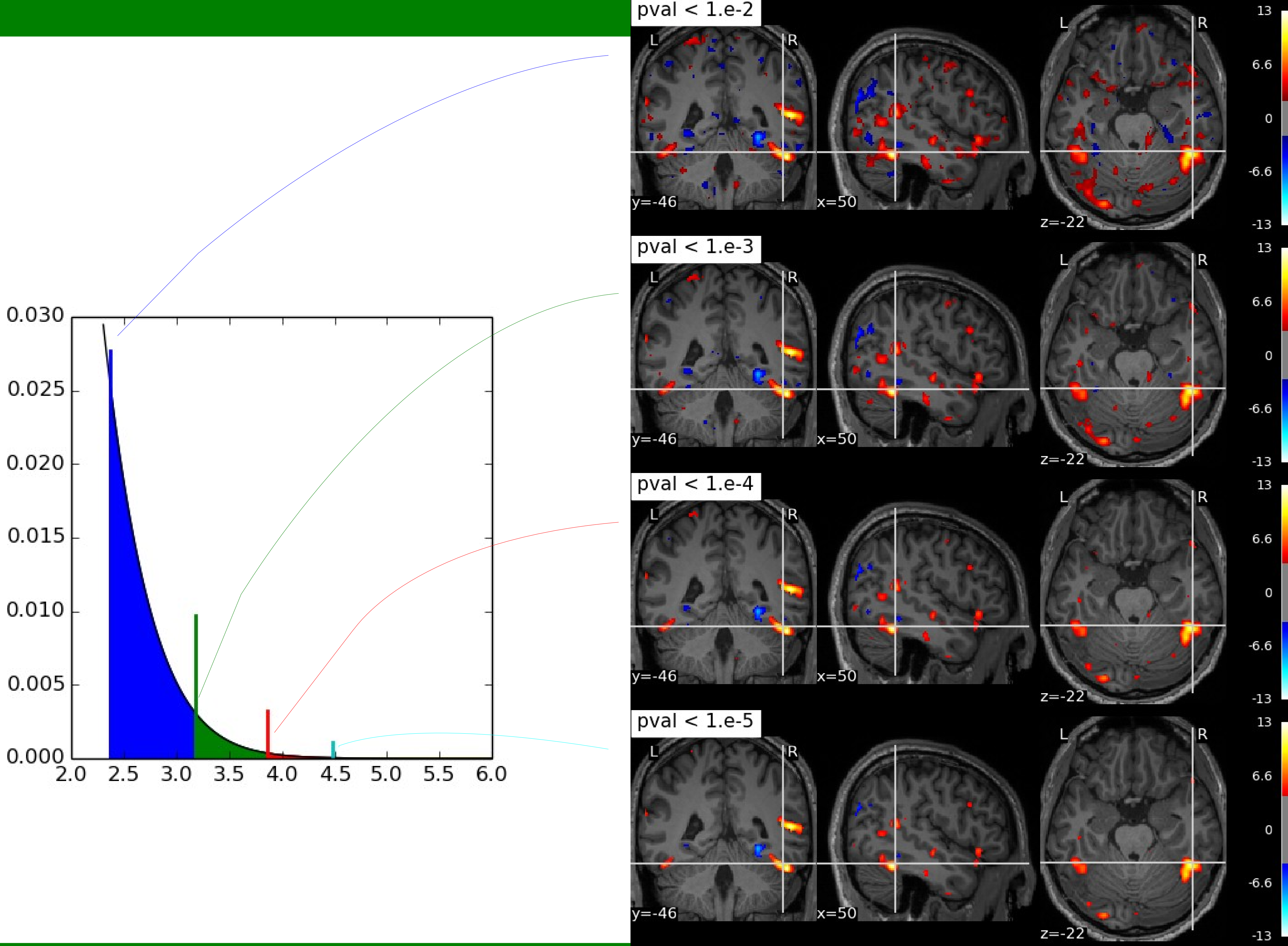
$$\mathbb{P}(\Lambda > \Lambda_\alpha | \boldsymbol{\beta}_c = 0) \leq \alpha$$

Statistical Analysis

- Under certain hypotheses, the LR statistic is equivalent to t and F statistic
$$\Lambda = \left(\frac{\text{rank}(\mathbf{c})}{\text{dof}} F + 1 \right)^{\frac{n}{2}}$$
- T-statistic $$t = \frac{\mathbf{c}^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}\sqrt{\mathbf{c}^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{c}}}$$
- F statistic $$F = \frac{1}{\hat{\sigma}} \text{Tr}(\hat{\boldsymbol{\beta}}^T \mathbf{c}\mathbf{c}^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{c}^T \hat{\boldsymbol{\beta}})$$
- Under the null hypothesis, these follow student and Fisher distributions with known degrees of freedom
  - exact specificity control
- But: LR is equivalent to F iff V is known, which is not the case [Dekker et al.2009]

Statistical analysis: classical tests





0.030
0.025
0.020
0.015
0.010
0.005
0.000
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
pval < 1.e-2
pval < 1.e-3
pval < 1.e-4
pval < 1.e-5
L
R
y=-46
x=50
z=-22
13
6.6
0
-6.6
-13

Multiple Comparison problem

- Specificity = α for one test.
- But we perform n tests (10^4 to 10^5)
- Probability of one false detection (assuming independence):

$$1 - (1-\alpha)^n \simeq n\alpha \text{ if } n\alpha \ll 1$$

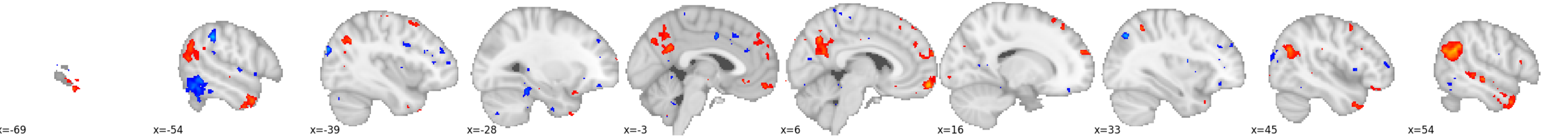
- Correct the significance of test by n: *Bonferroni correction*
- Very conservative inference, little power to detect true effects

Example of experimental result

Level 1 Statistics

Voxel significance				Coordinates in MNI referential			Cluster Size
p FWE corr (Bonferroni)	p FDR corr	Z	p uncorr	x (mm)	y (mm)	z (mm)	(voxels)
1.000	**0.036**	**4.45**	**0.000**	**-52**	**-56**	**-7**	**518**
1.000	0.039	4.40	0.000	-48	-53	-10	
1.000	0.039	4.39	0.000	-56	-51	-10	
1.000	0.040	4.36	0.000	-58	-50	-10	
0.037	**0.014**	**5.20**	**0.000**	**-54**	**-39**	**41**	**157**
1.000	0.067	3.78	0.000	-62	-36	42	
1.000	0.083	3.61	0.000	-52	-39	48	
0.118	**0.018**	**4.98**	**0.000**	**-2**	**-2**	**41**	**78**
1.000	0.066	3.80	0.000	2	4	39	
0.818	**0.029**	**4.59**	**0.000**	**-33**	**-36**	**-22**	**55**
0.145	**0.018**	**4.94**	**0.000**	**-39**	**-86**	**27**	**49**
0.021	**0.014**	**5.30**	**0.000**	**33**	**-71**	**42**	**35**
1.000	**0.042**	**4.31**	**0.000**	**3**	**39**	**24**	**31**
1.000	0.049	4.18	0.000	4	37	21	
1.000	0.054	4.04	0.000	-2	34	26	
1.000	**0.066**	**3.80**	**0.000**	**60**	**-59**	**-13**	**27**
1.000	**0.039**	**4.38**	**0.000**	**-50**	**-44**	**53**	**27**
1.000	**0.053**	**4.12**	**0.000**	**-44**	**10**	**3**	**19**

Threshold Z: 2.77 (fdr control at 0.100)
Cluster level p-value threshold: 1.0
Cluster size threshold: 15 voxels
Number of voxels: 996
Number of clusters: 10

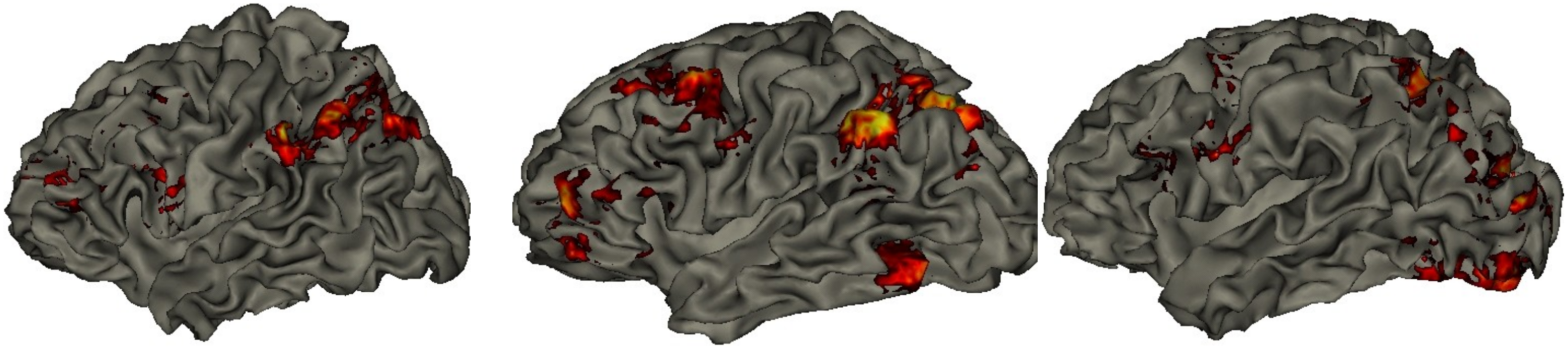


Conclusion

- Classical/frequentist tests: only the control under the null hypothesis is considered
  - Because it is hard to define the signal distribution under the converse
- Correcting the tests for whole brain analysis drastically reduces sensitivity
  - Use of spatial regularization

Next step

compare individual result
Build group-level models



Exercise

We make observations of random variables $(\beta_i)_{i=1..n}$, that are assumed to be instances of unique random variable β. For each $i \in [1..n]$, let $\hat{\beta}_i$ and σ_i^2 be the empirical estimate and variance of β_i. What is the linear combination of these variables $\hat{\beta} = \sum_{i=1}^{n} w_i \beta_i \; s.t. \sum_{i=1}^{n} w_i = 1 \; and \; w_i > 0 \; \forall i \in [1..n]$
such that $\hat{\beta}$ has the minimal variance ? What are the corresponding variance and effect estimates ?