Beyond Univariate Analyses: Multivariate Modeling of Functional Neuroimaging Data

F. DuBois Bowman

Department of Biostatistics and Bioinformatics
Center for Biomedical Imaging Statistics
Emory University, Atlanta, GA, 30322

How Not to Analyze Your Data: A Skeptical Introduction to Modeling Methods
OHBM 2013
Seattle, WA

June 16, 2013

EMORY
ROLLINS SCHOOL OF PUBLIC HEALTH
CBIS
Center for Biomedical Imaging Statistics

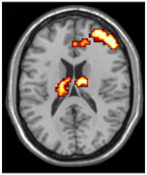
Outline

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

The Problem

Neuroactivation Studies

- Task-related designs
- Seek group-level inferences relating stimuli to neural response
  - Contrasts specify task-related changes (and possibly group differences) in neural activity
  - Estimation and hypothesis testing about group-level contrasts
- Multiple contrasts for each subject, derived from multiple tasks/effects
- Linear model framework (linear in parameters)



Univariate versus Multivariate Linear Models

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH
CBIS Center for Biomedical Imaging Statistics

Univariate Linear Models

- Involve a single dependent variable
- May involve one or more independent variables
    - Multiple regression

Multivariate Linear Models

- Involve multiple dependent variables
- Dependent variables are possibly correlated
    - Over voxels
    - Over time
    - Related stimuli/tasks
- May involve one or more independent variables

Common Univariate Analysis Framework

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Two-stage Model: Mass Univariate Approach

- First, fit a linear model separately for each subject (at each voxel)
  - Convolution with a HRF
  - Temporal correlations between scans: AR models (+ white noise)
    - Linear covariance structure
    - Pre-coloring/temporal smoothing [Worsley and Friston, 1995]
    - Pre-whitening [Bullmore et al, 1996; Purdon and Weisskoff, 1998]
    - Alternative structures available for PET [Bowman and Kilts, 2003]
- Second, fit linear model that combines subject-specific estimates
  - A two-stage (random effects) model
    - Simplifies computations*
    - Sacrifices efficiency
- For Inference: Compute t-statistics at each voxel and threshold
  - Consider a multiple testing adjustment (Bonferonni-type, FDR, RFT)

Common Univariate Analysis Framework

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Properties

- Two-stage (random effects) model
  - Simplifies computations
  - Sacrifices efficiency
- May assume independence between different regression coefficients
- Assumes independence between different brain locations

Data Example

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Working Memory in Schizophrenia Patients

- N=28 subjects: 15 schizophrenia patients and 13 healthy controls
- fMRI Tasks: Serial Item Recognition Paradigm (SIRP)
  - Encoding set: Memorize 1, 3, or 5 target digits.
  - Probing set: Shown single digit probes and asked to press a button:
    - with their index finger, if the probe matched
    - with their middle finger, if not.
  - Between conditions, subjects fixated on a flashing cross.
- 6 runs per subject: (177 scans per run for each subject)
  - 3 runs of working memory tasks on each of 2 days
- Objective: Compare working memory-related brain activity between patients and controls

Data from the Biomedical Informatics Research Network (BIRN) [1]: Potkin et al. (2002).

Statistical Modeling

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

General Linear Model: Stage I

$$\mathbf{Y}_i(v) = \mathbf{X}_{iv}\beta_i(v) + \mathbf{H}_{iv}\gamma_i(v) + \varepsilon_i(v)$$

$\mathbf{Y}_i(v)$	$S \times 1$	serial BOLD activity at voxel v.
$\mathbf{X}_{iv}$	$S \times q$	design matrix reflecting fixation and WM tasks.
$\beta_i(v)$	$q \times 1$	parameter vector linking experimental tasks.
$\varepsilon_i(v)$	$S \times 1$	random error about ith subject's mean.
$\mathbf{H}_{iv}$	$S \times m$	contains other covariates, e.g. high-pass filtering.
$\varepsilon_i(v)$	$\sim$	$\text{Normal}(\mathbf{0}, \tau_v^2\mathbf{V})$.

Statistical Modeling: Univariate

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

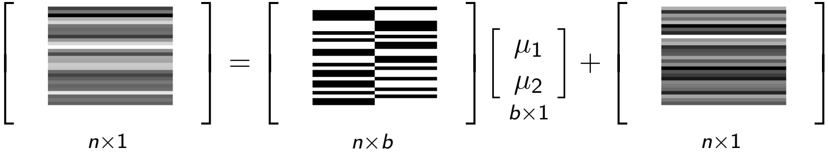
General Linear Model: Stage II (Contrast of Interest)

$$\mathbf{C}\boldsymbol{\beta}_{ij}(v) = \mu_j(v) + e_{ij}(v)$$

$\boldsymbol{\beta}_{ij}(v)$		stage I fixation and WM parameters; subject i, group j.
$\mathbf{C}$		contrast matrix (linear combinations of elements in $\boldsymbol{\beta}_{ij}(v)$).
$\mu_j(v)$		group-level mean (for group j).
$e_{ij}(v)$		random error.
$e_{ij}(v)$	$\sim$	Normal$(\mathbf{0}, \sigma^2(v))$.

Statistical Modeling: Univariate

Working Memory Data:



General Linear Model: Stage II (Matrix Model)

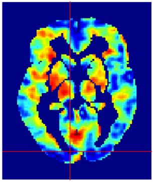
$$\begin{bmatrix} \mathbf{C}\beta_{11}(v) \\ \vdots \\ \mathbf{C}\beta_{n_c1}(v) \\ \mathbf{C}\beta_{12}(v) \\ \vdots \\ \mathbf{C}\beta_{n_p2}(v) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1(v) \\ \mu_2(v) \end{bmatrix} + \begin{bmatrix} e_{11}(v) \\ \vdots \\ e_{n_c1}(v) \\ e_{12}(v) \\ \vdots \\ e_{n_p2}(v) \end{bmatrix}$$

$$(\mathbf{I} \otimes \mathbf{C})\beta(v) = \mathbf{X}\mu(v) + \mathbf{e}(v)$$

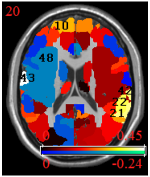
Statistical Modeling

Mass Univariate Approach

- May not fully acknowledge the correlations between
  - Multiple effects/contrasts
  - Effects/contrasts at different voxels



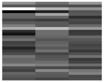
- Separately models contrasts of interest
  - Does not yield information on correlations between contrasts.
  - Does not enable comparisons or linear combinations of contrasts.



Statistical Modeling

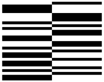
EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Working Memory Data:



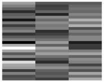
$n \times p$

$=$



$n \times q$

$$\begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{bmatrix}_{q \times p} +$$



$n \times p$

General Linear Multivariate Model: Stage II

$$\boldsymbol{\beta}(v) = \mathbf{X}\boldsymbol{\mu}(v) + \mathbf{e}(v)$$

- Multiple summary statistics (or contrasts) included for each subject
  - E.g. working memory load contrasts
- Rows contain data from different subjects
  - Each row assumed to have variance covariance matrix $\boldsymbol{\Sigma}$ reflecting correlations between summary statistics/contrasts
- Define $\theta(v) = \mathbf{C}\boldsymbol{\mu}(v)\mathbf{U}$, e.g. $(\mu_{13} - \mu_{11}) - (\mu_{23} - \mu_{21})$.

Statistical Modeling

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Contrast Variance:

- Define $\theta(v) = \mathbf{C}\boldsymbol{\mu}(v)\mathbf{U}$, e.g. $(\mu_{13} - \mu_{11}) - (\mu_{23} - \mu_{21})$
  - $\mathbf{C} = \begin{bmatrix} 1 & -1 \end{bmatrix}$
  - $\mathbf{U} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

$\mathrm{Var}(\hat{\theta})$

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}(v)) &= \mathrm{Var}(\hat{\theta}(v)') \\
&= \mathrm{Var}\left[\mathrm{vec}((\mathbf{C}\hat{\boldsymbol{\mu}}(v)\mathbf{U})')\right] \\
&= \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' \otimes \mathbf{U}'\boldsymbol{\Sigma}(v)\mathbf{U} \\
&= \left(\frac{1}{n_c} + \frac{1}{n_p}\right)(\sigma_1^2 + \sigma_3^2 - 2\sigma_{13}), \text{ for WM data.}
\end{aligned}$$

Application to Working Memory Data

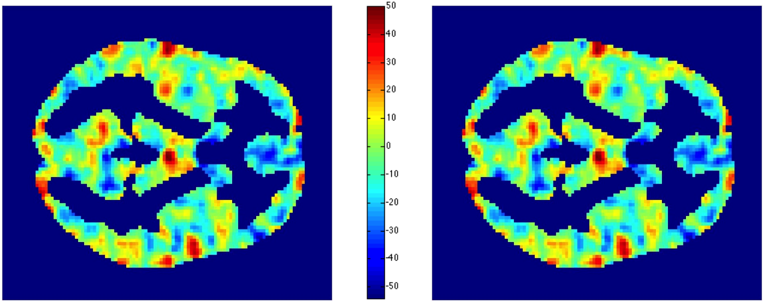
EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

- Stage I analysis produces estimates of visual fixation, WM load 1, WM load 3, and WM load 5 for each subject (FSL, SPM, etc).
- Compute contrasts of each WM load versus fixation for each subject.
  - (1) Load 1 vs. Fixation, (2) Load 3 vs. Fixation, and (3) Load 5 vs. Fixation.
- Fit second-stage univariate model (GLM) to estimate the group-level effects and associated variances.
  - Estimate final contrast to compare Load 3 vs Load 1 between controls and schizophrenia patients
  - Calculate test-statistic
- Fit second-stage multivariate model to estimate the group-level effects and associated variances.
  - Estimate final contrast to compare Load 3 vs Load 1 between controls and schizophrenia patients
  - Calculate test-statistic

Estimation

Contrast estimates:

GLMM	GLUM

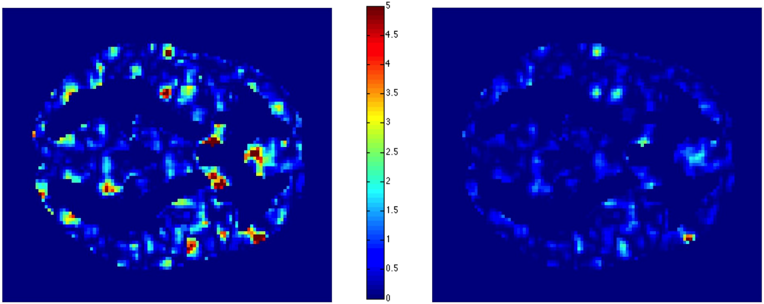


- Both methods produce unbiased estimates of regression coefficients and associated contrasts.
- $\theta = [\text{task3} - \text{task1}]_{\text{Controls}} - [\text{task3} - \text{task1}]_{\text{Patients}}$

Test Statistics

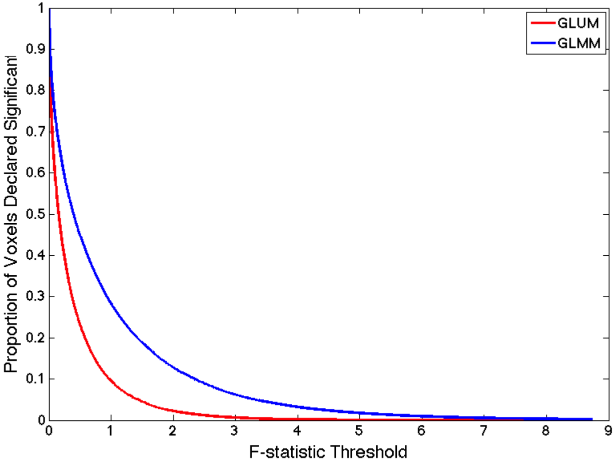
F-statistics:

GLMM	GLUM



- The GLMM often produces larger test statistics than the GLUM.
- $\theta = [\text{task3} - \text{task1}]_{\text{Controls}} - [\text{task3} - \text{task1}]_{\text{Patients}}$

Test Statistics

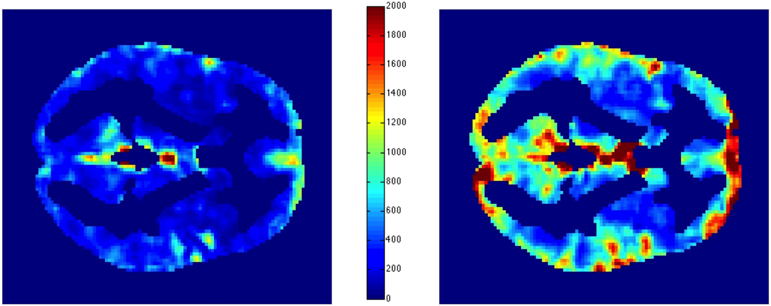


- This figure clearly reveals increased statistical power in GLMM relative to GLUM.

Variances

Contrast Variances:

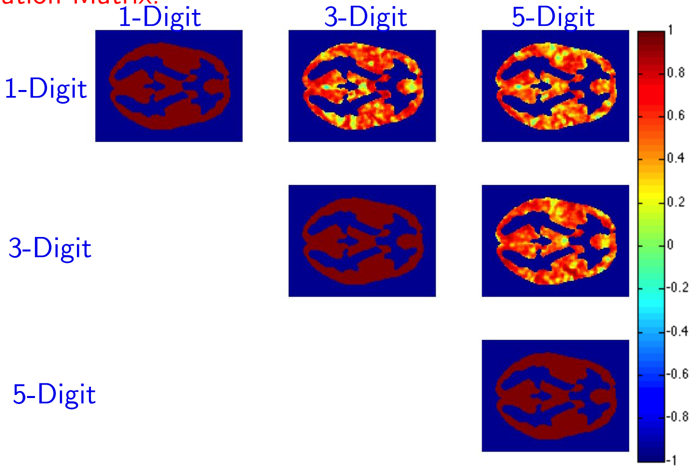
GLMM	GLUM



- The GLUM produces larger variances and will thus sacrifice statistical power.
- $\theta = [\text{task3} - \text{task1}]_{\text{Controls}} - [\text{task3} - \text{task1}]_{\text{Patients}}$

Task Correlations

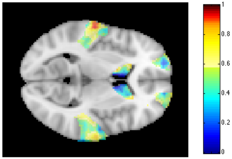
Correlation Matrix:



- The GLMM yields estimates of correlations between the three working memory loads (stage I contrasts).

Summary

- Mass univariate and multivariate linear models produce identical estimates of task-related changes.
- Multivariate modeling approaches consider dependencies between multiple dependent variables
  - Multiple effects/contrasts
  - Multiple voxels [Bowman et al., 2008; Zhang et al., 2012]
  - Multiple time points (e.g. longitudinal study)
- By accounting for correlations, multivariate methods generally
  - Increase efficiency (reduces variability)
  - Increase statistical power.
- Univariate approaches may have a deleterious effect on inference.



Acknowledgements

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | CBIS Center for Biomedical Imaging Statistics

Special Thanks*

- Phebe Brenne Kemmer, *Emory, Biostatistics and Bioinformatics*
- Anthony Pileggi, *Emory, Biostatistics and Bioinformatics*
- Lijun Zhang, PhD, *Emory, Biostatistics and Bioinformatics*

Research Team

- Shuo Chen, PhD, *Univ. of Maryland, Epidemiology and Biostatistics*
- Gordana Derado, PhD, *CDC*
- Ying Guo, PhD, *Emory, Biostatistics and Bioinformatics*
- Jian Kang, PhD, *Emory, Biostatistics and Bioinformatics*
- Wenqiong Xue, PhD, *Emory, Biostatistics and Bioinformatics*

- Qing He, *Emory, Biostatistics and Bioinformatics*
- Yize Zhao, *Emory, Biostatistics and Bioinformatics*