Statistical Analysis Programs in R for FMRI Data

Gang Chen, Ziad S. Saad, and Robert W. Cox

Scientific and Statistical Computing Core

NIMH/NIH/HHS/USA

http://afni.nimh.nih.gov/sscc/gangc

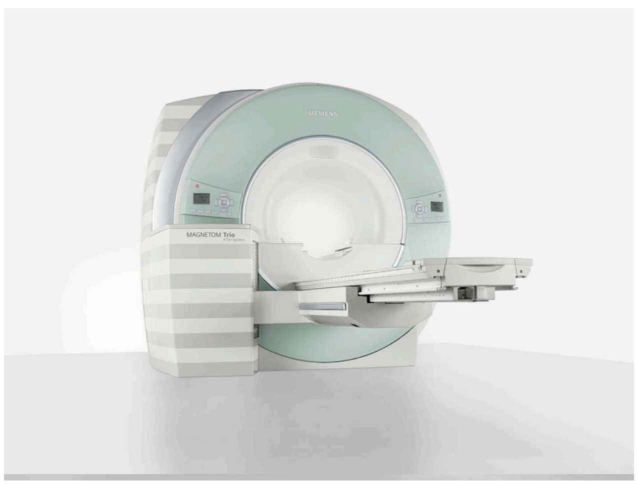
July 22, 2010

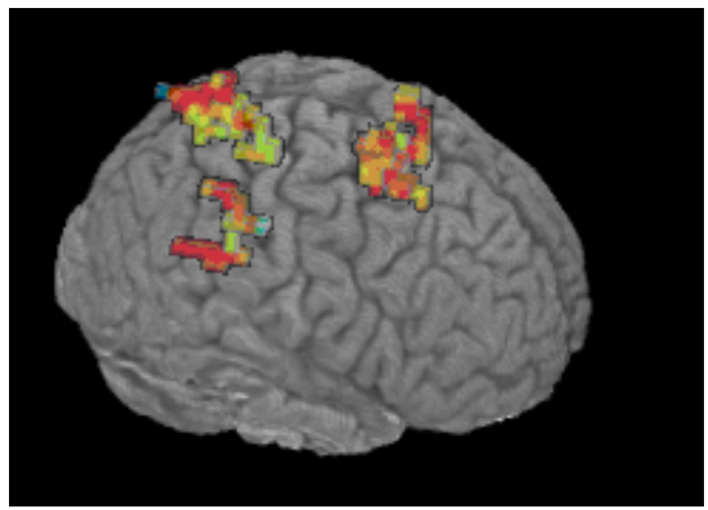
Overview

- What is FMRI?
- What kinds of analysis involved in FMRI data analyses
- Programs in R for FMRI data analyses (of NIfTI/AFNI data)
  - Group analysis
    - Mixed-effects meta analysis (MEMA): **3dMEMA**
    - Linear mixed-effects analysis (LME): **3dLME**
  - Connectivity analysis
    - Granger causality (vector autoregressive or VAR): **3dGC**, **1dGC**
    - Intra-class correlation analysis (ICC): **3dICC** and **3dICC_REML**
    - Structural equation modeling (SEM): **1dSEMr**
  - Data-drive analysis: Independent component analysis (ICA): **3dICA**
  - Kolmogorov-Smirnov test: **3dKS**
- Summary

FMRI in Neuroimaging

- Typical scanner: 3 Tesla = 60000 ✕ earth's magnetic field
- Measure changes in blood flow (hemodynamic response): BOLD signal
  - Indirect measure associated with neural activity during a task/condition
- Started in early 1990s; Little invasion, no radiation, *etc.*
- Interdisciplinary: physics, statistics, psychology, neuroanatomy, cognitive science, …
- Mind reading? Not there yet, but analyses produce colored blobs denoting activation regions in the brain





Data type in FMRI

- Brain volume
  - Anatomical: 3D
    - Typical spatial resolution: $1\times1\times1mm^3$; Dimensions: $256\times256\times128$ ~ 8 million voxels
  - Functional: 4D
    - Typical spatial resolution: $2.75\times2.75\times3.0mm^3$; Dimensions: $80\times80\times33$ ~ 20,000 voxels
    - Typical temporal resolution: ~2s; Dimension: a few hundred time points
  - Number of subjects: 10-20
- Surface
- ROI
- Behavioral

Analysis types in FMRI

- Individual subjects: time series regression
  - Voxel-wise or massively univariate model $y = X\beta + \varepsilon,\ \varepsilon \sim N(0,\sigma^2 V)$
  - σ^2 and V vary spatially (across voxels)
  - REML + GLSQ
  - Runtime: 1 minute or more
- Group analysis: summarizing across subjects
  - t-test, ANOVA, regression
  - Runtime: seconds
- Connectivity analysis: search for or test network in the brain
  - Correlation analysis, structural equation modeling, Granger causality, dynamic causal modeling, *etc*.
- Multivariate approach: data-driven
  - PCA/ICA, SVM, kernel methods, *etc*.

AFNI = **A**nalysis of **F**unctional **N**euro**I**mages

- Developed to provide an environment for FMRI data analyses
  - Started in 1994 by Bob Cox at MCW, Milwaukee , Wisconsin
  - **Open source** mainly in C, plus some R and Matlab
- Important principles in the development of AFNI:
  - Allow user to stay close to the data and view it in many different ways
  - Power to assemble pieces in different ways to make customized analyses
    - "With great power comes great responsibility"

      — **to understand the analyses and the tools**
  - Provide mechanism/tools, not policy/assembling line

Conventional group analysis in FMRI

- Take regression coefficient β's from each subject, and run *t*-test, AN(C)OVA, LME
  - One-sample *t*-test: $y_i = \alpha_0 + \delta_i$, for *i*th subject; $\delta_i \sim N(0, \tau^2)$
- Three assumptions
  - Within/intra-subject variability (standard error, sampling error) is relatively small compared to cross/between/inter-subjects variability
  - Within/intra-subject variability roughly the same across subjects
  - Normal distribution for cross-subject variability (no outliers)
- Violations prevalent, leading to suboptimal/invalid analysis
  - Common to see 40 - 100% variability due to within-subject variability
  - Non-uniform within/intra-subject variability across subjects
  - Not rare to see outliers

Mixed-Effects Meta Analysis

- For each effect estimate (β or linear combination of β's)
  - How good is the β estimate?
    - Reliability/precision/efficiency/certainty/confidence: standard error (SE)
    - Smaller SE → more accurate estimate
  - t-statistic of the effect
    - Signal-to-noise or effect vs. uncertainty: $t = \beta/\text{SE}$
    - SE contained in t-statistic: SE = β/t
  - Trust those β's with high reliability/precision (small SE) through weighting/compromise
    - β estimate with high precision (lower SE) has more say in the final result
    - β estimate with high uncertainty gets downgraded
  - One-sample model: $y_i = \alpha_0 + \delta_i + \varepsilon_i$, for ith subject
    - $\delta_i \sim N(0, \tau^2)$, $\varepsilon_i \sim N(0, \sigma_i^2)$, σ_i^2 known

New group analysis program: 3dMEMA

- Algorithms (MoM/REML + WLS) similar to R package **metafor** (Wolfgang Viechtbauer) with parallel computing using R package **snow**
- Runtime: a few minutes or more with 4 CPUs
- Analysis types
  - 1-, 2-, paired-sample test
  - Covariates: age, IQ, behavioral data, between-subjects factors, *etc.*
- Input: effect estimate + *t* from individual subjects
- Output
  - Group level: group effect + Z/*t*
  - Cross-subject heterogeneity + χ^2-test
  - Individual level: ICC + Z
- Assessing outliers with 4 estimated quantities
  - Cross-subject variance (heterogeneity) τ^2 at group level
  - χ^2-test for H_0: $\tau^2=0$ at group level
  - Intra-class correlation for each subject
  - *Z*-statistic for the residuals for each subject
- Outliers modeled through a Laplace distribution of cross-subject variability

Comparison: 3dMEMA vs. FLAME1+2

- Frequentist (REML) vs. Bayesian (MCMC)
- Runtime: a Mac OS X 10.6.2 with 2×2.66 GHz dual-core Intel Xeon. Group analysis: 10 subjects, 218379 voxels. FSL ver. 4.1.4

	3dMEMA with 4 parallel jobs	3dMEMA with 2 parallel jobs	3dMEMA with a single processor	Flame 1+2 (FSL)
Without modeling outliers	3	4.5	8	385
Modeling outliers	22.5	34.5	65	847

Linear Mixed-Effects Analysis

- $Y_i = X_i\beta + Z_i b_i + \varepsilon_i$, $b_i \sim N_q(0, \psi)$, $\varepsilon_i \sim N_{n_i}(0, \sigma^2\Lambda_i)$, $q=1$
- Parameters: β, ψ, and $\sigma^2\Lambda_i$
- Fixed/mean/systematic effects in population $X_i\beta$
- Random effects $Z_i b_i$
  - Across-subjects variability: deviation of each subject from mean effects $X_i\beta$
- Random effect ε_i
  - Within-subject variability (across multiple effects)

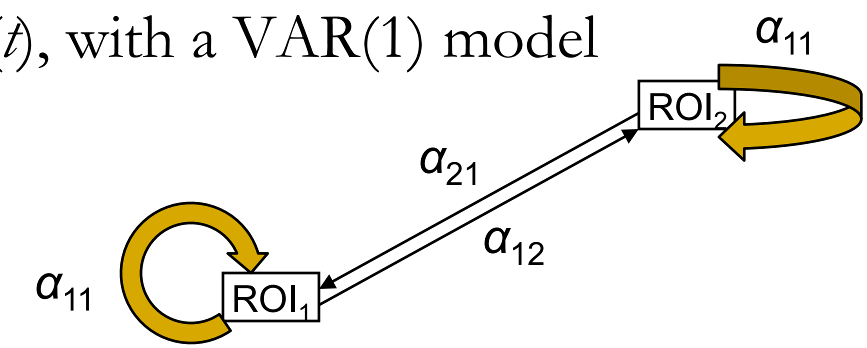
Linear Mixed-Effects Analysis: **3dLME**

- Use function **lme**() in R package **nlme** (Pinheiro *et al.*)
- Parallel computing using R package **snow** (Tierney *et al.*)
- Contrasts through R package **contrast** (Kuhn *et al.*)
- Runtime: a few minutes or more with 4 CPUs
- **3dLME** is more flexible than conventional approach
  - Popular ANOVA, paired-, one- and two-sample *t*-test: special cases of LME
    - ANOVA: compound symmetry in ψ
  - Capable to model various structures in ψ and $\sigma^2\Lambda_i$
  - Much easier to deal with missing data and covariates
  - Modeling subtle HRF shape through multiple basis functions
    - Zero intercept with H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ (k = # time points in HRF)

Granger Causality or VAR

- Granger causality: A Granger causes B if
  - time series at A provides statistically significant information about time series at B at some time delays (order)
- 2 ROI time series, $y_1(t)$ and $y_2(t)$, with a VAR(1) model

$$y_1(t) = \alpha_{10} + \alpha_{11} y_1(t-1) + \alpha_{12} y_2(t-1) + \varepsilon_1(t)$$
$$y_2(t) = \alpha_{20} + \alpha_{21} y_1(t-1) + \alpha_{21} y_2(t-1) + \varepsilon_2(t)$$



  - Matrix form: $Y(t) = \boldsymbol{\alpha} + AY(t\text{-}1) + \boldsymbol{\varepsilon}(t)$, where

$$Y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} \quad A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}$$

- n ROI time series, $y_1(t), \ldots, y_n(t)$, with VAR(p) model

$$Y(t) = \alpha + \sum_{i=1}^{p} A_i Y(t-i) + \varepsilon(t) \quad \alpha = \begin{bmatrix} \alpha_{10} \\ \vdots \\ \alpha_{n0} \end{bmatrix} \quad Y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{bmatrix} \quad A_i = \begin{bmatrix} \alpha_{11i} & \cdots & \alpha_{1ni} \\ \vdots & \ddots & \vdots \\ \alpha_{n1i} & \cdots & \alpha_{n1i} \end{bmatrix} \quad \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_n(t) \end{bmatrix}$$

GC in AFNI: **3dGC** and **1dGC**

- Exploratory approach: ROI search with **`3dGC`**
  - Not a solid approach; can explore possible ROIs in a network
  - Bivariate model: Seed vs. rest of brain
  - 3 paths: seed to target, target to seed, and self-effect
  - Use R packages **vars** (Bernhard Pfaff) and **snow** (Tierney *et al.*)
- Path strength significance testing in a network: **`1dGC`**
  - Assume all ROIs are known in the network
  - Multivariate model with pre-selected ROIs
  - Use R package **vars** for VAR modeling (Bernhard Pfaff)
  - Use R package **network** for plotting (Butts *et al.*)
  - Preserve path sign (+ or -), in addition to its direction, from individual subjects all the way to group level analysis

Intra-Class Correlation (ICC)

- Classical definition
  - Variability of a random variable relative to total variance
  - ICC varieties in *Shrout and* Fleiss (1979), Psychological Bulletin, Vol. 86, No.2, 420-428
    - Based on mean squares of variance in ANOVA framework
    - Problem: not rare to have **negative** ICC values, and difficult to interpret
  - Applied to FMRI data
    - Reliability of scanning sessions/sites
- Extended definition
  - Linear mixed-effects model

3dICC and **3dICC_REML**

- 3dICC
  - Use function **lm**() in R
  - Parallel computing using R package **snow** (Tierney *et al.*)
  - 2-way and 3-way random-effects ANOVA model
  - May get negative ICC values
- 3dICC_REML
  - Use function **lmer**() in R package **lme4** (Bates and Maechler)
  - No negative ICC values
  - Missing data allowed
  - No limit on # random variables

Miscellaneous Tools

- SEM or path analysis, analysis of covariance: **1dSEMr**
  - Causal model for a network of ROIs
  - Use R package **sem** (John Fox)
- Independent component analysis: **1dICA**
  - Use R package **fastICA** (Marchini *et al.*)
  - Spatial ICA
- Kolmogorov-Smirnov test: **3dKS**
  - Use R package **snow** (Luke Tierney *et al.*)

Summary

- Statistical analysis programs in R for FMRI data analysis of NIfTI/AFNI datasets
  - Mixed-effects meta analysis (MEMA): **3dMEMA**
  - Linear mixed-effects analysis (LME): **3dLME**
  - Granger causality (vector autoregressive or VAR): **3dGC**, **1dGC**
  - Intra-class correlation analysis (ICC): **3dICC** and **3dICC_REML**
  - Structural equation modeling (SEM): **1dSEMr**
  - Independent component analysis (ICA): **3dICA**
  - Kolmogorov-Smirnov test: **3dKS**
- All programs available for download with AFNI, and at

  http://afni.nimh.nih.gov/sscc/gangc

Acknowledgements

- **Karsten Tabelow**, Weierstrass Institute for Applied Analysis and Stochastics
- **Jarrod Hadfield**, Institute of Evolutionary Biology, School of Biological Sciences, King's Buildings, University of Edinburgh
- **Wolfgang Viechtbauer**, Department of Methodology and Statistics, School for Public Health and Primary Care, Maastricht University
- **John Fox**, Department of Sociology, McMaster University
- **Douglas Bates**, Department of Statistics, University of Wisconsin - Madison
- **Eugene Demidenko**, Section of Biostatistics & Epidemiology, Dartmouth Medical School
- **Harvey Xianggui Qu**, Department of Mathematics and Statistics, Oakland University
- **Bernhard Pfaff**, Invesco Asset Management Deutschland GmbH, Frankfurt Am Main, Germany
- **Patrick Brandt**, School of Economic, Political and Policy Sciences, University of Texas, Dallas
- **Chris Sims**, Department of Economics, Princeton University
- **Luke Tierney**, Department of Statistics and Actuarial Science, University of Iowa