Don't Fall in Love with Your Model: Model Selection for Graphical Models

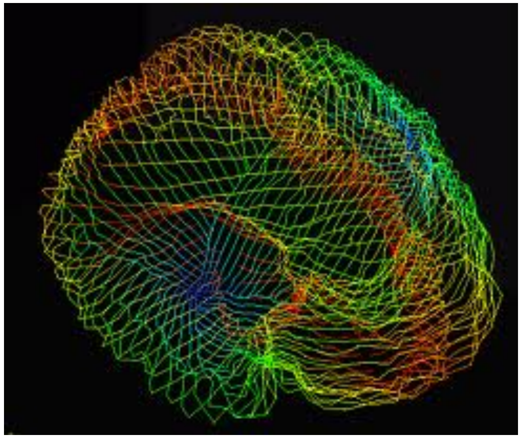
Joseph Ramsey
Carnegie Mellon University

Or Optimistically--

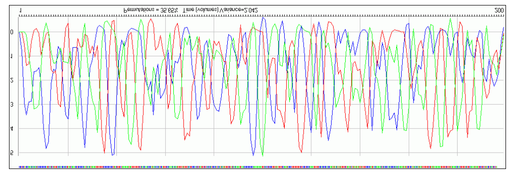
Prospects for Computer Aided Model Specification in fMRI

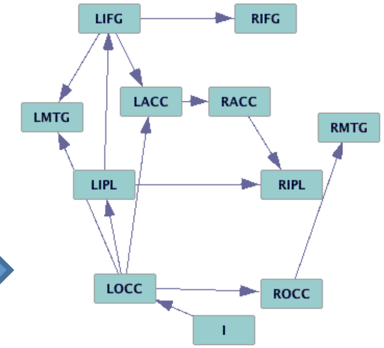
Joseph Ramsey

Carnegie Mellon University



Goals





1. From Imaging data, to extract as much information as we can, as accurately as we can, about which brain regions influence which others in the course of psychological tasks.

2. To generalize over tasks

3. To specialize over groups of people.

What Are the Brain Variables?

In current studies, from

20,000 + ……………………………….3

voxels ROIs

ROI = Region of interest

Question: How sensitive are causal inferences to brain variable selection?

Search Complexity: How Big is the Set of Possible Explanations?

For graphical models:

$X \quad\quad Y$

$$\left.\begin{array}{l} X \longrightarrow Y \\ X \longleftarrow Y \\ X \rightleftarrows Y \\ X \longleftrightarrow Y \end{array}\right\}$$

For N variables:

$$8^{\binom{N}{2}}$$

Statistical Complexity

- Graphical models are untestable unless parameterized into statistical models.
- Incomplete models of associations are likely to fail tests.
- Multiple testing problems.
- Multiple subjects/Missing ROIs.
- No fast scoring method for mixed ancestral graphs that model feedback and latent common causes.

Measurement Complexity

- Sampling rate is slower than causal interaction speed.
- Indirect measurement creates spurious associations of measured variables:

$$\begin{tikzcd} N1 \arrow[r] \arrow[d] & N2 \arrow[r] \arrow[d] & N3 \arrow[d] \\ X1 & X2 & X3 \end{tikzcd} \qquad \begin{tikzcd} X1 \arrow[rr] \arrow[d] & & X3 \\ X2 \arrow[urr] & & \end{tikzcd}$$

Neural N, measured X   Regression of X3 on X1, X2

Specification Strategies

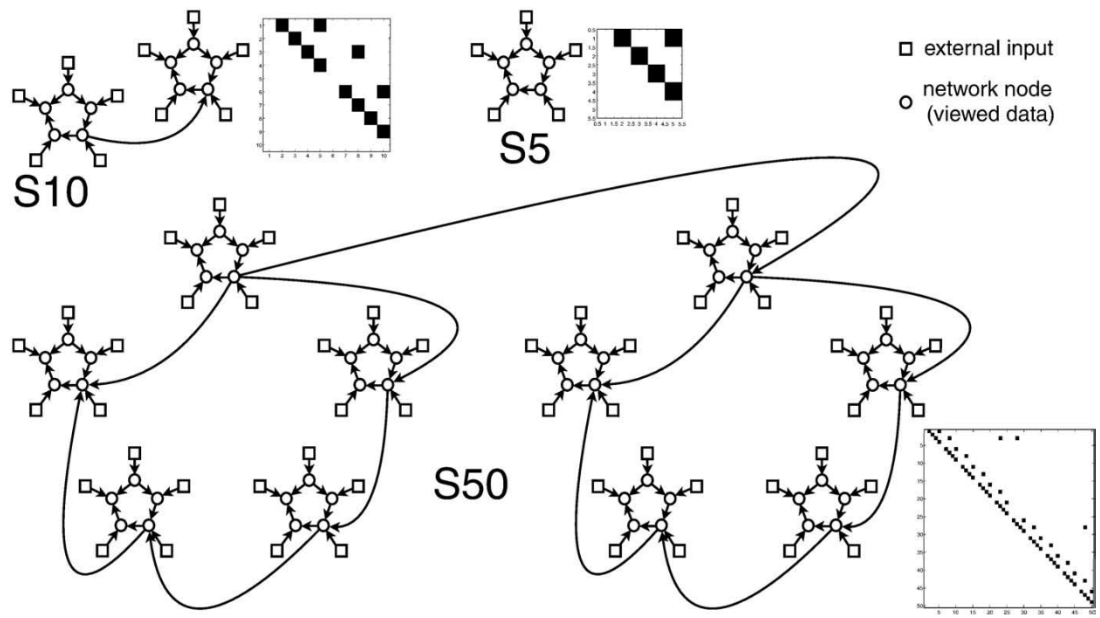
1. Guess a model and test it.
2. Search the model space or some restriction of it.
   a. Search for the full parameterized structure
   b. Search for graphical structure alone
   c. Search for graphical features (e.g, adjacencies)

What Evidence of What Works, and Not?

- Theory.
  - Limiting correctness of algorithms (PC, FCI, GES, LiNGAM, etc.)
- Prior Knowledge
  - Do automated search results conform with established relationships?
- Animal Experiments (Limited)
- Simulation Studies

Brief Review: Smith's Simulation Study

- 5 to 50 variables
- 28 simulation conditions, 50 subjects/condition.
- 38 search methods
- Search 1 subject at a time.



Smith's Results

- Adjacencies:
  - Partial Correlation methods (GLASSO) and several "Bayes Net" methods from CMU get ~ 90% correct in most simulations.
- Edge Directions
  - Smith: "None of the methods is very accurate, with Patel's τ performing best at estimating directionality, reaching nearly 65% d-accuracy, all other methods being close to chance." (p. 883)
  - Most of the adjacencies for Patel's τ are false.

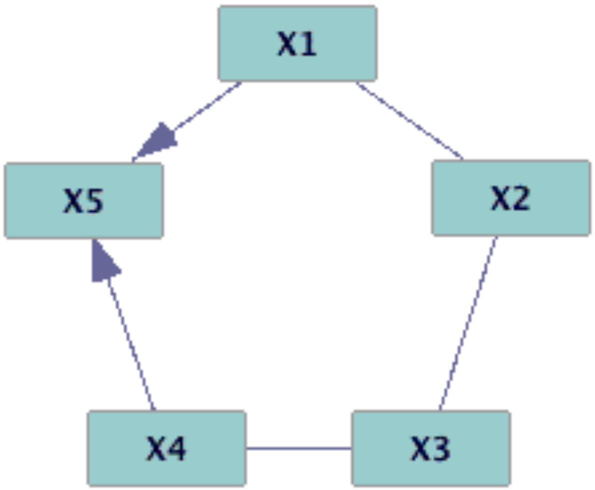
Take Away Conclusion?

- Nothing works!
- Methods that get adjacencies (90%) cannot get directions of influence.
- Methods that get directions (60% - 70%) for normal session lengths cannot tell true adjacencies from false adjacencies.
- Even with unrealistically long sessions (4 hours), the best method gets 90% accuracy for directions but finds very few adjacencies.

Idea…

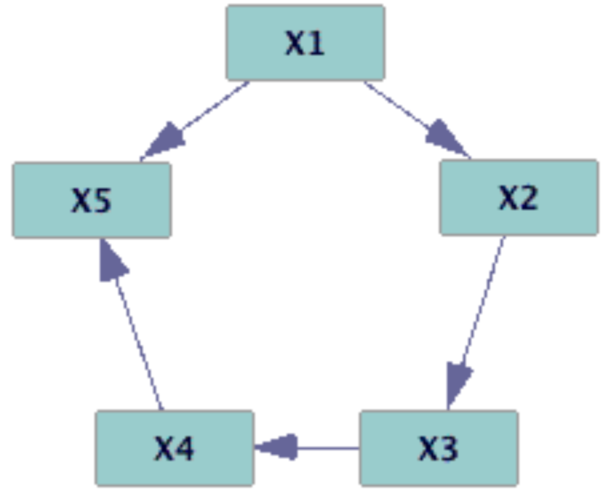
- If we could:
  - Increase sample size (effectively) by using data from multiple subjects
  - Focus on a method with strong adjacencies
  - Combine this with a method with strong orientations
- We may be able to do better (Ramsey, Hanson and Glymour, Accepted, *NeuroImage*)

What *can* we learn?



Linear Models,
Covariance Data,
Pattern/CPDAG

(1)



Truth!
Linear Models, Non-Gaussian Noises (LiNG),
Directed Graph

(2)

IMaGES

- Adaptation for multiple subjects of GES, a Bayes net method tested by Smith, et al.
- Iterative model construction using Bayesian scores separately on each subject at each step; edge with best average score added.
- Tolerates ROIs missing in various subjects.
- Seeks feed forward structure only.
- Finds adjacencies between variables with latent common causes.
- Forces sparsity by penalized BIC score to avoid triangulated variables (see Measurement Complexity)

Comparisons Using Smith's Data

Averages for GES (run on 50 single subjects) versus IMAGES (run on 10 subjects, drawn 50 times with replacement)

#Vars	#Edges	Adjacencies		Directions		
		# False +	# False -	#Correct	#Not	Accuracy
1. 5	5	.06/0	0.38/0	1.2/1.78	.66/.26	65%/88%
4. 50	61	6.8/0.1	14.6/.98	26/42	18/7	62%/85%

Similar IMaGES dominance against GES and other methods for simulation conditions not made deterministic, or with insignificant connectivities, or corrupted ROIS

IMaGES/LOFS

- Smith (2011): “Future work might look to optimize the use of higher-order statistics specifically for the scenario of estimating directionality from fMRI data.”
- LiNGAM orients edges by non-Normality of higher moments of the distributions of adjacent variables.
- LOFS uses the IMaGES adjacencies, and the LiNGAM idea for directing edges (with a different score for non-Normality, and without independent components).
- Unlike IMaGES, LOFS can find cycles.

Comparison using Smith's Data

Averages for IMaGES versus IMaGES/LOFS (run on 10 subjects, drawn 50 times with replacement)

	#Vars	#Edges	Adjacencies		Directions		
			#False +	# False -	#Correct	#Not	Accuracy
1.	5	5	0	0	1.8/4.9	0.3/0.0	88%/100%
4.	50	61	0.1	0.98	42/53	7/3.6	85%/94%

For all "uncorrupted" simulations, with 5 or 10 variables, IMaGES/LOFS has

- almost perfect adjacencies
- most orientation accuracies above 90%.
- correctly identifies a 5 cycle and the 2 cycle for the only relevant simulation set with at least 10 subjects.

Remaining Problems for Any Known Method

- Real data are messier than simulations
- Discovering nearly canceling 2 cycles is hard (but maybe not impossible).
- Reliability of search may be worse with event designs than with block designs
- Search results may not be stable over slightly varying search parameters.
- Can latent latents be identified?
- Subjects that differ in *causal* structures will yield poor results for multi-subject methods.
- Sensitivity to how ROIs are formed?

Conclusion

- Machine Learning methods taken off the shelf from other disciplines aren't going to work on fMRI data.
- Even sensible algorithms for analyzing fMRI data in simulation draw conclusions that are sensitive to many factors, in simulation and in practice.
- What we need in search algorithms aren't push-button programs but rather good robotic colleagues.
  - Give good advise.
  - Say why they're giving that advise.
  - Replaceable when better robotic colleagues come along.

Thanks!

- Software available at http://www.phil.cmu.edu/tetrad.

- S. M. Smith, K L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, M. W. Woolrich (2011), Network modelling methods for fMRI, *NeuroImage*.
- J.D. Ramsey, S.J. Hanson, C. Hanson, Y.O. Halchenko, R.A. Poldrack, and C. Glymour(2010), Six Problems for causal inference from fMRI, *NeuroImage*.
- J.D. Ramsey, S.J. Hanson, C. Glymour. Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. Accepted, *NeuroImage*.

- Thanks to the James S. McDonnell Foundation.