Network analysis: Dynamic Causal Modeling and Granger Causality

Luis Hernandez-Garcia, Ph.D.
UM FMRI course
Motivation

- Brain as a network
  - Inputs, outputs, causal relationships
- FMRI can localize brain activity
- FMRI can identify connectivity (networks)
- Can FMRI be used to characterize the hierarchy of the network? Top Down or Bottom-Up?
- Does activity in A cause activity in B?
Today’s outline

- Networks and FMRI
- Dynamic Causal Modeling
- Granger Causality
- … but what is Causality?
- Ascendancy
- Evaluation of methods.
Elements of Brain Networks

- Nodes
- Edges: driving inputs
  - Endogenous connections
  - Exogenous Inputs
- Modulation
- Feedback

Hernandez-Garcia, FMRI Course 2015
Temporal Resolution in FMRI

- We can image quite fast (<50ms./slice)
  - with new multi-band techniques you can do a whole brain in less than 500 ms.
  - just barely enough to catch neuronal shifts in activation

- HOWEVER - we observe much slower vascular effects, not neuronal ones.

- In a linear system, a shift in neuronal events corresponds to the same shift in vascular events…
  - Is the BOLD effect linear enough?
  - Vascular responses are variable throughout the brain
High temporal resolution: mental chronometry

Temporal Resolution in FMRI

- **Sensitivity of statistical tests to temporal shifts**
Granger Causality

A test to determine the directionality of a connection between a pair of nodes.

Disclaimer: “Granger-causality” is not “Causality”
Autoregressive systems (background)

- Part of the “current” sample can be predicted from previous samples (history)

\[ A[n] = \sum_{m=1}^{M} \delta_{a,m} A[n - m] + \epsilon_a[n] \]

- Examples: drifts in FMRI scans, price of oil

...
Granger Causality

What if the current sample of a given process (A) can also be predicted in part by the history of another process (B)?

\[
A[n] = \sum_{m=1}^{M} \delta_{a,m} A[n - m] + \sum_{m=1}^{M} \delta_{ab,m} B[n - m] + \varepsilon_a[n]
\]

\[
B[n] = \sum_{m=1}^{M} \delta_{b,m} B[n - m] + \sum_{m=1}^{M} \delta_{ba,m} A[n - m] + \varepsilon_b[n]
\]


Hernandez-Garcia, FMRI Course 2015
Granger Causality

Are the cross-terms significant?

1. Fit the models with and without the cross terms
2. Use the residuals to compute F statistics

\[
F_{B \rightarrow A} = \frac{RSS_{0a}}{RSS_{1a}} \quad F_{A \rightarrow B} = \frac{RSS_{0b}}{RSS_{1b}}
\]

Inference through “boot-strapping”
Multivariate Granger Causality

- Multiple time courses
- Multiple influences of previous samples

\[ X(t) = \sum_{n=1}^{p} A(n)X(t-n) + E(t) \]

Multiple ROI’s

Time-shifted ROI’s

Influence coefficients

Variations on the theme

- Add instantaneous terms into the formulation. “SEM - Granger hybrid”


- Partial Directed Coherence. Similar analysis - frequency domain.

Limitations of Granger Causality

Different HRF’s in different parts of the brain could appear as false Granger Causal relationships.


Heavily dependent on sampling and noise characteristics.

- Adding noise to one channel can “reverse” the direction of the causality.
Dynamic Causal Modeling

Accounts for:
- Endogenous connections
- Modulation
- Direct Inputs
- Feedback


Hernandez-Garcia, FMRI Course 2015
Dynamic Causal Modeling

Fit of a priori network model:

- “state space” models include feedback, external modulation
- Vascular models
- …Lots of parameters to estimate! (Bayesian Estimation needed)
Dynamic Causal Modeling

State-Space Modeling in a nutshell:

- Let $z$ be the values in the different nodes.
  - $Z$ is a vector with a value for each node
  - $Z$ is a function of time
- Those values will change as a function of the inputs from other nodes and from external inputs $[u(t)]$ … etc.

\[
\frac{d\tilde{z}(t)}{dt} = A \cdot \tilde{z}(t) + B \cdot u(t) \cdot \tilde{z}(t) + C \cdot \tilde{u}(t)
\]
\[
\frac{d\tilde{z}(t)}{dt} = A \cdot \tilde{z}(t) + B \cdot u(t) \cdot \tilde{z}(t) + C \cdot \tilde{u}(t)
\]
Further complications: we don’t observe $z(t)$ directly, but rather $z(t) \ast \text{hrf}(t)$

Dynamic Causal Modeling

- The whole thing can be expressed as a bilinear model

\[
y = h(\theta, u) + X\beta + W
\]

\[
\theta = \begin{bmatrix} \theta_c \\ \theta_h \end{bmatrix}
\]

- Our mission is to estimate \( \theta \)
Estimation in DCM

- Lots of parameters to estimate!
- Estimate hemodynamic parameters and DCM parameters separately.

Bayesian Estimation:

- \( P(M \mid Y) = \frac{P(Y \mid M) \cdot P(M)}{P(Y)} \)
- Maximize Likelihood / Likelihood Ratio.
“Likelihood” is the probability of observing the data, given a set of parameters for a model.

\[ P(Y \mid M) \]

**Example -**

- Imagine a linear model and assume that the observed data follows a normal distribution.
- We can write an equation for the probability of observing the data, given an arbitrary set of parameters.
- If we guess the parameters at random, the residuals will follow Gaussian distribution.

\[ Y = X\beta + \varepsilon \]

\[ L(\beta) = P(Y \mid \beta) \]

\[ L = \frac{1}{\left(2\pi|\sigma^2 I|\right)^{n/2}} \exp \left( -\frac{1}{2} \frac{\|Y - X\beta\|^2}{\sigma^2} \right) \]
Maximizing the Likelihood

- Sometimes it’s a simple, analytical max/min problem: figure out derivative, find the zeros, ... etc.

- Usually ... it’s not so simple. Use iterative techniques to find the solution that minimizes the residuals, maximizes likelihood function.
Bayesian Estimation

Bayes Theorem:

\[ P(M|y) = \frac{P(y|M) \cdot P(M)}{P(y)} \]

Find M that maximizes posterior probability \( P(M|y) \)
Some practical points about DCM implementation (SPM)

- Outputs are the estimated distributions of the parameters.
  - Parameter estimates
  - Likelihood of parameters
  - Covariance of the errors

- Need at least ten data points per parameter to estimate (heuristic).

- Model comparison is a powerful tool for inference.
  - Trade-off between complexity and accuracy.

- Look at posterior covariance matrices.
  - High off-diagonal terms indicate that some parameters are correlated, redundant?
Limitations of DCM

- Very complete model, but ...
- Many parameters. Can you trust it?
- Confounds between hemodynamic and neural parameters.
- Dependence on sampling rate.
- Priors are important!
Examples of Clinical Applications of DCMs

**Major depression**

**Autism**

**Schizophrenia**
About Causality …
How do we infer Cause and Effect?

- Big picture: Philosophy, Cognition, Social Sciences, Policy, Artificial Intelligence…
- We infer causal relationships by association measures and exploring “counterfactuals”.
- Experimentally: Interact with the system. What changes? Is function preserved? (TMS, ablations, modulation …)

"… we may define a cause to be an object, followed by another, and where all objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed …"
— David Hume, An Enquiry Concerning Human Understanding.
"Footprints" of Causal Relationship

- Causality in time series (signal processing)
  - Correlation / Statistical Dependence
  - **Timing**: Effect always after cause

- Science in general uses “counterfactuals”
  - Necessity ?
  - Sufficiency ?
How do we infer Cause and Effect?

The First Lawsuits Against Cigarette Manufacturers

When the first reports emerged linking cigarettes to cancer emerged in the 1950s, plaintiffs began suing cigarette manufacturers. Plaintiffs in these early cases — usually smokers with lung cancer — typically employed several legal theories in their lawsuits:

- negligent manufacture - the tobacco companies failed to act with reasonable care in making and marketing cigarettes
- product liability - the tobacco companies made and marketed a product that was unfit to use
- negligent advertising - the tobacco companies failed to warn consumers of the risks of smoking cigarettes
- fraud, and
- violation of state consumer protection statutes (most of which prohibit unfair and deceptive business practices).

Tobacco manufacturers responded in full force, fighting each lawsuit and refusing to settle out of court. They relied on several defense strategies, arguing that:

- Tobacco was not harmful to smokers.
- Smokers’ cancer was caused by other factors.
- Smokers assumed the risk of cancer when they decided to smoke.

The tobacco companies prevailed in all of these early lawsuits. (To learn more about product liability, negligence, and fraud, see Nolo's articles Toxic Torts: Legal Theories of Liability and Defective Product Claims: Theories of Liability.)
Causal Relationships?

Newton  Leibnitz  Fourier  Einstein

Descartes  Fermat  Taylor
Ascendancy

- Looks into the questions of sufficiency and necessity

- General idea:
  - Identify neuronal events in two nodes
  - Is one node active every time the other one is?
    A subset of the time?
  - Is that statistically significant?
Ascendancy

- Begin by identifying events: indicator function (very tricky stuff !!!)

- Look for coincidences in events (Relax timing requirements – it’s all blurred anyway)

- Is activity in one node a subset of the activity in another node? Ascendant.

- You can think of this as a measure of “necessity”
Ascendancy

\[ \tau_{AB} = \begin{cases} 
1 - \frac{P(AB) + P(\overline{AB})}{P(AB) + P(\overline{AB})} & \text{, if } P(AB) > P(\overline{AB}) \\
\frac{P(AB) + P(\overline{AB})}{P(AB) + P(\overline{AB})} - 1 & \text{, otherwise}
\end{cases} \]

- Use Bayesian statistics to infer significance
- (Dirichlet prior distributions)

Hernandez-Garcia, FMRI Course 2013
Limitations of Ascendancy

- Event detection, timing
- Physiological confounds
- Interpretation: we’re only looking at two nodes at a time
- Bayesian statistics require assumptions
Comparing Network Analysis Strategies


Strategies:

- Coherence.
- Lags: Granger, DCM.
- Bayes Net: Graph Theory.
- Conditional Probability.
Comparing Strategies


Fig. 2. The main network topologies fed into the FMRI data simulations. For each network graph the corresponding connection matrix is shown. Where an element in the upper diagonal of the matrix implies a directed connection from a lower-numbered node to a higher-numbered one.
Conclusions

- Tread carefully
- DCM is most popular (requires a priori network)
- Many approaches: None of these methods are perfect. (In fact, they’re all pretty lousy for FMRI data)
  - Importance of good ROIs!
  - Correlation based and DAG methods are very good at isolating relationships.
  - Ascendancy is very good at identifying pairwise directionality.
  - Lag based methods break down very easily in FMRI
- Open discussion?
Other references