Statistical Challenges for Neuroimaging Data Analysis
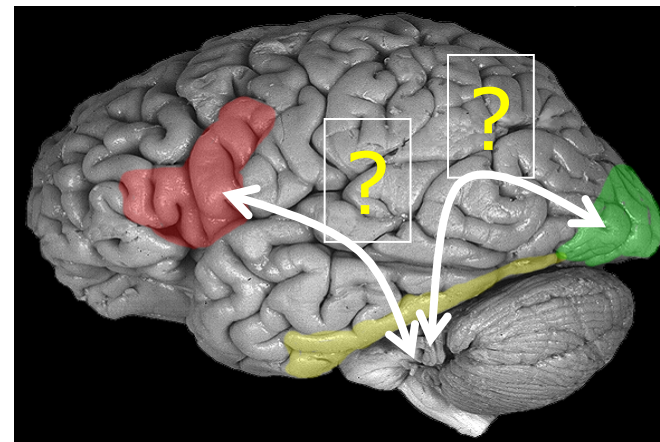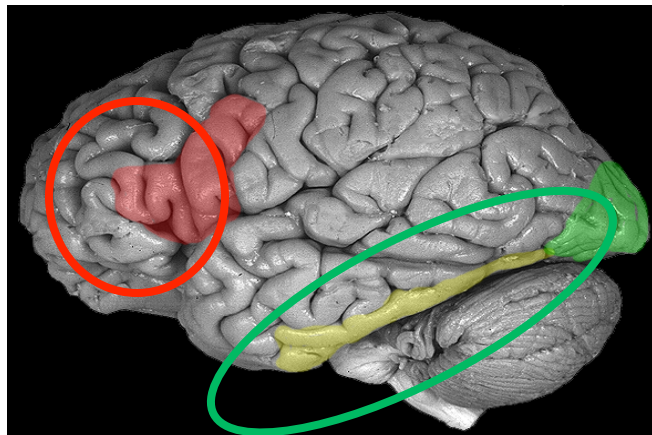## Lecture 2. Brain Connectivity Analysis

WNAR 2018 @ Edmonton
Linglong Kong
lkong@ualberta.ca

UNIVERSITY OF ALBERTA

"uplifting the whole people"

— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Brain Connectivity

➢ **Functional Segregation:** Human brain mapping has been primarily used to provide maps that show which regions of the brain are activated by specific tasks.

➢ Recently, there has been an interest in augmenting this type of analysis with brain connectivity studies which describe how various regions interact and how interactions depend on experimental conditions.

Functional Connectivity

➢ Functional connectivity is a statement about observed associations among regions and/or performance and physiological variables.

➢ It does not comment on how these associations are mediated.

➢ Functional connectivity analysis is usually performed using data-driven transformation methods which make no assumptions about the underlying biology.

Functional Connectivity

➢ Methods:

◇ Seed analyses

◇ Psychophysiological interaction analyses

◇ Principle Components Analysis

◇ Partial Least Squares

◇ Independent Components Analysis

Effective Connectivity

➢ Effective connectivity analysis is performed using statistical models which make anatomically motivated assumptions and restricts inference to networks comprising of a number of pre-selected regions of interest.

➢ These methods are hypothesis driven rather than data-driven and most applicable when it is possible to specify the relevant functional areas.

Effective Connectivity

➢ Methods:

✧ Structural Equation Modeling

✧ Granger Causality

✧ Dynamic Causal Modeling

• Note that Granger causality does not rely on an a priori specified structural model.

Functional Connectivity

Levels of Analysis

➢ Functional connectivity can be applied at different levels of analysis, with different interpretations at each.

➢ Connectivity across time can reveal networks that are dynamically activated across time.

➢ Connectivity across trials can identify coherent networks of task related activations.

➢ Connectivity across subjects can reveal patterns of coherent individual differences.

➢ Connectivity across studies can reveal tendencies for studies to co-activate within sets of regions.
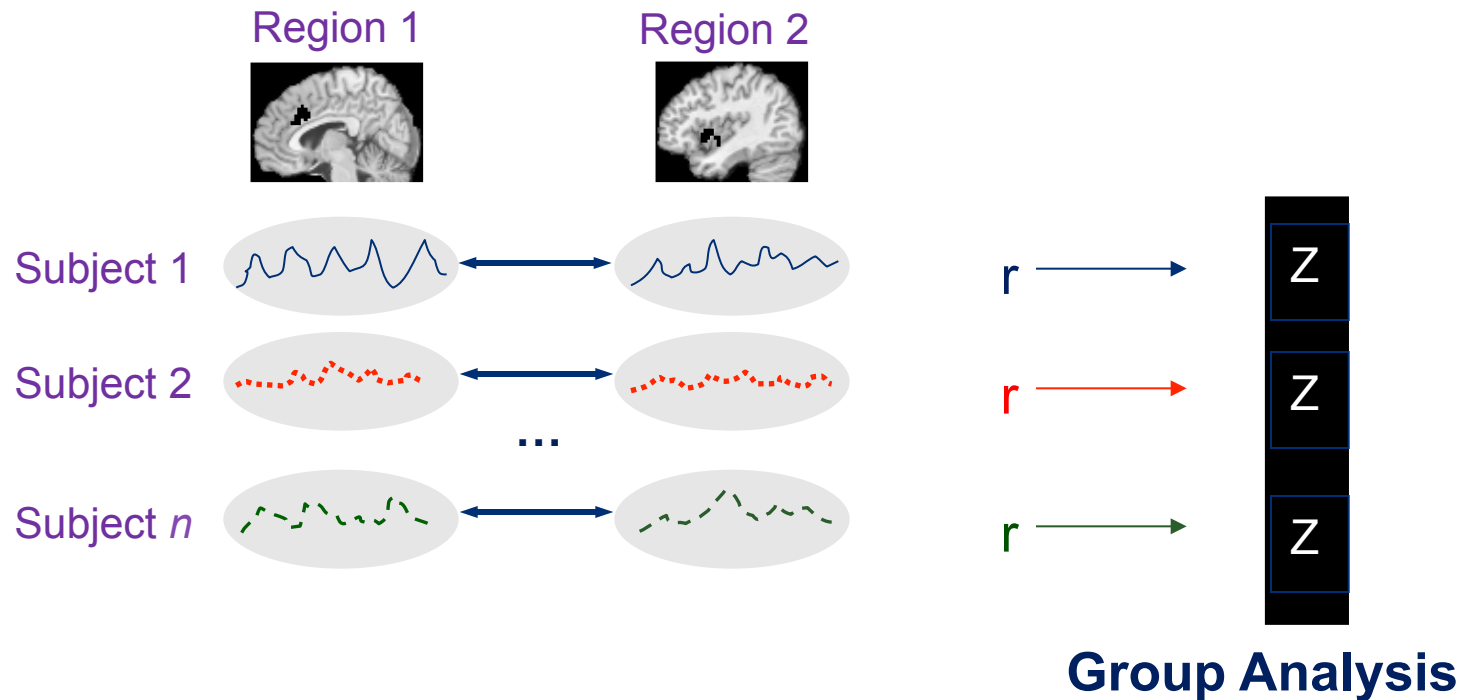
Bivariate Connectivity

➢ Simple functional connectivity

◇ Region A is correlated with Region B.

◇ Provides information about relationships among regions.

◇ Can be performed on time series data within a subject, or individual differences (contrast maps, one per subject).

Time Series Connectivity

➤ Calculate the cross-correlation between time series from two separate brain regions.

"uplifting the whole people"
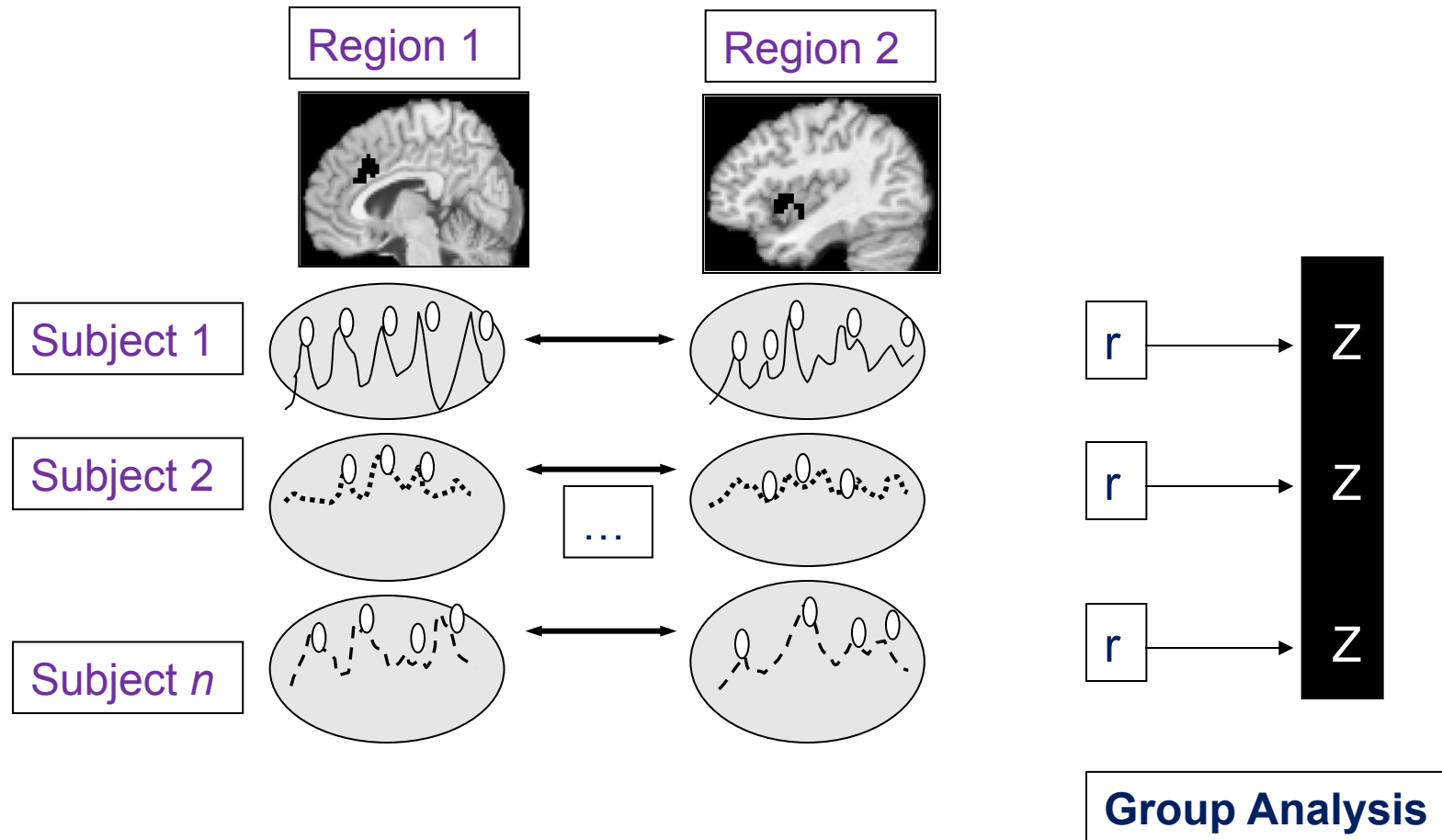— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Issues

➢ One of the main problems with time series connectivity is the fact that there may be different hemodynamic lags in different regions:

 ✧ Time series from different regions may not match up, even if neural activity patterns match up.

 ✧ If lags are estimated from data, temporal order may be caused by vascular (uninteresting) or neural (interesting) response.

Beta Series

➢ The beta series approach can be used to minimize issues of inter-region neurovascular coupling.

➢ Procedure:

    ✧ Fit a GLM to obtain separate parameter estimates for each individual trial.

    ✧ Compute the correlation between these estimates across voxels.

Beta Series

Mediation

```
┌─────┐      ┌─────┐      ┌─────┐
│  A  │ ───→ │  M  │ ───→ │  B  │
└─────┘      └─────┘      └─────┘
```
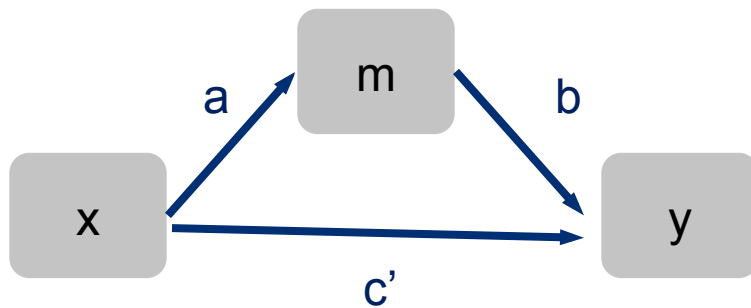
➢ **Mediation (Baron & Kenny, 1986)**
  ✧ The relationship between regions A and B is mediated by M
  ✧ Can identify functional pathways spanning > 2 regions
  ✧ Can be performed on time series data within a subject, or individual differences (contrast maps, one per subject)
  ✧ Also: Test of whether task-related activations in B are mediated, or explained, by M.

```
┌──────┐      ┌─────┐      ┌─────┐
│ Task │ ───→ │  M  │ ───→ │  B  │
└──────┘      └─────┘      └─────┘
```

Demonstrating Mediation

Full model, with mediator

$$m = i_m + ax + e_m$$
$$y = i_y + bm + c'x + e'_y$$

Reduced model, without mediator

$$y = i_{y'} + cx + e_y$$

Baron and Kenny (1986) – conjunction of 3 effects:

1) *c* effect: There is a relationship to be mediated
2) *a* effect: initial variable related to mediator
3) *b* effect: mediator relates to outcome

UNIVERSITY OF ALBERTA

"*uplifting the whole people*"
— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Decomposition of Effects

➢ The mediation framework allows us to decompose the total effect of x on y as follows:

Total effect = Direct effect + Mediated effect

➢ Does *m* explain some of the x-y relationship?

   ✧ Test c – c', which is equivalent to significance of a*b product.

Test of Mediation

➤ Sobel test:

$$Z = \frac{ab}{(b^2 se(a)^2 + a^2 se(b)^2)^{1/2}}$$

✧ Z ~ N(0, 1), standard normal distribution

✧ Assumes a, b are normally distributed

✧ Usually conservative (p-values higher than needed)
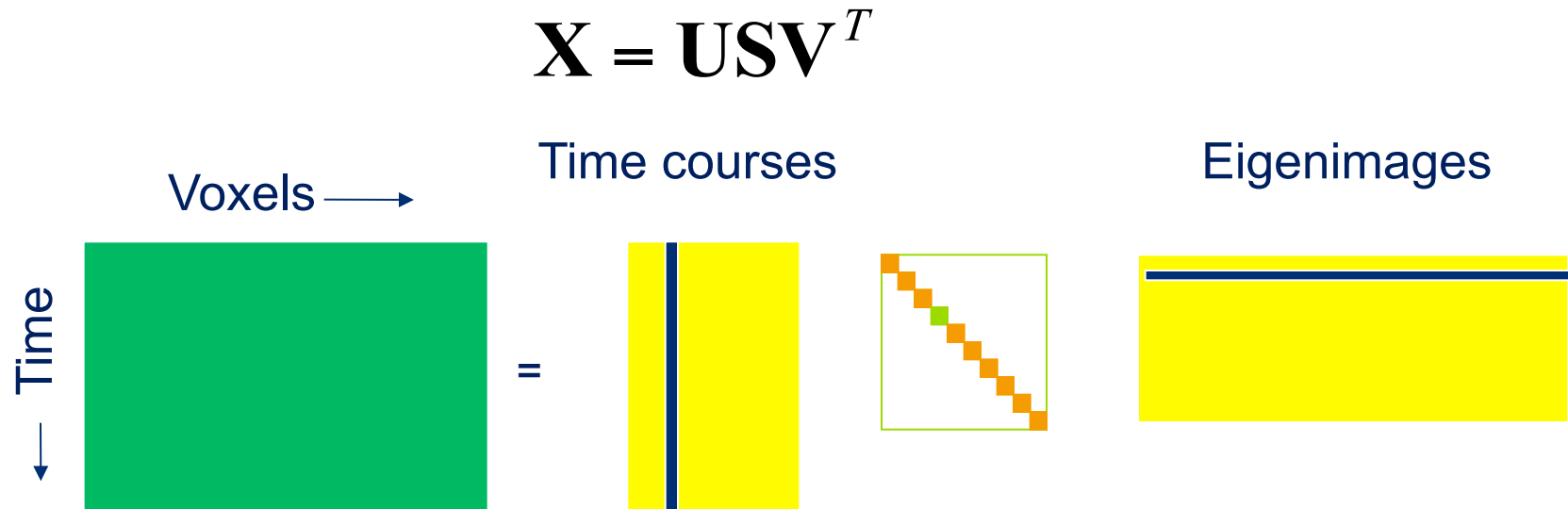
• Bootstrap test

Multivariate Methods

➤ We often use multivariate methods to study functional connectivity.

➤ When using multivariate methods observations at each voxel are considered jointly.

➤ This has the potential to allow for better understanding of how different brain regions interact with one another.

Principal Components Analysis

➢ Principal components analysis involves finding spatial modes, or eigenimages, in the data.

➢ Spatial modes are the patterns that account for most of the variance-covariance structure in the data.

➢ The eigenimages are obtained using singular value decomposition (SVD), which decomposes the data into two sets of orthogonal vectors that correspond to patterns in space and time.

Principal Components Analysis

$$\mathbf{X} = \mathbf{USV}^T$$

Voxels ⟶

Time courses

Eigenimages

Time

=

$$\mathbf{X} = s_1\mathbf{u}_1\mathbf{v}_1^T + s_2\mathbf{u}_2\mathbf{v}_2^T + \ldots + s_N\mathbf{u}_N\mathbf{v}_N^T$$

Each column of U correspond to the time-dependent profiles associated with each eigenimage.

Each column of V defines a distributed brain region that can be displayed as an image (eigenimages).
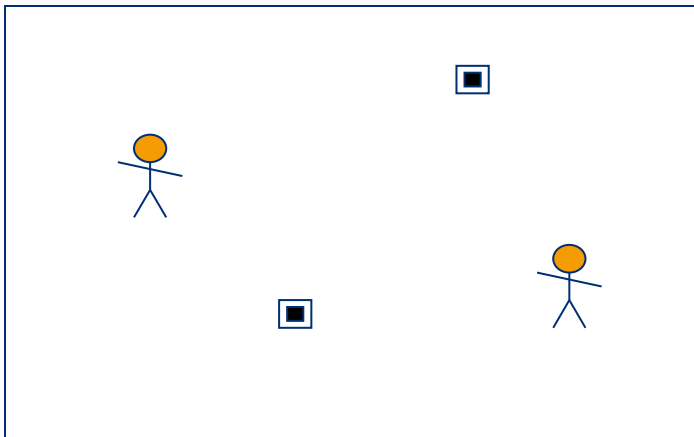
Independent Components Analysis

➤ Independent Components Analysis (ICA) is a family of techniques used to extract independent signals from some source signal.

➤ ICA provides a method to blindly separate the data into spatially independent components.

➤ The key assumption is that the data set consists of p spatially independent components, which are linearly mixed and spatially fixed.

Cocktail Party Problem

Two people are talking simultaneously in a room with two microphones

Speakers: $s_1(t)$ and $s_2(t)$.

Microphones: $x_1(t)$ and $x_2(t)$

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

$$\mathbf{X = AS}$$

Problem Formulation

➢ We want to solve:

$$X = AS$$

where **A** is the mixing matrix, **S** is the source matrix and **X** is the data matrix. Both **A** and **S** are unknown.

➢ Assume that Cov(**X**)=**I** and **A** is orthogonal.

➢ Find **A** such that **S**=**A**^{T}**X**.

Assumptions

➢ ICA is able the solve this problem by exploiting some key assumptions.

➢ Assumptions:

   ✧ Linear mixing of sources.

   ✧ The components s_i are statistically independent.

   ✧ The components s_i are non-Gaussian.

Mutual Information

➢ One approach is to minimize the mutual information between different components.

➢ This is equivalent to minimizing the Kullback-Leibler divergence between the joint density and the product of the marginal densities.

➢ The Kullback-Leibler divergence is a measure of similarity between two density functions.

ICA for fMRI

➢ It is assumed that the fMRI data can be modeled by identifying sets of voxels whose activity both vary together over time and are maximally different from the activity in other sets.

➢ Decompose the data set into a set of spatially independent component maps with a set of corresponding time-courses.

Overview

Voxels →

Time ↓

Time Courses

Spatially independent Components

Data

=

Mixing Matrix

Components

$$X = AS$$

where the matrix S contains statistically independent maps in its rows each with an internally consistent time-course contained in the associated column of the mixing matrix A.

Use an ICA algorithm to find **A** and **S**.

Comments

➢ Unlike PCA which assumes an orthonormality constraint, ICA assumes statistical independence among a collection of spatial patterns.

➢ Independence is a stronger requirement than orthonormality.

➢ However, in ICA the spatially independent components are not ranked in order of importance as they are when performing PCA.

Effective Connectivity

Effective Connectivity

➢ Effective connectivity is the influence one neuronal system exerts over another.

➢ Effective connectivity depends on two models:

   ✧ A neuroanatomical model that describes which areas are connected.

   ✧ A mathematical model that describes how areas are connected.
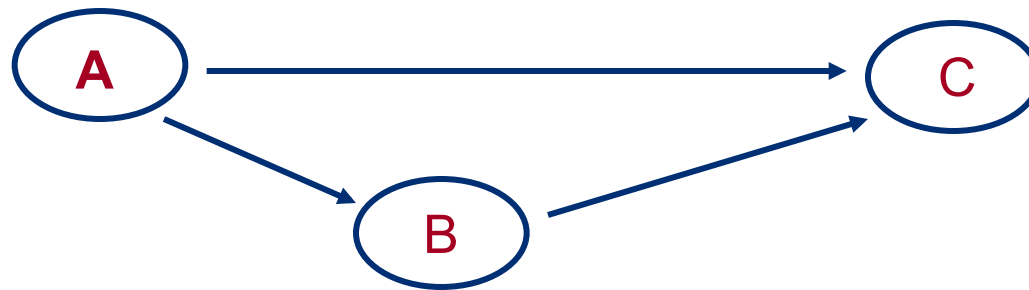
Structural Equation Modelling

➢ Structural equation Modeling (SEM) was first applied to imaging data by McIntosh and Gonzalez-Lima (1991).

➢ SEM allows for the analysis of more complicated models consisting of many different ROIs.

➢ Instead of considering variables individually the emphasis in SEM lies on the variance-covariance structure of the data.

SEM

Structural Equation Models comprise a set of regions and a set of directed connections

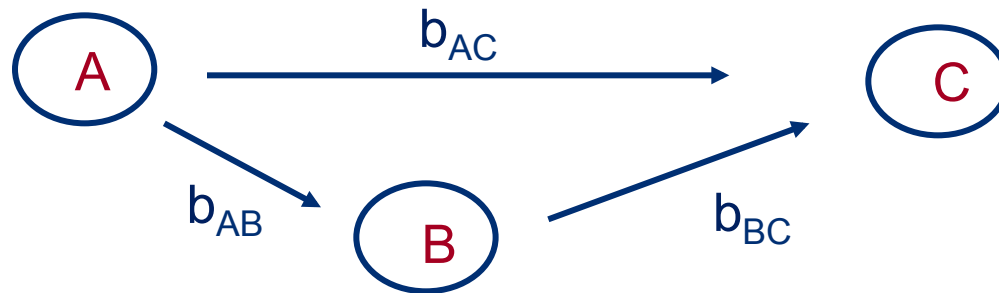A **causal relationship** is attributed to the connections. An arrow from A to B implies A causes B.

Note: **Causal relationships are assumed** *apriori*.

SEM

Further define path coefficients between the various nodes.

- **The coefficients imply a set of correlations among the regions.**
- **A path coefficient is the expected change in activity of one region given a unit change in the region influencing it.**
- **The path coefficient indicates the average influence across the time interval measured.**

SEM

➢ The covariance of the data represents how the activities in two or more regions are related.

➢ In SEM we seek to minimize the difference between the observed covariance matrix and the one implied by the structure of the model.

➢ The parameters of the model are adjusted to minimize the difference between the observed and modeled covariance matrix.

Set-Up

➢ Consider a network consisting of N different regions, where the activity at time t is given by the vector y_t which is of length N.

➢ Further suppose the data consists of T separate time points.

➢ We can write the full data as $$\mathbf{Y} = (y_1, \ldots y_T)$$

➢ Next, assume that the network activity is independent from sample to sample, i.e. y_i is independent of y_j for all $i \neq j$.

➢ This is not particularly realistic, but heuristic corrections exist.

Set-Up

Under this assumption we can write the likelihood of the data as

$$p(Y \mid \theta) = \prod_{t=1}^{T} p(y_t \mid \theta)$$

where θ are the parameters of the SEM.

Further suppose that

$$y_t \sim N(0, \Sigma(\theta))$$

The covariance matrix is a function of the connectivity parameters contained in θ.

Model

The form of $\Sigma(\theta)$ is specified by how the activity in various regions are related to one another, i.e.

$$y_t = M y_t + e_t$$

where *M* now describes the set of path coefficients.

The M_{ij} term of the matrix represents a connection between regions *i* and *j*.

The noise term e_t is normally distributed with mean 0 and covariance matrix $R = diag\{\sigma_1^2, ... \sigma_N^2\}$.

Solution

We can rewrite y_t as

$$y_t = (I - M)^{-1} e_t$$

Hence, we can write the covariance matrix of y_t as

$$\Sigma(\theta) = (I - M)^{-1} R \left((I - M)^{-1} \right)^T$$

The parameters θ are the unknown elements of the matrices M and R.

Find the parameters θ that maximize the likelihood function:

$$y_t \sim N(0, \Sigma(\theta))$$

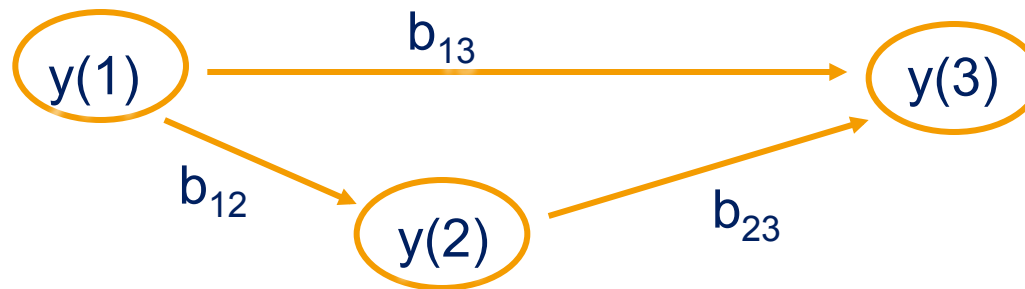This gives us the desired path coefficients.

Inference

➢ All inference regarding the path coefficients rests on the use of nested or stacked models.

➢ A more complex model is compared to a simpler model nested within the first model.

➢ If the complex model fits a particular dataset significantly better, then the additional parameters of this model are needed in the subsequent analyses.

➢ Given a constrained model, which is defined by the omission of a pathway, hypothesis testing may be construed as evidence for or against the pathway by nesting it in a free model where the pathway is included.

➢ If the difference in goodness of fit is highly unlikely to have occurred by chance, the connection can be declared active.
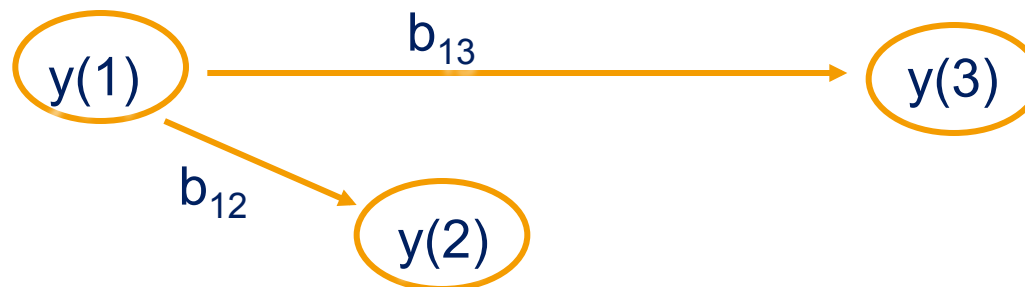
Example

Fixed

$y(1)$ —— b_{13} ——> $y(3)$

$y(1)$ —— b_{12} ——> $y(2)$ —— b_{23} ——> $y(3)$

Constrained

$y(1)$ —— b_{13} ——> $y(3)$

$y(1)$ —— b_{12} ——> $y(2)$

$$H_0 : b_{23} = 0$$

UNIVERSITY OF ALBERTA

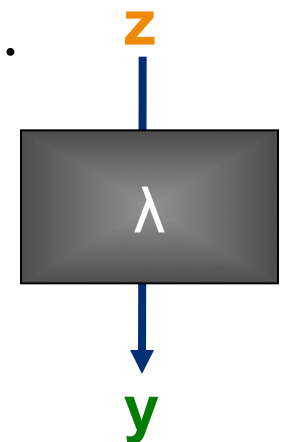*"uplifting the whole people"*
— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Likelihood Ratio Test

➢ The likelihood ratio test (LRT) is a statistical test of the goodness-of-fit between two models.

➢ A relatively more complex model is compared to a simpler model to determine whether it fits the dataset significantly better.

➢ If so, the additional parameters in the more complex model need to be included.
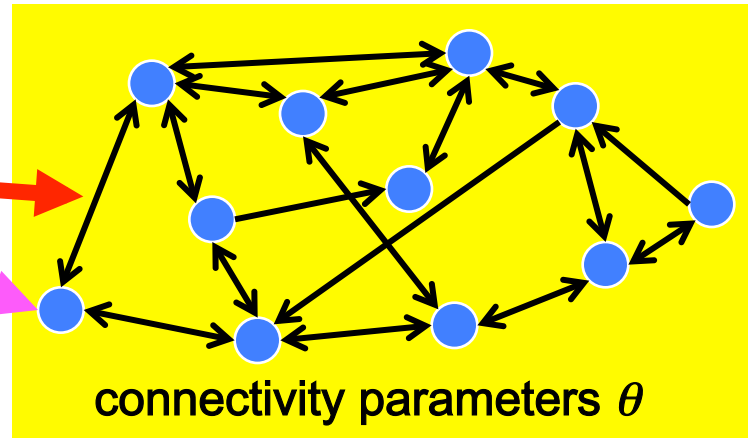
Neuronal Interactions

➢ It is important to note that the measurements used in each of the connectivity studies we have described so far are hemodynamic in nature and this limits an interpretation of the results at the level of neuronal interactions.

➢ Dynamic Casual Modeling (Friston *et. al.*) is an attempt to move the connectivity analysis to the neuronal level.

➢ The modelled neuronal dynamics (z) is transformed into area-specific BOLD signals (y) by a hemodynamic forward model (λ).

➢ The aim of DCM is to estimate <u>parameters at the neuronal level</u> (computed separately for each area) such that the modelled BOLD signals are maximally similar to the experimentally measured BOLD signals.

z

λ

y

UNIVERSITY OF
ALBERTA

"uplifting the whole people"
— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

Dynamic Casual Modeling

➢ Dynamic Causal Modeling (DCM) is an attempt to model neuronal interactions using hemodynamic time series.

➢ DCM treats the brain as a deterministic nonlinear dynamic system that is subject to inputs and produces outputs.

➢ It makes inference about the coupling among brain areas and how the coupling is influenced by changes in experimental context.

Input *u(t)*

System = **a set of elements which interact in a spatially and temporally specific fashion**

connectivity parameters θ

System state *z(t)* $\dot{z} = F(z, u, \theta)$

Dynamic Casual Modeling

➢ DCM is based on a neuronal model of interacting cortical regions, supplemented with a forward model of how neuronal activity is transformed into a measured response.

➢ Effective connectivity is parameterized in terms of the coupling among unobserved neuronal activity in different regions.

➢ We can estimate these parameters by perturbing the system and measuring the response.

State-space Model

- ➢ DCM is a state-space model.
- ➢ The initial formulation of DCM did not consider noise and was therefore a deterministic state-space model stated in terms of ordinary differential equations.
- ➢ More recent versions of DCM have been in terms of stochastic differential equations.

State changes of a system are dependent on:
- **the current state**
- **external inputs**
- **its connectivity**
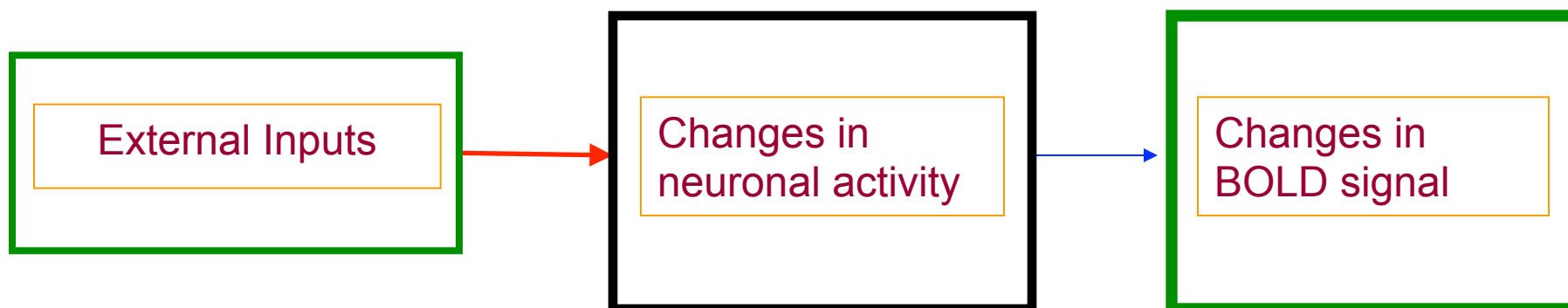- **time constants & delays**

System state *z(t)*

$$\dot{z} = F(z, u, \theta)$$

Dynamic Casual Modeling

In DCM a distinction is made between the neuronal level and the hemodynamic level.

Experimental inputs cause changes in effective connectivity expressed at the neuronal level which in turn cause changes in the observed hemodynamics.

Dynamic Casual Modeling

> DCM uses a bilinear model for the neuronal level and an extended Balloon model for the hemodynamic level.

> In a DCM model we have J experimental inputs and N outputs (one for each region).

> Each region has five state variables, four corresponding to the hemodynamic model and a fifth corresponding to neuronal activity.

Neurodynamics

Define the neuronal states as:

$$z = \left(z_1, \ldots z_N \right)^T$$

The effective connectivity model is described by:

$$\dot{z} = F(z, u, \theta)$$

where F is a non-linear function describing the influences that z and u exert upon changes in the neuronal states.

Bilinear form

The model consists of a bivariate nonlinear function.

We can approximate such a function using a bilinear approximation:

$$f(x,u) \approx ax + bxu + cu$$

where

$$a = \frac{\partial f}{\partial x} \qquad b = \frac{\partial^2 f}{\partial x \partial u} \qquad c = \frac{\partial f}{\partial u}$$

Neurodynamics

The effective connectivity model

$$\dot{z} = F(z, u, \theta)$$

can be rewritten in bilinear form as:

$$\dot{z}_t = \left(A + \sum_{j=1}^{J} u_t(j) B^j \right) z_t + C u_t$$

where z_t is the neuronal activity at time t (latent) and $u_t(j)$ is the jth of J inputs at time t (known).

Neurodynamics

➢ The matrix A represents the first order connectivity among regions in the absence of input. It specifies which regions are connected and whether these connections are uni- or bidirectional.

➢ The matrix C represents the extrinsic influence of inputs on neuronal activity. It specifies which inputs are connected to which regions.

➢ The matrices B_j represent the change in coupling induced by the jth input. It specifies which intrinsic connections are changed by which inputs.

Interpretation

➢ The units of connection are per unit time and therefore correspond to rates.

➢ A strong connection means an influence that is expressed quickly or with a small time constant.

Hemodynamics

➢ The neuronal activities in each region cause changes in blood volume and deoxyhmoglobin that, in turn, cause changes in the observed BOLD response.

➢ The hemodynamics are described using an extended Balloon model, which involves a set of hemodynamic state variables, state equations and hemodynamic parameters θ^h.

Comments

➢ DCM models interactions at the neuronal rather than the hemodynamic level.

➢ It is therefore more biologically accurate than the other models described today.

➢ However, it is quite computationally demanding. It is limited to 8 regions in SPM.

Granger Causality

➢ Granger causality is a technique that was originally developed in economics that has recently been applied to connectivity studies.

➢ It does not rely on the a priori specification of a structural model, but rather is an approach for quantifying the usefulness of past values from various brain regions in predicting current values in other regions.

➢ Granger causality provides information about the temporal precedence of relationships among two regions, but it is in some sense a misnomer because it does not actually provide information about causality.

Set Up

➢ Let x and y be two time courses of length *N* extracted from two brain regions.

➢ Each time course is modeled using a linear autoregressive model of the *M*th order

$$x[n] = \sum_{m=1}^{M} a[i]x[n-m] + \varepsilon_x[n]$$

$$y[n] = \sum_{m=1}^{M} b[i]y[n-m] + \varepsilon_y[n]$$

➢ Here ε_x and ε_y are both white noise, while a and b are model parameters.

Set Up

➢ Using these models one can test whether the history of x has any predictive value on the current value of y (and vice versa).

➢ If the model fit is significantly improved by the inclusion of the cross-autoregressive terms, it provides evidence that the history of one of the time courses can be used to predict the current value of the other and a "Granger-causal" relationship is inferred.

Measuring Influence

➢ Geweke has proposed a measure of linear dependence $F_{x,y}$ between x[n] and y[n] which implements Granger causality in terms of vector autoregressive models.

➢ The term $F_{x,y}$ can be decomposed into the sum of three components:

$$F_{x,y} = F_{x \rightarrow y} + F_{y \rightarrow x} + F_{x \cdot y}$$

➢ $F_{x,y}$ is a measure of the total linear dependence between x and y.

◈ If nothing about the current value of x (or y) can be explained by a model containing all values of y (or x) then $F_{x,y}$ will be 0.

Measuring Influence

➢ $F_{x \to y}$ and $F_{y \to x}$ are measures of linear directed influence from x to y and y to x, respectively.

  ✧ If past values of x improve the prediction of the current value of y, then $F_{x \to y} > 0$.

  ✧ A similar interpretation holds for $F_{y \to x}$.

➢ $F_{x \cdot y}$ is a measure of the undirected instantaneous influence between the series.

  ✧ The improvement in the prediction of the current value of x (or y) by including the current value of y (or x) in a linear model already containing the past values of x and y.

Computation

➢ If past values of x improve on the prediction of the current value of y, then $F_{x \to y}$ is large.

➢ A similar interpretation, but in the opposite direction, holds for $F_{y \to x}$.

➢ The difference between the two terms can be used to infer which regions history is more influential on the other. This difference is referred to as Granger Causality.