Welcome!
#pod-031
Week #3, Day 3

(Reviewed by: Deepak)

neuromatch academy

facebook Reality Labs

Agenda

- Tutorial 1 (Causality)
  - 3 exercises
- Tutorial 2 (Correlations)
  - 1 exercise + 1 bonus
- Tutorial 3 (Simultaneous fitting)
  - 1 Exercise
- Tutorial 4 (Instrumental variables)
  - 3 exercises + 2 bonus

Objective

- Definitions of causality

- estimating causality with 4 different methods and understand when they fail:
  1. Perturbations
  2. Correlations
  3. Simultaneous fitting/regression
  4. Instrumental variables

Tutorial #1 Explanations

Objective

1. Simulate a neural system
    a. Simulate system of neurons: estimate causal effects in neurons of bigger networks

2. Understand perturbation as a method of estimating causality

Defining causality

- *"A causes B".*
- *take two neurons. What does it mean to say that neuron A causes neuron B to fire?*
  The interventional definition of causality says that:

$$(A \text{ causes } B) \Leftrightarrow (\text{If we force } A \text{ to be different, then } B \text{ changes})$$

- *To determine if A causes B to fire, inject current into neuron A and see what happens to B.*

Mathematical definition of causality

Over many trials, avg causal effect $\delta_{A \to B}$
(avg change in B when A=1 vs A=0).

$$\delta_{A \to B} = E[B|A=1] - E[B|A=0]$$

Sophisticated conditional effects
— A only effects B when it's not effective

Relation to a randomized controlled trial (RCT)

If you randomly give 100 people a drug and 100 people a placebo, the effect is the difference in outcomes.

Randomised controlled trial for 2 neurons

→ Neuron A synapsing on neuron B.

$$B = A + \mathcal{E}$$

↳ std normal noise: $\mathcal{E} \sim \mathcal{N}(0, 1)$

represents the activities of the neurons

#TASK: Lecture 4 and confirm that B changes

Causal effects in bigger neurons

Our system has N interconnected neurons that affect each other over time. Each neuron at time $t+1$ is a function of the activity of the other neurons from the previous time t.

Neurons affect each other nonlinearly: each neuron's activity at time $t+1$ consists of a linearly weighted sum of all neural activities at time t, with added noise, passed through a nonlinearity:

In our system, neurons will receive connections from only 10% of the whole population on average.

n dimensional vector

n×n causal ground truth connectivity matrix

$$\vec{x}_{t+1} = \sigma\left(A\vec{x}_t + \varepsilon_t\right)$$

@ every timestep, activity vector $\vec{x}$ is updated as per-

Sigmoid nonlinearity

ε_t : randomness

$\varepsilon_t \sim N(\vec{0}, I_n)$.

$\vec{x}$: represents n neuron system @ timestep t

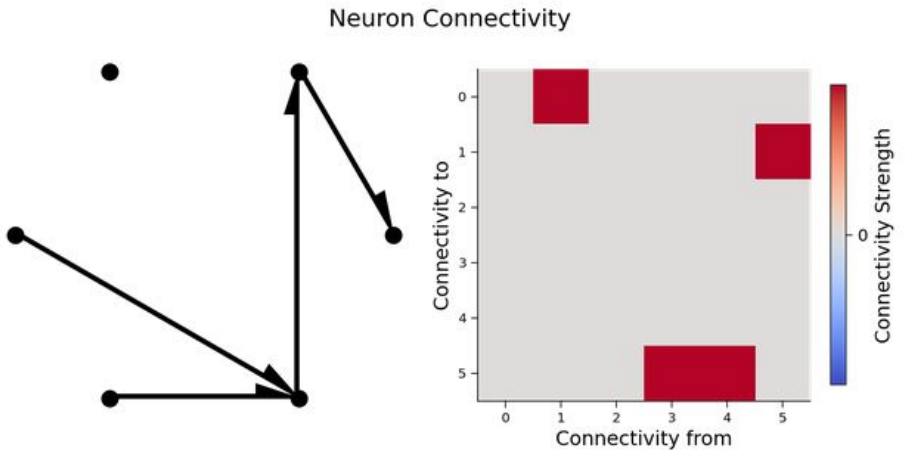
$\vec{x}_0 = \vec{0}$.

$$A = \begin{bmatrix} \cdots \\ \vdots \\ \cdots \end{bmatrix}_{n \times n}$$

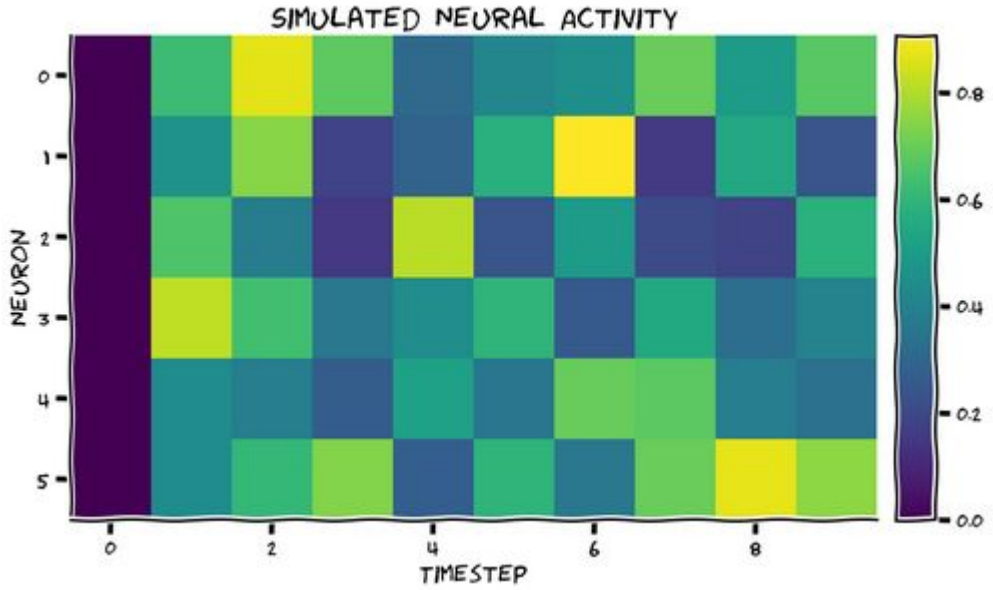
Causal effect of neuron i on j

Visualize true connectivity

Create a connectivity matrix between 6 neurons and visualize it in two different ways: as a graph with directional edges between connected neurons and as an image of the connectivity matrix.



Simulated neural activity



Random perturbation in our system of neurons

To get the causal effect of each neuron upon each other neuron. The ground truth of the causal effects is the connectivity matrix A.

$$\delta_{A \to B} = E[B \mid A=1] - E[B \mid A=0]$$

randomly set system state to 0/1 and observe outcome after 1 time step } repeat N times

≅ intervening in activity every other timestep

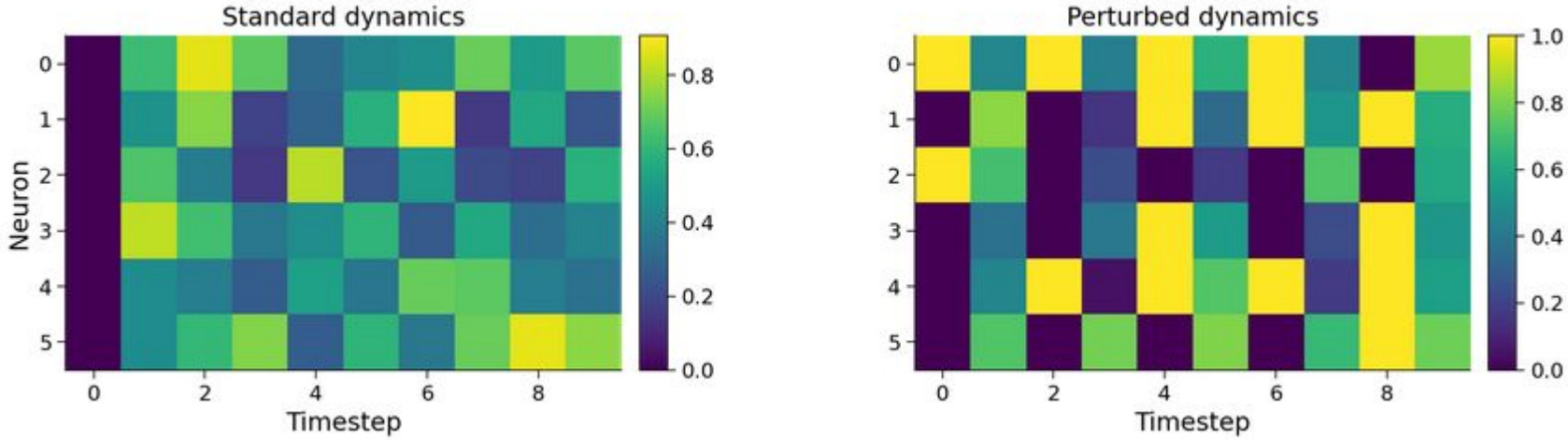
avg diffs of activity of neuron j in 2 conditions

$$\delta_{x^i \to x^j} \sim \frac{1}{N} \sum_i^N \left[x^j_{t+1} \mid x^i_t = 1 \right] - \frac{1}{N} \sum_i^N \left[x^j_{t+1} \mid x^i_t = 0 \right]$$

dynamics

This means that at every other timestep, every neuron's activity is changed to either 0 or 1.

Visually comparing the dynamics:

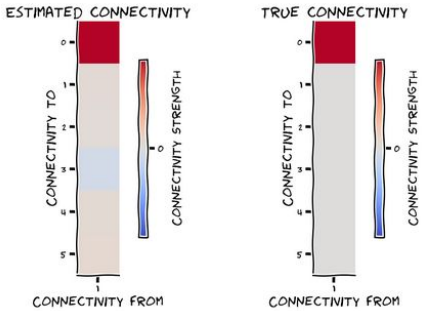


Using perturbed dynamics to recover connectivity

Despite perturbing every neuron at every other timestep, we compute the causal effect of a single neuron.

Exclusively use the timesteps without perturbation for x_j $t+1$ and the timesteps with perturbation for x_j t.

Quantify how close our estimated connectivity matrix is to our true connectivity matrix by correlating them (almost perfect correlation between our estimates and the true connectivity).



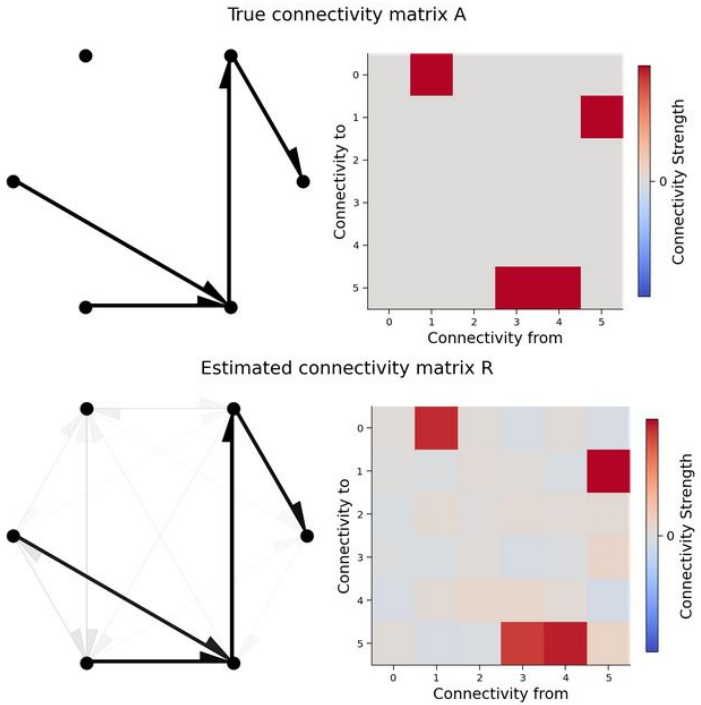
Interpretation of A

A is not the matrix of causal effects but rather the dynamics matrix.

A and the effect matrix both are 0 everywhere except where there is a directed connection. So they should have a correlation of

1 if we estimate the effects correctly. (Their scales, however, are different. This in part because the nonlinearity σ squashes the values of x to [0,1].)

Measuring how perturbations recover the entire connectivity matrix



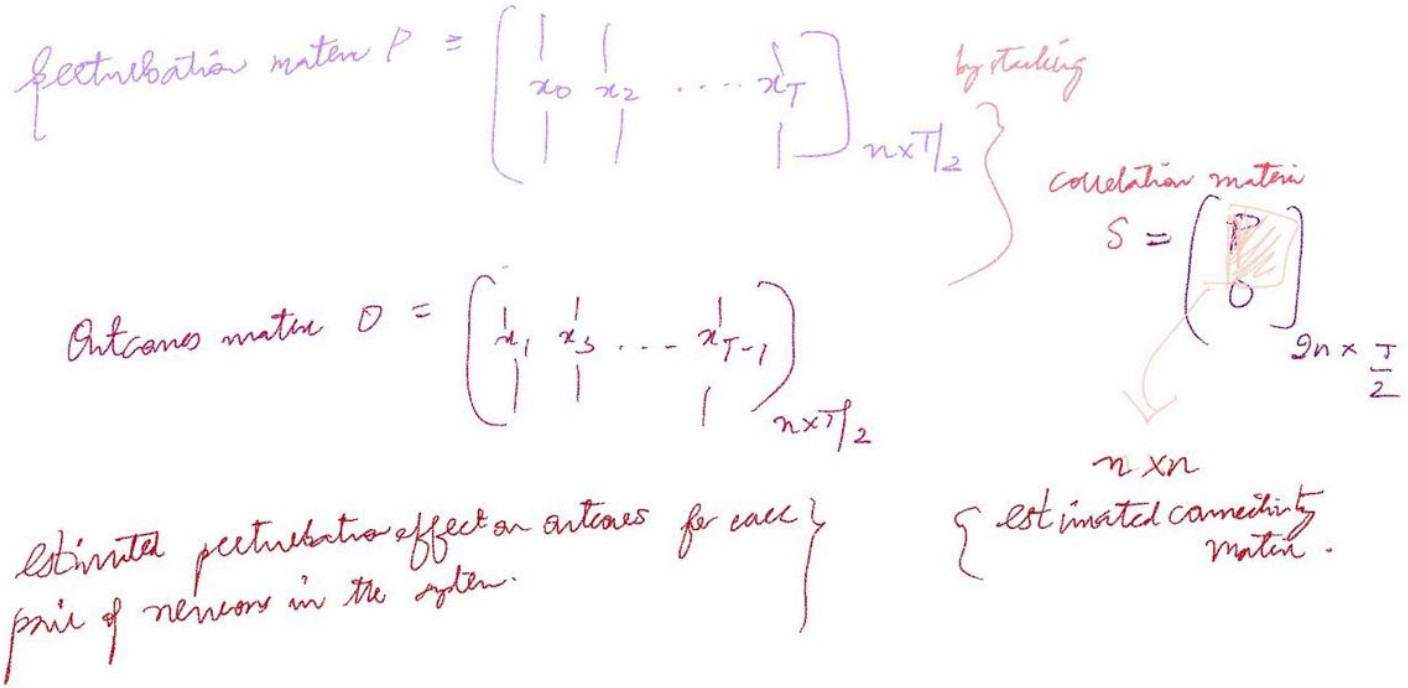
Resources

- *Causal Inference for Statistics, Social, and Biomedical Sciences* by Imbens and Rubin
- *Causal Inference: What If* by Hernan and Rubin
- *Mostly Harmless Econometrics* by Angrist and Pischke
- https://www.nature.com/articles/s41562-018-0466-5 for application to neuroscience

Appendix

Computation of the estimated connectivity matrix

method gives an estimated connectivity matrix that is the proportional to the result obtained with differences in means, and differs only in a proportionality constant that depends on the variance of x



Summary

we implemented and explored the dynamical system of neurons

We also learned about the "gold standard" of measuring causal effects through random perturbations.

random perturbations are often not possible

Tutorial #2
Explanations

Objectives

Alternative methods to attempt to measure causality. We will:

- *estimate connectivity from observations assuming* ***correlations approximate causation***
- *Works only when the network is small*

Tutorial 2 setting

Often, we can't force neural activities or brain areas to be on or off. We just have to observe.

when is correlation a "good enough" substitute for causation? Sometimes.

Try to approximate causation with correlation

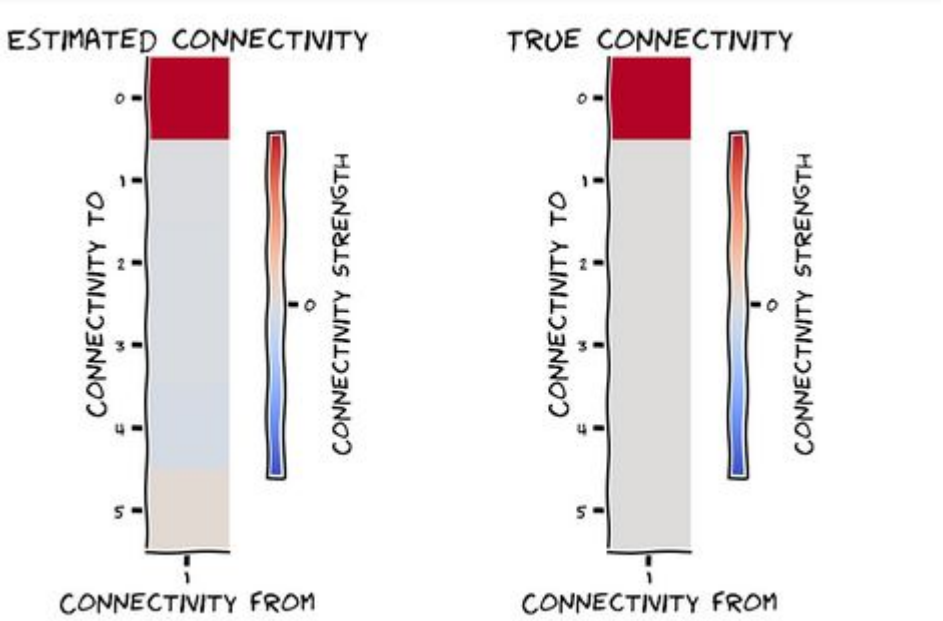
In small systems, correlation can look like causation.

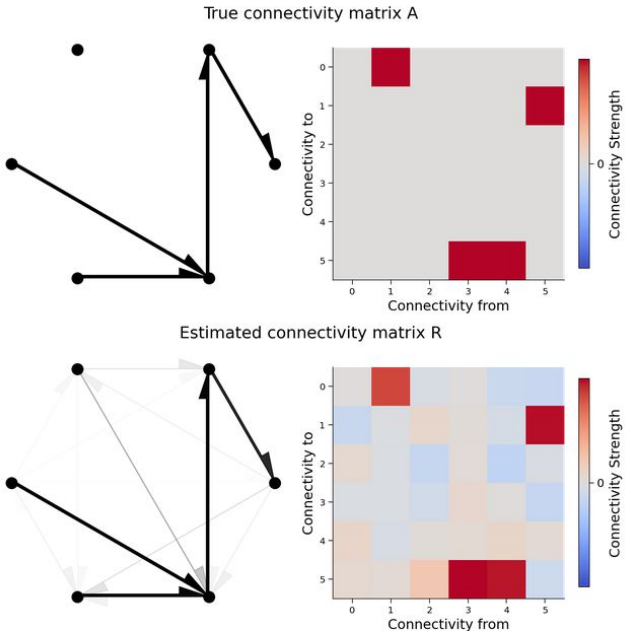
$$C = \vec{x}_t \vec{x}_{t+1}^T$$

Recover true connectivity matrix by correlating neural state at each timestep with previous state.

estimate the connectivity matrix of a single neuron by calculating the correlation coefficients with every other neuron at the next timestep. That is, correlating two vectors:

1) the activity of a selected neuron at time t
2) The activity of all other neurons at time t+1.


ESTIMATED CONNECTIVITY
CONNECTIVITY TO
CONNECTIVITY STRENGTH
CONNECTIVITY FROM
TRUE CONNECTIVITY
CONNECTIVITY TO
CONNECTIVITY STRENGTH
CONNECTIVITY FROM

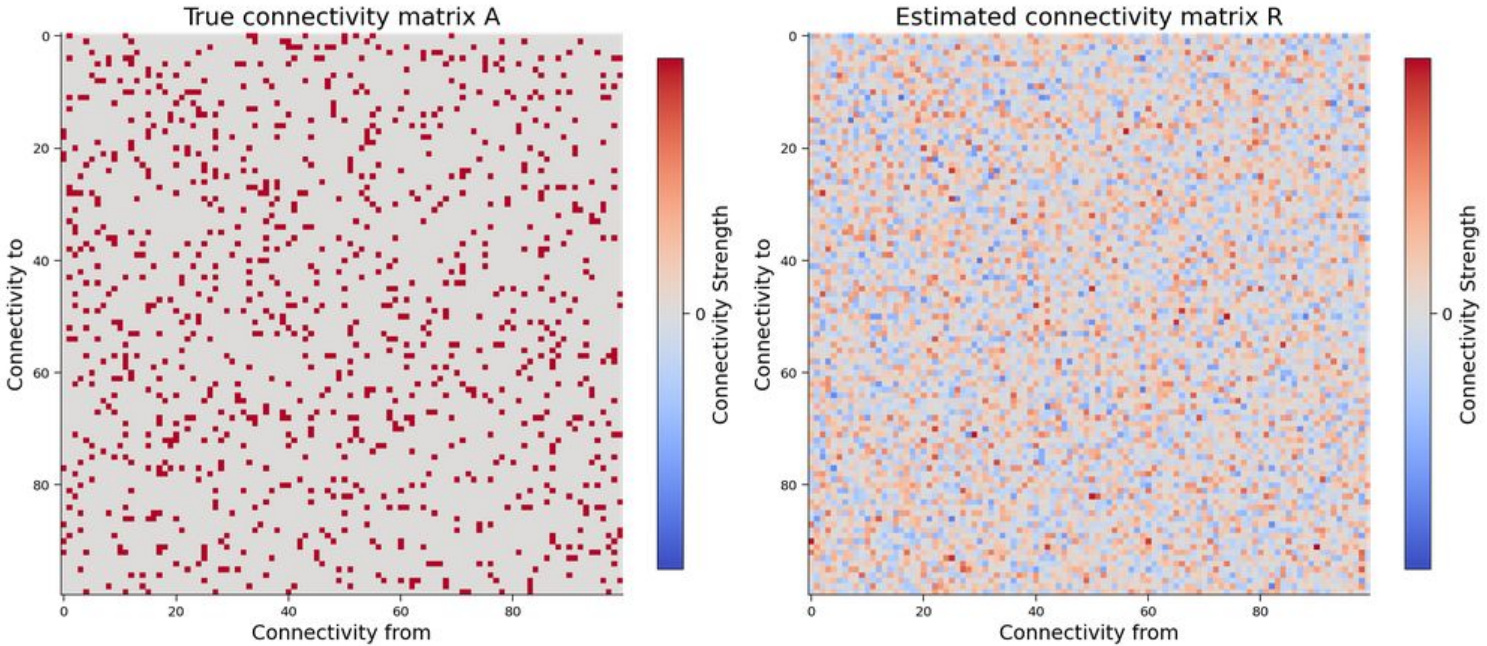

True connectivity matrix A
Connectivity to
Connectivity from
Connectivity Strength
Estimated connectivity matrix R
Connectivity to
Connectivity from
Connectivity Strength

Large systems

Failure of correlation in complex systems

As our system becomes more complex however, correlation fails to capture causality.

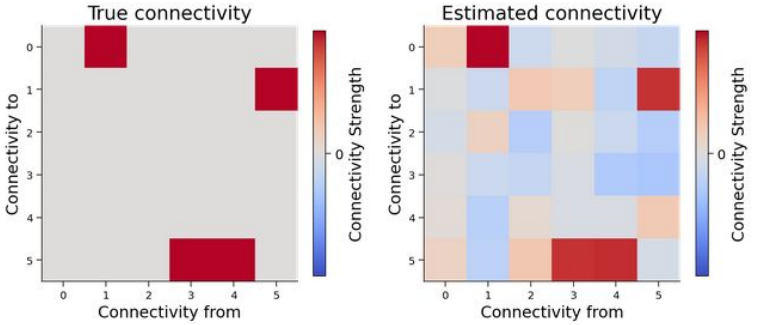
Correlation matrix of A and R: 0.3629329824177816



Connectivity estimation as a function of number of neurons

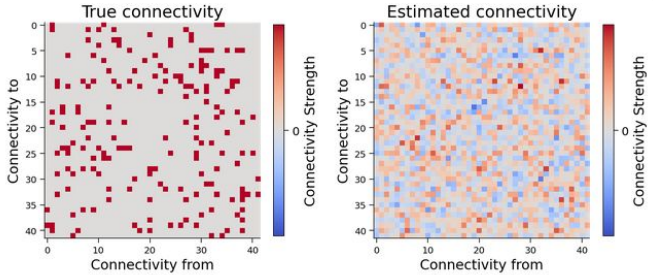
Systematically vary the number of neurons and plot the resulting changes in correlation coefficient between the true and estimated connectivity matrices.

n_neurons 6



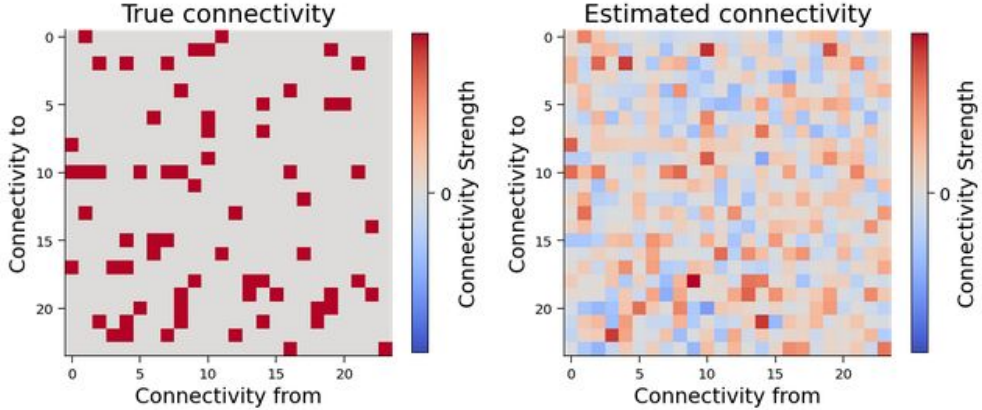
Correlation : 0.91

n_neurons 42



Correlation : 0.43

n_neurons 24

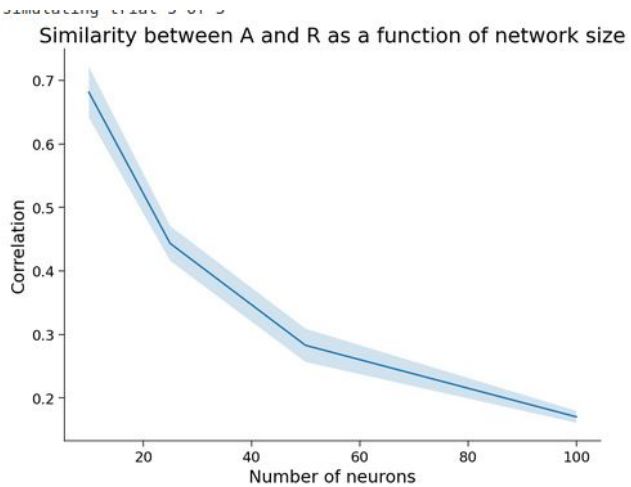


Correlation : 0.61

Variability due to randomness

Of course there is some variability due to randomness in A.

Average over a few trials and find the relationship.

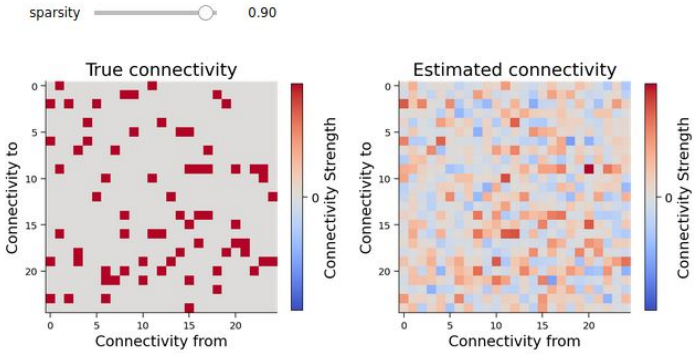


Connectivity estimation as a function of the sparsity of A

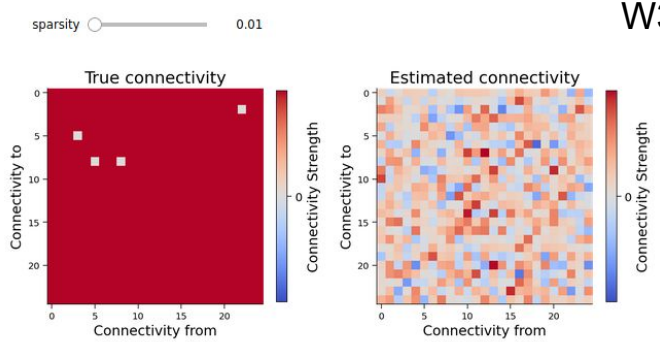
correlation only fails for large systems for certain types of A?

Examine connectivity estimation as a function of the sparsity of A.

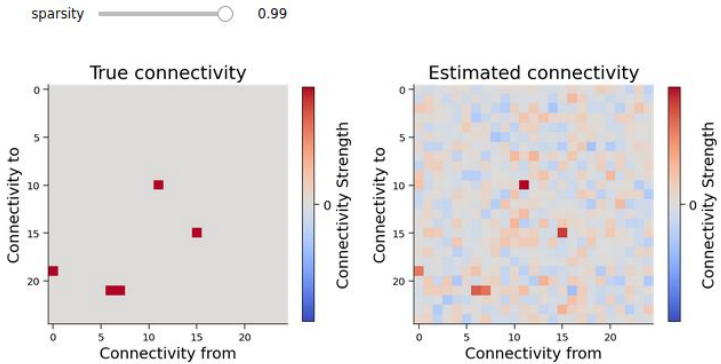
Does connectivity estimation get better or worse with less sparsity?



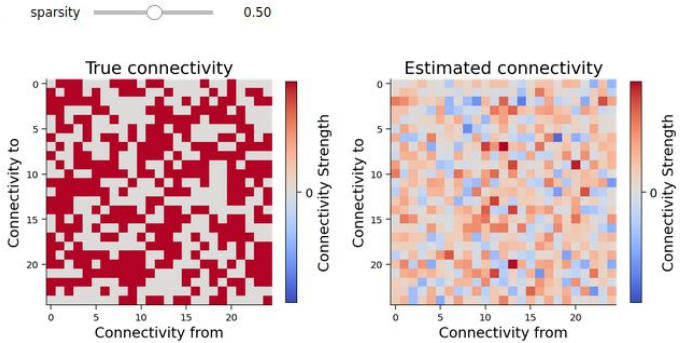
Correlation : 0.62



Correlation : 0.03



Correlation : 0.5



Correlation : 0.33

FOOD FOR THOUGHT

- Do they imply a causal relationship, in its interventional definition? (*regulates, mediates, generates, modulates, shapes, underlies, produces, encodes, induces, enables, ensures, supports, promotes, determines*)

A underlies B } carries / bears
not necessarily causal

A produces B } by definition, causal

A encodes B } conceals
not necessarily causal

A induces B } by definition causal

A enables B } entitle / empower / authorise
not necessarily causal.

A ensures B } Secure / warrant / guarantee
- not causal here

A supports B } bear / carry / reinforce
- not particularly causal

A promotes B } assist / aid / support / nurtures
- not causal

A determines B } control / regulate / affect
- not causal

A regulates B } - controls / balances / modulates / manages

Not necessarily causal

A mediates B } arbitrates

not necessarily causal

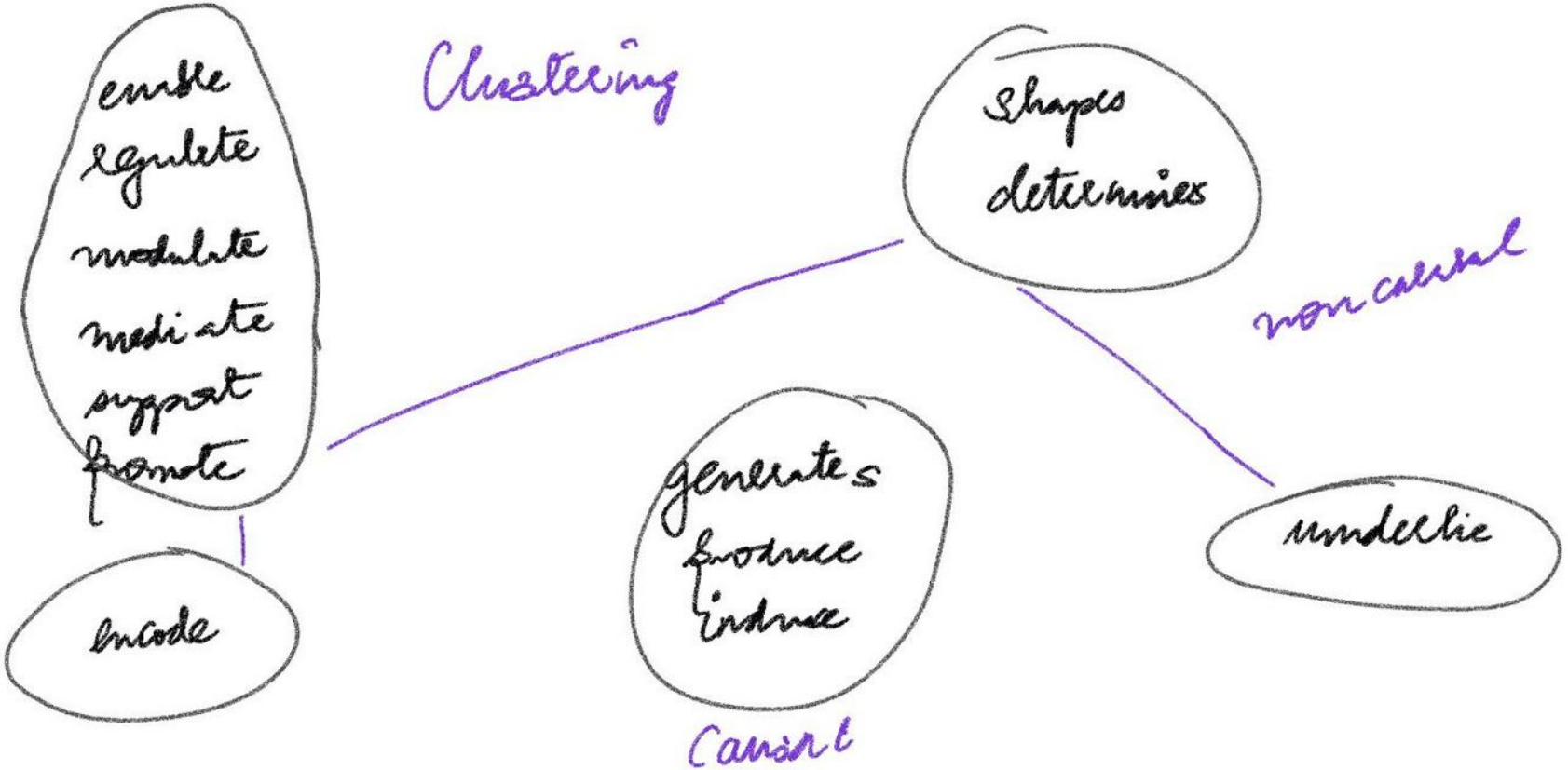
A generates B } by definition causal

A modulates B } regulates

not necessarily causal

A shapes B } forms / frames / models / sculpts / casts

Causal


Clustering
enable
regulate
modulate
mediate
support
promote
encode
generates
produce
induce
Causal
shapes
determines
non causal
underlie

Food for thought

- What dimensionality would you (very roughly) estimate the brain to be? Would you expect correlations between neurons to give you their connectivity? Why?

High dimensional

So, the correlation-connectivity theorem does not work.

Summary

for large systems correlation ≠ causation.

But what about when we coarsely sample the large system? Do we get better at estimating the effective causal interaction between groups (=average of weights) from the correlation between the groups?

The answer appears to be no: as the number of neurons per group increases, we don't see any significant increase in our ability to estimate the causal interaction between groups.

Appendix

Correlation as similarity metric

Pearson correlation coefficients: measure similarity between our estimated connectivity matrix R and the ground truth connectivity A

Note: This is not strictly correct usage of Pearson correlations as elements of A are not normally distributed (they are in fact binary).

We use Pearson correlations as they are quick and easy to compute within the Numpy framework and provide qualitatively similar results to other correlation metrics. Other ways to compute similarities:

- Spearman rank correlations, which does not require normally distributed data
- dichotomizing our estimated matrix R
- by the median and then running concordance analysis, such as computing Cohen's kappa
- some measure of the similarity between A and R => Element-wise comparisons are one way to do this.

Low resolution systems

A common situation in neuroscience is that you observe the *average* activity of large groups of neurons. (Think fMRI, EEG, LFP, etc.) We're going to simulate this effect, and ask if correlations work to recover the average causal effect of groups of neurons or areas.

In a big system in which correlations fail to estimate causality, can you at least recover average connectivity between groups?

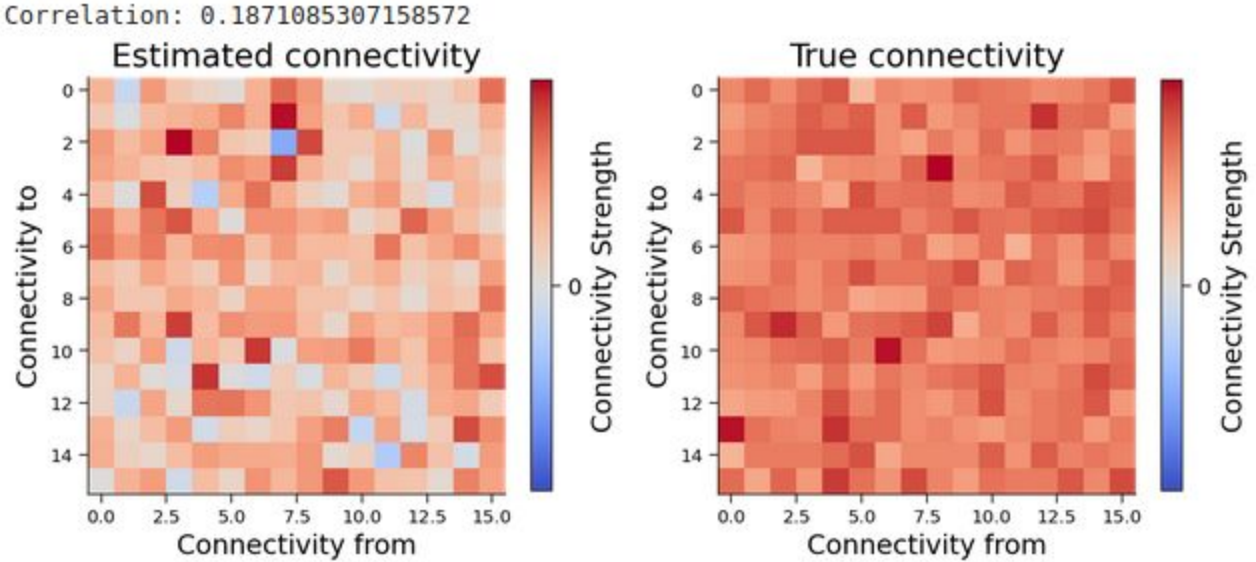
Assumption: Connectivity is random. In real brains, the neurons that are averaged have correlated input and output connectivities. This will improve the correspondence between correlations and causality for the average effect because the system has a lower true dimensionality. However, in real brains the system is also order of magnitudes larger than what we examine here, and the experimenter never has the fully-observed system.

Note: We observe the average activity of groups of the system and not the system itself.

Compute average activity across groups and compare resulting connectivity to the truth

New matrix coarse_X that has 16 groups, each reflecting the average activity of 16 neurons (since there are 256 neurons in total).

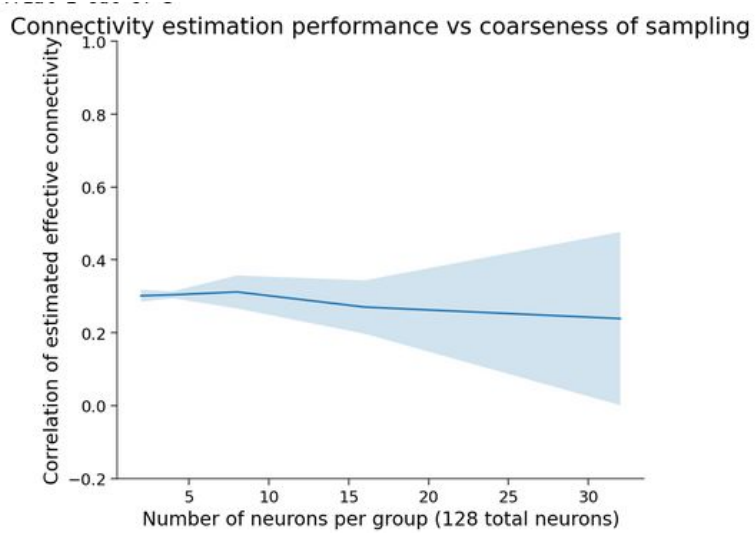
True coarse connectivity is defined as the average of the neuronal connection strengths between groups.


Correlation: 0.1871085307158572
Estimated connectivity
True connectivity
Connectivity to
Connectivity from
Connectivity Strength
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
2
4
6
8
10
12
14

Connectivity estimation performance vs sample coarseness

How close is the estimated coarse connectivity matrix to the truth?

Look at the estimation quality for different levels of coarseness when averaged over 3 trials.



Tutorial #3
Explanations

Objective

- Advanced (but also controversial) techniques for estimating causality from observational data:
  - conditional probabilities (**regression**)
- Explore limitations and failure modes
  - understand the problem of **omitted variable bias**

These methods rely on fitting a function to our data directly, instead of trying to use perturbations or correlations. Since we have the full closed-form equation of our system, we can try these methods and see how well they work in estimating causal connectivity when there are no perturbations.

Regression approach

Correlation only implies causation when there no hidden confounders. This aligns with intuition that correlation only implies causality when no alternative variables could explain away a correlation.

A confounding example: Suppose you observe that people who sleep more do better in school. It's a nice correlation. But what else could explain it? Maybe people who sleep more are richer, don't work a second job, and have time to actually do homework. If you want to ask if sleep causes better grades, and want to answer that with correlations, you have to control for all possible confounds.

A confound is any variable that affects both the outcome and your original covariate. In our example, confounds are things that affect both sleep and grades.

Controlling for a confound: Confonds can be controlled for by adding them as covariates in a regression. But for coefficients to be causal effects:

1. **All** confounds are included as covariates
2. Regression assumes the same mathematical form of how covariates relate to outcomes (linear, GLM, etc.)
3. No covariates are caused by both the treatment (original variable) and the outcome. These are colliders;
   In the real world it is very hard to guarantee these conditions are met. In the brain it's even harder (as we can't measure all neurons).

Recovering connectivity by model fitting

Our system is a closed system, too, so there are no omitted variables. The regression coefficients should be the causal effect.

$$\vec{x}_{t+1} = \sigma\left(A\vec{x}_t + \varepsilon_t\right)$$

sigmoid nonlinearity $\left\{ \sigma(x) = \frac{1}{1+e^{-x}} \right.$

linear regression to determine A

(regression approach to estimate causal influence of all neurons to neuron #1)

$$\sigma^{-1}\left(\vec{x}_{t+1}\right) = A\vec{x}_t + \varepsilon_t$$

inverse sigmoid transformation

$\left\{ \right.$ logit transformation $\sigma^{-1}(x) = \log\left(\frac{x}{1-x}\right)$

let W be the $\vec{x}_t$ values,
up to second to last timestep $T-1$

$$W = \begin{bmatrix} | & | & & | \\ \vec{x}_0 & \vec{x}_1 & \cdots & \vec{x}_{T-1} \\ | & | & & | \end{bmatrix}_{n \times (T-1)}$$

let Y be the $\vec{x}_{t+1}$ values for selected neuron,
indexed by i starting from 2nd timestep upto last.

$$Y = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,T} \end{bmatrix}_{1 \times (T-1)}$$

Lasso a.k.a. ***L1 regularization*** *causes the coefficients to be sparse, containing mostly zeros – to simulate brain's sparsity*

following model : $\sigma^{-1}(y^T) = W^T V$

W: $n \times 1$ coefficient matrix of regression

Estimated connectivity matrix between selected neuron vs rest of the neurons

Observations

Using regression, our estimated connectivity matrix has a correlation of 0.865 with the true connectivity matrix.

With correlation, our estimated connectivity matrix has a correlation of 0.703 with the true connectivity matrix.

Multiple regression is better than simple correlation for estimating connectivity.

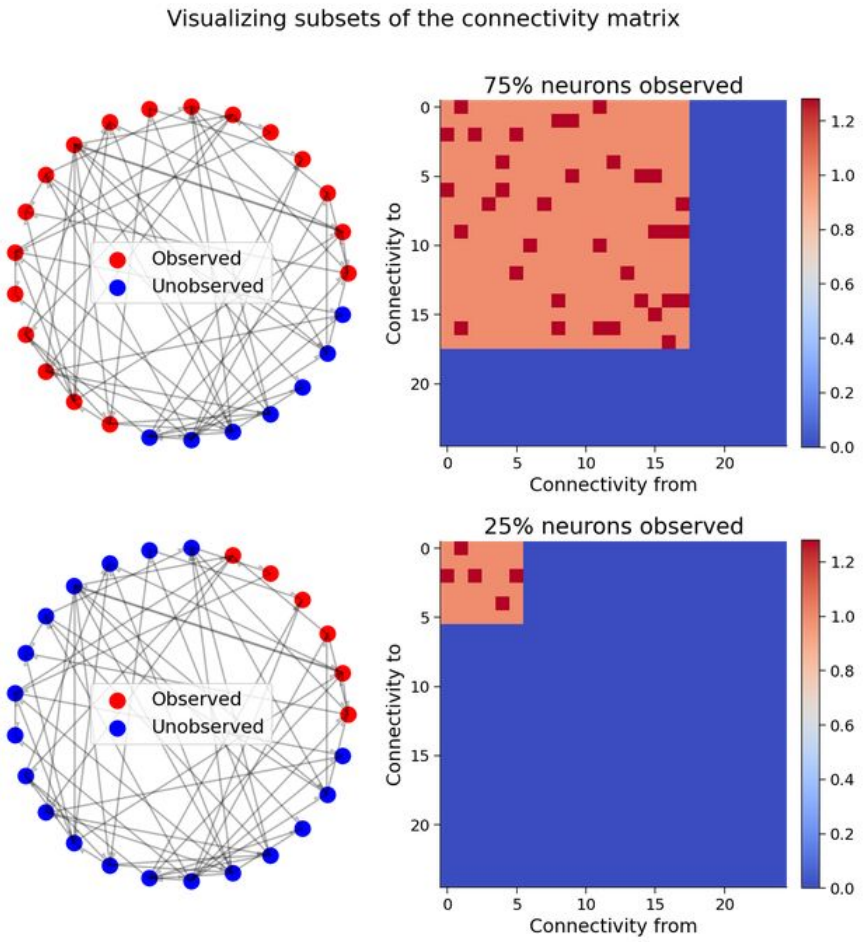
Omitted Variable Bias

If we are unable to observe the entire system, **omitted variable bias** becomes a problem. If we don't have access to all the neurons, and so therefore can't control for them, can we still estimate the causal effect accurately?

Visualizing subsets of the connectivity matrix

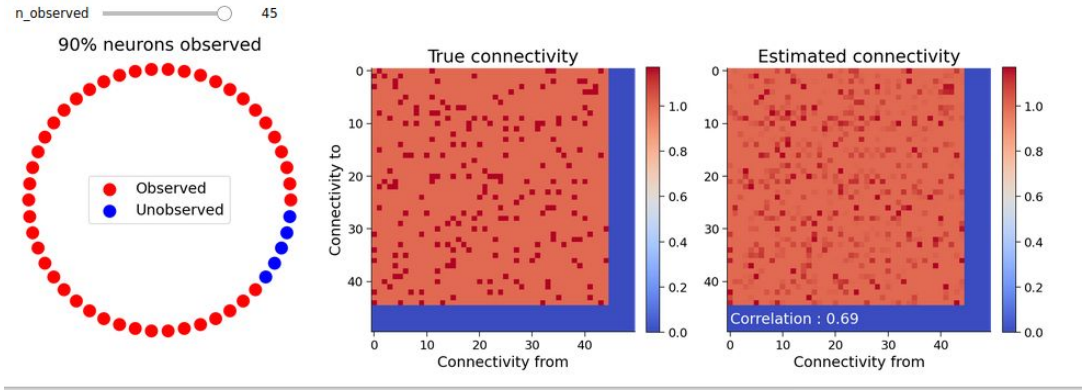
Visualize different subsets of the connectivity matrix when we observe 75% of the neurons vs 25%.

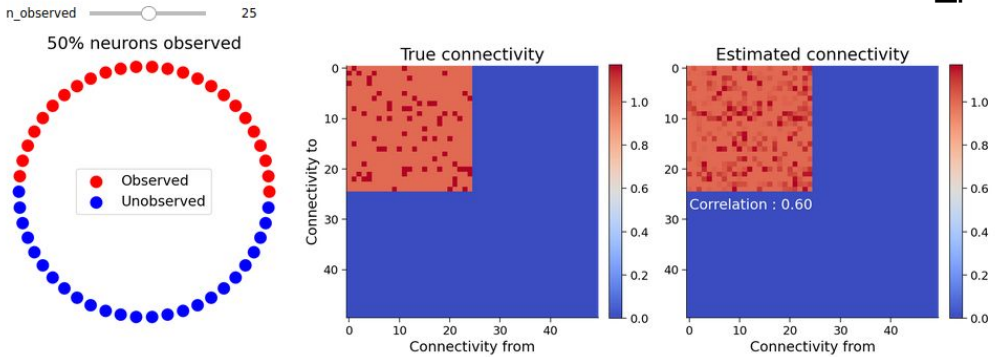
Meaning of entries in our connectivity matrix: A[i,j]=1 means a connectivity **from** neuron i **to** neuron j with strength 1.

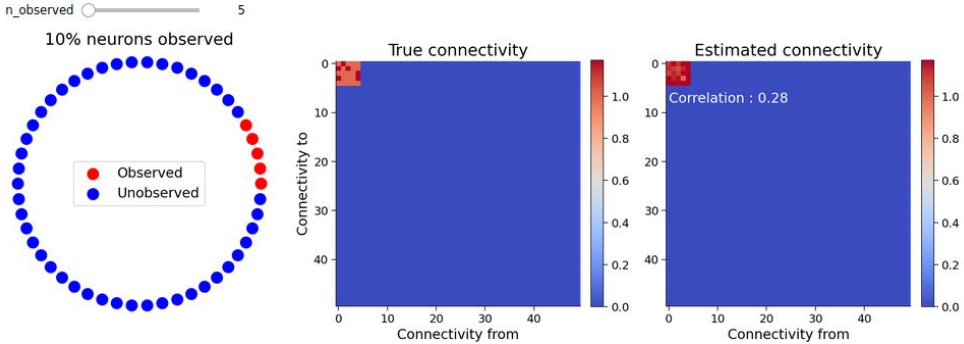

Visualizing subsets of the connectivity matrix
Observed
Unobserved
75% neurons observed
Connectivity to
Connectivity from
Observed
Unobserved
25% neurons observed
Connectivity to
Connectivity from

Regression performance as a function of the number of observed neurons

Change the number of observed neurons in the network and inspect the resulting estimates of connectivity

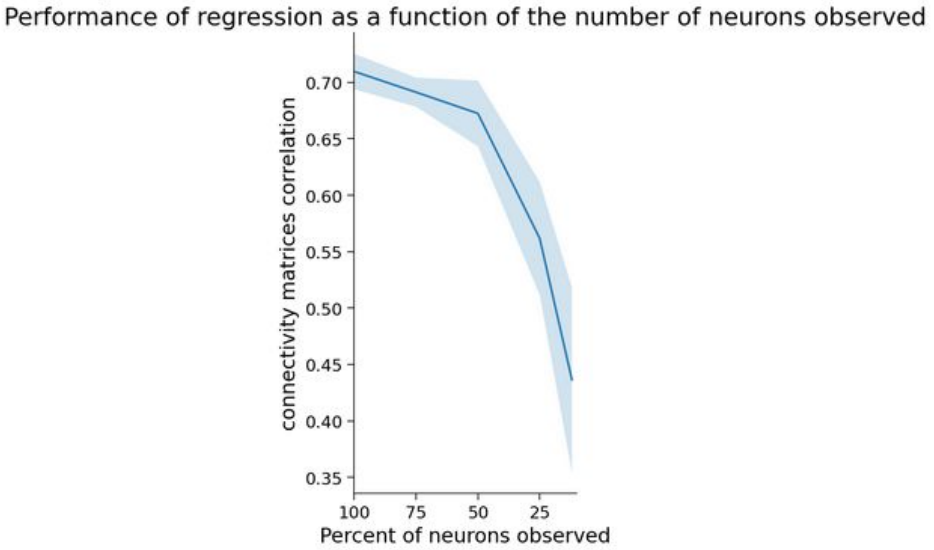



n_observed
25
50% neurons observed
Observed
Unobserved
True connectivity
Estimated connectivity
Correlation : 0.60
Connectivity to
Connectivity from


n_observed
5
10% neurons observed
Observed
Unobserved
True connectivity
Estimated connectivity
Correlation : 0.28
Connectivity to
Connectivity from

Performance vs number of observed neurons

Inspect a plot of the correlation between true and estimated connectivity matrices vs the percent of neurons observed over multiple trials. What is the relationship that you see between performance and the number of neurons observed?



Tutorial #4
Explanations

Agenda

Even sophisticated techniques such as simultaneous fitting fail to capture causality in the presence of omitted variable bias.

Techniques to obtain valid causal measurements when we can't perturb the system:

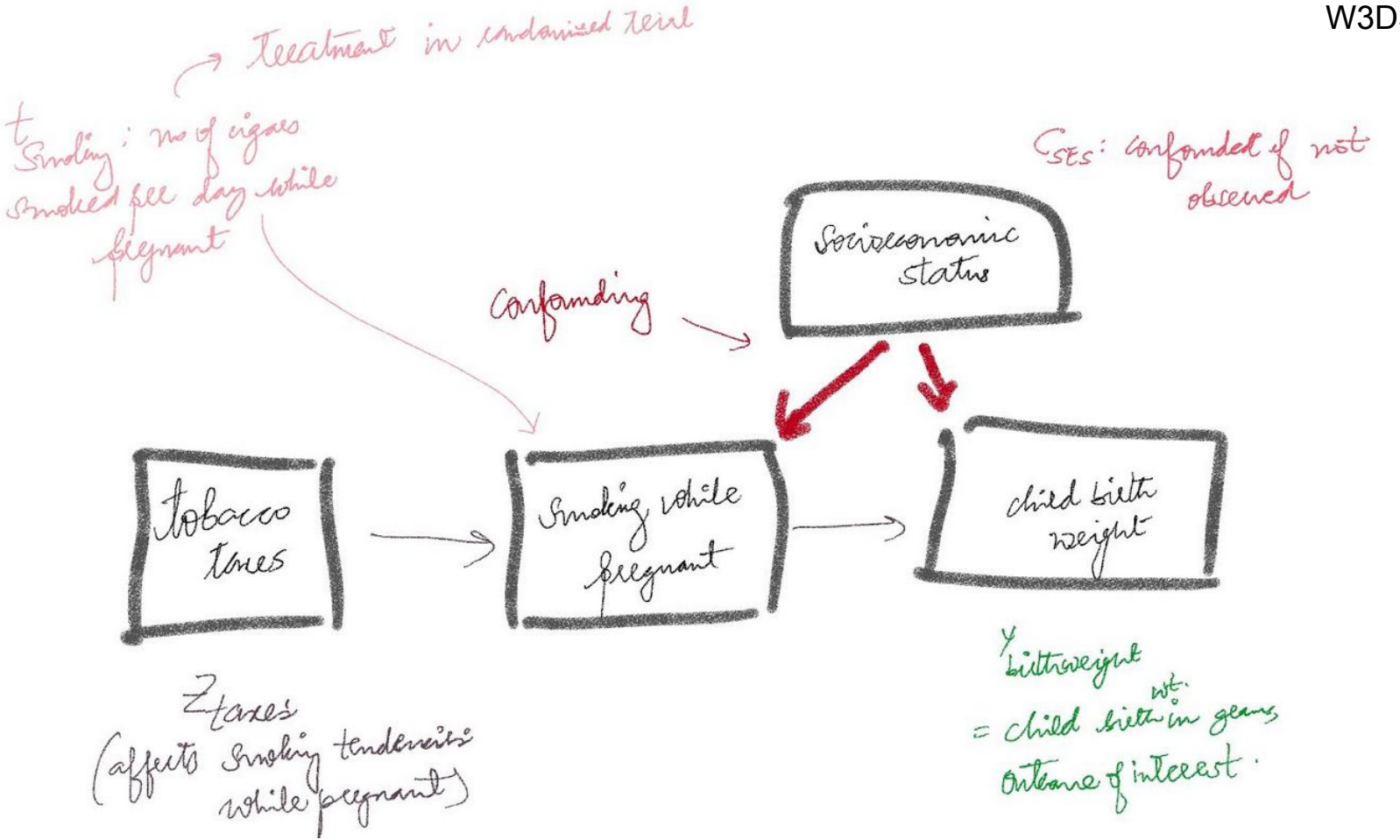
- **Instrumental variables,** a method that does not require experimental data for valid causal analysis
- Explore benefits of instrumental variable analysis and limitations
    - Addresses **omitted variable bias** seen in regression
    - Less efficient in terms of sample size than other techniques
    - Requires a particular form of randomness in the system in order for causal effects to be identified

Instrumental Variables

If there is randomness naturally occurring in the system that we can observe, this in effect becomes the perturbations we can use to recover causal effects. This is called an **instrumental variable**. At high level, an instrumental variable must

1. be observable
2. effect a covariate you care about
3. **not** effect the outcome, except through the covariate

Note: It's rare to find these things in the wild


Treatment in randomized trial
t
Smoking: no of cigars smoked per day while pregnant
C_SES: confounded if not observed
Socioeconomic status
Confounding
Tobacco taxes
Smoking while pregnant
child birth weight
Z
taxes
(affects smoking tendencies while pregnant)
Y
birthweight
= child birth wt. in grams
Outcome of interest.

Iv techniques

A classic example is estimating the effect of smoking cigarettes while pregnant on the birth weight of the infant. There is a (negative) correlation, but is it causal? Unfortunately many confounds affect both birth weight and smoking. Wealth is a big one.

Here the instrumental variable is **state taxes on tobacco**. These

1. Are observable
2. Affect tobacco consumption
3. Don't affect birth weight except through tobacco

By using the power of IV techniques, you can determine the causal effect without exhaustively controlling for everything.

Correlations and confounders

CSES is correlated with both Tsmoking and Ybirthweight, so CSES is a potential confounder if not included in analysis.

If it is difficult to observe and quantify CSES, so we do not have it available to regress against. (**omitted variable bias**)

- Ztaxes is correlated with Tsmoking but is uncorrelated with CSES
- Ztaxes doesn't affect Ybirthweight except through Tsmoking (ie Ztaxes doesn't affect or is affected by CSES)
- Z taxes is also observable

$$Y_{birthweight} = 3000 + C_{SES} - 2T_{smoking}$$

+ C_{SES} is negatively correlated with $T_{smoking}$

Causal effect to estimate coefficient -2 for $T_{smoking}$

$\Rightarrow$ if mother smokes additional cigar per day while pregnant baby is 2g lighter @ birth

How IV works, at high level

The easiest way to imagine IV is that the instrument is **an observable source of "randomness"** that affects the treatment.

The key is that we need to extract **the component of the treatment that is due only to the effect of the instrument.**

It is simply the predicted value of T found in a regression that has only the instrument Z as input.

Once we have the unconfounded component in hand, getting the causal effect is as easy as regressing the outcome on T^.

instrumental variable estimation: 2 stage least squares

2 regressions

① $\hat{T}_{Smoking}$ by regressing $T_{Smoking}$ on Z_{taxes} fitting $\hat{\alpha}$ parameter

estimates unconfounded component of $T_{Smoking}$

$$\hat{T}_{Smoking} = \hat{\alpha} Z_{taxes}$$

② regress $Y_{birthweight}$ on $\hat{T}_{Smoking}$ to obtain $\hat{\beta}$ estimate of causal effect

uses unconfounded component $\hat{T}_{Smoking}$ to estimate effect of smoking on $\hat{Y}_{birthweight}$

$$\hat{Y}_{birthweight} = \hat{\beta} \hat{T}_{Smoking}$$

Compute regression stage 1

Run the regression of Tsmoking on Ztaxes to compute T^ smoking. We will then check whether our estimate is still confounded with

CSES by comparing the correlation of CSES with Tsmoking vs T^smoking

Results: correlation between T and C of -0.483 and between T^ and C of 0.009.

Regress again and obtain estimated causal effect on number of cigars (T on birthweight Y)

Results: obtain estimated causal effect of -1.984 which is close to true causal effect of -2

Dynamics of neurons –

$$\vec{x}_{t+1} = \sigma\left(A\vec{x}_t + \eta\vec{z}_{t+1} + \varepsilon_t\right)$$

additional source of noise (random binary variable)

strength of it

$\vec{z}_t \sim \text{Bernoulli}(0.5)$

All about z (IV)

For each neuron I, we are trying to figure out whether I is connected to (causally affects) the other neurons in our system *at the next time step*. So for timestep T, we want to determine whether $\vec{X}_{I,T}$ affects all the other neurons at $\vec{X}_{T+1}$.

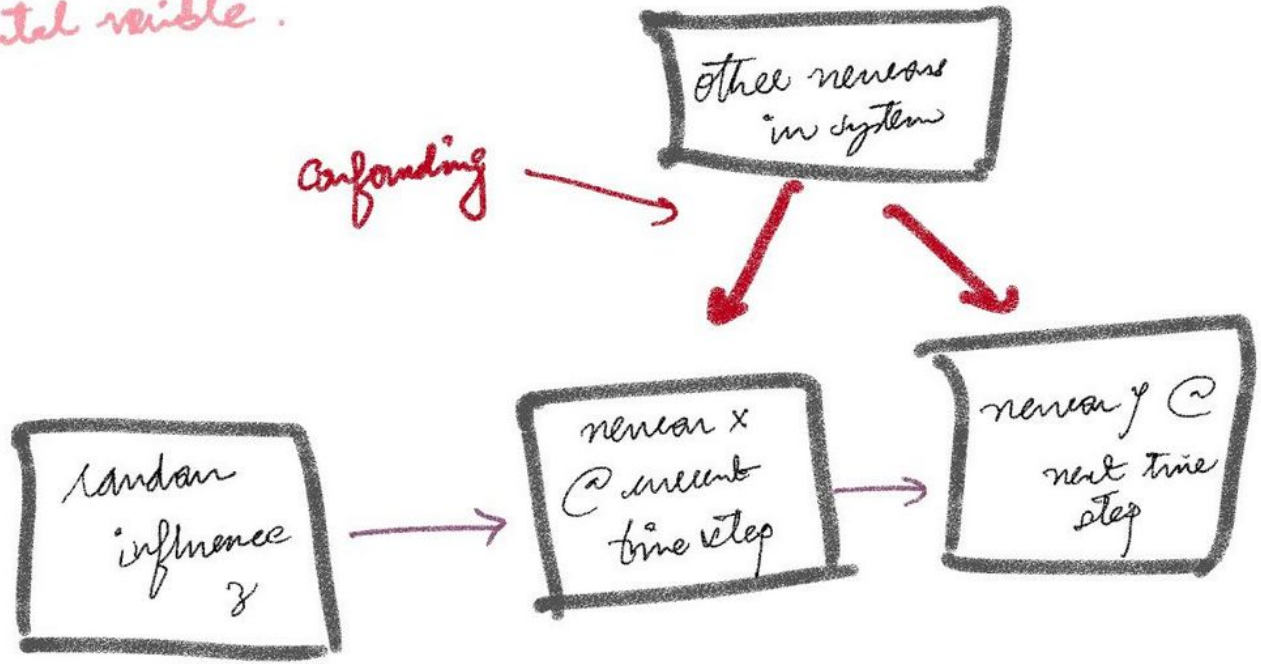
For a given neuron I, $\vec{Z}_{I,T}$ satisfies the 3 criteria for a valid instrument.

What could Z be, biologically?

Imagine Z to be some injected current through an *in vivo* patch clamp. It affects each neuron individually, and only affects dynamics through that neuron.

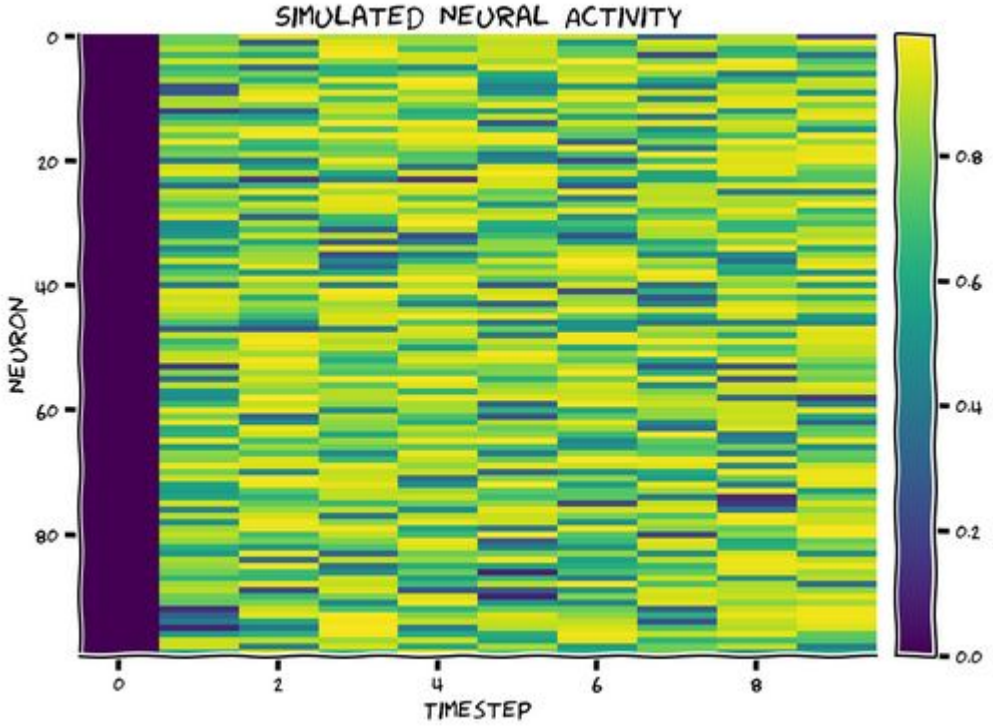
Don't have to control Z yourself - it can be observed. So if you mess up your wiring and accidentally connect the injected voltage to an AM radio, no worries. As long as you can observe the signal the method will work.

Instrumental variable.

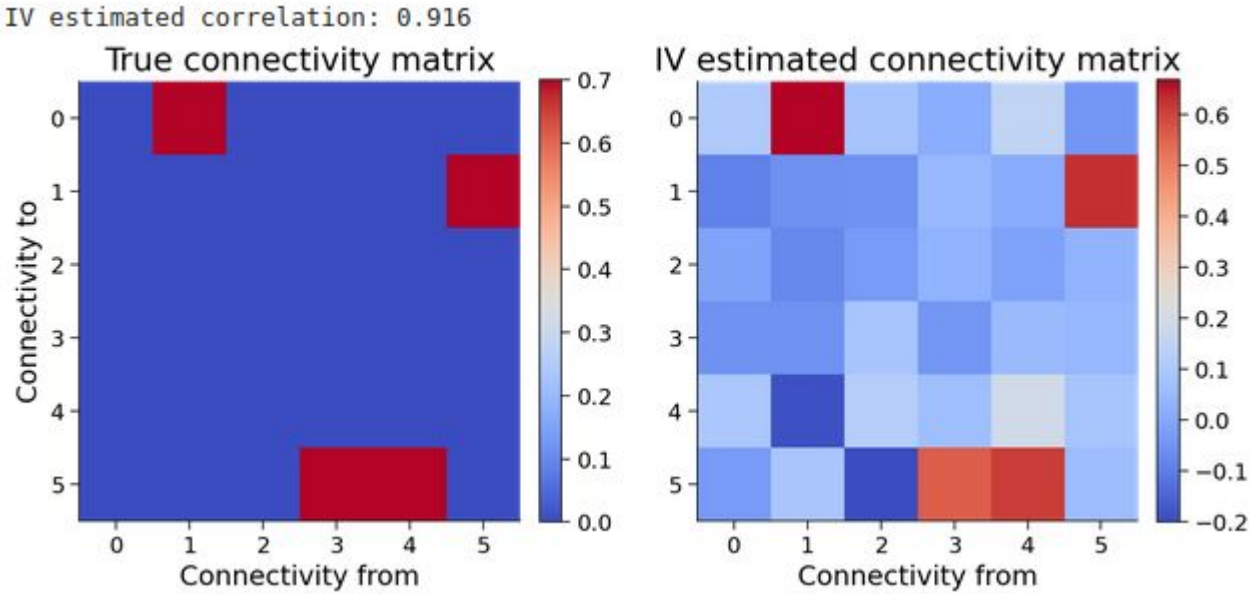


Simulate a system with IV

Modify the function that simulates the neural system, but update rule includes the effect of the instrumental variable z.

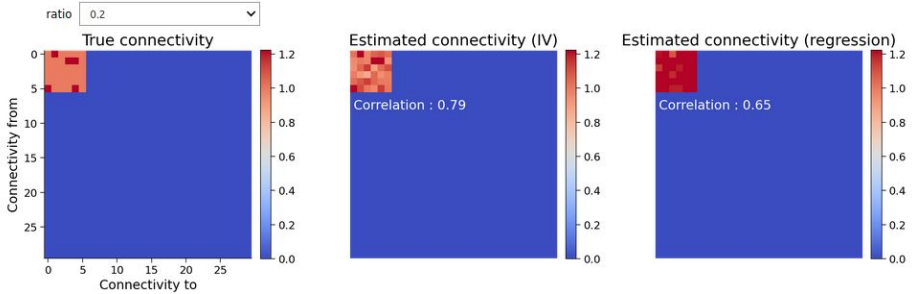


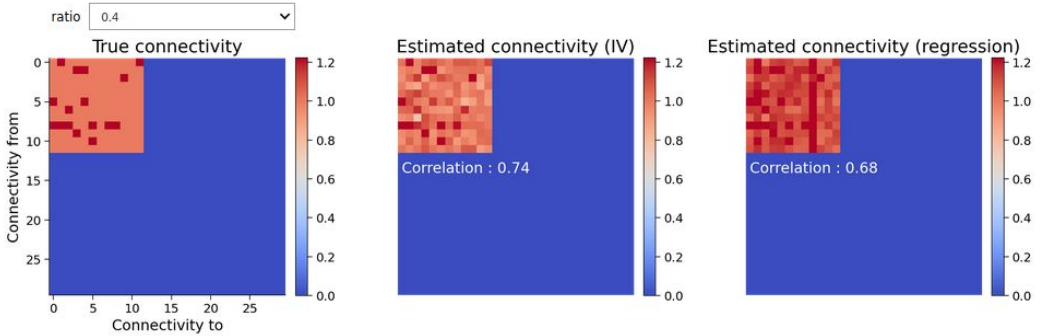
IV estimates to recover the connectivity matrix

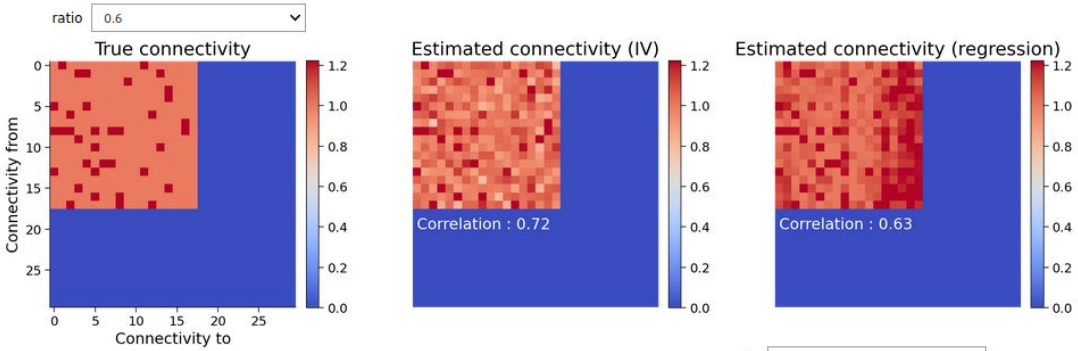


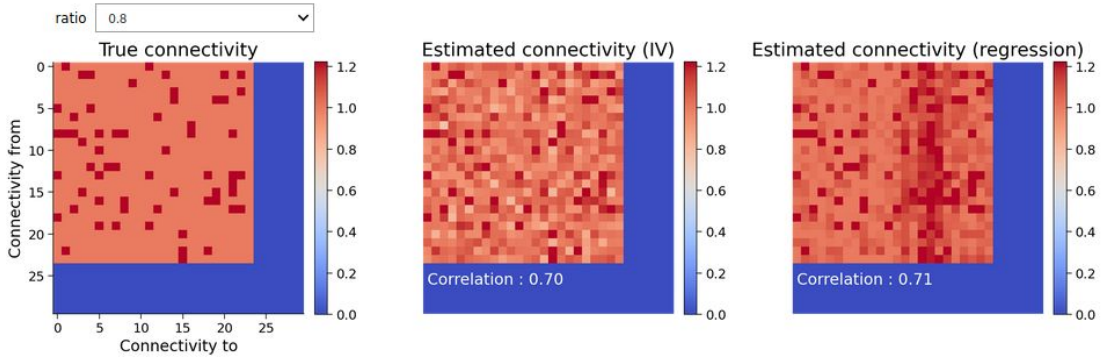
Estimating connectivity with IV vs regression on a subset of observed neurons

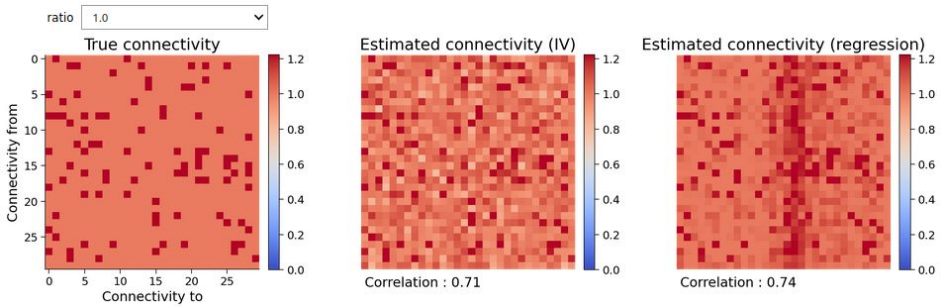
Observe ratio of observed neurons and look at the impact on the quality of connectivity estimation using IV vs regression



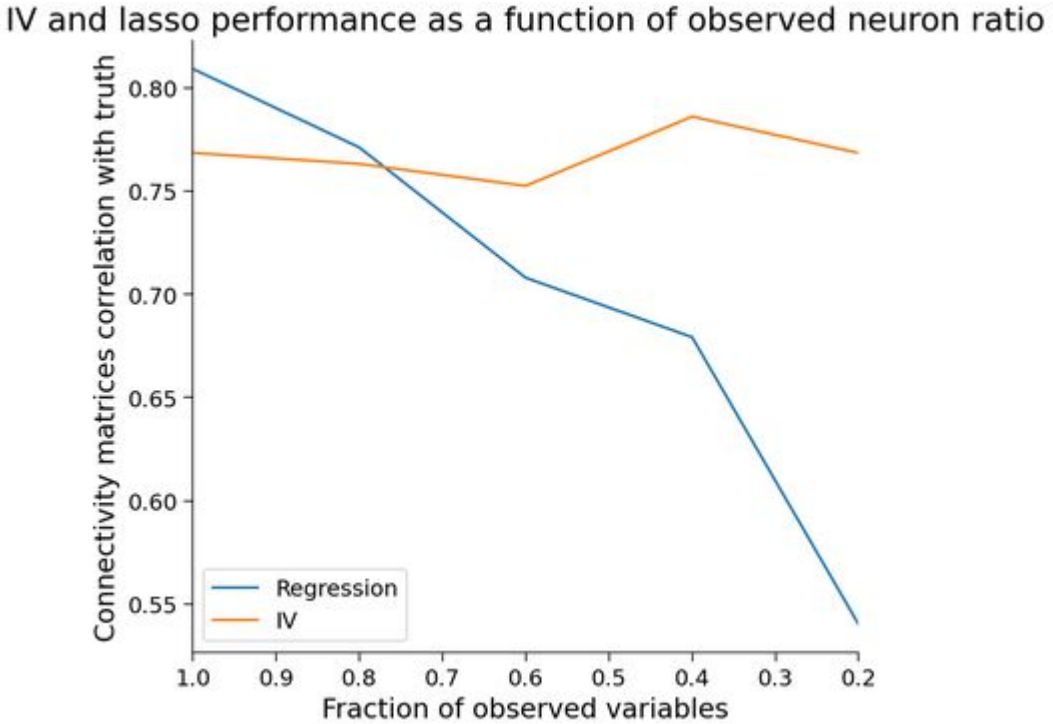



ratio 0.6
True connectivity
Estimated connectivity (IV)
Estimated connectivity (regression)
Connectivity from
Connectivity to
Correlation : 0.72
Correlation : 0.63


ratio 0.8
True connectivity
Estimated connectivity (IV)
Estimated connectivity (regression)
Connectivity from
Connectivity to
Correlation : 0.70
Correlation : 0.71


ratio 1.0
True connectivity
Estimated connectivity (IV)
Estimated connectivity (regression)
Connectivity from
Connectivity to
Correlation : 0.71
Correlation : 0.74

Visualize the performance of regression and IV as a function of the observed neuron ratio



IV analysis

IVs handle omitted variable bias (when the instrument is strong and we have enough data).

The costs of IV analysis

- we need to find an appropriate and valid instrument
- Because of the 2-stage estimation process, we need strong instruments or else our standard errors will be large

Summary

- Explored instrumental variables and how we can use them for causality estimates
- Compared IV estimates to regression estimates

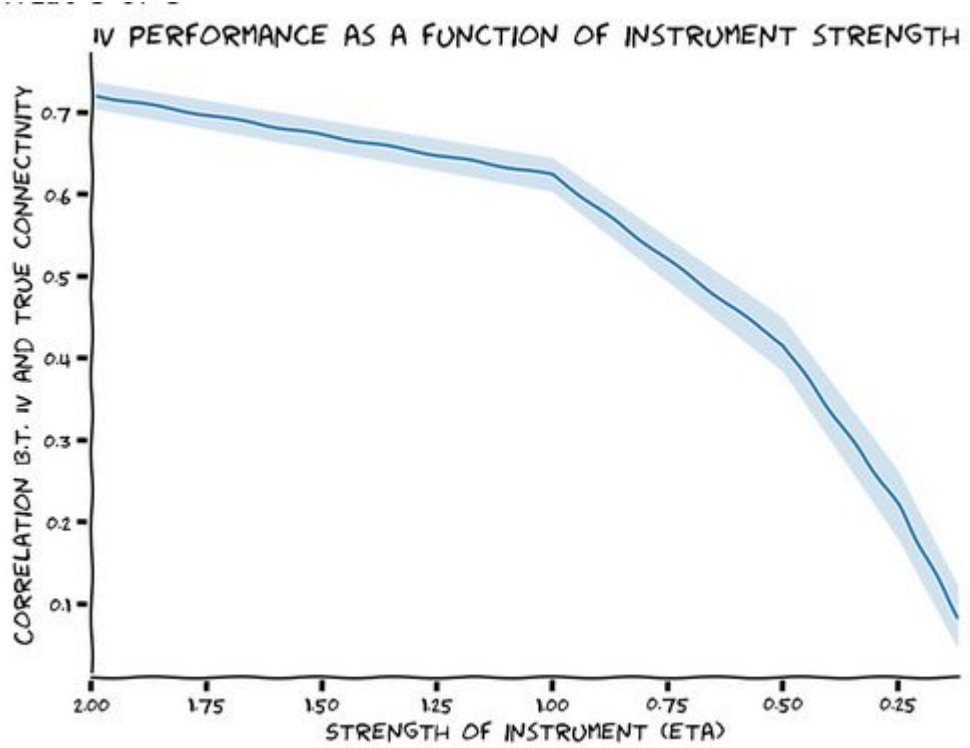
Appendix

Exploring instrument strength

how the strength of the instrument η

affects the quality of estimates with instrumenta

variables.



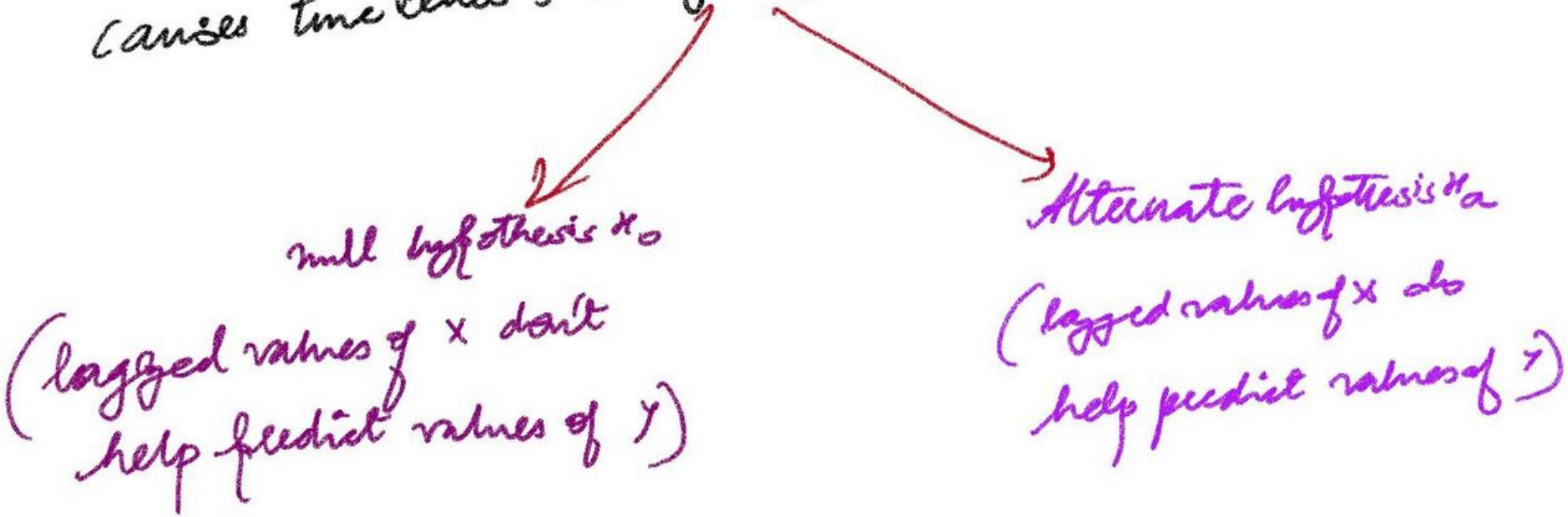
Granger Causality

But, like the simultaneous fitting, this method still fails in the presence of unobserved variables.

Evaluate the Granger causality between our neurons:

The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another, first proposed in 1969.

testing whether time series X - granger causes time series Y through hypothesis test



mechanically accomplished by fitting autoregressive models for y_t

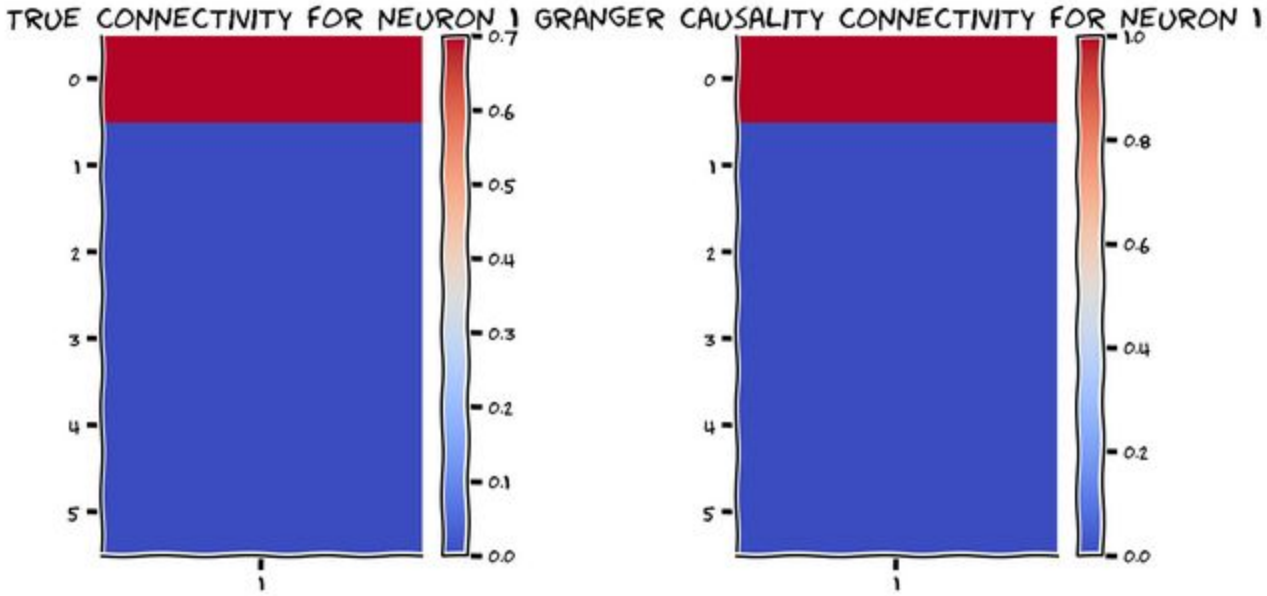
fail to reject hypothesis if none of x_{t-k} terms are retained as significant in the regression.

for simplicity: only 1 time lag.

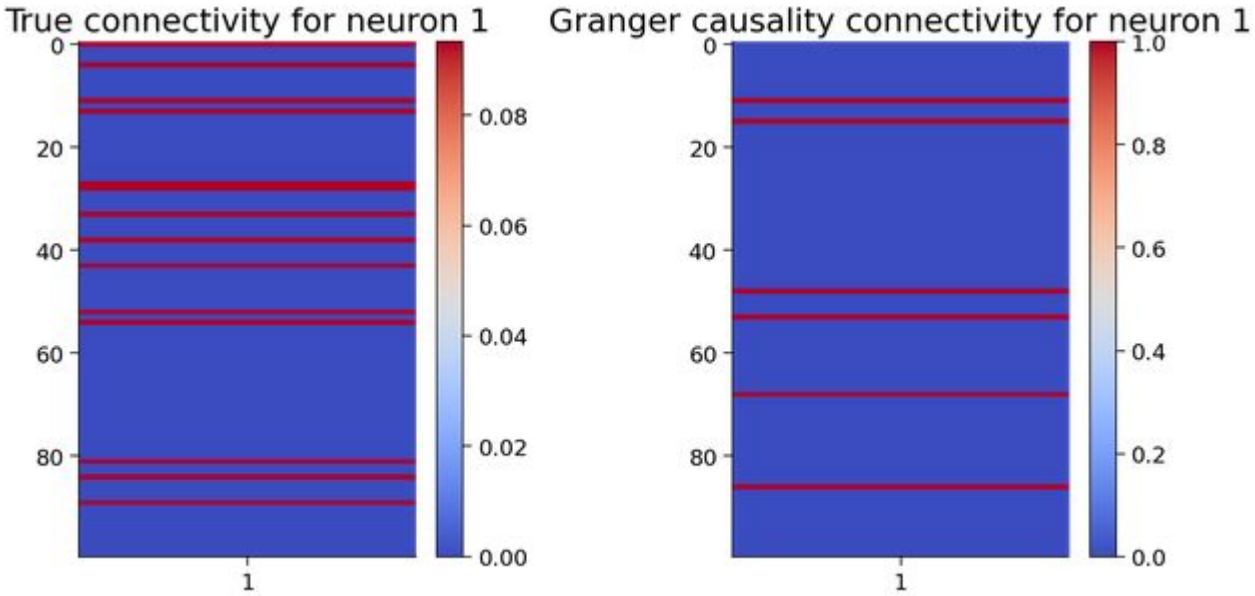
$$H_0: y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

$$H_a: y_t = a_0 + a_1 y_{t-1} + b_1 x_{t-1} + \varepsilon_t$$

When we have a small system, we correctly identify the causality of neuron 1.



Granger causality in large systems (100 neurons)



Note

Considere bivariate Granger causality -- for each pair of neurons A,B, does one Granger-cause the other?

Conditional Granger Causality is a technique that allows for a multivariate system, where we test whether A Granger-causes B conditional on the other variables in the system.

Even after controlling for variables in the system, conditional Granger causality will also likely perform poorly as our system gets larger. Plus, measuring the additional variables to condition on may be infeasible in practical applications, which would introduce omitted variable bias.

As our estimation procedures become more sophisticated, they also become more difficult to interpret. We always need to understand the methods and the assumptions that are made.