Neuromatch Academy: Network Causality - Summary Sheet[1]

Interventions

Overview

Here we will describe causality, the tools we use to ask if and how a variable influences other variables. Causal questions are everywhere in neuroscience. How do neurons influence one another? How does a drug affect neurons? How does a stimulus affect behavior? We will talk about how we can answer questions of a causal kind. Causal questions are important all across neuroscience. For example, model fitting, machine learning, and dimensionality reduction, are often used to argue for or against causal models. For example, a regression may be used to argue that a brain region influences another brain region based on fMRI data. Today's materials give us a better understanding of the problems that come with the approach. There are tight links between causality and Bayesian statistics where Bayesian techniques are used for the estimation of causality (see e.g. the work of Judea Pearl). Causality is often seen as the bedrock of science, today's materials above all produce clarity about what it is.
Causality approaches are central across neuroscience. When we run experiments, we often randomly assign them to treatment groups vs control. Alternatively we stimulate animals at random points of time. These methods are all versions of randomized perturbations and probably constitute a good part of all of neuroscience. We also use model fitting frequently to drive arguments about how brains work. This is common for spike data, EEG data, imaging data etc. Lastly, we should be able to sometimes use instrumental variable techniques to estimate the effects of e.g. treatments with drugs. These materials are simultaneously at the heart of the field and are frequently ignored.

Interventions

Introduction

How do we know if a relationship is causal? What does that mean? And how can we estimate causal relationships within neural data?

The methods we'll learn today are very general and can be applied to all sorts of data, and in many circumstances. Causal questions are everywhere!

Defining and estimating causality

Let's think carefully about the statement "**A causes B**". To be concrete, let's take two neurons. What does it mean to say that neuron A causes neuron B to fire?
The *interventional* definition of causality says that:

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

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

A mathematical definition of causality:
Over many trials, the average causal effect $\delta_{A\to B}$ of neuron A upon neuron B is the average change in neuron B's activity when we set $A = 1$ versus when we set $A = 0$.

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

where $E[B|A = 1]$ is the expected value of B if A is 1 and $E[B|A = 0]$ is the expected value of B if A is 0.
Note that this is an average effect. While one can get more sophisticated about conditional effects (A only effects B when it's not refractory, perhaps), we will only consider average effects today.
Relation to a randomized controlled trial (RCT): The logic we just described is the logic of a randomized control trial (RCT). If you randomly give 100 people a drug and 100 people a placebo, the effect is the difference in outcomes.

Randomized controlled trial for two neurons

Let's pretend we can perform a randomized controlled trial for two neurons. Our model will have neuron A synapsing on Neuron B:

$$B = A + \epsilon$$

where A and B represent the activities of the two neurons and ϵ is standard normal noise $\epsilon \sim \mathcal{N}(0, 1)$.
To confirm if A is perturbed that B changes.
We can calculate the a difference in means of '0.990719' (so very close to one)

Interventions

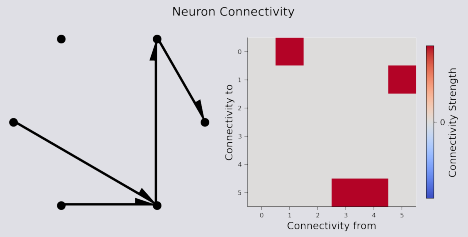
Simulating a system of neurons

Can we still estimate causal effects when the neurons are in big networks? This is the main question we will ask today. Let's first create our system, and the rest of today we will spend analyzing it. Here we introduce a big causal system (interacting neurons) with understandable dynamical properties and how to simulate it.
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:

$$\vec{x}_{t+1} = \sigma(A\vec{x}_t + \epsilon_t),$$

- $\vec{x}_t$ is an n-dimensional vector representing our n-neuron system at timestep t
- σ is a sigmoid nonlinearity
- A is our $n \times n$ "causal ground truth connectivity matrix" (more on this later)
- ϵ_t is random noise: $\epsilon_t \sim N(\vec{0}, I_n)$
- $\vec{x}_0$ is initialized to $\vec{0}$

A is a connectivity matrix, so the element A_{ij} represents the causal effect of neuron i on neuron j. In our system, neurons will receive connections from only 10% of the whole population on average.
We will create the true 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.
Check your understanding: do you understand how the left plot relates to the right plot below?



[1] 't Hart et al., (2022). Neuromatch Academy: a 3-week, online summer school in computational neuroscience. Journal of Open Source Education, 5(49), 118. https://doi.org/10.21105/jose.00118

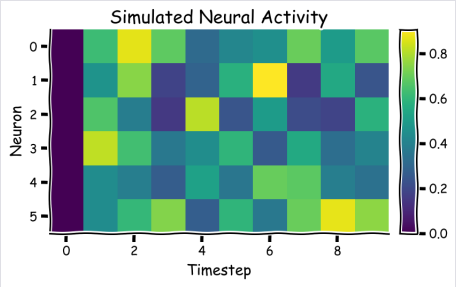
Interventions

System simulation

To simulation a system of six neurons we using a sigmoid σ function so that at every timestep the activity vector x is updated according to:

$$\vec{x}_{t+1} = \sigma(A\vec{x}_t + \epsilon_t).$$

giving the plot:



Random perturbation in our system of neurons

We want to get the causal effect of each neuron upon each other neuron. The ground truth of the causal effects is the connectivity matrix A.
Remember that we would like to calculate:

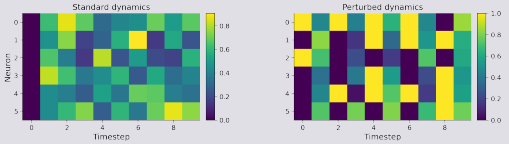
$$\delta_{A \to B} \quad = \quad E[B|A=1] - E[B|A=0]$$

We'll do this by randomly setting the system state to 0 or 1 and observing the outcome after one timestep. If we do this N times, the effect of neuron i upon neuron j is:

$$\delta_{x^i \to x^j} \quad \approx \quad \frac{1}{N} \sum_{t=0,\ t \text{ even}}^{N} [x^j_{t+1} | x^i_t = 1]$$

$$-\frac{1}{N} \sum_{t=0,\ t \text{ even}}^{N} [x^j_{t+1} | x^i_t = 0]$$

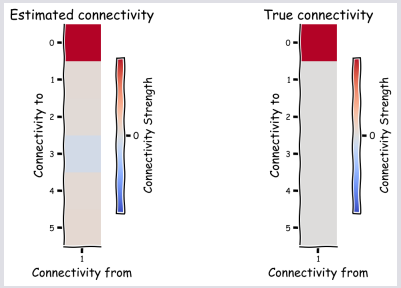
This is just the average difference of the activity of neuron j in the two conditions.
We are going to calculate the above equation, but imagine it like *intervening* in activity every other timestep.



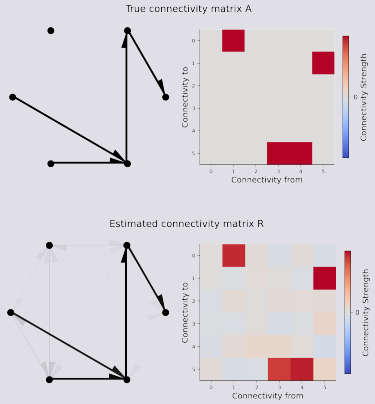
Interventions

Recovering connectivity from perturbed dynamics

Recall that we perturbed every neuron at every other timestep. Despite perturbing every neuron, in this exercise we are concentrating on computing the causal effect of a single neuron (we will look at all neurons effects on all neurons next). We want to exclusively use the timesteps without perturbation for x^j_{t+1} and the timesteps with perturbation for x^j_t in the formulas above.



We can quantify how close our estimated connectivity matrix is to our true connectivity matrix by correlating them. We should see almost perfect correlation between our estimates and the true connectivity - do we?
Note on interpreting A: Strictly speaking, A is not the matrix of causal effects but rather the dynamics matrix. So why compare them like this? The answer is that 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 is in part because the nonlinearity σ squashes the values of x to $[0, 1]$.)



We can again calculate the correlation coefficient between the elements of the two matrices. If it is almost 1 we have done a good job recovering the true causality of the system!

Correlations

Recovering connectivity from perturbed dynamics

Here, we implemented and explored the dynamical system of neurons we will be working with throughout all of the tutorials today. We also learned about the "gold standard" of measuring causal effects through random perturbations. As random perturbations are often not possible, we will now turn to alternative methods to attempt to measure causality. We will:

- Learn how to estimate connectivity from observations assuming **correlations approximate causation**
- Show that this only works when the network is small

Often, we can't force neural activities or brain areas to be on or off. We just have to observe. Maybe we can get the correlation between two nodes – is that good enough? The question we ask here is **when is correlation a "good enough" substitute for causation?**

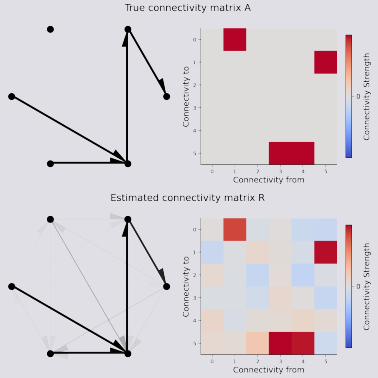
The answer is not "never", actually, but "sometimes".

Try to approximate causation with correlation

In small systems, correlation can look like causation. Let's attempt to recover the true connectivity matrix (A) just by correlating the neural state at each timestep with the previous state:

$$C = \vec{x_t} \vec{x_{t+1}}^{\top}.$$

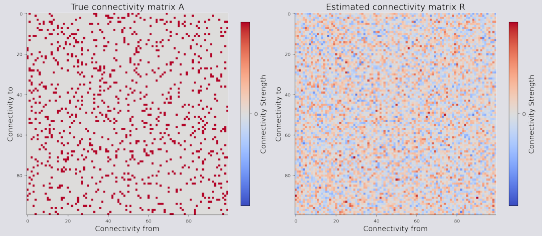
To calculate the connectivity matrix of a single neuron by calculating the correlation coefficients with every other neuron which correlate two vectors: 1) the activity of a selected neuron at time t 2) The activity of all other neurons at time $t+1$.
The matrix answer is the same as the summation form. Furthermore the estimated vs true connectivity look the same using the matrix calculation.



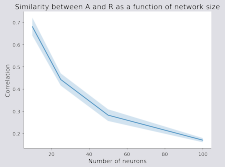
Correlations

Large systems

As our system becomes more complex however, correlation fails to capture causality. Let's jump to a much bigger system. Instead of 6 neurons, we will now use 100 neurons. How does the estimation quality of the connectivity matrix change?

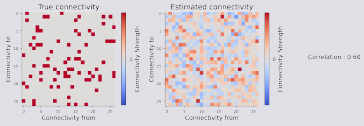


Correlation as a function of network size



Connectivity estimation as a function of the sparsity of A

You may rightly wonder if correlation only fails for large systems for certain types of A. Does connectivity estimation get better or worse with less sparsity?



Summary

Now for the takeaway. We know that for large systems correlation $\neq$ 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?
From our simulation above, 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.

Simultaneous fitting/regression

Objectives

We have explored correlation as an approximation for causation and learned that correlation $\neq$ causation for larger networks. However, computing correlations is a rather simple approach, and you may be wondering: will more sophisticated techniques allow us to better estimate causality? Can't we control things?
Here we'll use some common advanced (but controversial) methods that estimate causality from observational data. 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. Specifically, we will:

- Learn about more advanced (but also controversial) techniques for estimating causality
- conditional probabilities (**regression**)
- Explore limitations and failure modes
- understand the problem of **omitted variable bias**

Regression approach

You may be familiar with the idea that correlation only implies causation when there are no hidden *confounders*. This aligns with our 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: Confounds can be controlled for by adding them as covariates in a regression. But for your coefficients to be causal effects, you need three things:

1. All confounds are included as covariates
2. Your 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; we won't introduce it today (but Google it on your own time! Colliders are very counterintuitive.)

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). Luckily today we simulated the system ourselves.

Simultaneous fitting/regression

Fitting a General Linear Model (GLM)

We will use a regression approach to estimate the causal influence of all neurons to neuron 1. Specifically, we will use linear regression to determine the A in:

$$\sigma^{-1}(\vec{x}_{t+1}) = A\vec{x}_t + \epsilon_t, \tag{1}$$

where σ^{-1} is the inverse sigmoid transformation, also sometimes referred to as the **logit** transformation: $\sigma^{-1}(x) = \log(\frac{x}{1-x})$.
Let W be the $\vec{x}_t$ values, up to the second-to-last timestep $T-1$:

$$W = \begin{bmatrix} | & | & \dots & | \\ \vec{x}_0 & \vec{x}_1 & \dots & \vec{x}_{T-1} \\ | & | & \dots & | \end{bmatrix}_{n\times(T-1)} \tag{2}$$

Let Y be the $\vec{x}_{t+1}$ values for a selected neuron, indexed by i, starting from the second timestep up to the last timestep T:

$$Y = \begin{bmatrix} x_{i,1} & x_{i,2} & \dots & x_{i,T} \end{bmatrix}_{1\times(T-1)} \tag{3}$$

You then fit the following model:

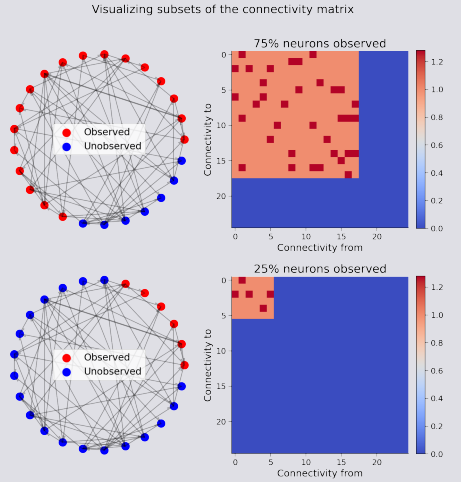
$$\sigma^{-1}(Y^T) = W^T V \tag{4}$$

where V is the $n \times 1$ coefficient matrix of this regression, which will be the estimated connectivity matrix between the selected neuron and the rest of the neurons.
We can see that multiple regression is better than simple correlation for estimating connectivity.

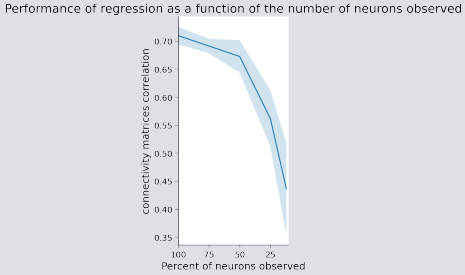
Simultaneous fitting/regression

Partially Observed Systems

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 them, can we still estimate the causal effect accurately?
We first visualize different subsets of the connectivity matrix when we observe 75% of the neurons vs 25%.
Recall the meaning of entries in our connectivity matrix: $A[i,j] = 1$ means a connectivity from neuron i to neuron j with strength 1.



Next, we will 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?



Instrumental Variables

Objectives

We have seen that even more sophisticated techniques such as simultaneous fitting fail to capture causality in the presence of omitted variable bias. So what techniques are there for us to obtain valid causal measurements when we can't perturb the system? Here we will:

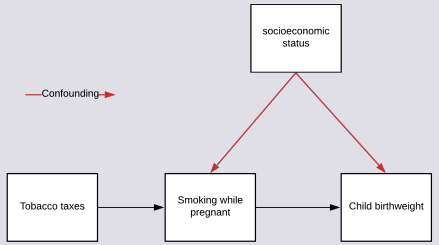
- learn about **instrumental variables**, a method that does not require experimental data for valid causal analysis
- explore benefits of instrumental variable analysis and limitations
- addresses **omitted variable bias** seen in regression
- less efficient in terms of sample size than other techniques
- 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. Affect a covariate you care about
3. **Not** affect the outcome, except through the covariate

It's rare to find these things in the wild, but when you do it's very powerful.



Instrumental Variables

A non-neuro example of an Instrumental Variables (IV)

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. Instead of controlling everything imaginable, one can find an IV. 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.
Let's represent our tobacco example above with the following notation:

- Z_{taxes}: our tobacco tax *instrument*, which only affects an individual's tendency to smoke while pregnant within our system
- T_{smoking}: number of cigarettes smoked per day while pregnant, our "treatment" if this were a randomized trial
- C_{SES}: socioeconomic status (higher means wealthier), a *confounder* if it is not observed
- $Y_{\text{birthweight}}$: child birthweight in grams, our outcome of interest

Let's suppose we have the following function for our system:

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

with the additional fact that C_{SES} is negatively correlated with T_{smoking}.
The causal effect we wish to estimate is the coefficient -2 for T_{smoking}, which means that if a mother smokes one additional cigarette per day while pregnant her baby will be 2 grams lighter at birth. We've provided a covariance matrix with the desired structure in the code cell below, so please run it to look at the correlations between our variables.
Correlation between SES status and cigarettes: -0.483
Correlation between SES status and birth weight: 0.740
We see what is exactly represented in our graph above: C_{SES} is correlated with both T_{smoking} and $Y_{\text{birthweight}}$, so C_{SES} is a potential confounder if not included in our analysis.
Let's say that it is difficult to observe and quantify C_{SES}, so we do not have it available to regress against. This is another example of the **omitted variable bias**.
What about Z_{taxes}?
Correlation between taxes and cigarettes: 0.519
Correlation between taxes and SES status: 0.009
Perfect! We see that Z_{taxes} is correlated with T_{smoking} (2) but is uncorrelated with C_{SES} (3). Z_{taxes} is also observable (1), so we've satisfied our three criteria for an instrument:

1. Z_{taxes} is observable
2. Z_{taxes} affects T_{smoking}
3. Z_{taxes} doesn't affect

$Y_{\text{birthweight}}$ except through T_{smoking} (ie Z_{taxes} doesn't affect or is affected by C_{SES})

Instrumental Variables

How IV works

The easiest way to imagine IV is that the instrument is **an observable source of "randomness"** that affects the treatment. In this way it's similar to the interventions we talked about in Tutorial 1.
But how do you actually use the instrument? The key is that we need to extract **the component of the treatment that is due only to the effect of the instrument**. We will call this component $\hat{T}$.

$$\hat{T} \leftarrow \text{The unconfounded component of } T$$

Getting $\hat{T}$ is fairly simple. 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 $\hat{T}$.

IV estimation using two-stage least squares

The fundamental technique for instrumental variable estimation is **two-stage least squares**.
We run two regressions:

1. The first stage gets $\hat{T}_{\text{smoking}}$ by regressing T_{smoking} on Z_{taxes}, fitting the parameter $\hat{\alpha}$:

$$\hat{T}_{\text{smoking}} = \hat{\alpha} Z_{\text{taxes}} \tag{5}$$

2. The second stage then regresses $Y_{\text{birthweight}}$ on $\hat{T}_{\text{smoking}}$ to obtain an estimate $\hat{\beta}$ of the causal effect:

$$\hat{Y}_{\text{birthweight}} = \hat{\beta} \hat{T}_{\text{smoking}} \tag{6}$$

The first stage estimates the **unconfounded component** of T_{smoking} (ie, unaffected by the confounder C_{SES}), as we discussed above.
Then, the second stage uses this unconfounded component $\hat{T}_{\text{smoking}}$ to estimate the effect of smoking on $\hat{Y}_{\text{birthweight}}$.

Compute regression stage 1
Let's run the regression of T_{smoking} on Z_{taxes} to compute $\hat{T}_{\text{smoking}}$. We will then check whether our estimate is still confounded with C_{SES} by comparing the correlation of C_{SES} with T_{smoking} vs $\hat{T}_{\text{smoking}}$.
The result of the correlation between T and C of '-0.483' and between $\hat{T}$ and C of '0.009'.

Least squares regression stage 2
Now let's implement the second stage! We will again use a linear regression model with an intercept. We will obtain the estimated causal effect of the number of cigarettes (T) on birth weight (Y).

The result of estimated causal effect of '-1.984', This is quite close to the true causal effect of -2!

Instrumental Variables

IVs in our simulated neural system

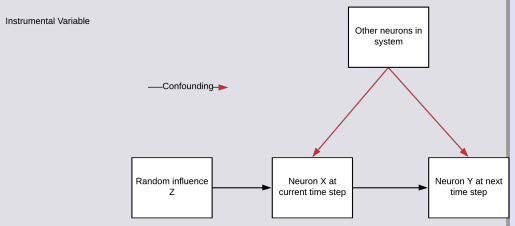
Now, say we have the neural system we have been simulating, except with an additional variable $\vec{z}$. This will be our instrumental variable.
We treat $\vec{z}$ as a source of noise in the dynamics of our neurons:

$$\vec{x}_{t+1} = \sigma(A\vec{x}_t + \eta\vec{z}_{t+1} + \epsilon_t) \quad (7)$$

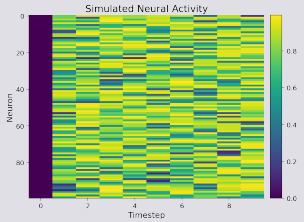
where η is what we'll call the "strength" of our IV, and $\vec{z}_t$ is a random binary variable, $\vec{z}_t \sim Bernoulli(0.5)$
Remember that for each neuron i, we are trying to figure out whether i is connected to (causally affects) the other neurons in our system *at the next time step*. So for timestep 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.
The cool thing about IV is that you 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.



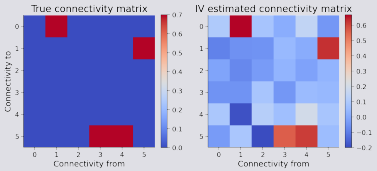
Simulate a system with IV



Instrumental Variables

Estimate IV for simulated neural system

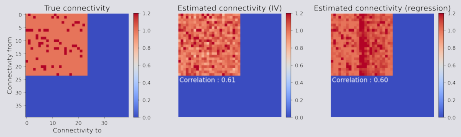
Since you just implemented two-stage least squares, let's see how our IV estimates do in recovering the connectivity matrix.



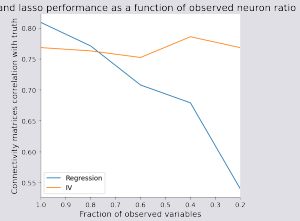
The IV estimates seem to perform pretty well! In the next section, we will see how they behave in the face of omitted variable bias.

IVs and omitted variable bias

Changing the ratio of observed neurons and look at the impact on the quality of connectivity estimation using IV vs regression. The plots below are for a ratio of 0.6.



We can also visualize the performance of regression and IV as a function of the observed neuron ratio below.



We see that 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