Connectivity as a probabilistic model

- The simplest model to study connectivity is a **linear equation**:

$$y = \beta^T x$$

where *y* is the outcome variable that we want to predict and *x* are known variables that affect the outcome.

Connectivity as a probabilistic model

- The simplest model to study connectivity is a **linear equation**:

$$y = \beta^T x$$

where y is the outcome variable that we want to predict and x are known variables that affect the outcome.

- The real world involves **uncertainty**. To deal with this uncertainty, we resort to a probability distribution:

$$p(x, y)$$

Benefits of probabilistic modeling

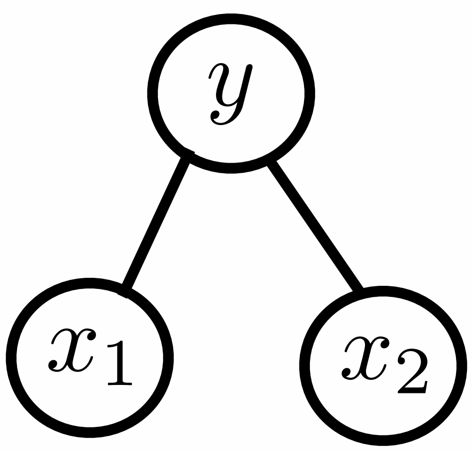
- General I: Less **assumptions** than the linear-model framework (no linear connections, no Gaussianity of the error residuals, etc.).
- General II: It can encompass **connectivities** and other **variable dependencies** in the same model.

Benefits of probabilistic modeling

- General I: Less **assumptions** than the linear-model framework (no linear connections, no Gaussianity of the error residuals, etc.).

- General II: It can encompass **connectivities** and other **variable dependencies** in the same model.

- There is a one-to-one correspondence with **graphs**.



$$p(y, x_1, x_2) = p(x_1|y)p(x_2|y)p(y)$$

Difficulties of probabilistic modeling

- Even for a **binary** alphabet, model requires one parameter for each combination

$$y, x_1, \dots, x_n$$

resulting in 2^{n+1} parameters.

- This is **impractical** from both a **computational** (how do we store this large list?) and from a **statistical** (how do we efficiently estimate the parameters from limited data?) point of view.

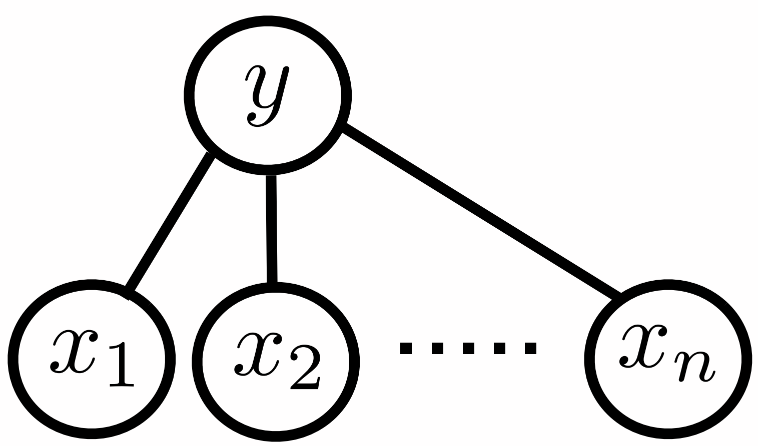
- As exponentially-sized objects, we need **simplifying assumptions** about their structure.

The conditional-independence assumption

- Variables might be **independent** conditioned on the outcome of one variable.

- Model probability as a **product of factors**:

$$p(y, x_1, \ldots, x_n) = p(y) \prod_{i=1}^{n} p(x_i|y)$$



- The number of model parameters to estimate becomes **linear**.

Outline

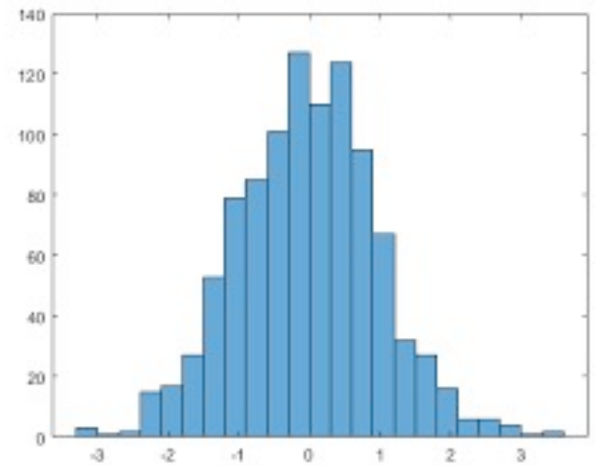
- Step back to Probability basics
- Information-theoretic measures
- Toy model example

Random variables

- A **random variable** *X* can be regarded as a real-valued function of random outcomes (e.g., the number of tails that appear in 10 coin tosses).

- *X* can take values in **discrete** or **continuous** alphabets.

- Random variables are defined by its underlying **probability measure** (CDFs, PDFs, and PMFs)



- Simplest characterization of random variables is via its **expectation** and **variance**.

Conditional probability, chain rule and independence

1. **Condition probability**: What is the probability distribution over *Y*, when we know that *X* must take on a certain value *x*?

$$P_{Y|X}(x|y) = P_{X,Y}(x,y)/P_X(x) \qquad\qquad P_X(x) \neq 0$$

Conditional probability, chain rule and independence

1. **Condition probability**: What is the probability distribution over *Y*, when we know that *X* must take on a certain value *x*?

$$P_{Y|X}(x|y) = P_{X,Y}(x,y)/P_X(x) \qquad\qquad P_X(x) \neq 0$$

2. The **chain rule** can be applied for random variables as

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P_{X_1}(x_1)P_{X_2|X_1}(x_2|x_1)\ldots P_{X_n|X_{n-1}}(x_n|x_1,\ldots,x_n)$$

Conditional probability, chain rule and independence

1. **Condition probability**: What is the probability distribution over *Y*, when we know that *X* must take on a certain value *x*?

$$P_{Y|X}(x|y) = P_{X,Y}(x,y)/P_X(x) \qquad\qquad P_X(x) \neq 0$$

2. The **chain rule** can be applied for random variables as

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P_{X_1}(x_1)P_{X_2|X_1}(x_2|x_1)\ldots P_{X_n|X_{n-1}}(x_n|x_1,\ldots,x_n)$$

3. Two random variables *X* and *Y* are **independent** if

$$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

for all values of *x* and *y.*

Mutual information

1. The **mutual information** of two variables is a measure of the **mutual influence** between two random variables. It determines how similar the joint distribution is to the products of factored marginal distribution:

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Mutual information

1. The **mutual information** of two variables is a measure of the **mutual influence** between two random variables. It determines how similar the joint distribution is to the products of factored marginal distribution:

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

2. **Mutual information** and **independence**:

$$\boxed{I(X;Y) = 0 \quad \text{if and only if } X \text{ and } Y \text{ are independent}}$$

Conditional Mutual information

1. The **conditional mutual information** of two variables is the expected value of the mutual information of two random variables given the value of a third.

$$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} P_Z(z) \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y|Z}(x,y|z) \log \frac{P_{X,Y|Z}(x,y|z)}{P_{X|Z}(x|z) P_{Y|Z}(y|z)}$$

Conditional Mutual information

1. The **conditional mutual information** of two variables is the expected value of the mutual information of two random variables given the value of a third.

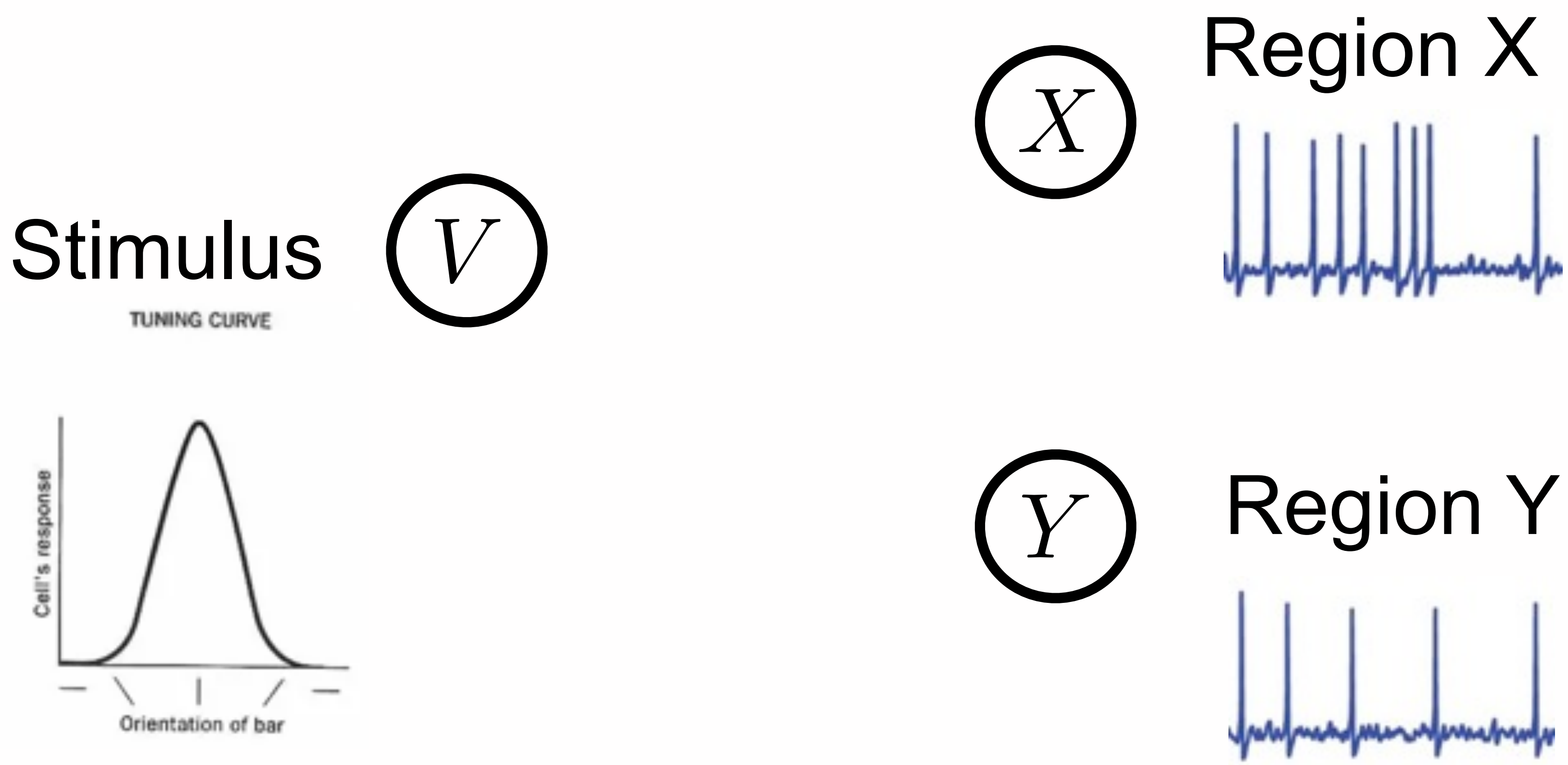
$$I(X;Y|Z) = \sum_{z \in \mathcal{Z}} P_Z(z) \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y|Z}(x,y|z) \log \frac{P_{X,Y|Z}(x,y|z)}{P_{X|Z}(x|z) P_{Y|Z}(y|z)}$$

2. **Conditional mutual information** and **independence**:

$I(X;Y|Z) = 0$ if and only if *X* and *Y* are conditional independent given *Z*.

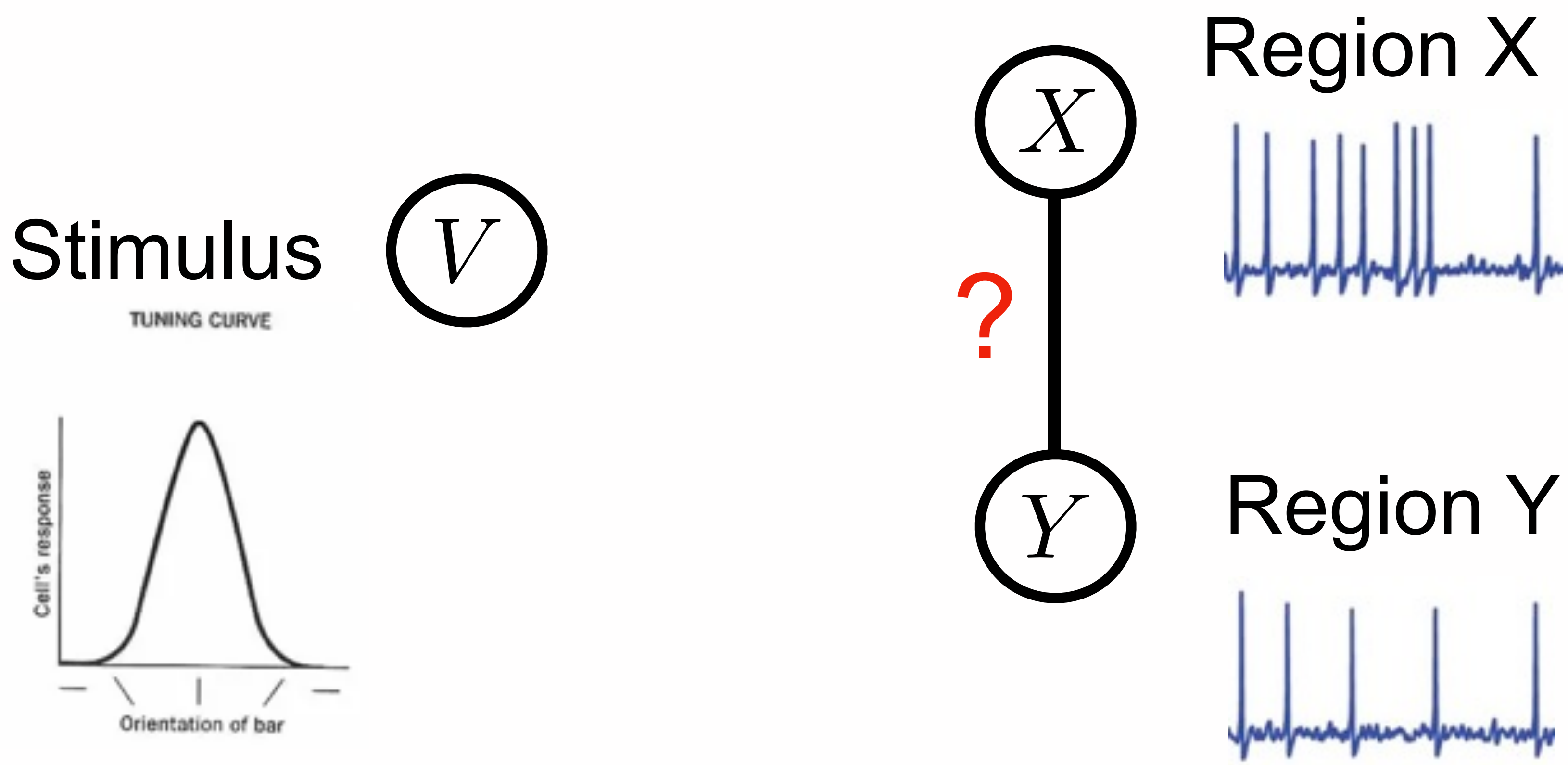
Toy model example

•<u>Cognitive study</u>: stimulus *V* is presented with repetitions, is **non-linearly** encoded by sensory neurons, and the spike trains from two regions in the **sensory pathway** are **simultaneously recorded** (*X* and *Y*) during short time epochs.



Toy model example

•<u>Cognitive study</u>: stimulus *V* is presented with repetitions, is **non-linearly** encoded by sensory neurons, and the spike trains from two regions in the **sensory pathway** are **simultaneously recorded** (*X* and *Y*) during short time epochs.



•Q: How can we assess whether these two regions are connected?

Problem specification

3 **discrete variables**. Assume:

- Stimulus V is discrete uniform
- Sequence X and Y are binomial with no Gaussian approximation.

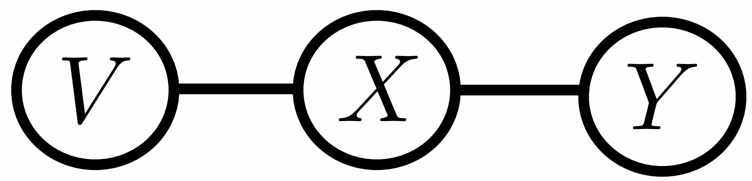
Problem specification

3 **discrete variables**. Assume:

- Stimulus V is discrete uniform
- Sequence X and Y are binomial with no Gaussian approximation.

3 **potential models**:

- Connectivity chain 1



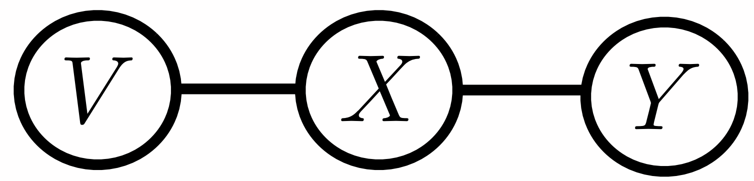
Problem specification

3 **discrete variables**. Assume:

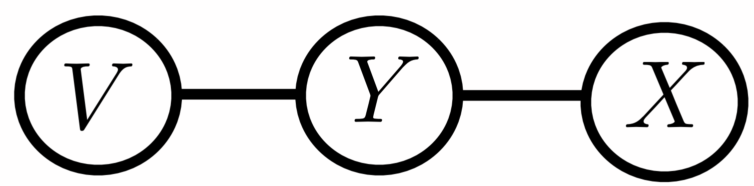
- Stimulus V is discrete uniform
- Sequence X and Y are binomial with no Gaussian approximation.

3 **potential models**:

- Connectivity chain 1



- Connectivity chain 2



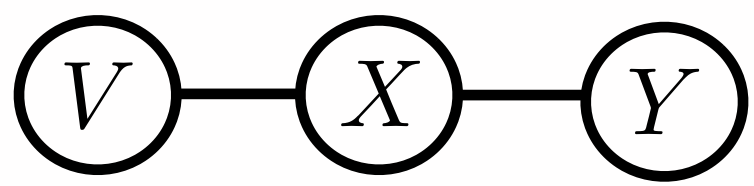
Problem specification

3 **discrete variables**. Assume:

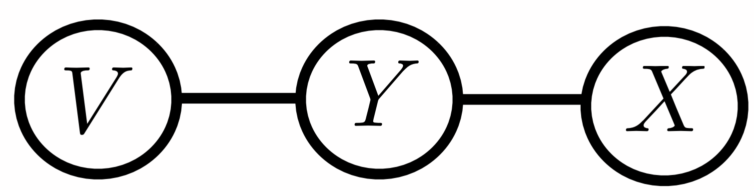
- Stimulus V is discrete uniform
- Sequence X and Y are binomial with no Gaussian approximation.

3 **potential models**:

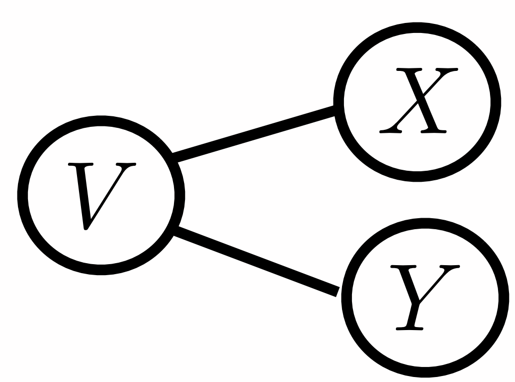
- Connectivity chain 1



- Connectivity chain 2



- Common input



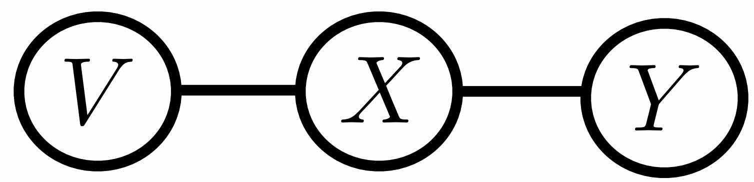
Problem specification

- Q: Given simultaneous observables of *V*, *X* and *Y*, what is the **more likely model** that explains the **interaction between both regions**?

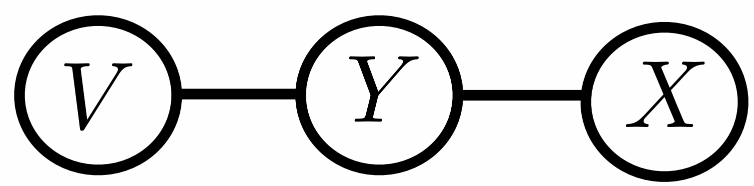
Problem specification

•Q: Given simultaneous observables of *V*, *X* and *Y*, what is the **<u>more likely model</u>** that explains the **interaction between both regions**?

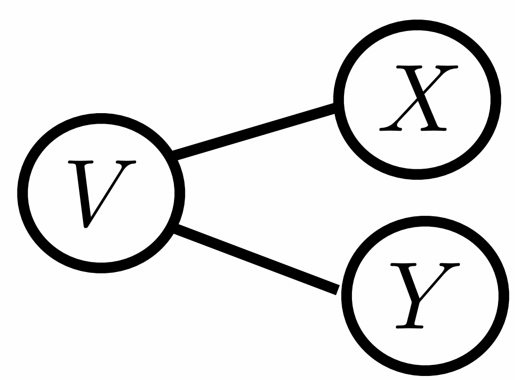
- Connectivity chain 1 (**Effective** connectivity *X*-*Y*)



- Connectivity chain 2 (**Effective** connectivity *X*-*Y*)



- Common input (**No effective** connectivity *X*-*Y*)



Required ingredients

- Measure type?

The relationship between the variables might be **non-linear** and assume non-directionality in the time domain**.** Hence, use of **information-theoretic** measures (e.g. mutual information instead of Pearson or Partial correlation)

Required ingredients

•Measure type?

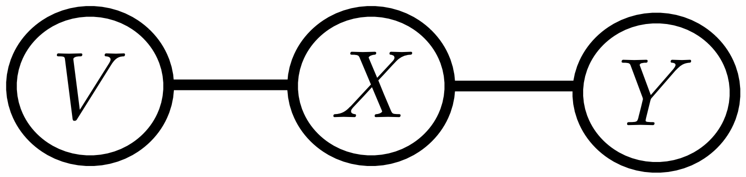
The relationship between the variables might be **non-linear** and assume non-directionality in the time domain**.** Hence, use of **information-theoretic** measures (e.g. mutual information instead of Pearson or Partial correlation)

•Third-node effects?

The mutual information across the three variables will be **signficant** regardless of the **3 models**. Hence, to discriminate the likelihood of each model, use of **conditional mutual information.**

Model conditions

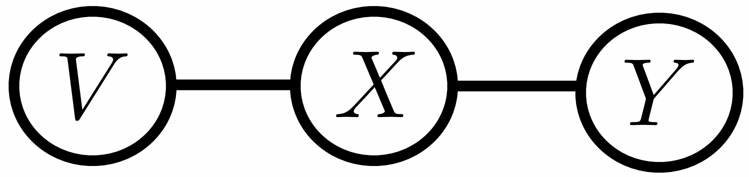
- **Connectivity chain 1**



$I(V;X|Y) > 0$ $\quad I(X;Y|V) > 0$ $\quad \boxed{I(V;Y|X) = 0}$

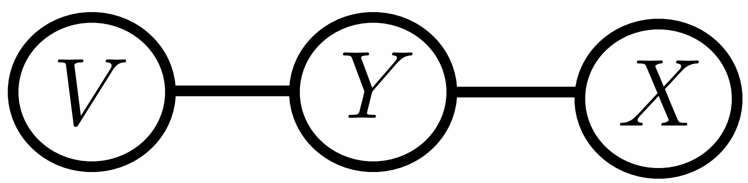
Model conditions

- **Connectivity chain 1**



$$I(V;X|Y) > 0 \quad I(X;Y|V) > 0 \quad \boxed{I(V;Y|X) = 0}$$

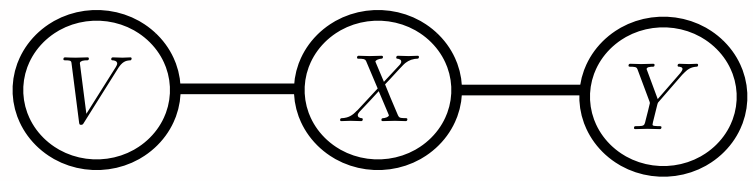
- **Connectivity chain 2**



$$I(V;Y|X) > 0 \quad I(X;Y|V) > 0 \quad \boxed{I(V;X|Y) = 0}$$

Model conditions

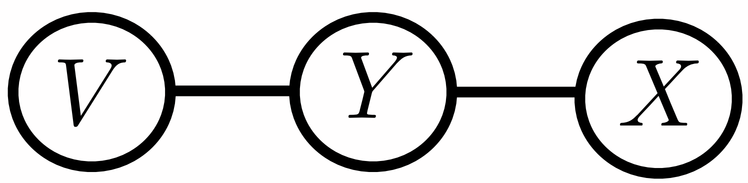
- **Connectivity chain 1**



$I(V;X|Y) > 0 \quad I(X;Y|V) > 0$

$$\boxed{I(V;Y|X) = 0}$$

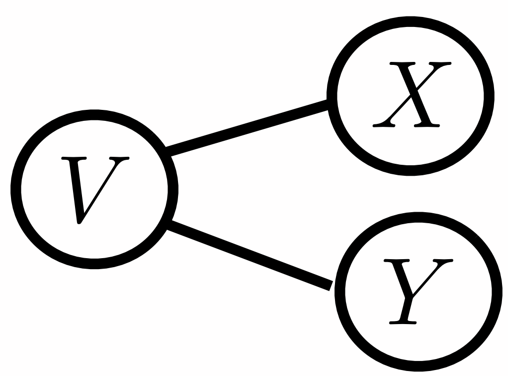
- **Connectivity chain 2**



$I(V;Y|X) > 0 \quad I(X;Y|V) > 0$

$$\boxed{I(V;X|Y) = 0}$$

- **Common input**

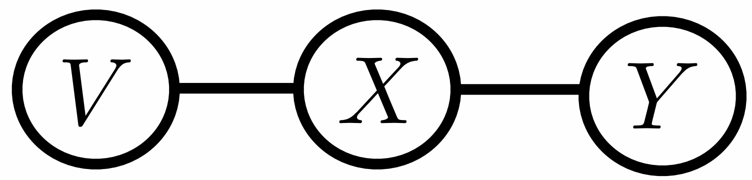


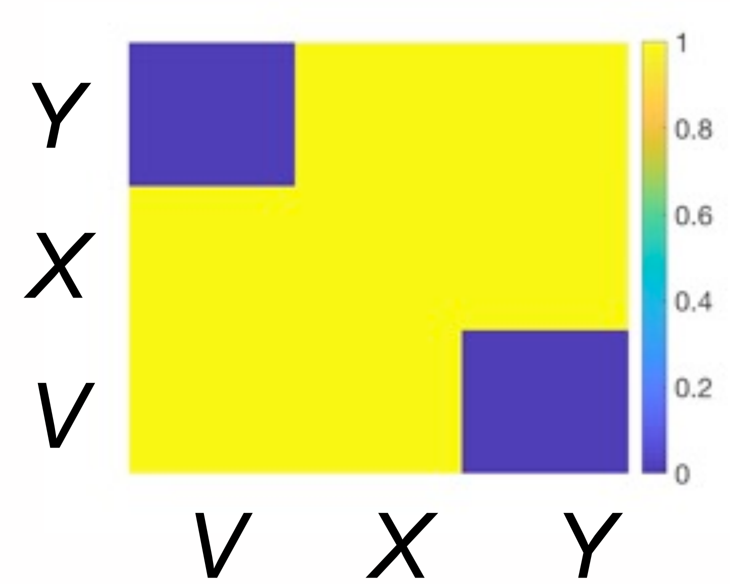
$I(V;X|Y) = 0 \quad I(V;Y|X) > 0$

$$\boxed{I(X;Y|V) = 0}$$

Simulation: Results

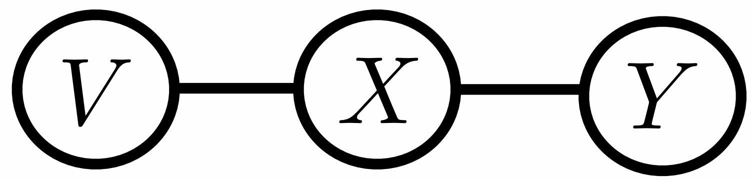
- Connectivity chain 1

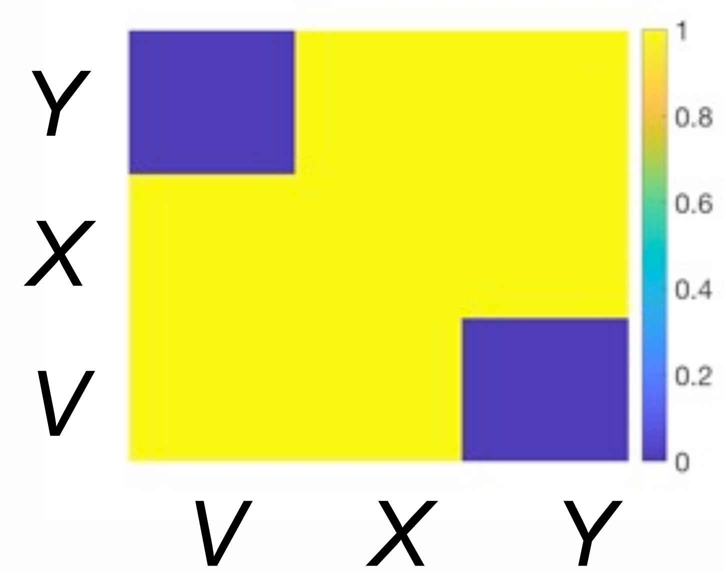




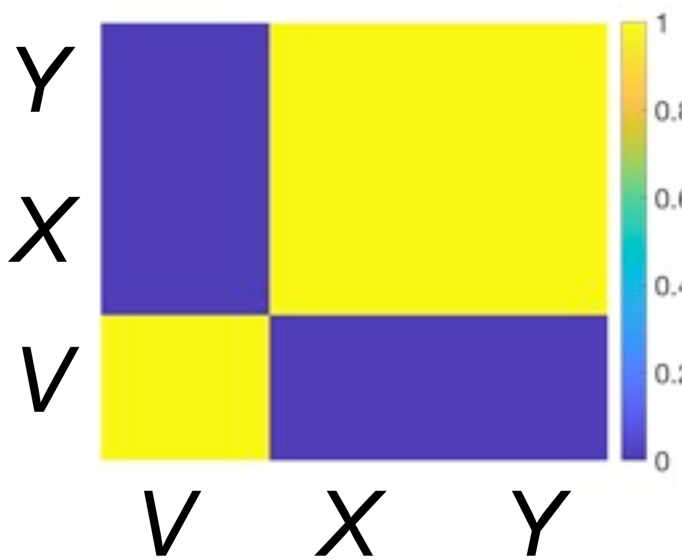
Simulation: Results

- Connectivity chain 1

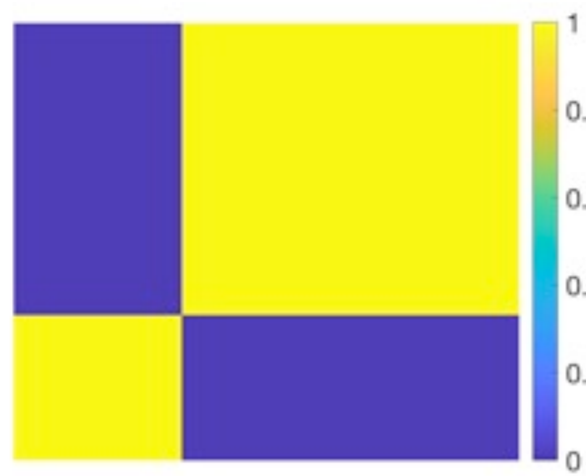




Pearson

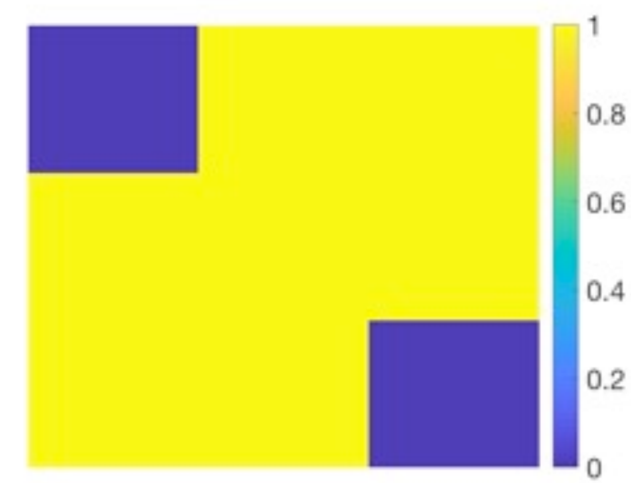


Partial



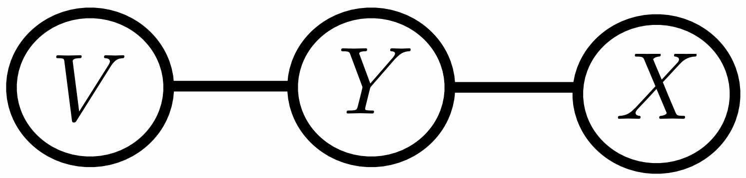
MI

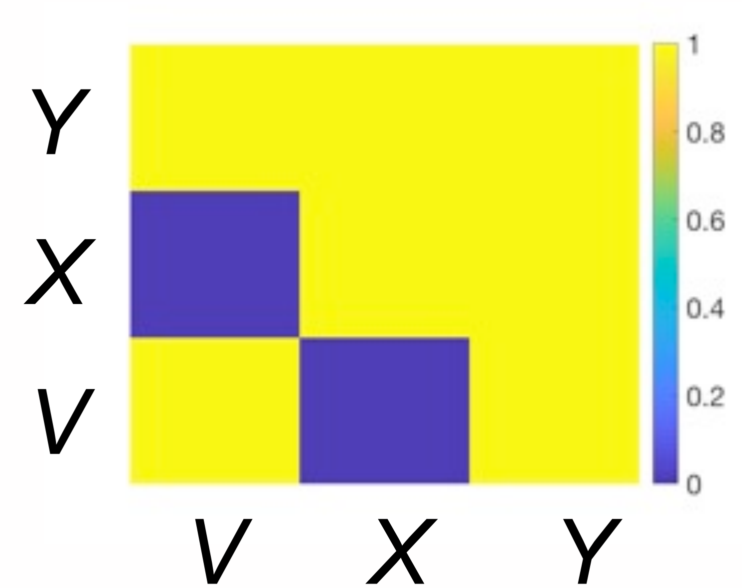
Cond MI



Simulation: Results

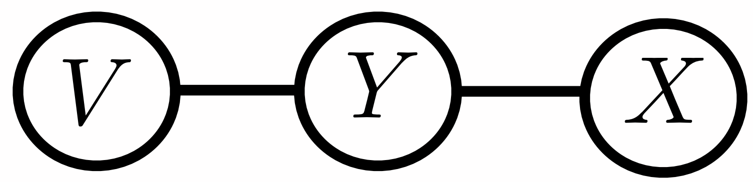
- Connectivity chain 2

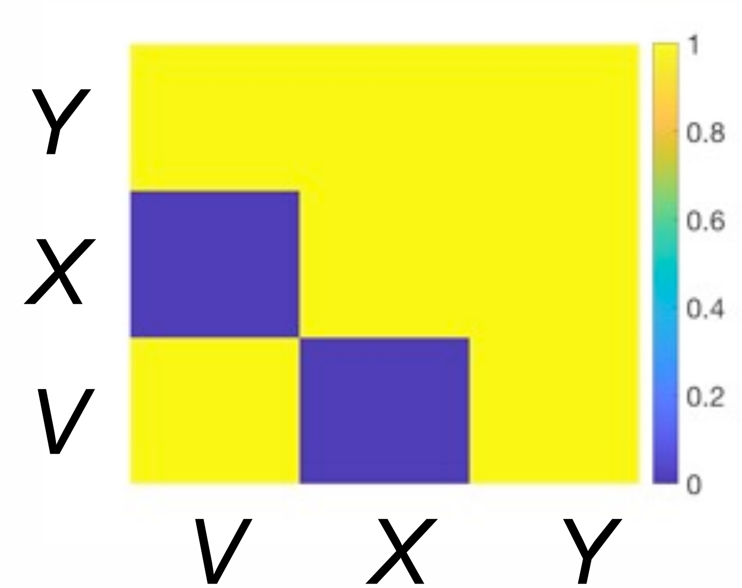




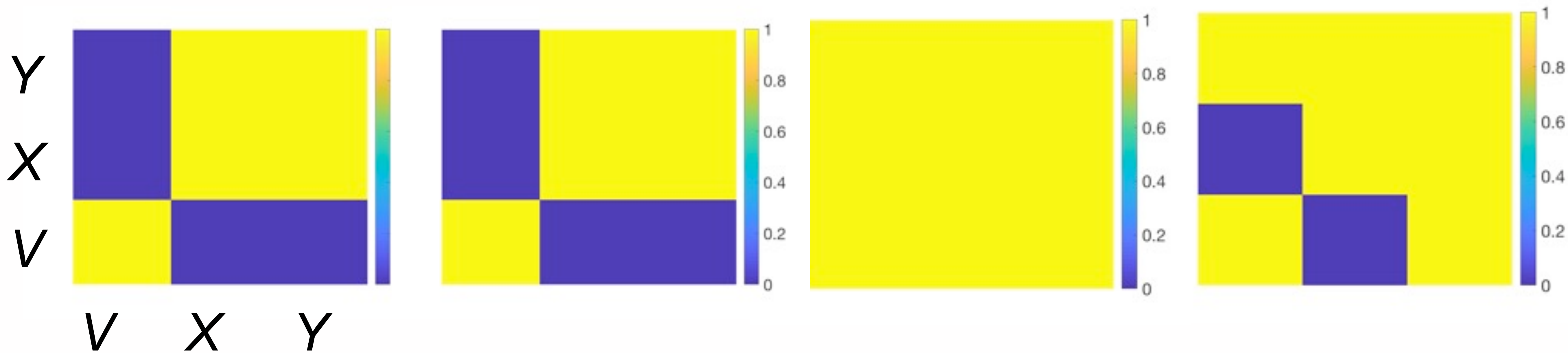
Simulation: Results

- Connectivity chain 2



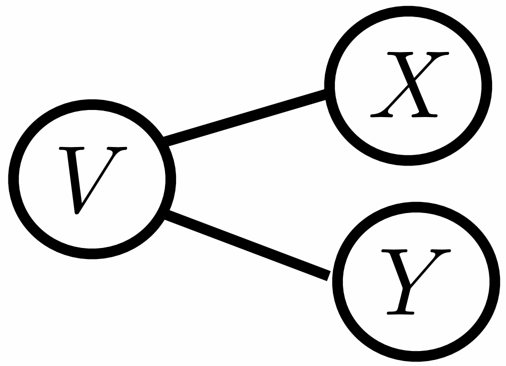


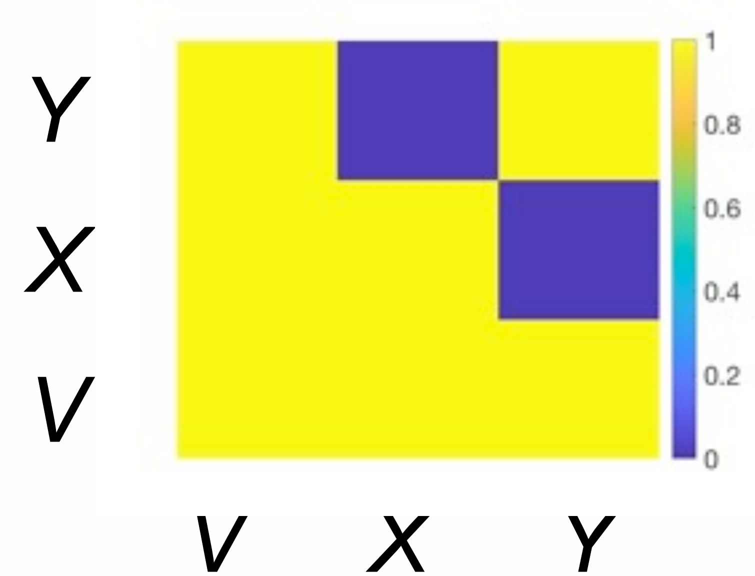
Pearson **Partial** **MI** **Cond MI**



Simulation: Results

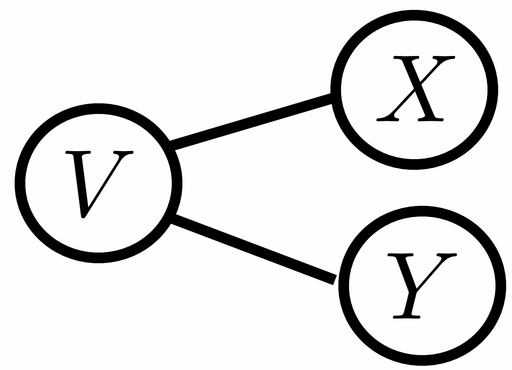
- Common input

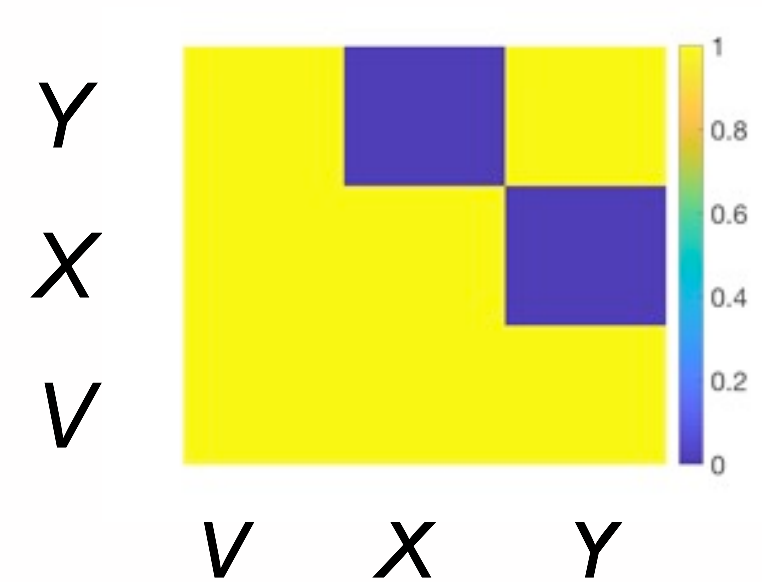




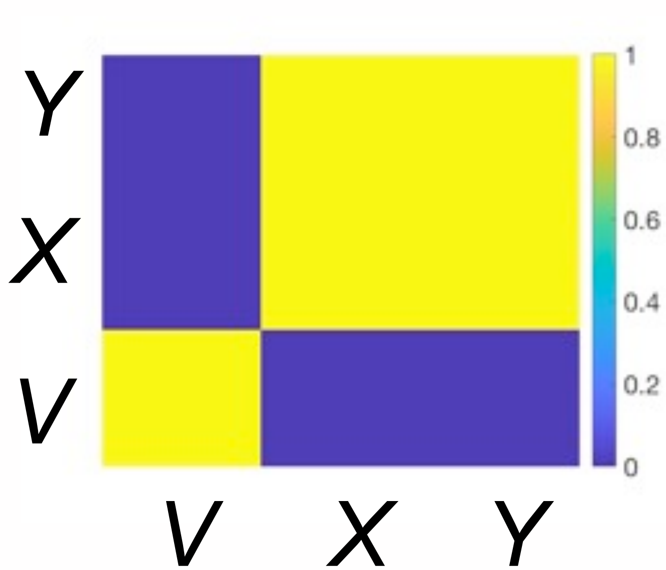
Simulation: Results

- Common input

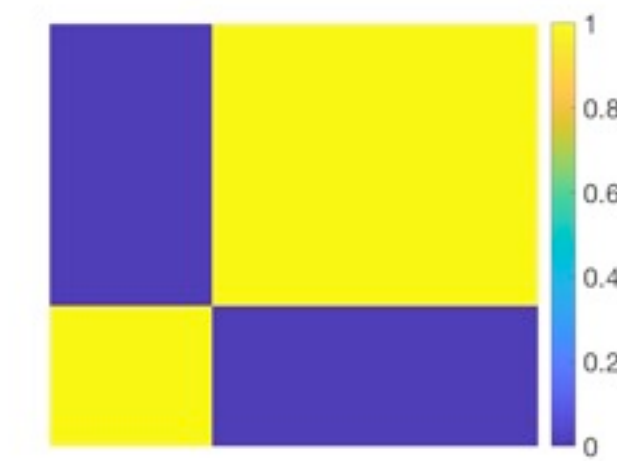




Pearson

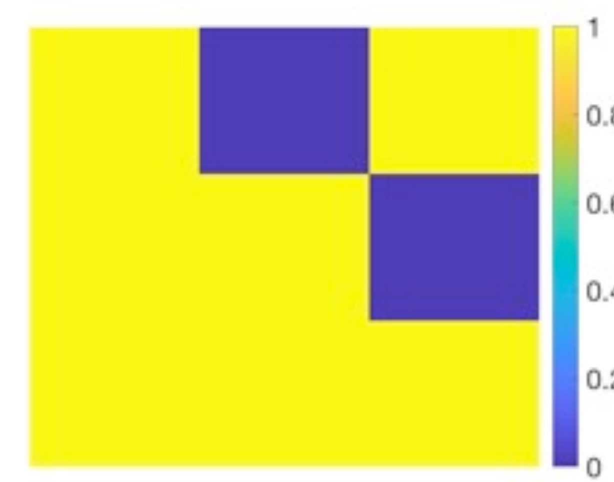


Partial



MI

Cond MI



Conclusions-Take home message

•**Probabilistic modeling** and **measures** (e.g., mutual information, etc.) offers a general framework to analyze connectivity problems with (1) as few **model assumptions** as possible, (2) incorporating variables of **different kind** and (3) can be formally represented by **graphs**.

Conclusions-Take home message

- **Probabilistic modeling** and **measures** (e.g., mutual information, etc.) offers a general framework to analyze connectivity problems with (1) as few **model assumptions** as possible, (2) incorporating variables of **different kind** and (3) can be formally represented by **graphs**.

- A priori **computationally expensive** but allows assumptions on **variable dependencies.**

Conclusions-Take home message

•**Probabilistic modeling** and **measures** (e.g., mutual information, etc.) offers a general framework to analyze connectivity problems with (1) as few **model assumptions** as possible, (2) incorporating variables of **different kind** and (3) can be formally represented by **graphs**.

•A priori **computationally expensive** but allows assumptions on **variable dependencies.**

•**Conditional mutual information** is a measure that correctly determines intrinsic connectivity pathways upon appropriate conditioning of variables.

Conclusions-Take home message

•**Probabilistic modeling** and **measures** (e.g., mutual information, etc.) offers a general framework to analyze connectivity problems with (1) as few **model assumptions** as possible, (2) incorporating variables of **different kind** and (3) can be formally represented by **graphs**.

•A priori **computationally expensive** but allows assumptions on **variable dependencies.**

•**Conditional mutual information** is a measure that correctly determines intrinsic connectivity pathways upon appropriate conditioning of variables.

•<u>Recommendation to use</u>: At the last stages of the analysis, after **dimensionality reduction**, non-linear model based on the remained few **interpretable variables**.

In next Chapters....

•**Directionality/Causality** analysis (Transfer entropy, etc.)**.**

•**Surrogate** tests for non-linear measures**.**

•Models integrating **scalar** and **time series** variables (e.g. model spike trains as binary sequences).

•Other **non-linear models** and measures (oscillators, synchronization) nor covered here.

Reading

- T.A. Cover and J. Thomas, “Elements of Information Theory”.

- DJC Mackay, “Information theory, inference and learning algorithms”.

How to use the non-linear code

1. Model generation

```
[vecV,vecX,vecY]=generate_model(P_source_max,epsilon,num_tria
ls,sequence_length,num_model, encoding)
```

Inputs: **P_source_max**: spiking probability for selected stimulus outcome.
epsilon: crossover probability between spikes and ‘0’ of *X* and *Y*.
num_trials: number of trials
sequence length: spike train length of *X* and *Y*.
num_model: 1 (chain 1), 2 (chain 2), 3 (common input).
encoding: linear /non linear:

Outputs: Column vectors with **observables** from the 3 variables
pval_cond,MI_sur_cond]=conditional_mutual_information(vecX, vecY, vecZ,

How to use the non-linear code

2. Connectivity estimation

- Mutual information

```
[MI0, pvalue, MI_sur,P_X_Y, Perm_mat]=mutual_information(vecX,
vecY, Nsur)
```

- Conditional mutual information

```
[MI, pvalue, MI_sur, MI_cond,
pval_cond,MI_sur_cond]=conditional_mutual_information(vecX,
vecY, vecZ, max_sur
```

- Quantization of the conditioned variable

```
[vecX2, vecY2]=equal_bin_quantization(vecX, vecY, num_levels,
num_trials)
```

How to use the non-linear code

3. Model Results

Figure 1

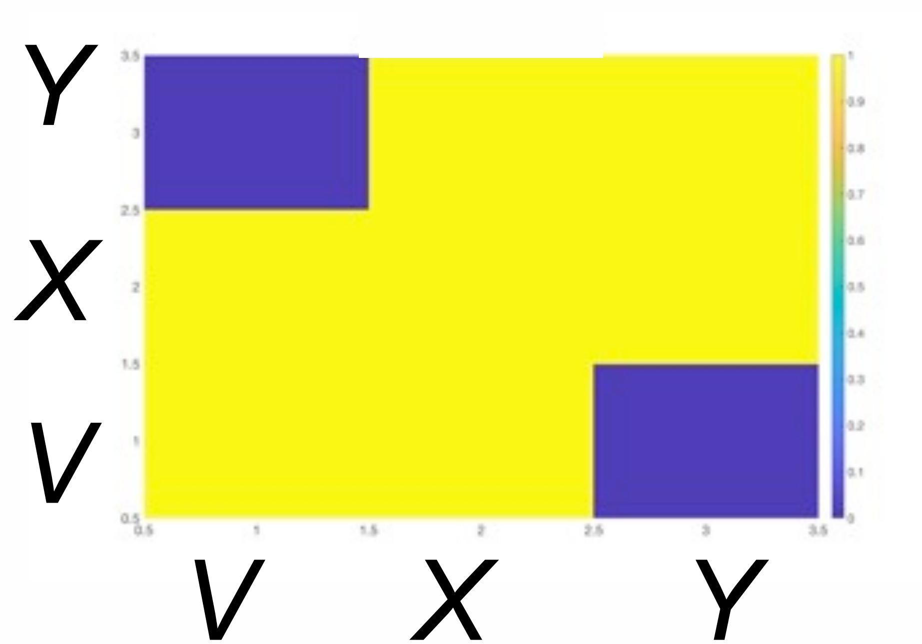


Figure 2

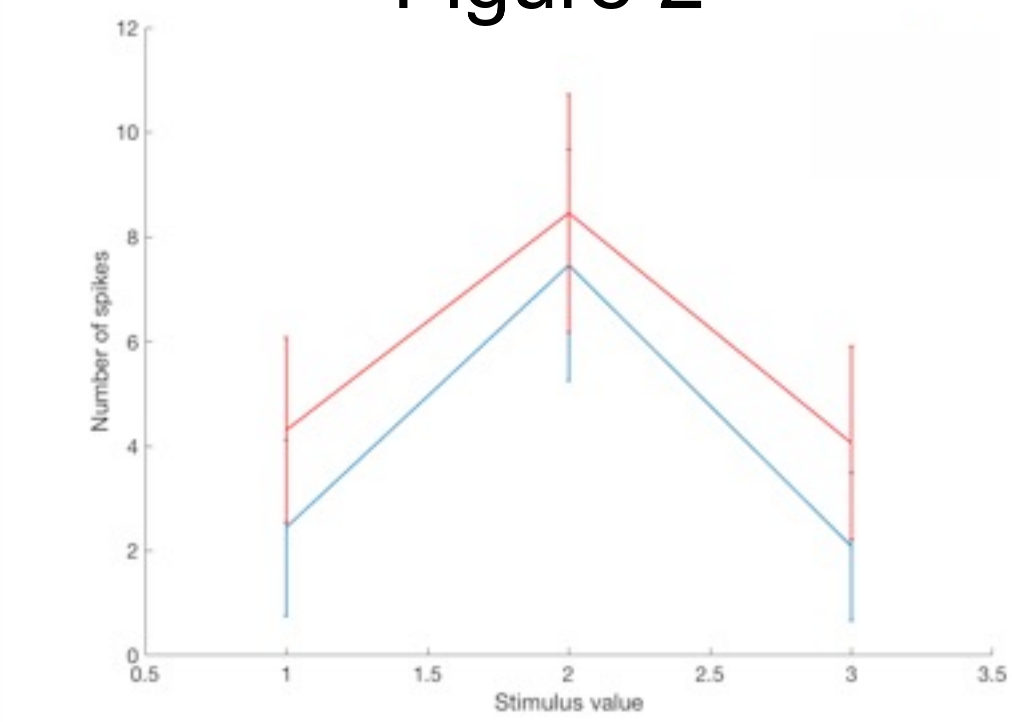
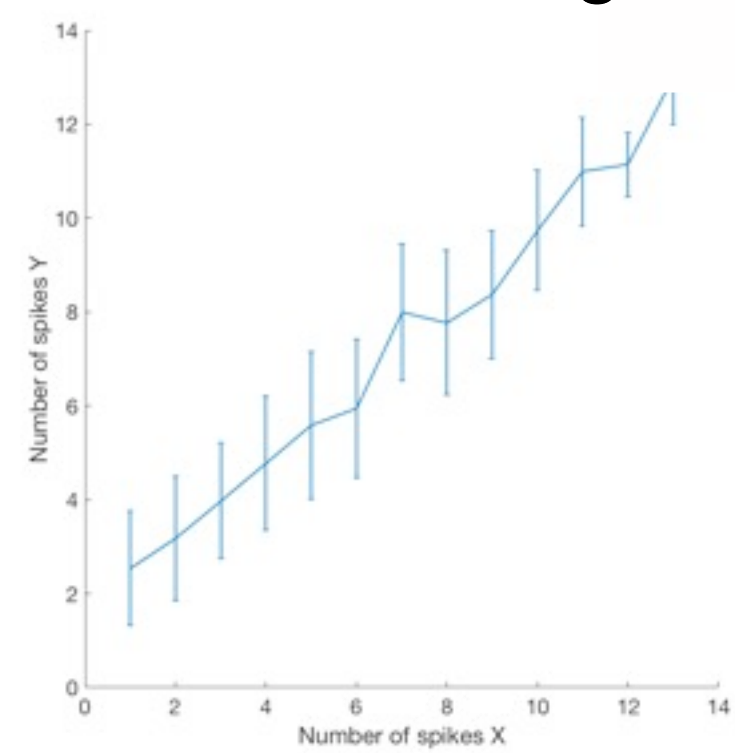


Figure 3



How to use the non-linear code

4. Result Figures

Figure 4
(Unweighted)

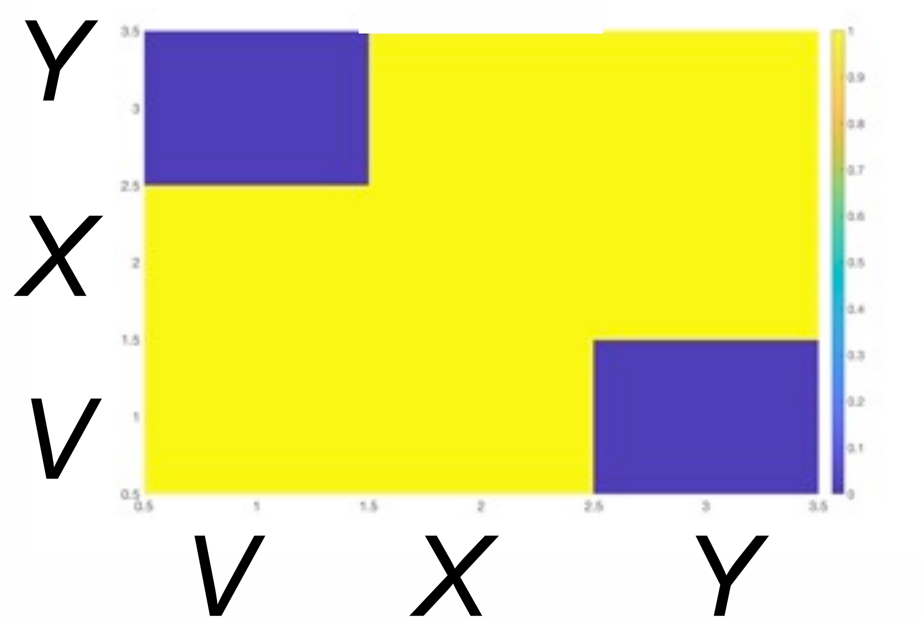


Figure 5
(Weighted)