Success Concepts for Causal Discovery

Konstantin Genin

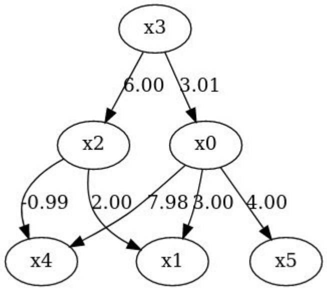
Group Leader: "Epistemology and Ethics of ML"
Cluster of Excellence: "Machine Learning for Science"

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Joint Work

with **Conor Mayo-Wilson**
Department of Philosophy
University of Washington

The LiNGAM Model



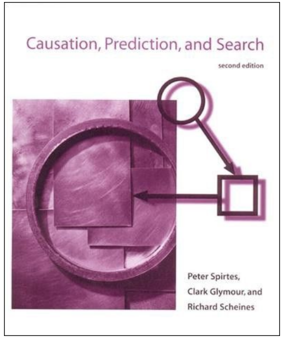
Theorem (Shimizu et al., 2006)**.** When

1. noise terms are **independent** and **non-**Gaussian,
2. functional relationships are **linear** and **a-cyclic** and
3. there are no unobserved confounders,

it is possible to converge (pointwise) to the **DAG** generating the data.

Shimizu, Shohei, Patrik O. Hoyer, Aapo Hyvärinen, Aapo, and Antti Kerminen. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research* 7, no. 72 (2006): 2003–30.

The Linear Gaussian Model



Theorem (Spirtes et al., 2001)**.** When

1. noise terms are **independent** and **Gaussian**,
2. functional relationships are **linear** and **a-cyclic** and
3. there are no unobserved confounders,

it is possible to converge (pointwise) to the **Markov equivalence class** of the DAG generating the data.

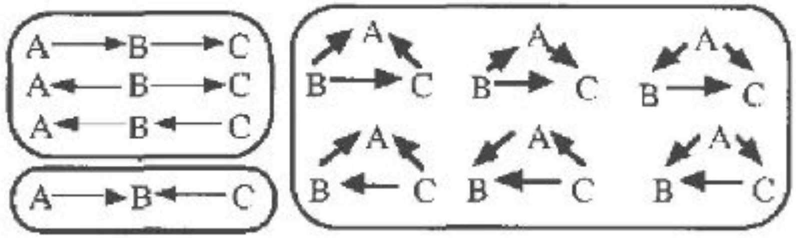


Figure 2: Three Acyclic Markov Equivalence Classes

LiNGAM + Confounding - Unfaithfulness

Theorem (Salehkaleybar et al., 2020)**.** When

1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** unobserved confounders, but
4. there are no cancelling paths (**faithfulness**),

then causal **ancestry** relationships between **observed** variables are **identified**.

Salehkaleybar, Saber, et al. (2020) "Learning Linear Non-Gaussian Causal Models in the Presence of Latent Variables." *Journal of Machine Learning Research* 21.39: 1-24.

Hoyer, Patrik O., et al. (2008) "Estimation of causal effects using linear non-Gaussian causal models with hidden variables." *International Journal of Approximate Reasoning* 49.2: 362-378.

Pitfalls of Pointwise

But identifiability does not imply the existence of discovery algorithms.

Pitfalls of Pointwise

Moreover, **pointwise convergence** is compatible with all kinds of short run behavior.

Pitfalls of Pointwise

If noise is Gaussian, causal conclusion can **flip** arbitrarily often as data accumulate.

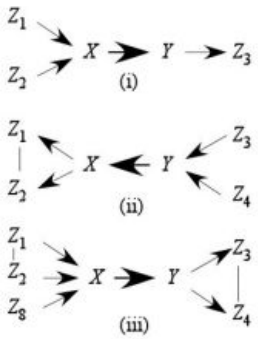


Figure 1: Causal Flips

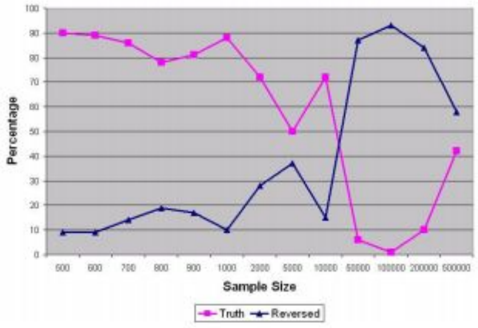


Figure 5: FCI Algorithm

Kelly, Kevin T, and Conor Mayo-Wilson (2010). “Causal Conclusions That Flip Repeatedly and Their Justification,” Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010). https://arxiv.org/abs/1203.3488v1

Uniform Convergence is Impossible

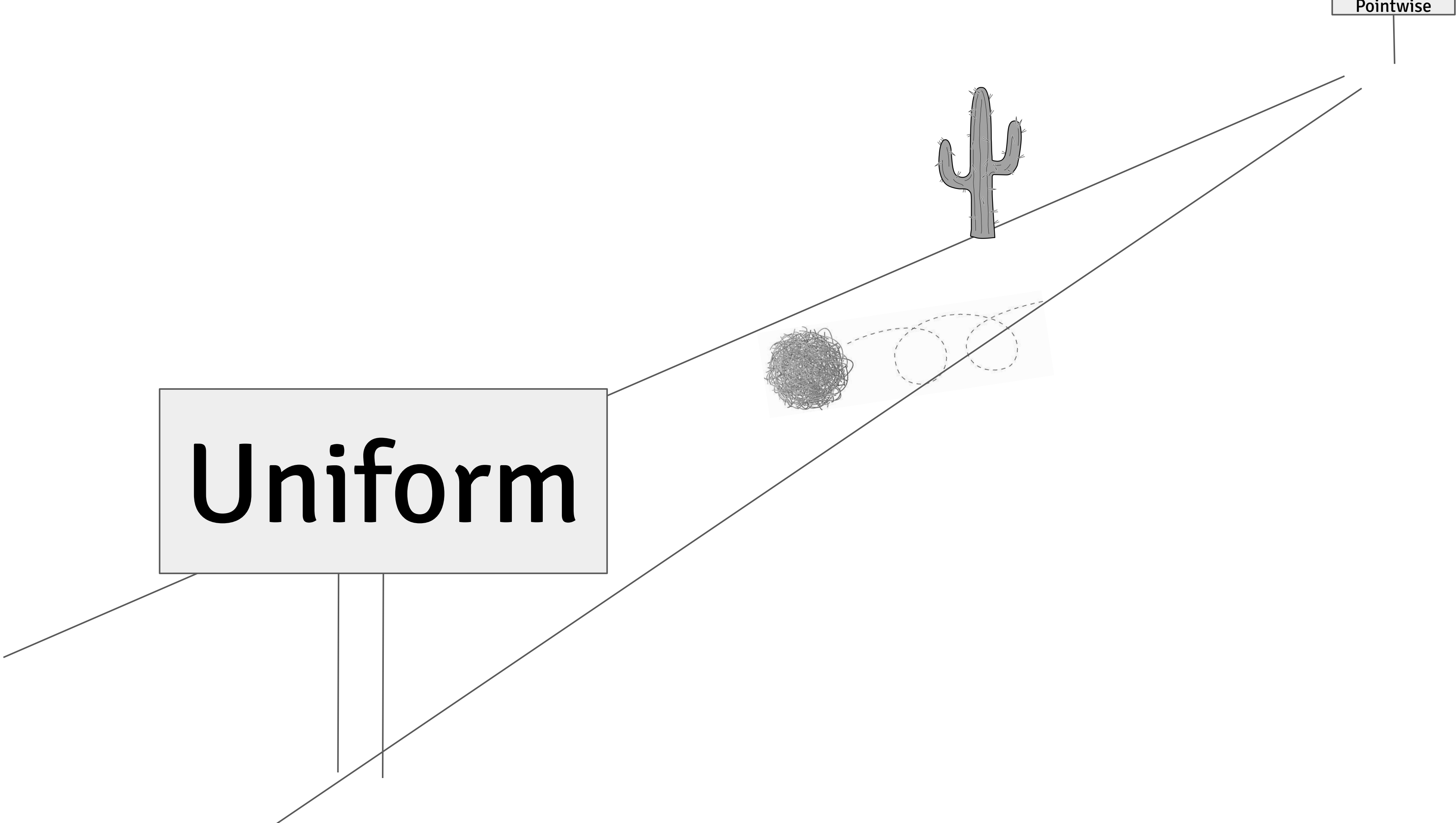
But **uniform** convergence to the true DAG is provably **impossible** in the LiNGAM framework.

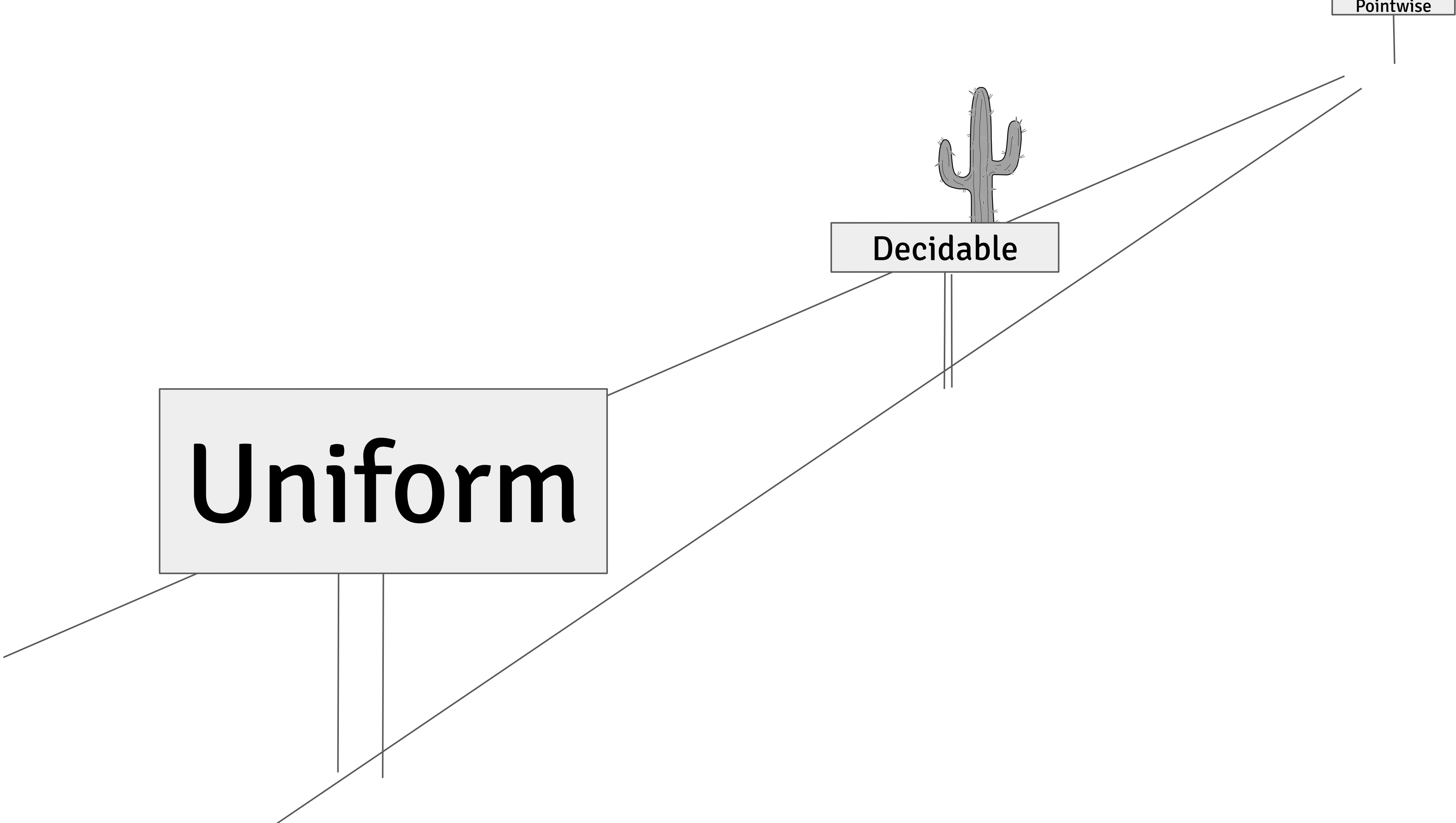
Uniform Convergence is Impossible

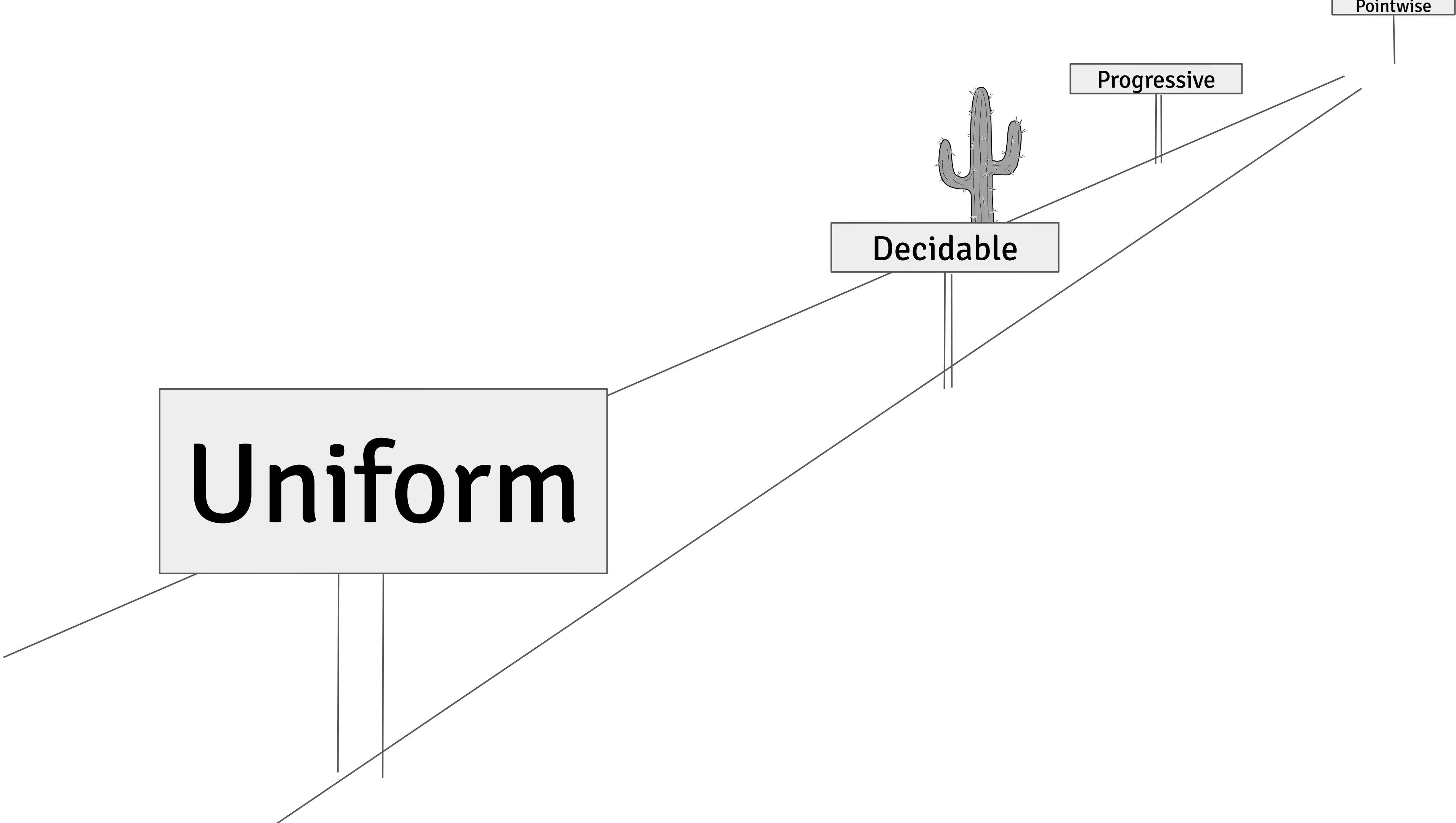
But **uniform** convergence to the true DAG is provably **impossible** in the LiNGAM framework.

And assumptions strong enough make uniform convergence feasible are not plausible.

Uhler, Caroline, et al. "Geometry of the faithfulness assumption in causal inference." *The Annals of Statistics* (2013): 436-463.


Pointwise
Uniform


Pointwise
Decidable
Uniform


Pointwise
Progressive
Decidable
Uniform

To Do

1. Define the success concepts.
2. Provide (topological) criteria for achieving the success concepts.
3. Apply to the case of LiNGAM with and without confounders.
4. Introduce a LiNGAM variant: the “Flamingo”.

I ● Topology

Roughly: qualitative geometry that abstracts from metrical concepts in favor of qualitative notions of **separation** or **arbitrary** closeness.

Often: the study of geometric properties preserved by stretching but not “cutting” and “gluing”

I Topology

- A point *w* is a **limit point** of a region *A* if there are points in *A* getting arbitrarily close to *w.*
- Two regions *A* and *B* are well separated if neither contains limit points of the other.
- “Cutting” separates regions that weren’t separated and “gluing” creates limit points that weren’t there before.

I Topology

- The **topological closure** of *A*, written cl(*A*), is the result of adding all of the limit points of *A* to *A.*

Why does topology matter?

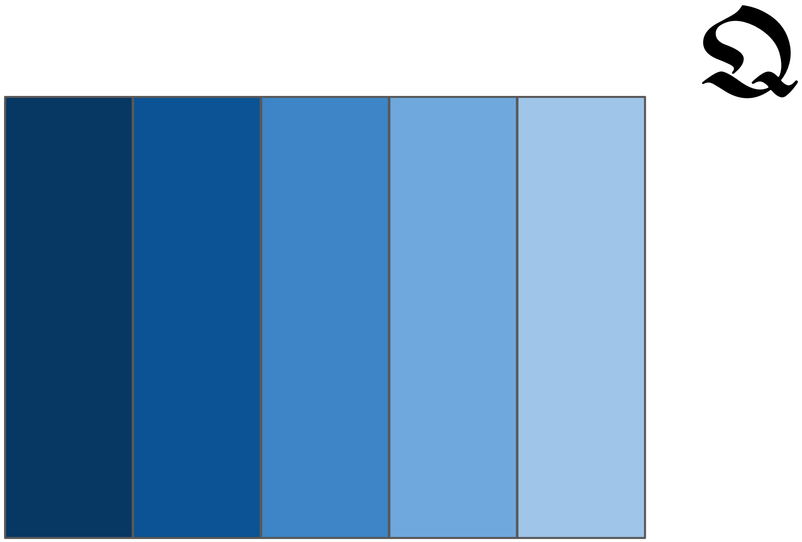
Qualitative relations of separation are important for understanding how “well resolved” possibilities are by data and, therefore, how hard causal discovery problems really are.

Statistical Questions

Let $\mathcal{M}$ be a set of causal models, each a potential data-generating mechanism.

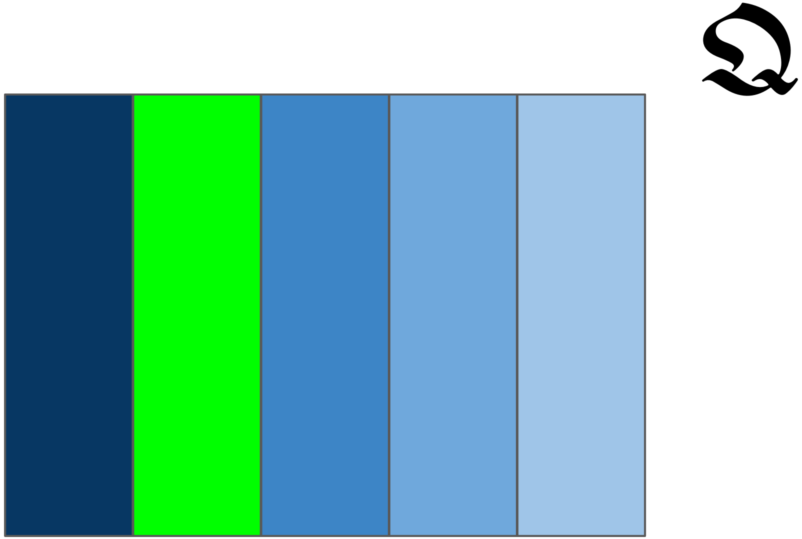
Statistical Questions

A **question** $\mathfrak{Q}$, partitioning $\mathscr{M}$ into a countable set of **answers.**



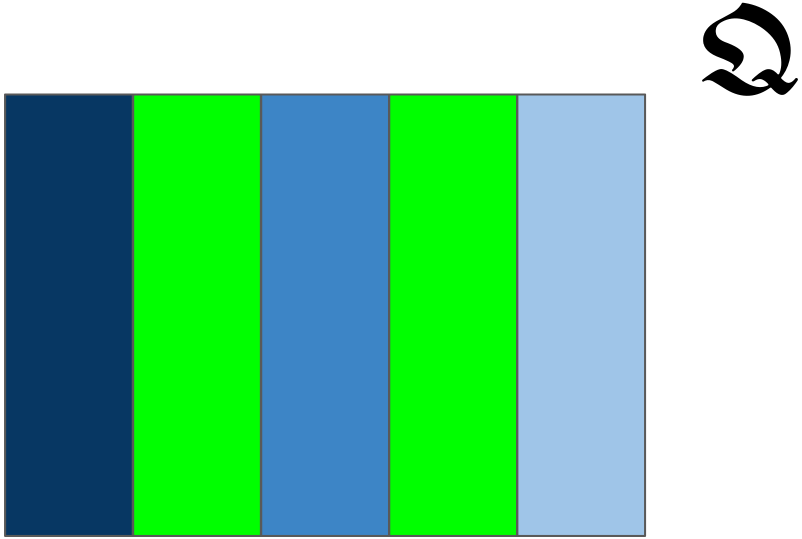
Statistical Questions

A **relevant response** is a **union** of answers.



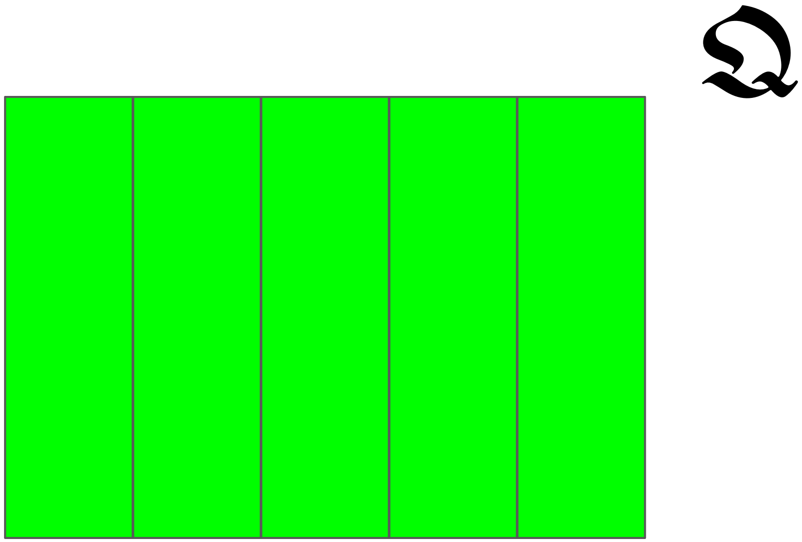
Statistical Questions

A **relevant response** is a **union** of answers.



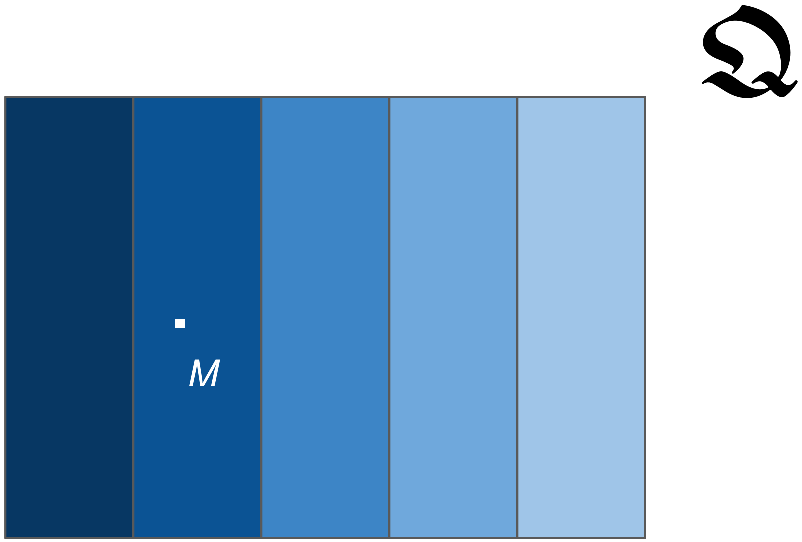
Statistical Questions

A **relevant response** is a **union** of answers.



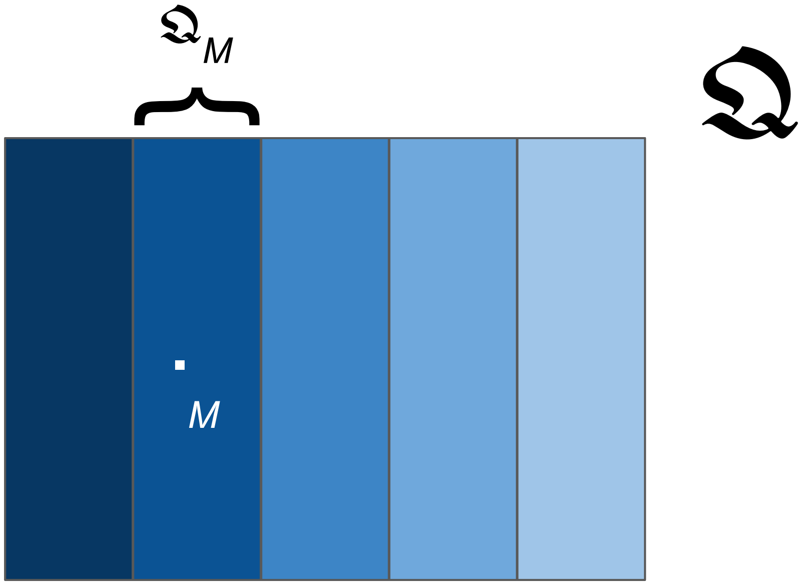
Statistical Questions

If $M \in \mathcal{M}$,



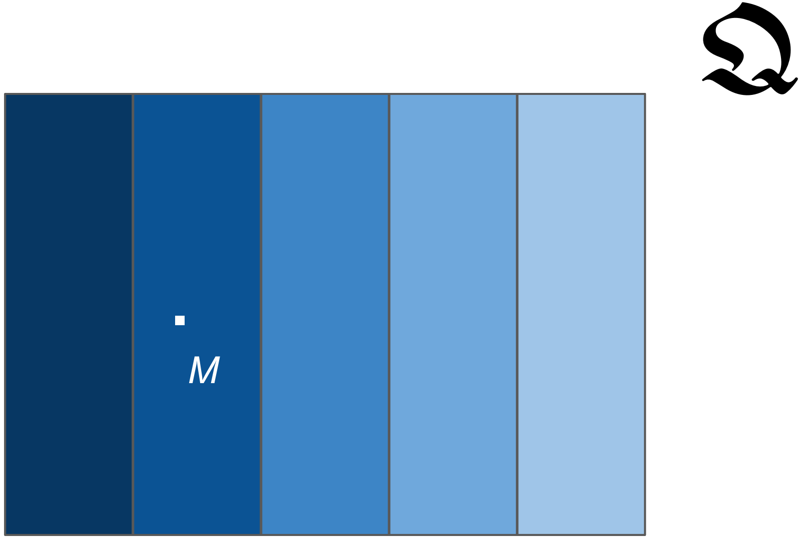
Statistical Questions

If $M \in \mathcal{M}$, let $\mathfrak{Q}_M$ be the answer true in M.



Statistical Questions

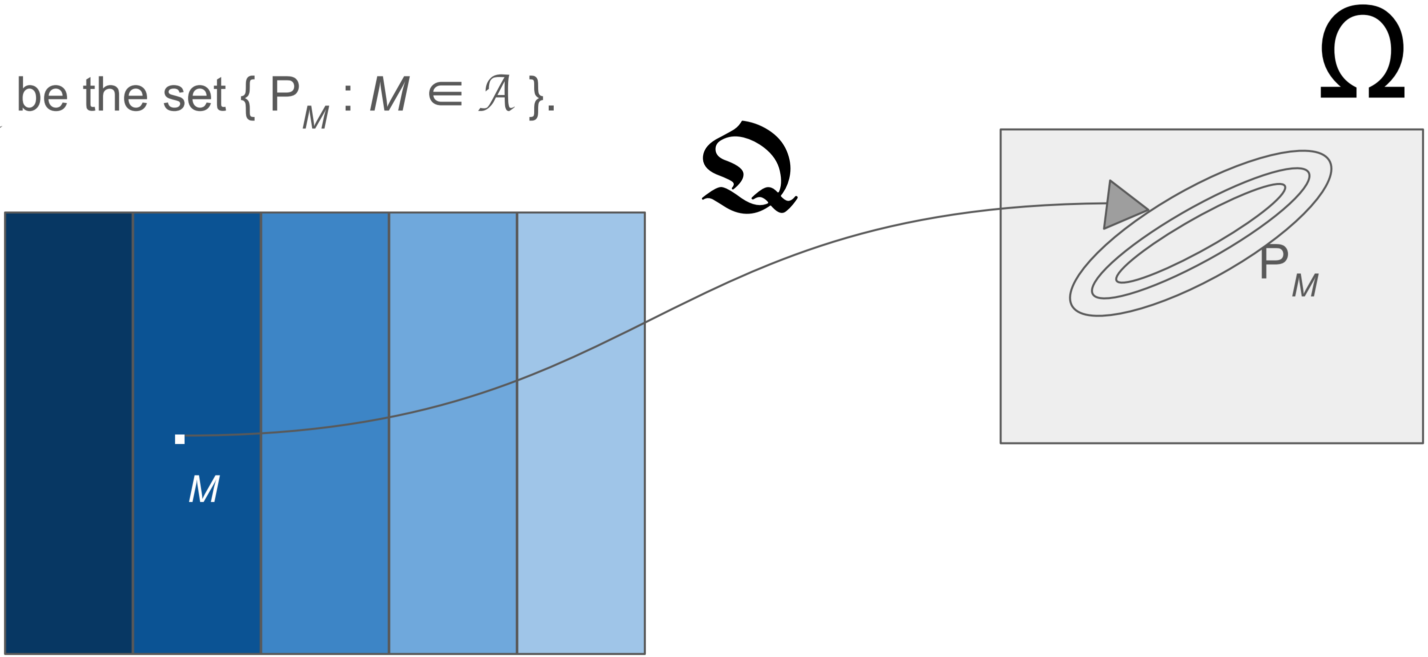
If $M \in \mathcal{M}$,



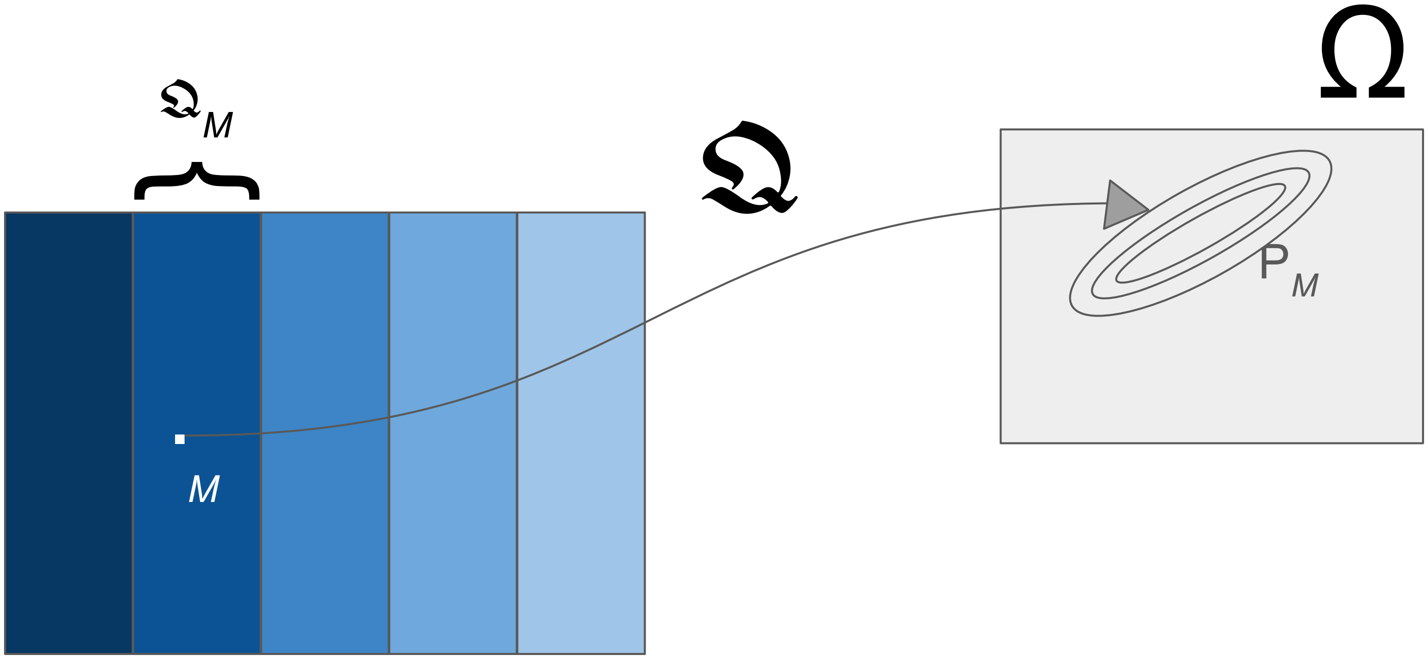
Statistical Questions

If $M \in \mathcal{M}$, let P_M be the distribution induced by M over observables.

If $\mathcal{A} \subseteq \mathcal{M}$, let $P_{\mathcal{A}}$ be the set $\{ P_M : M \in \mathcal{A} \}$.



Statistical Questions



Weak Topology

- The distributions P_n converge in the weak topology to P if for all "nice" events *A*, $P_n(A) \rightarrow P(A)$.

- *A* is nice if $P(\partial A) = 0$.

- Convergence in the weak topology is equivalent to convergence in distribution.

Statistical Methods

A set of measurable functions (T_n) is a **method** if each one is a function from samples of size *n* to **relevant responses** (unions of answers).

Note: a method can **suspend judgment** by outputting $\bigcup \mathfrak{Q}$.

Uniform Decidability

A method (T_n) **uniformly decides** $\mathfrak{Q}$ iff for all $\varepsilon>0$ there is a sample size n such that for all $M \in \mathcal{M}$ and

- $P_M(T_m = \mathfrak{Q}_M) > 1- \varepsilon$ for all $m \geq n$.

Topological Criterion for Uniform Decidability

Theorem.

A question is uniformly decidable only if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

- cl($P_{\mathcal{A}}$) is disjoint from cl($P_{\mathcal{B}}$).

Decidability in the Limit

A method (T_n) **decides $\mathfrak{Q}$ in the limit** iff for all M $\in \mathcal{M}$,

- $P_M(T_n = \mathfrak{Q}_M) \longrightarrow 1$ as $n \longrightarrow \infty$.

Decidability in the Limit

A method (T_n) **decides $\mathfrak{Q}$ in the limit** iff for all M $\in \mathcal{M}$,

- $P_M(T_n = \mathfrak{Q}_M) \longrightarrow 1$ as $n \longrightarrow \infty$.

A question $\mathfrak{Q}$ is **decidable in the limit** if some method decides it in the limit.

Topological Criterion for Limiting Decidability

Theorem. (Dembo & Peres, 1994)

A question is decidable in the limit if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

- $P_{\mathcal{A}}$ is disjoint from $P_{\mathcal{B}}$;
- $P_{\mathcal{A}}$ is a countable union of closed sets in the weak topology.

Dembo, Amir, and Yuval Peres (1994). “A Topological Criterion for Hypothesis Testing.” *Annals of Statistics* 22(1): 106–17.
https://doi.org/10.1214/aos/1176325360.

Decidability

A method (T_n) is an **α-decision procedure** for $\mathfrak{Q}$ iff it decides $\mathfrak{Q}$ in the limit and

- for all sample sizes *n*, $P_M(\mathfrak{Q}_M \nsubseteq T_n) < \alpha$.

A question $\mathfrak{Q}$ is statistically **decidable** iff it has an α-decision procedure for all α > 0.

Decidability

A method (T_n) is an **α-decision procedure** for $\mathfrak{Q}$ iff it decides $\mathfrak{Q}$ in the limit and

- for all sample sizes *n*, $P_M(\mathfrak{Q}_M \not\subseteq T_n) < \alpha$.

A question $\mathfrak{Q}$ is statistically **decidable** iff it has an α-decision procedure for all $\alpha > 0$.

Note: It may be that $P_M(T_n = \bigcup \mathfrak{Q}) \approx 1$ for arbitrarily large *n*.

Topological Criterion for Decidability

Theorem. (Genin & Kelly, 2017)

A question is decidable if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

$P_{\mathcal{A}}$ is disjoint from the (weak topology) closure of $P_{\mathcal{B}}$.

Genin, Konstantin, and Kevin T. Kelly. (2017) "The Topology of Statistical Verifiability." *Electronic Proceedings in Theoretical Computer Science* 251: 236–50. https://doi.org/10.4204/EPTCS.251.17.

Three Varieties of Decidability

	Output probably **correct** at every sample size.	Output probably **informative** after known sample size.	Output probably **correct** & **informative** after some (potentially unknown) sample size.
Uniformly Decidable	✔	✔	✔
Decidable	✔	✖	✔
Decidable in the Limit	✖	✖	✔

Three Varieties of Decidability

	Output probably **correct** at every sample size.	Output probably **informative** after known sample size.	Output probably **correct** & **informative** after some (potentially unknown) sample size.
Uniformly Decidable	✔	✔	✔
Decidable	✔	✖	✔
Decidable in the Limit	✖	✖	✔

Progressive Solvability

A method (T_n) is an **α-progressive solution** for $\mathfrak{Q}$ iff for all M $\in \mathcal{M}$,

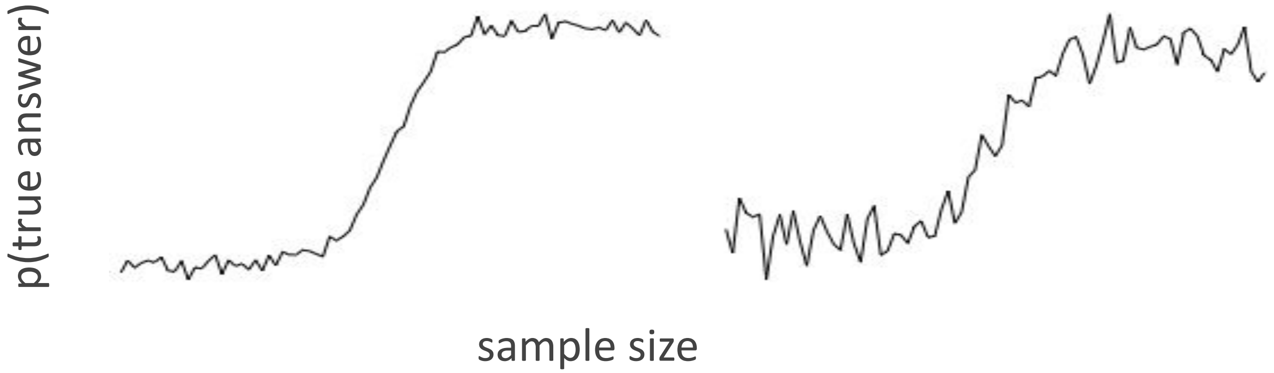
- (T_n) decides $\mathfrak{Q}$ in the limit;
- $P_M(T_{n2} = \mathfrak{Q}_M) + \alpha > P_M(T_{n1} = \mathfrak{Q}_M)$ for $n_2 > n_1$.

A question $\mathfrak{Q}$ is **progressively solvable** if it has an α-progressive solution for every α>0.

Progressive Methods

A **method** for answering a scientific question is α-**progressive** iff

- the **chance** that it outputs the **true** answer **never drops** by more than α.



Topological Criterion for Progressive Solvability

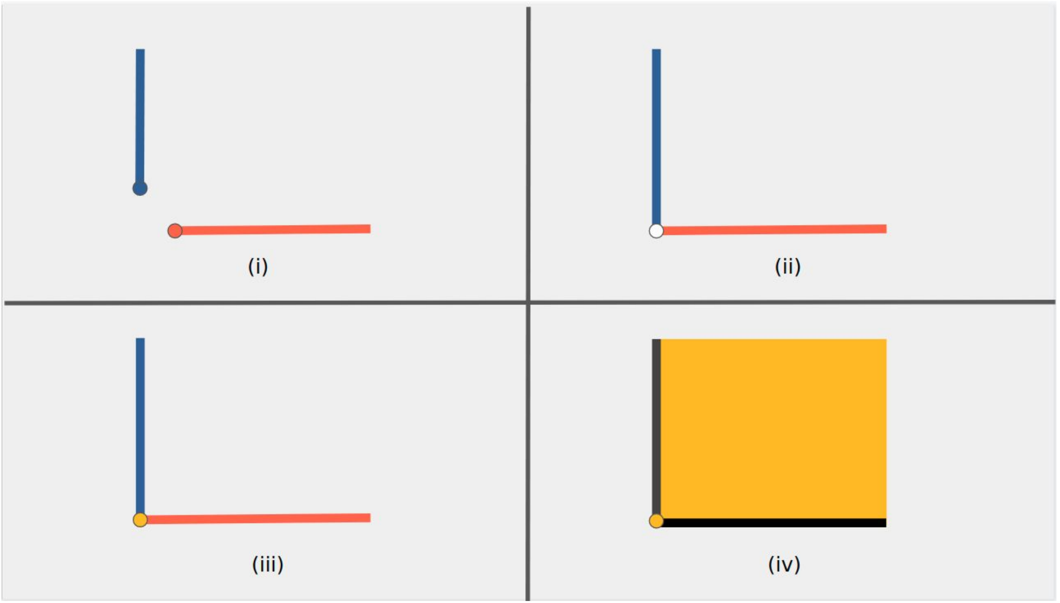
Theorem. (Genin, 2018)

A question is progressively solvable if there exists an enumeration $\mathcal{A}_1, \mathcal{A}_2, \ldots$ of the answers to $\mathfrak{Q}$ s.t. $\mathcal{A}_j \cap cl(\mathcal{A}_i)=\emptyset$ for $i < j$.

Genin, Konstantin, and Kevin T. Kelly. (2017) “The Topology of Statistical Verifiability.” *Electronic Proceedings in Theoretical Computer Science* 251: 236–50. https://doi.org/10.4204/EPTCS.251.17.

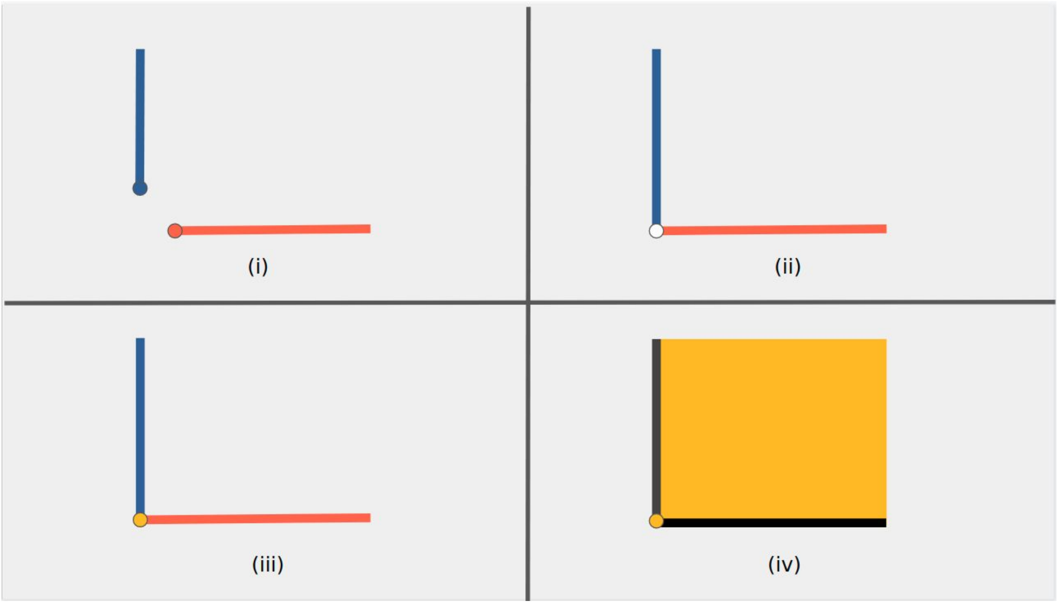
An illustration

Suppose we are interested in the (absolute) bias of two coins, one **red** and one **blue**.



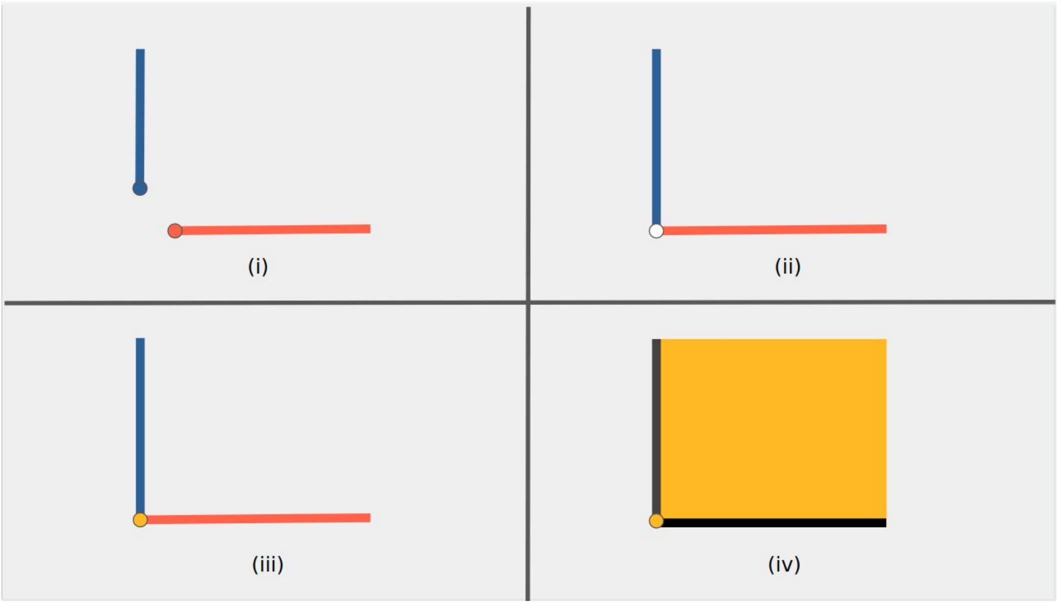
An illustration

(i) background knowledge determines that **exactly** 1 coin is biased and there is a non-zero lower bound on the degree of bias, if any.



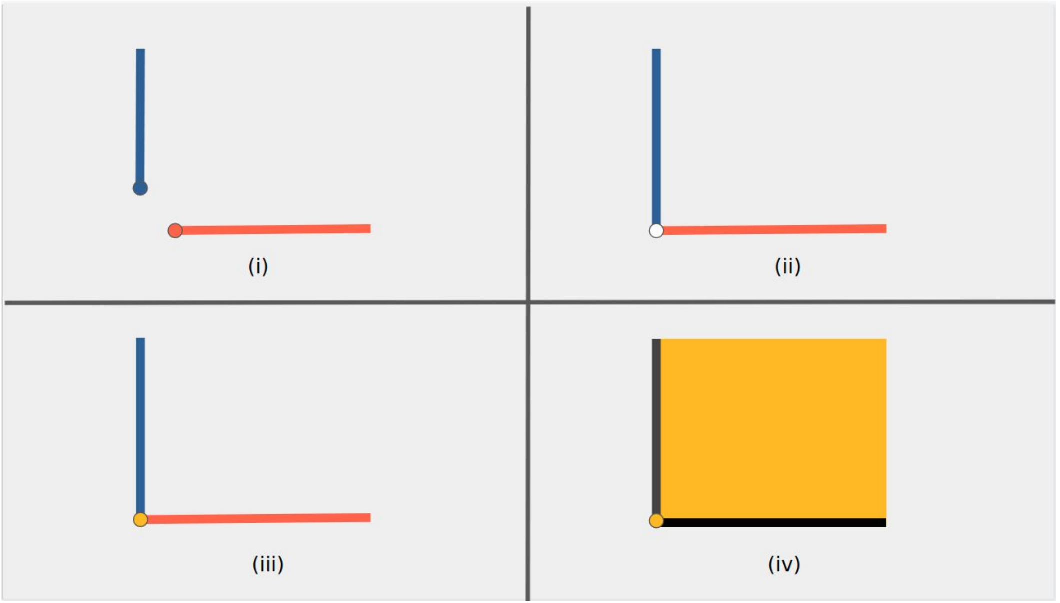
An illustration

This is fortunate situation in which there exists a uniform procedure deciding between the **red** and **blue** hypotheses.



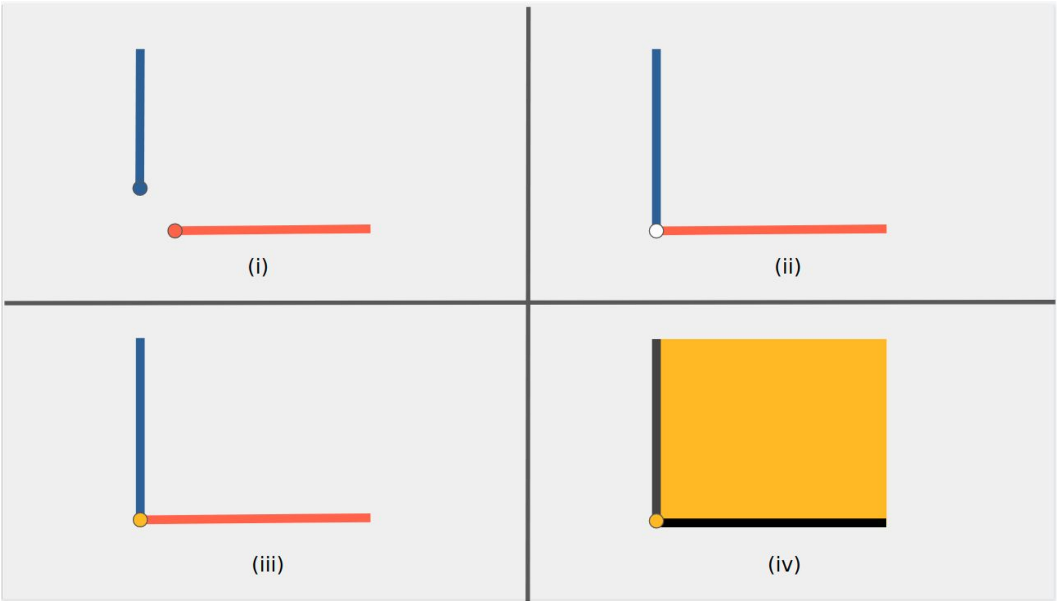
An illustration

(i) cl(**blue**) ∩ cl(**red**) = ∅.



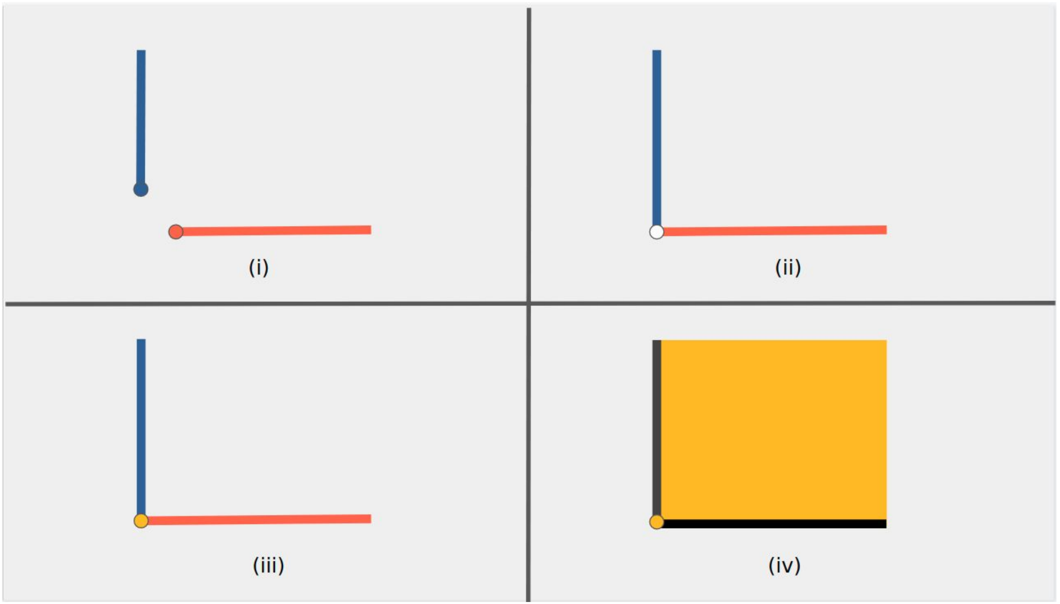
An illustration

(ii) background knowledge determines that exactly 1 coin is biased but the bias can be **arbitrarily close** to zero.



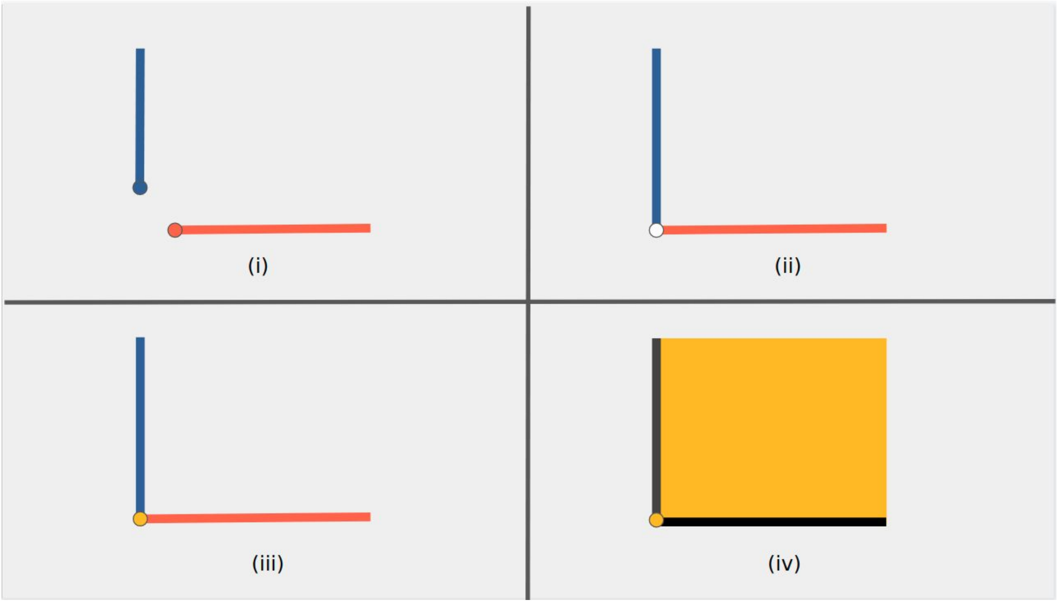
An illustration

Uniform decision procedures no longer exist.



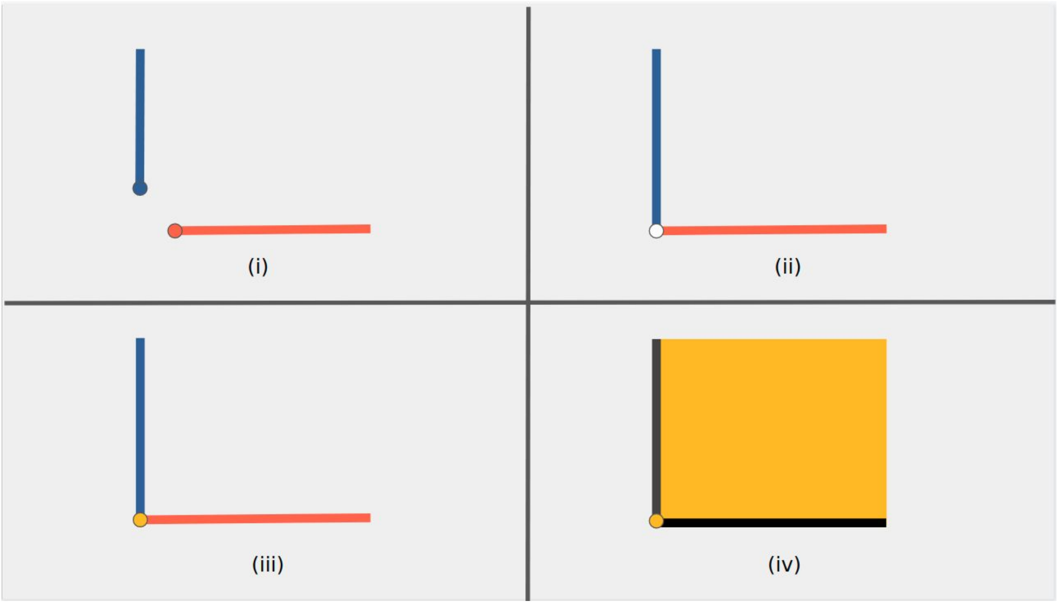
An illustration

But a decision procedure does exist: suspend judgement until a confidence interval rules out either **red** or **blue**.



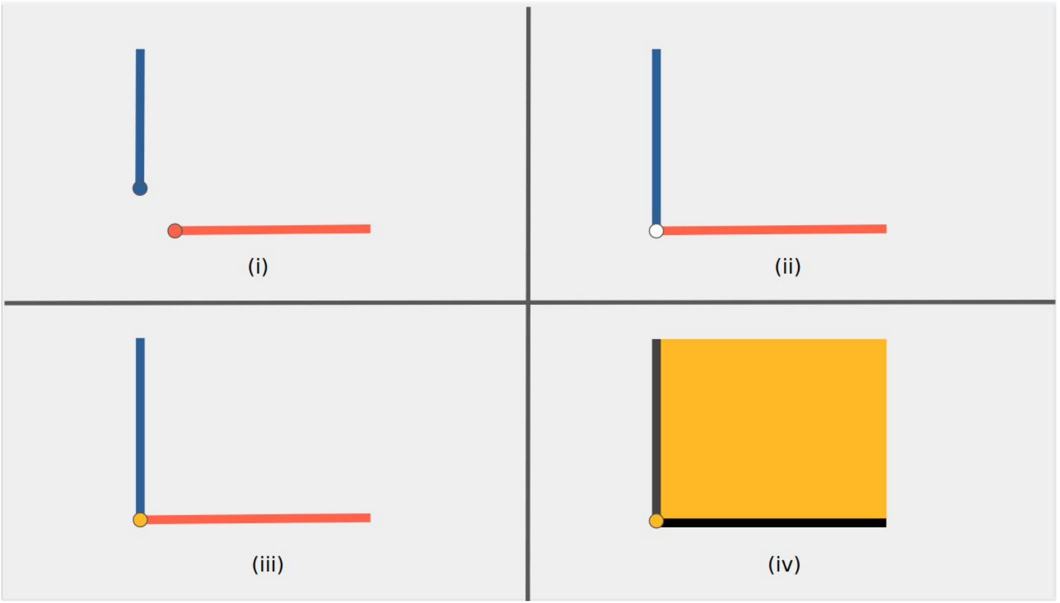
An illustration

(ii) cl(**blue**) ∩ **red** = Ø and cl(**red**) ∩ **blue**= Ø.



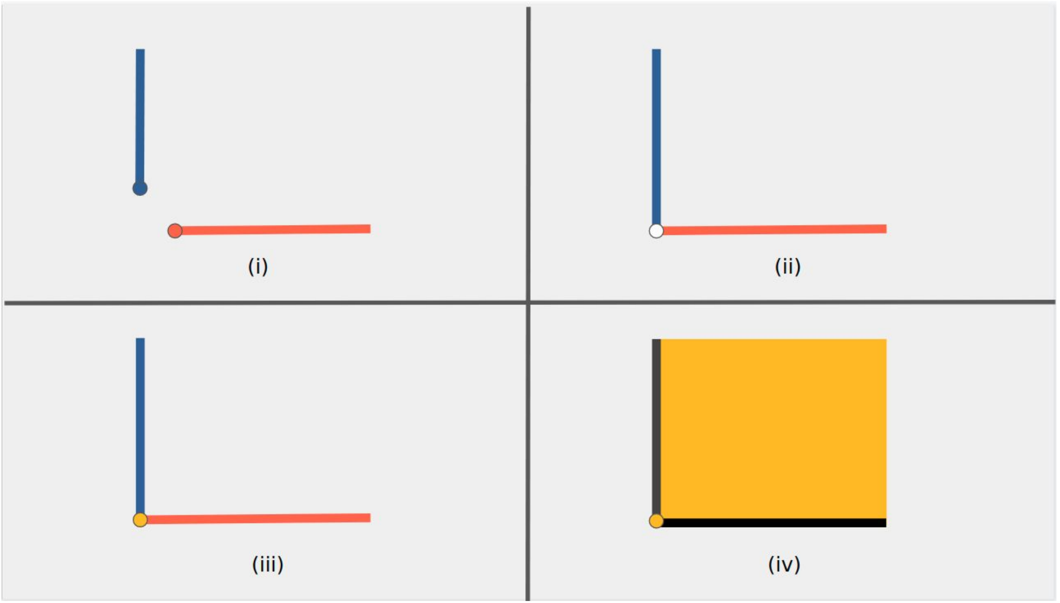
An illustration

(iii) background knowledge determines that **at most** 1 coin is biased and the bias can be arbitrarily close to zero.



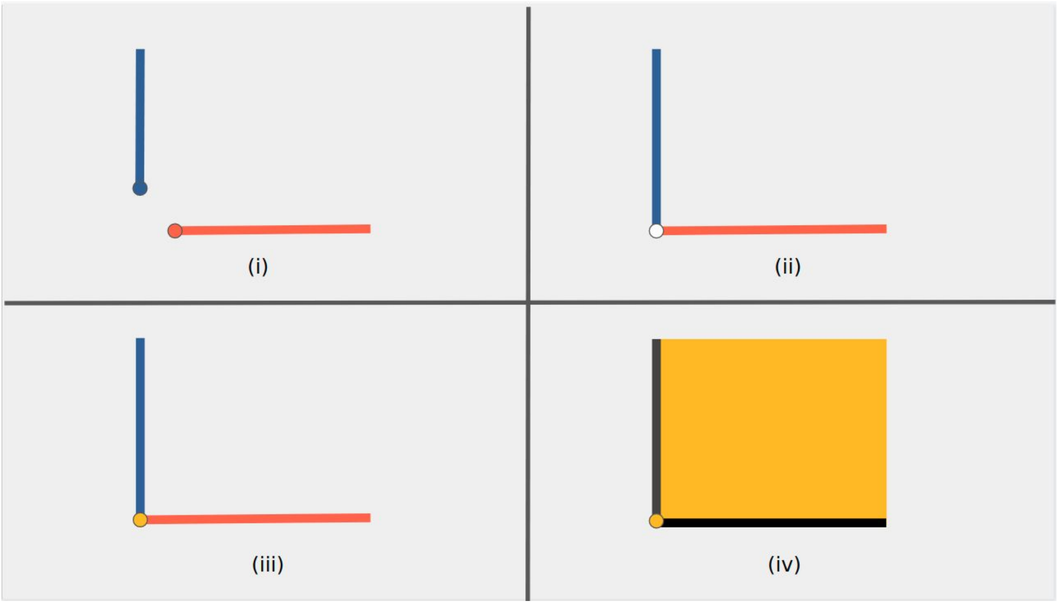
An illustration

This problem is no longer decidable.



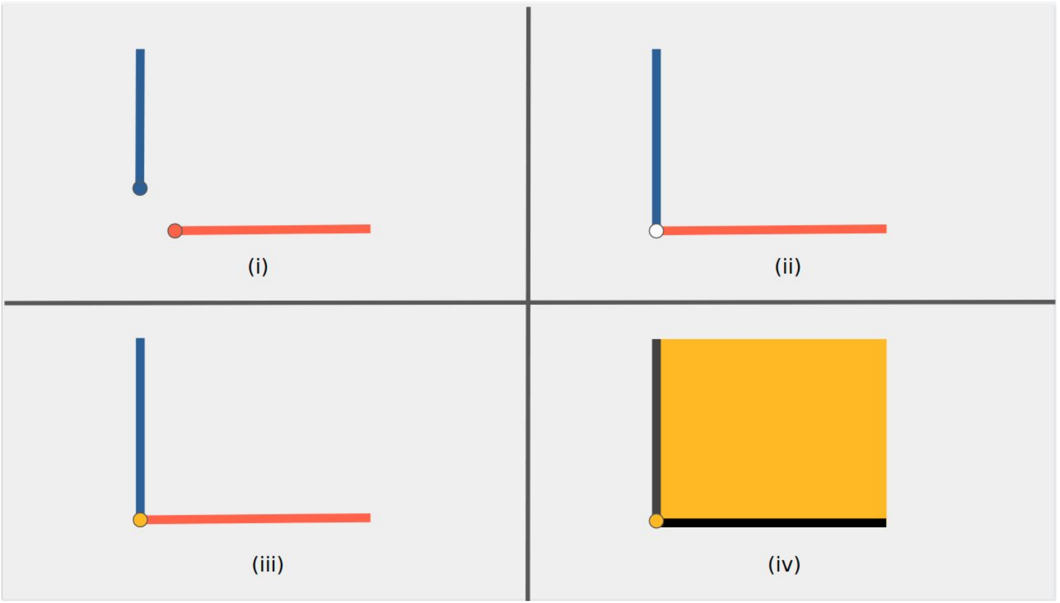
An illustration

If we want to converge to the right answer in the **yellow** possibility, we must open ourselves to error in nearby **blue** and **red** possibilities.



An illustration

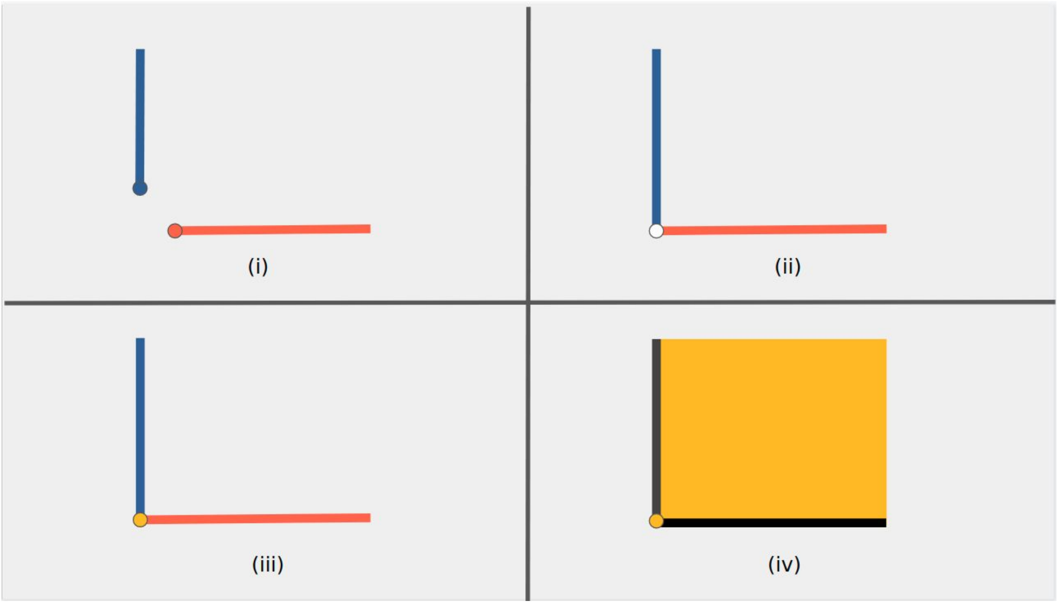
However, it is **progressively** solvable: conjecture **yellow** until it is inconsistent with the confidence interval.



An illustration

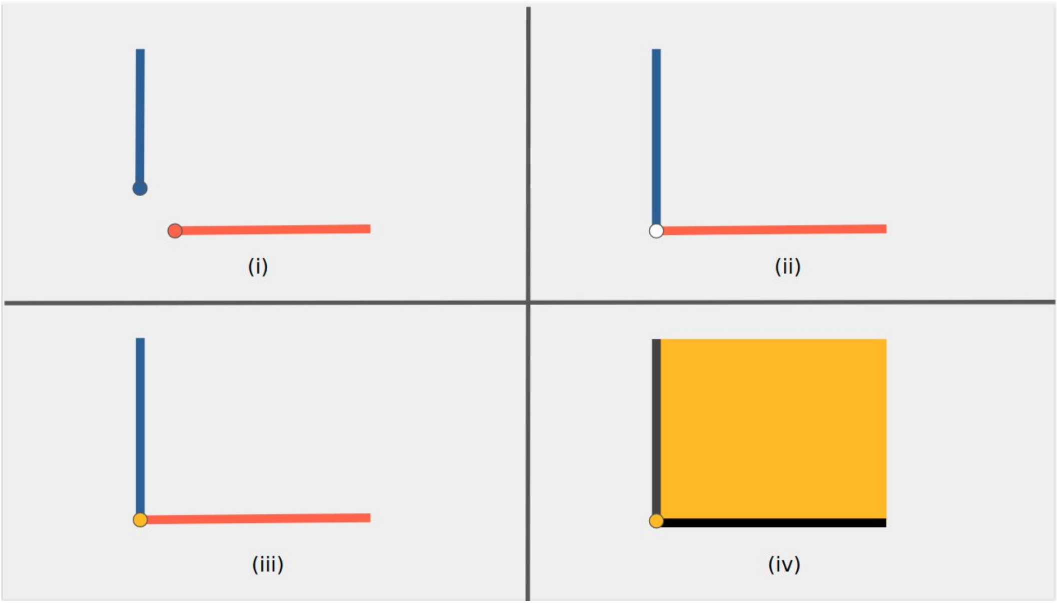
(iii) yellow, blue, red is a suitable enumeration:

blue ∩ cl(**yellow**)= Ø and **red** ∩ cl(**blue** U **yellow**)= Ø.



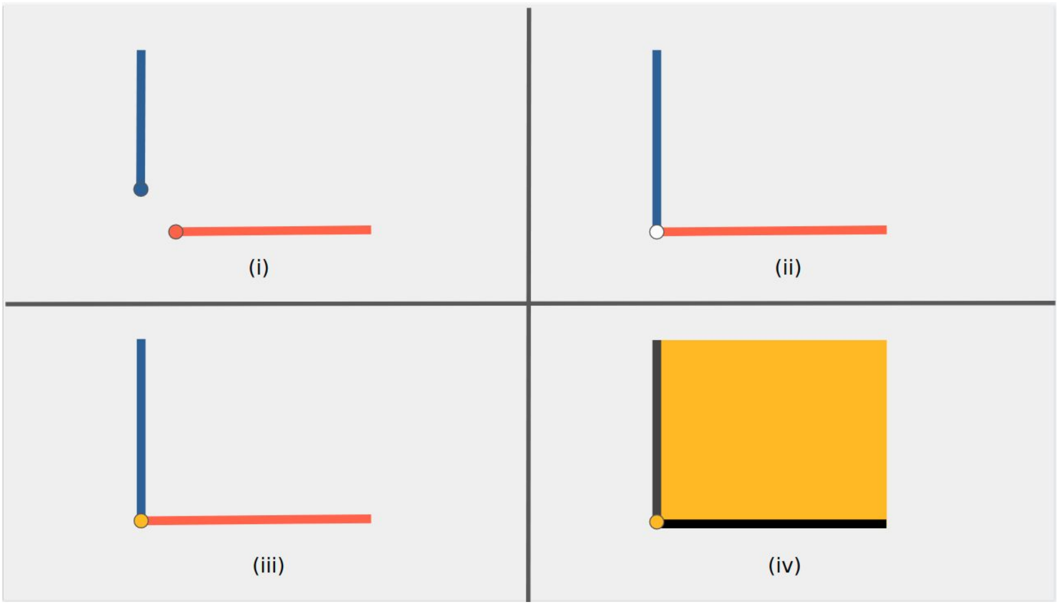
An illustration

(iv) Any combination of biases is possible. Question: are an **even** or **odd** number of coins biased?



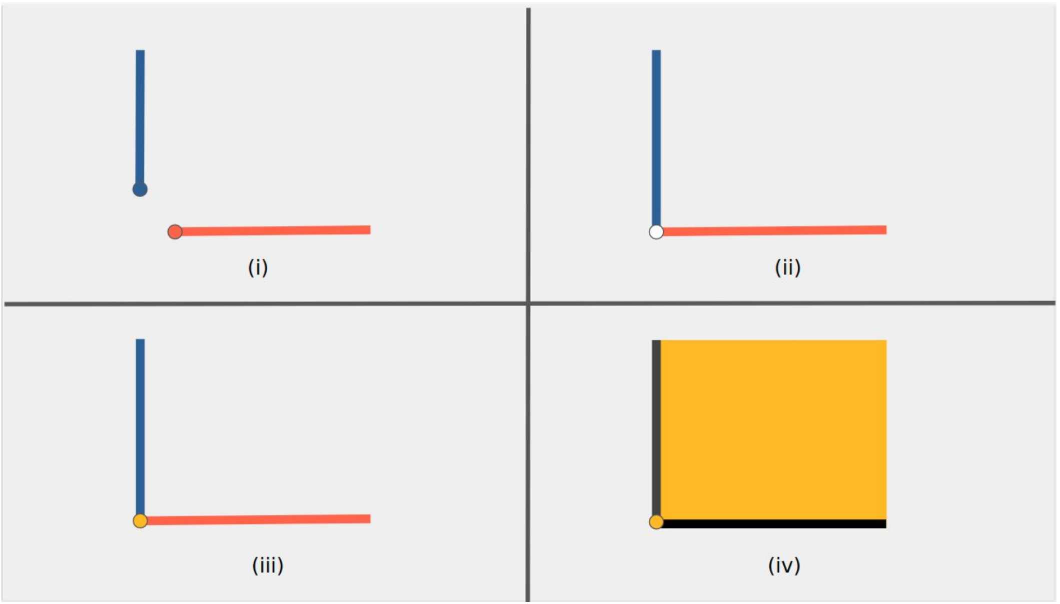
An illustration

Not even progressively solvable.



An illustration

(iv) black $\cap$ cl(**yellow**)$\neq \emptyset$ and **yellow** $\cap$ cl(**black**)$\neq \emptyset$.



The LiNGAM Model

Theorem (Genin and Mayo-Wilson, 2020). If

A. noise terms are **independent** and **non-**Gaussian,
B. functional relationships are **linear** and **a-cyclic** and
C. there are no unobserved confounders, then

then

1. $\mathfrak{Q} = \{\mathcal{M}^{i\to j}, \mathcal{M}^{i\leftarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{ioj}, \mathcal{M}^{i\to j}, \mathcal{M}^{i\leftarrow j}\}$ is progressively solvable, but not decidable.

Genin, Konstantin and Mayo-Wilson, Conor. “Statistical Decidability in Linear, Non-Gaussian Models.” *Causal Discovery & Causality-Inspired Machine Learning, NeurIPS 2020.*

Topological Criterion for Decidability

Theorem. (Genin & Kelly, 2017)

A question is decidable if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

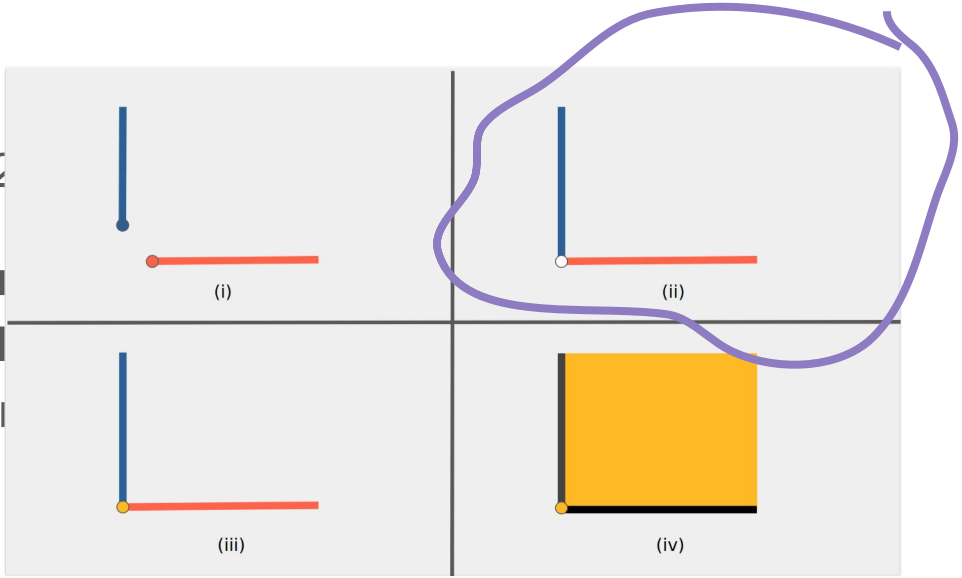
$P_{\mathcal{A}}$ is disjoint from the (weak topology) closure of $P_{\mathcal{B}}$.

Genin, Konstantin, and Kevin T. Kelly. (2017) "The Topology of Statistical Verifiability." *Electronic Proceedings in Theoretical Computer Science* 251: 236–50. https://doi.org/10.4204/EPTCS.251.17.

The LiNGAM Model

Theorem (Genin and Mayo-Wilson, 2

A. noise terms are **independe**
B. functional relationships are l
C. there are no unobserved co



then

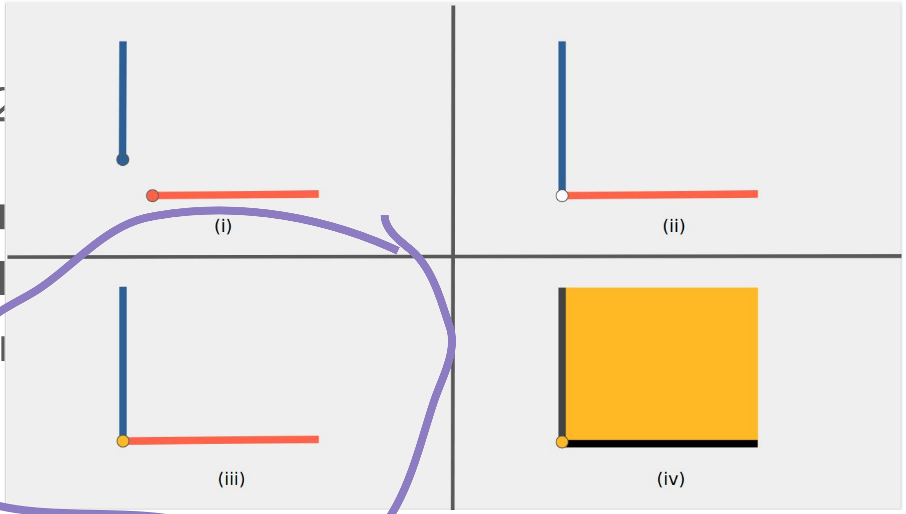
1. $\mathfrak{Q} = \{\mathcal{M}^{i\rightarrow j}, \mathcal{M}^{i\leftarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{i\circ j}, \mathcal{M}^{i\rightarrow j}, \mathcal{M}^{i\leftarrow j}\}$ is progressively solvable, but not decidable.

Genin, Konstantin and Mayo-Wilson, Conor. “Statistical Decidability in Linear, Non-Gaussian Models.” *Causal Discovery & Causality-Inspired Machine Learning, NeurIPS 2020.*

The LiNGAM Model

Theorem (Genin and Mayo-Wilson, 2

A. noise terms are **independe**
B. functional relationships are l
C. there are no unobserved co



then

1. $\mathfrak{Q} = \{\mathcal{M}^{i\to j}, \mathcal{M}^{i\leftarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{i\circ j}, \mathcal{M}^{i\to j}, \mathcal{M}^{i\leftarrow j}\}$ is progressively solvable, but not decidable.

Genin, Konstantin and Mayo-Wilson, Conor. “Statistical Decidability in Linear, Non-Gaussian Models.” *Causal Discovery & Causality-Inspired Machine Learning, NeurIPS 2020.*

LiNGAM + Confounding - Unfaithfulness

Theorem (Salehkaleybar et al., 2020)**.** When

1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** unobserved confounders, but
4. there are no cancelling paths (**faithfulness**),

then causal ancestry relationships between **observed** variables are **identified**.

Salehkaleybar, Saber, et al. (2020) "Learning Linear Non-Gaussian Causal Models in the Presence of Latent Variables." *Journal of Machine Learning Research* 21.39: 1-24.

LiNGAM + Confounding - Unfaithfulness

Theorem (Salehkaleybar et al., 2020)**.** When

1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** unobserved confounders, but
4. there are no cancelling paths (**faithfulness**),

then causal ancestry relationships between **observed** variables are **identified**.

But how *identified* are they, really?

Salehkaleybar, Saber, et al. (2020) "Learning Linear Non-Gaussian Causal Models in the Presence of Latent Variables." *Journal of Machine Learning Research* 21.39: 1-24.

Good News

Theorem (Genin, 2021)**.** When

1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** unobserved confounders, but
4. there are no cancelling paths (**faithfulness**),

then causal ancestry relationships between **observed** variables are **decidable in the limit**.

Topological Criterion for Limiting Decidability

Theorem. (Dembo & Peres, 1994)

A question is decidable in the limit if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

- $P_{\mathcal{A}}$ is disjoint from $P_{\mathcal{B}}$;
- $P_{\mathcal{A}}$ is a countable union of closed sets in the weak topology.

Dembo, Amir, and Yuval Peres (1994). “A Topological Criterion for Hypothesis Testing.” *Annals of Statistics* 22(1): 106–17. https://doi.org/10.1214/aos/1176325360.

Bad News

Theorem (Genin, 2021)**.** When

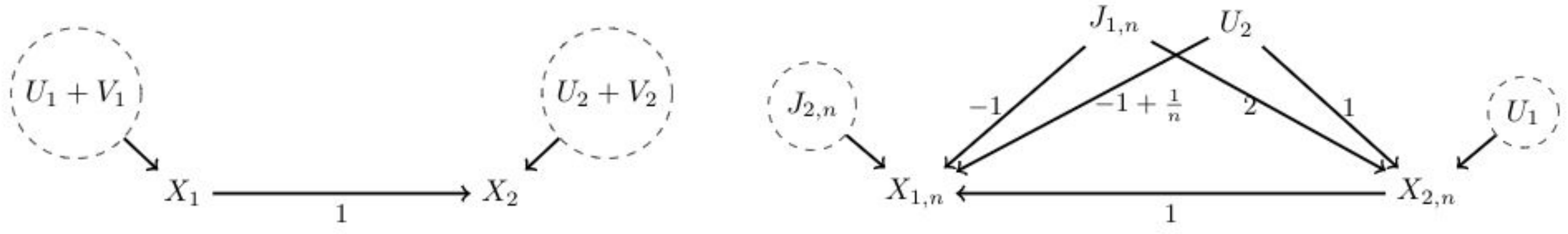
1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** confounders, but
4. there are no cancelling paths (**faithfulness**),

then causal ancestry relationships between **observed** variables are **not decidable**.

Bad News

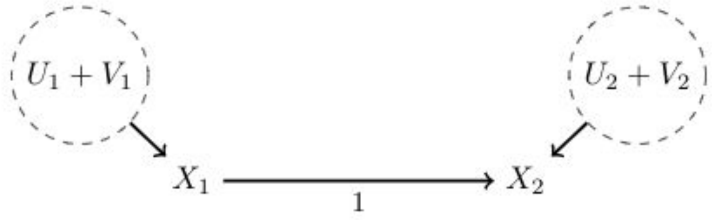
Flipping returns when we allow for unobserved confounders.

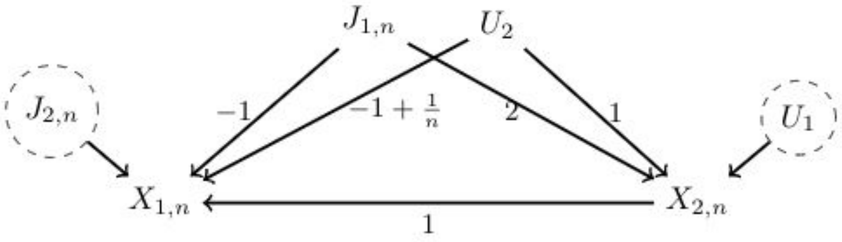
Although causal orientation is a solvable problem (assuming faithfulness), it is no longer decidable.



Bad News

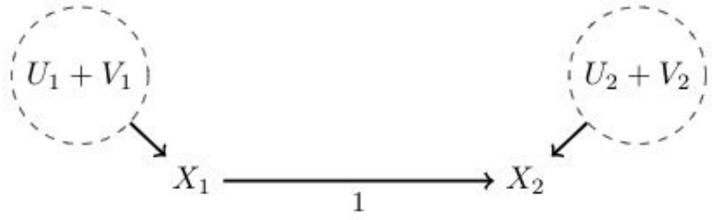
Let Z_1, Z_2 be independent, Gaussian.

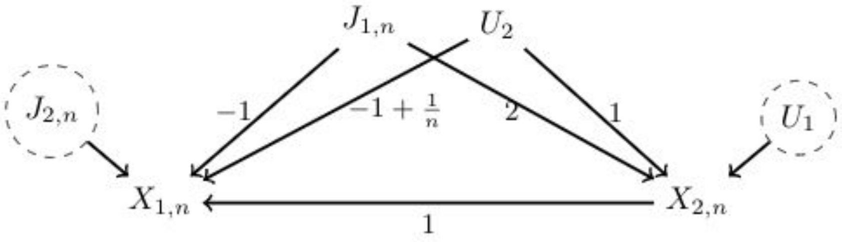




Bad News

Let $V_1 = Z_1 + Z_2$ and $V_2 = Z_1 - Z_2$.   Then $U_1 + V_1$ and $U_2 + V_2$ are independent and non-Gaussian.

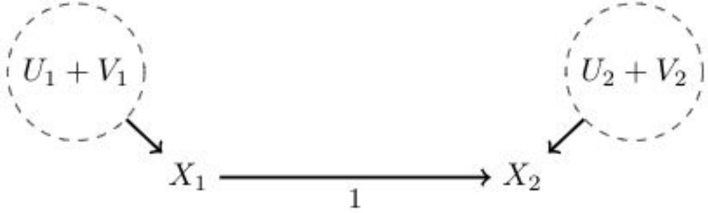


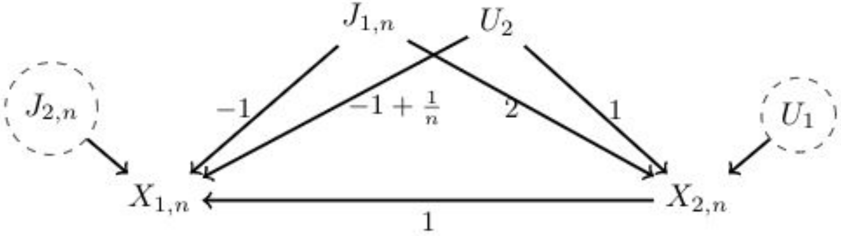


Bad News

Let $J_{1,n} = Z_1 + (1/n)W_1$ *and* $J_{2,n} = Z_2 + (1/n)W_2$.

Then the rhs are faithful, confounded LiNGAMs and $(X_{1,n},X_{2,n}) \Rightarrow (X_1,X_2)$.





Topological Criterion for Decidability

Theorem. (Genin & Kelly, 2017)

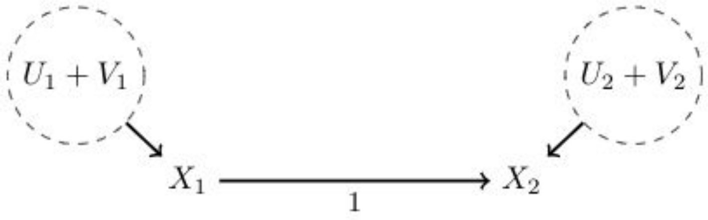
A question is decidable if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

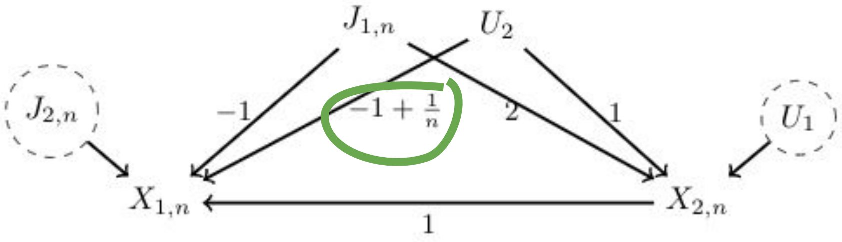
$P_{\mathcal{A}}$ is disjoint from the (weak topology) closure of $P_{\mathcal{B}}$.

Genin, Konstantin, and Kevin T. Kelly. (2017) “The Topology of Statistical Verifiability.” *Electronic Proceedings in Theoretical Computer Science* 251: 236–50. https://doi.org/10.4204/EPTCS.251.17.

Possible Escape Routes

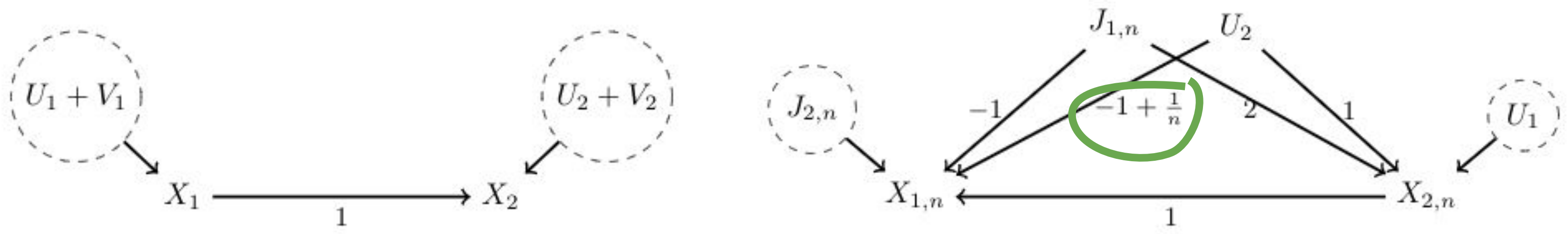
Route 1: Strengthen Faithfulness Assumption.





Possible Escape Routes

Route 1: Strengthen Faithfulness Assumption.

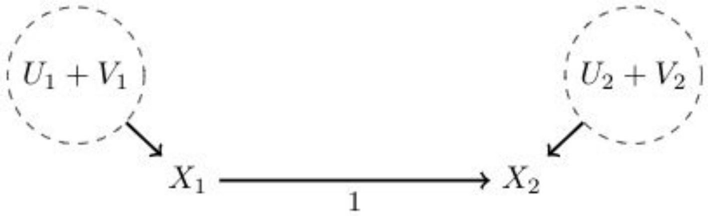


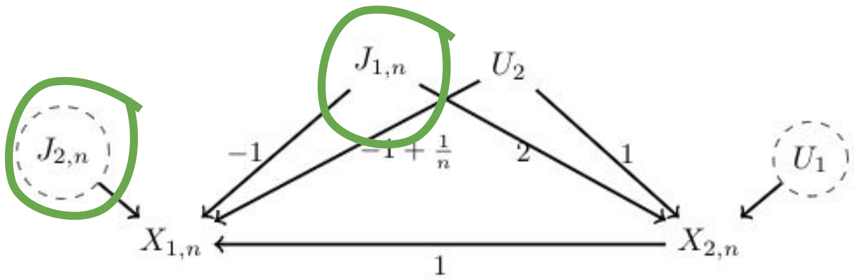
Uhler, Caroline, et al. "Geometry of the faithfulness assumption in causal inference." *The Annals of Statistics* (2013): 436-463.

Possible Escape Routes

Route 2: Strengthen Non-Gaussianity Assumption.

Recall: $J_{1,n} = Z_1 + (1/n)W_1$ *and* $J_{2,n} = Z_2 + (1/n)W_2$.

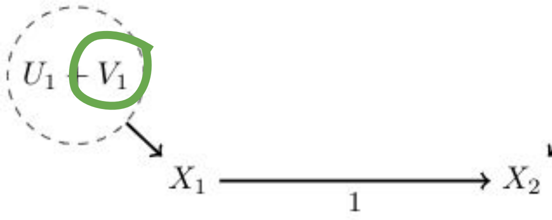


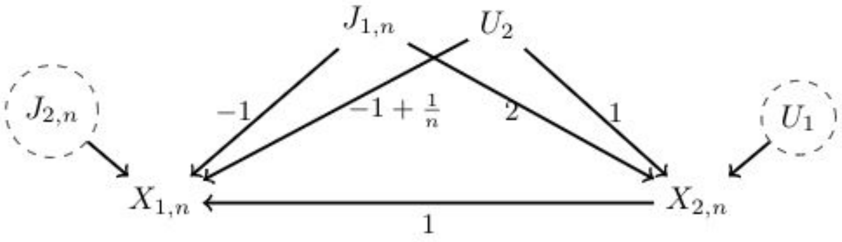


Possible Escape Routes

Route 3: No Gaussian Components.

Recall V_1=Z_1 + Z_2 and V_2=Z_1 - Z_2.

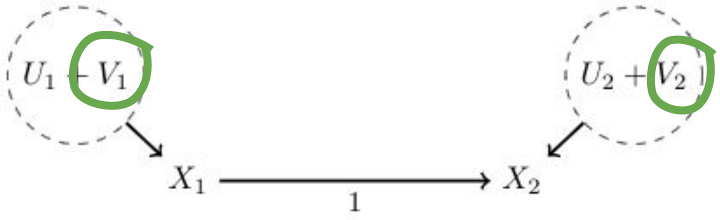


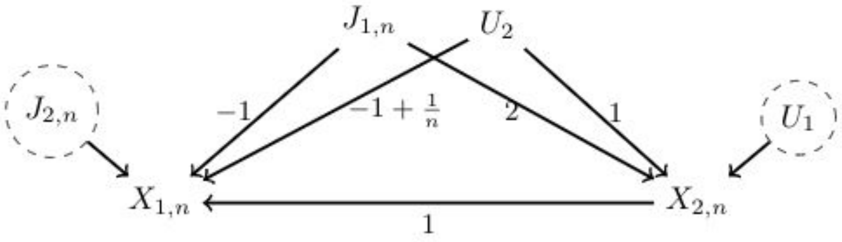


Possible Escape Routes

Route 3: No Gaussian Components.

X has Gaussian components if $X = Y + Z$, with $Y \perp Z$ and Z Gaussian.





FLAMNGCo Model

A “flamingo” is a **F**aithful, **L**inear, **A**cyclic **M**odel with **N**o **G**aussian **Co**mponents.

More precisely: no **linear combination** of exogenous noise terms has a Gaussian component.

Decidability Returns

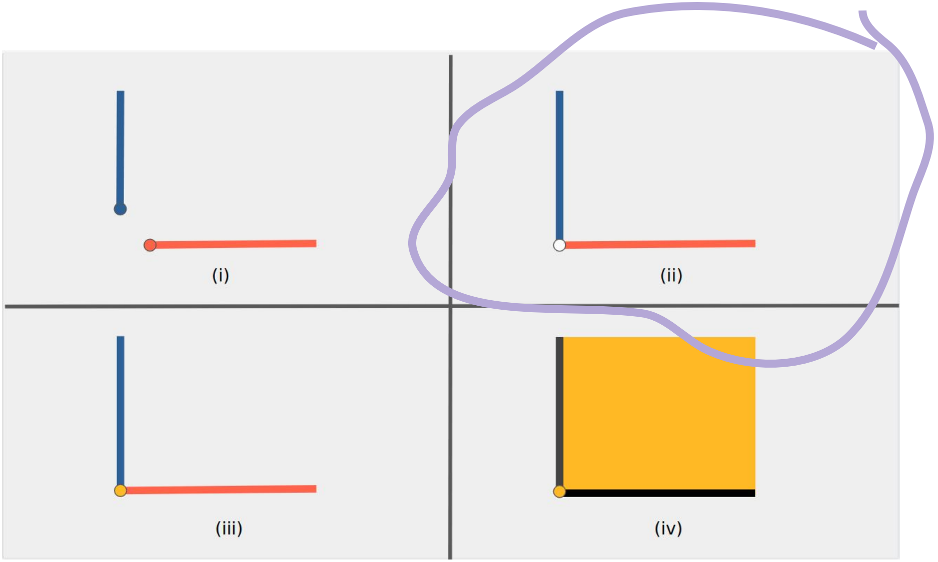
Theorem (Genin and Mayo-Wilson, 2022). If

A. There are no cancelling paths (**faithfulness)**,
B. functional relationships are **linear** and **a-cyclic**,
C. there **may** be unobserved confounders, but
D. **no** linear combination of noise terms has a **Gaussian component**,

then

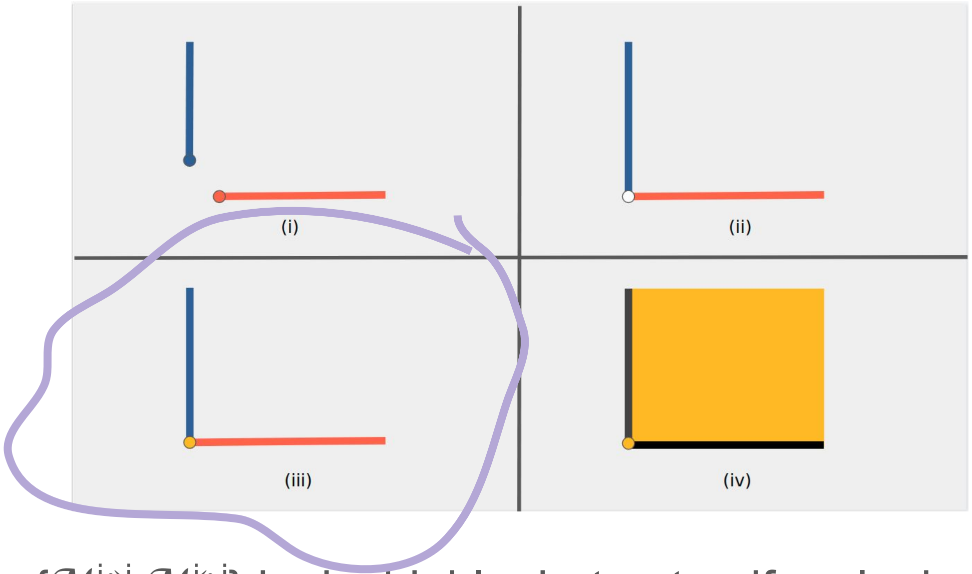
1. $\mathfrak{Q} = \{\mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{ioj}, \mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is progressively solvable, but not decidable.

FLAMNGCo Model, or: Decidability Returns



1. $\mathfrak{Q} = \{\mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{i \circ j}, \mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is progressively solvable, but not decidable.

FLAMNGCo Model, or: Decidability Returns



1. $\mathfrak{Q} = \{\mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is decidable, but not uniformly decidable;
2. $\mathfrak{Q} = \{\mathcal{M}^{i \circ j}, \mathcal{M}^{i \rightsquigarrow j}, \mathcal{M}^{i \leftsquigarrow j}\}$ is progressively solvable, but not decidable.

Thank You!

Questions:

konstantin.genin@uni-tuebingen.de