Statistical Decidability in Confounded, Linear Non-Gaussian Models

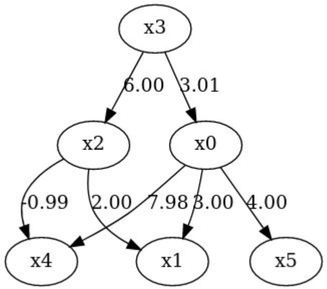
Konstantin Genin

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

NeurIPS 2021
35th Conference on Neural Information Processing Systems

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

The LiNGAM Model



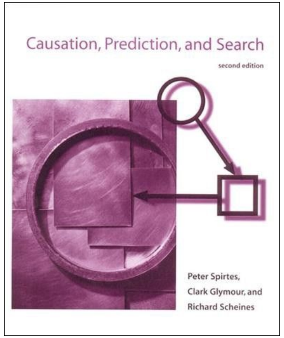
Theorem (Shimizu et al., 2006)**.** When

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.

Pitfalls of Pointwise

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

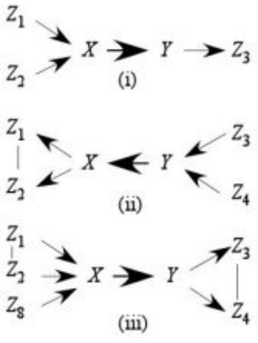


Figure 1: Causal Flips

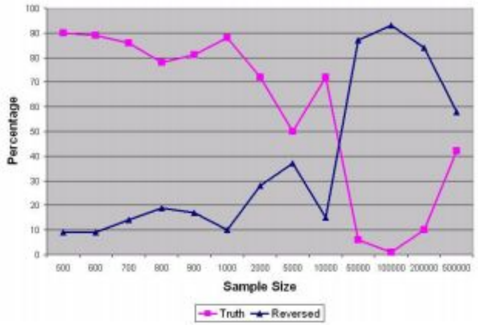


Figure 5: FCI Algorithm

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

Pitfalls of Pointwise

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

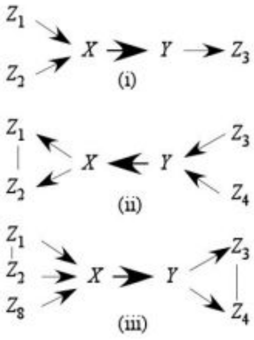


Figure 1: Causal Flips

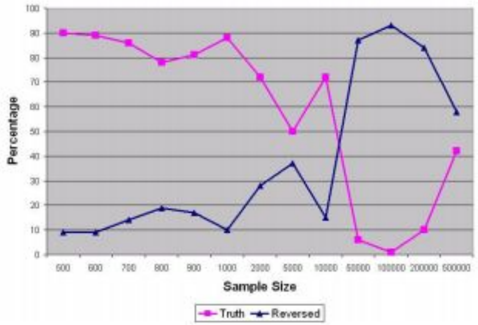
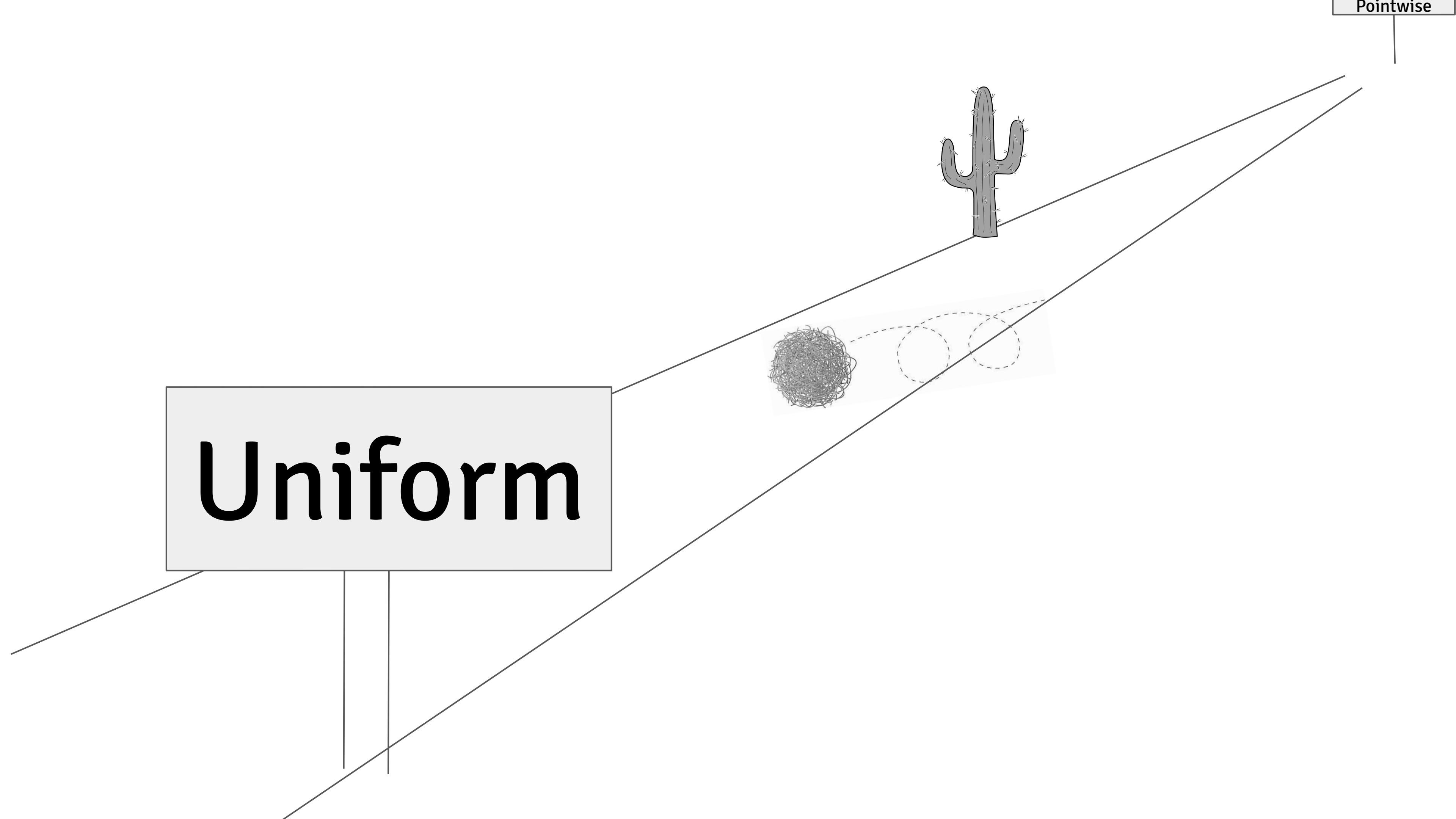


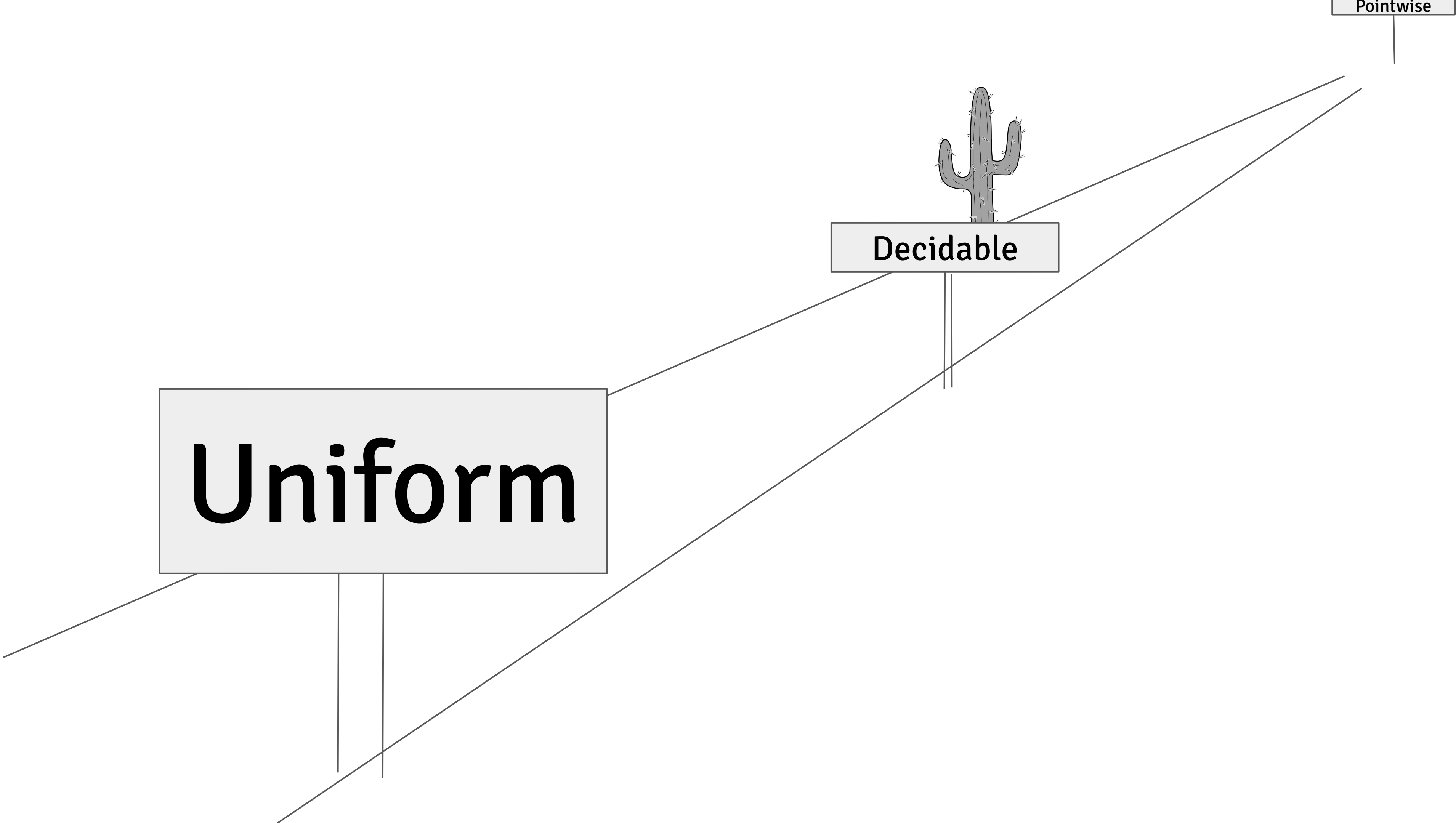
Figure 5: FCI Algorithm

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.


Pointwise
Uniform

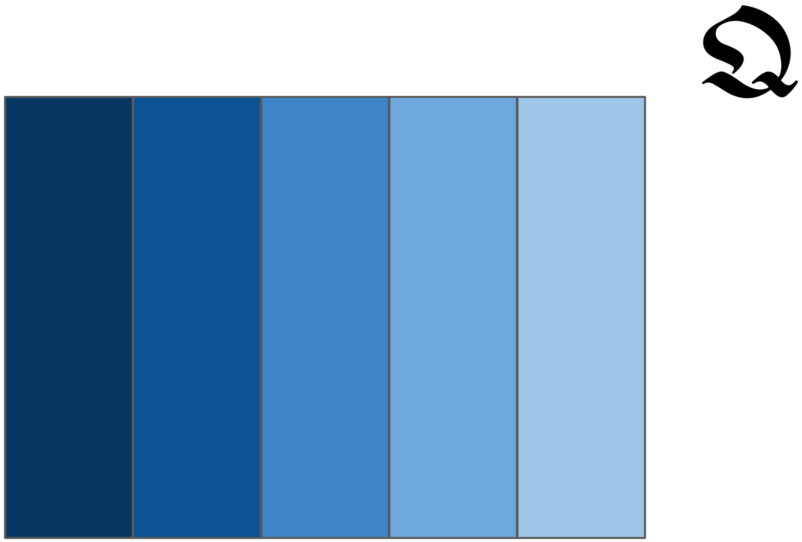

Pointwise
Decidable
Uniform

Statistical Questions

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

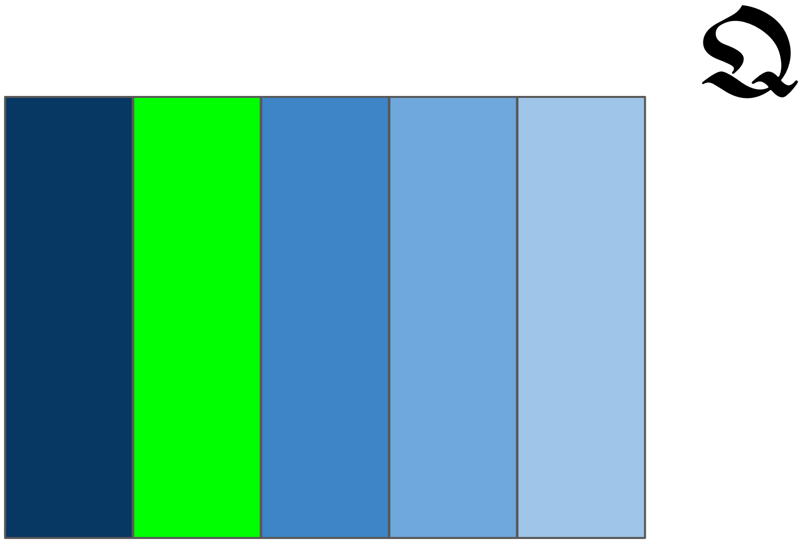
Statistical Questions

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



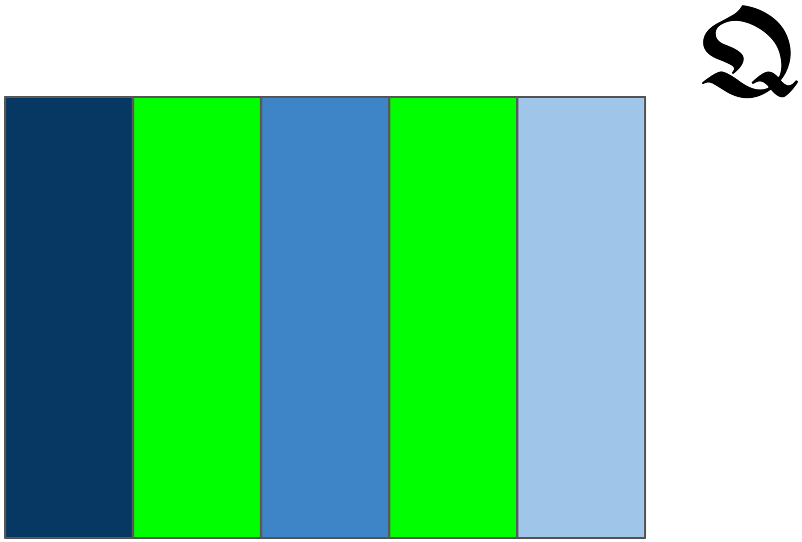
Statistical Questions

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



Statistical Questions

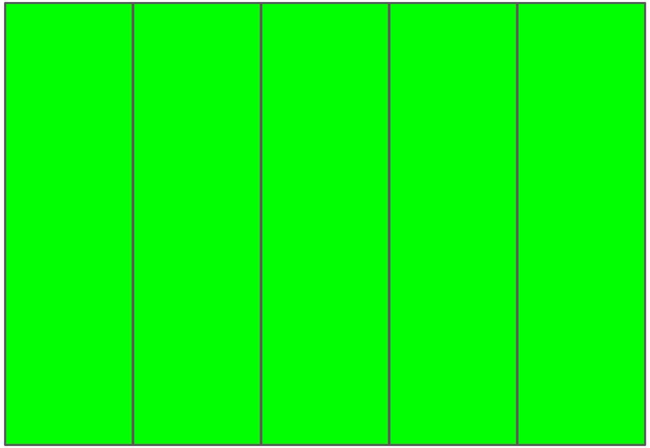
A **relevant response** is a **union** of answers.



Statistical Questions

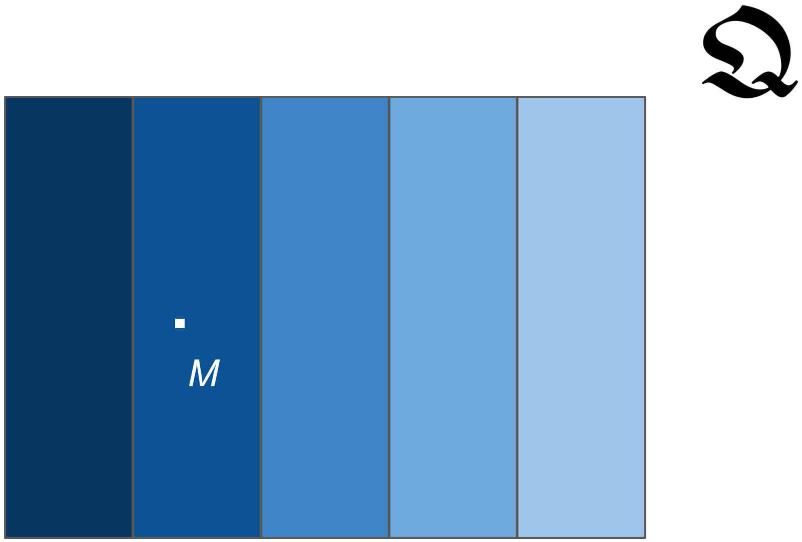
A **relevant response** is a **union** of answers.

$\mathfrak{Q}$



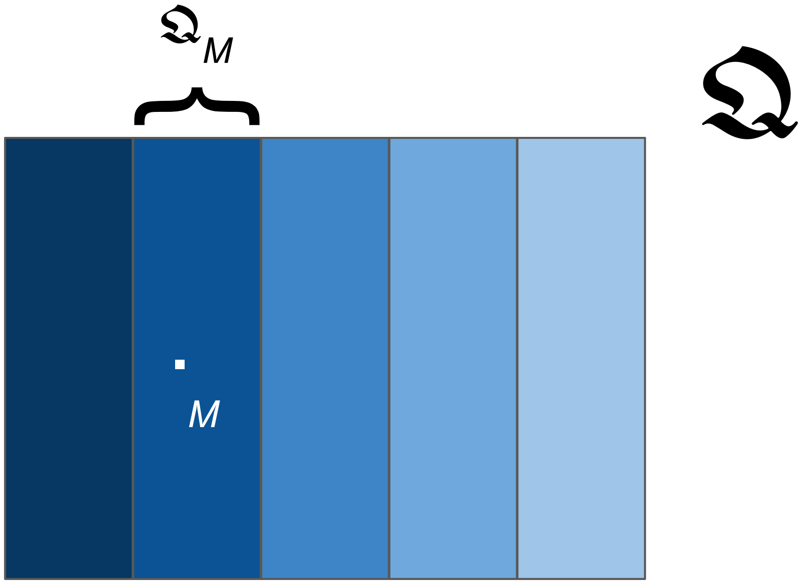
Statistical Questions

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



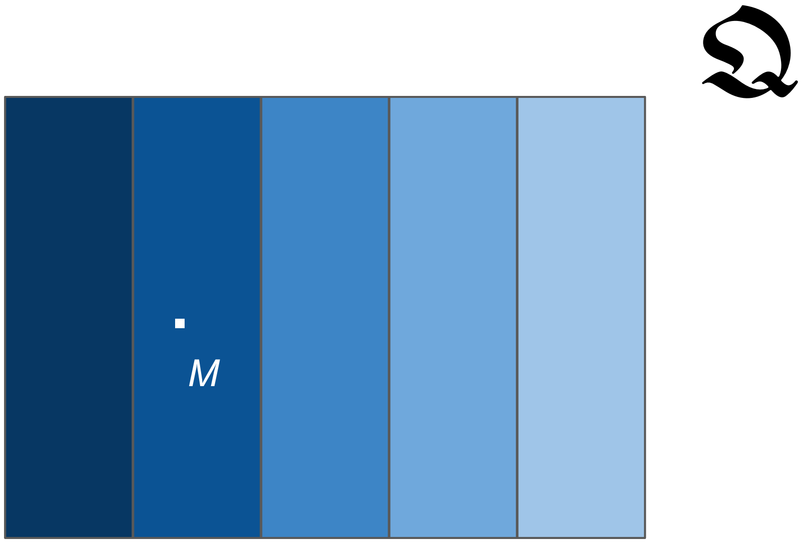
Statistical Questions

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



Statistical Questions

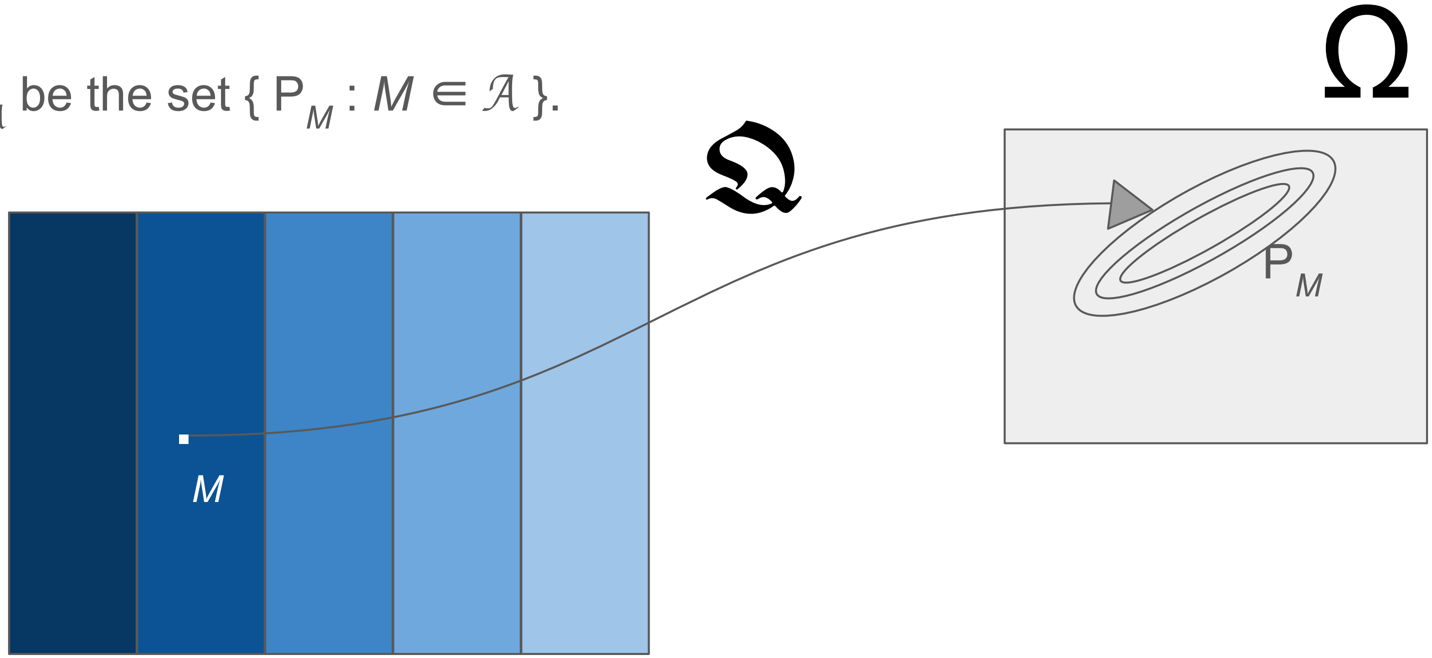
If $M \in \mathcal{M}$,



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 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 $\bigcup \mathfrak{Q}$.

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 $\alpha > 0$.

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*, $P_M(\mathfrak{Q}_M \nsubseteq T_n) < \alpha$.

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

Note: It may be that $P_M(T_n = \cup\mathfrak{Q}) \approx 1$ 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 **unknown** sample size.
Uniformly Decidable	✔	✔	✔
Decidable	✔	✖	✔
Decidable in the Limit	✖	✖	✔

The LiNGAM Model

Theorem (Genin and Mayo-Wilson, 2020)**.** When

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

then causal orientation is **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

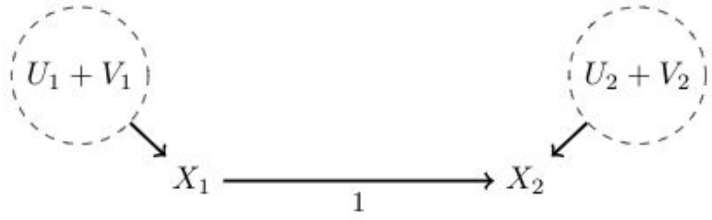
1. noise terms are **independent** and **non-**Gaussian**,**
2. functional relationships are **linear** and **a-cyclic**,
3. there **may be** confounders, but
4. there are no cancelling paths (**faithfulness**),

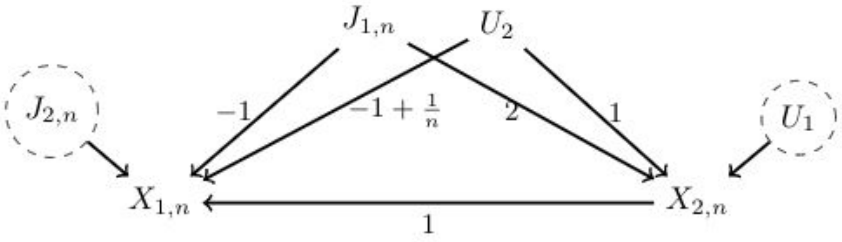
then causal ancestry relationships between **observed** variables are **not decidable**.

Bad News

Flipping returns when we allow for unobserved confounders.

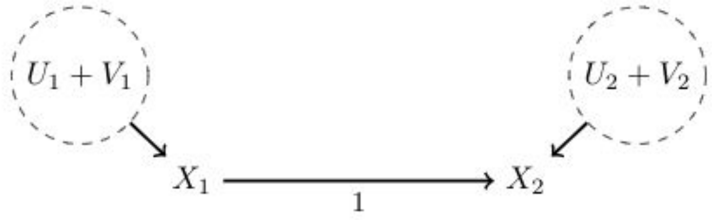
Although causal orientation is a solvable problem (assuming faithfulness), it is no longer decidable.

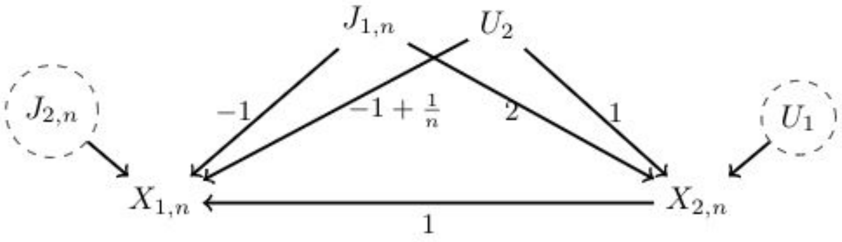




Bad News

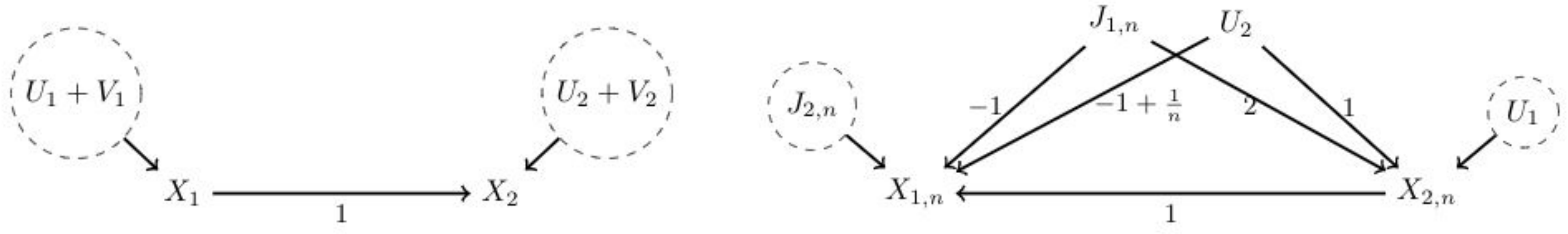
Let Z_1, Z_2 be independent, Gaussian.





Bad News

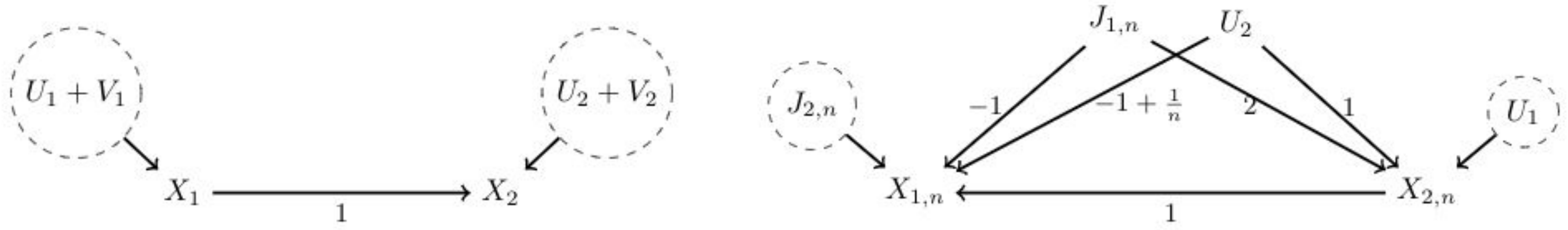
Let $V_1 = Z_1 + Z_2$ and $V_2 = Z_1 - Z_2$. Then $U_1 + V_1$ and $U_2 + V_2$ are independent and non-Gaussian.



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)

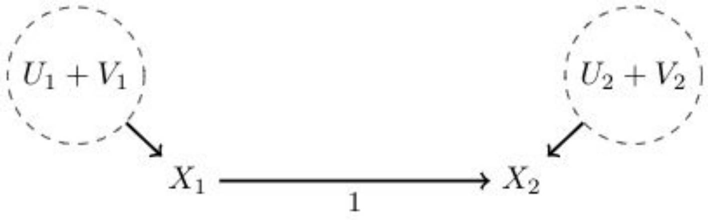
A question is decidable if for answers $\mathcal{A}$, $\mathcal{B}$ in $\mathfrak{Q}$,

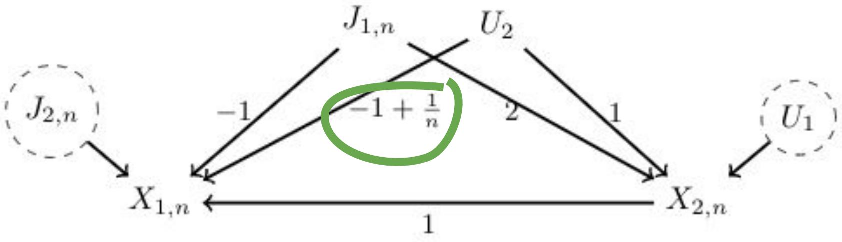
$P_{\mathcal{A}}$ is disjoint from the (weak topology) closure of $P_{\mathcal{B}}$.

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

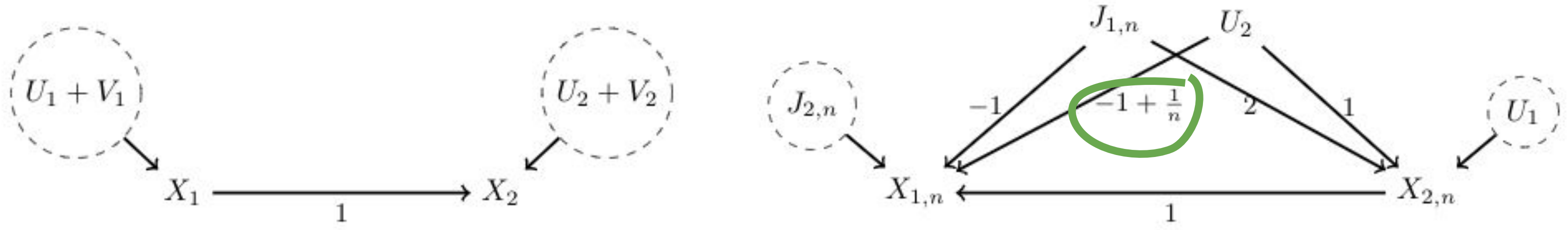
Route 1: Strengthen Faithfulness Assumption.





Possible Escape Routes

Route 1: Strengthen Faithfulness Assumption.

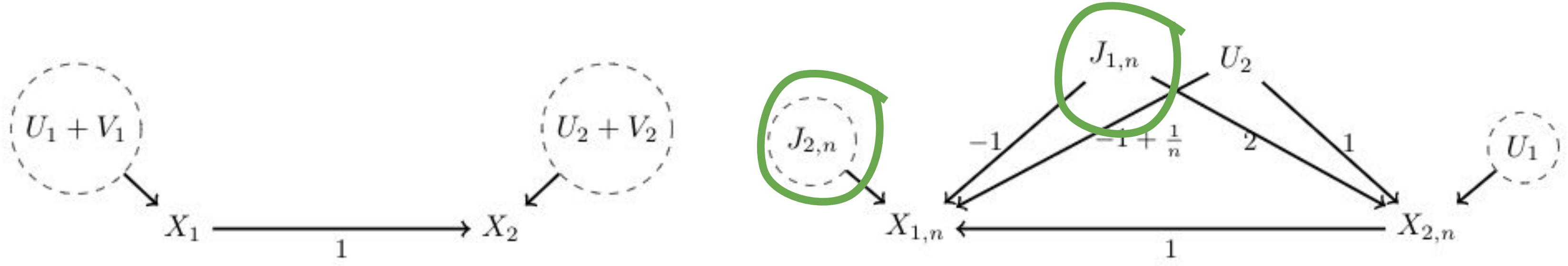


Uhler, Caroline, et al. "Geometry of the faithfulness assumption in causal inference." *The Annals of Statistics* (2013): 436-463.

Possible Escape Routes

Route 2: Strengthen Non-Gaussianity Assumption.

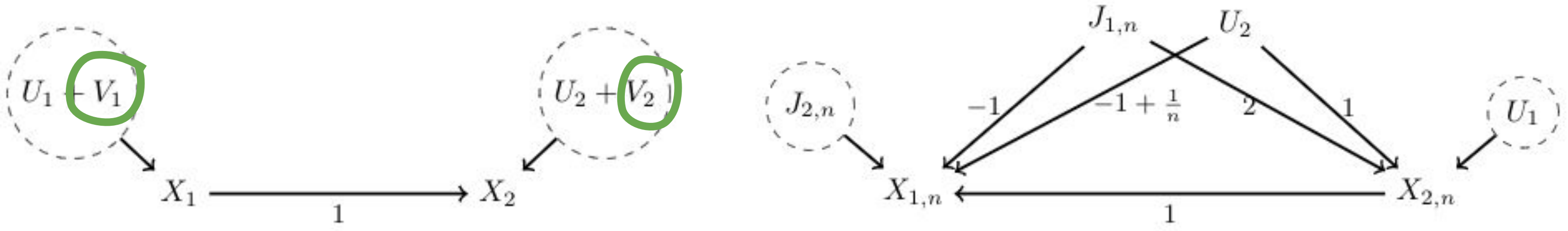
Recall: $J_{1,n} = Z_1 + (1/n)W_1$ *and* $J_{2,n} = Z_2 + (1/n)W_2$.



Possible Escape Routes

Route 3: No Gaussian Components.

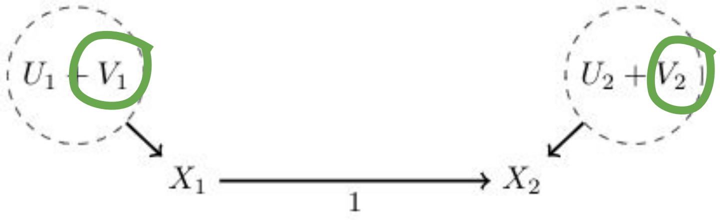
Recall $V_1 = Z_1 + Z_2$ and $V_2 = Z_1 - Z_2$.

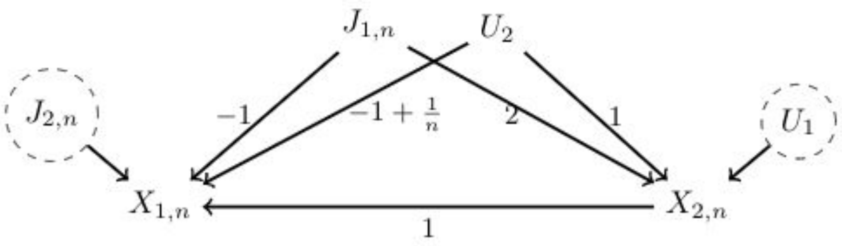


Possible Escape Routes

Route 3: No Gaussian Components.

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





Thank You!

Questions:
konstantin.genin@uni-tuebingen.de