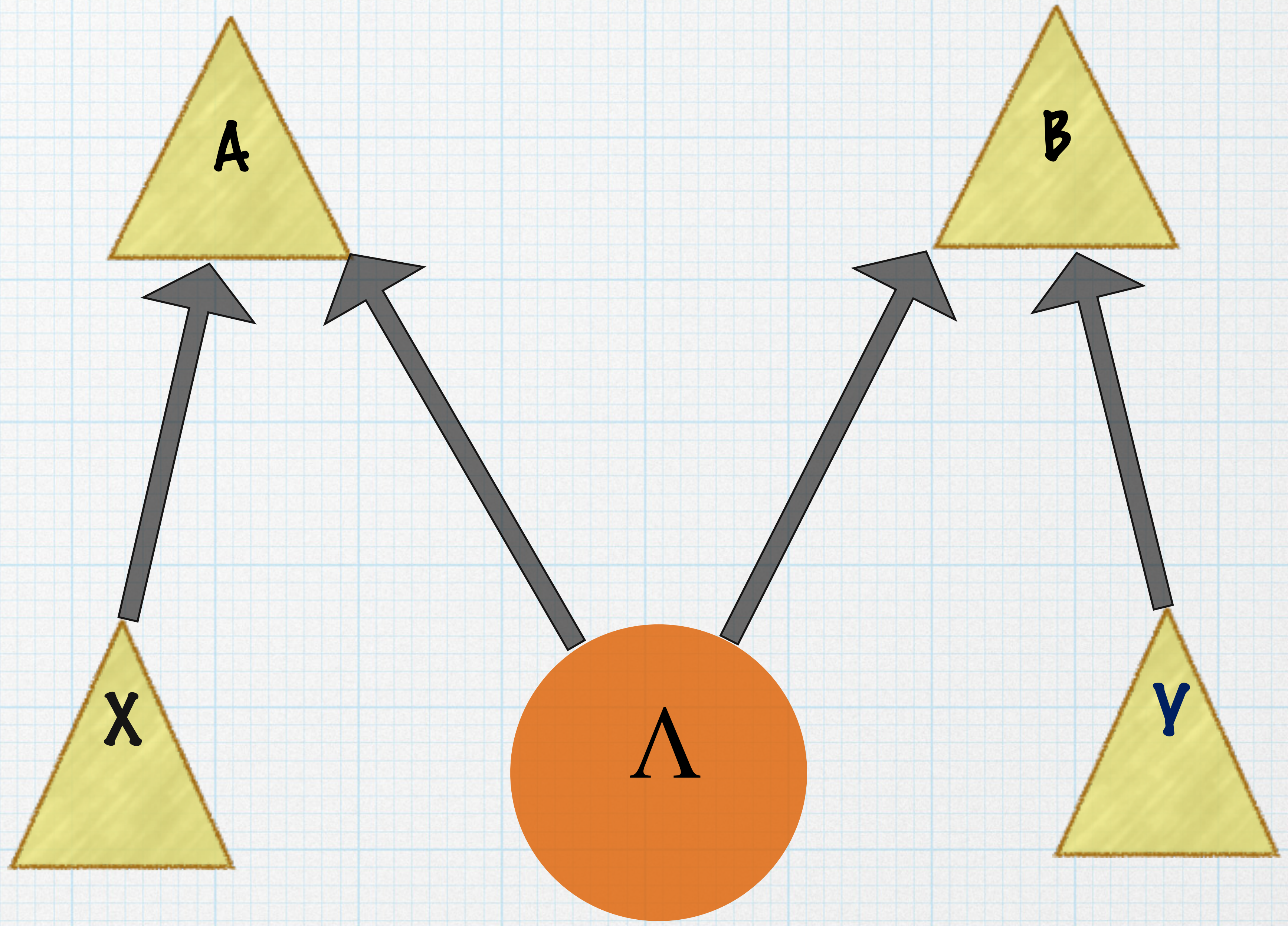
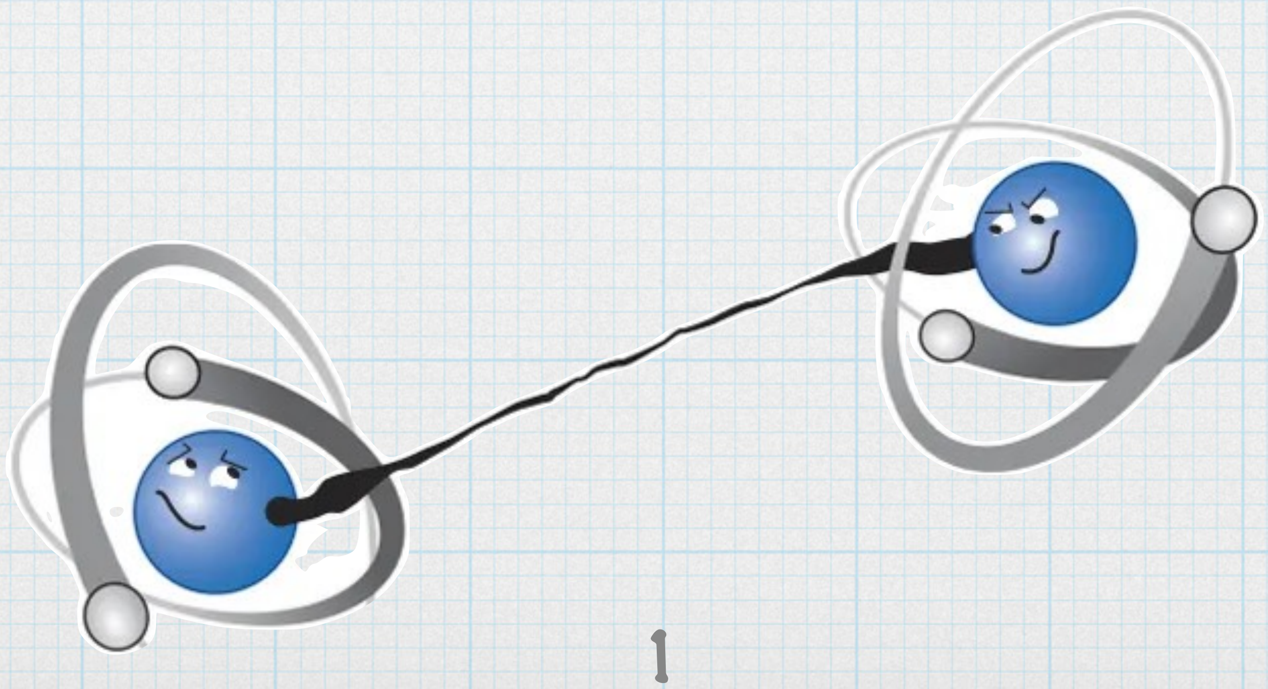
Which causal scenarios might support Non-Classical correlations



- Shashaank Khanna, Marina Maceil Ansanelli, Matthew F. Pusey, Elie Wolfe

PERIMETER INSTITUTE FOR THEORETICAL PHYSICS



UNIVERSITY of York

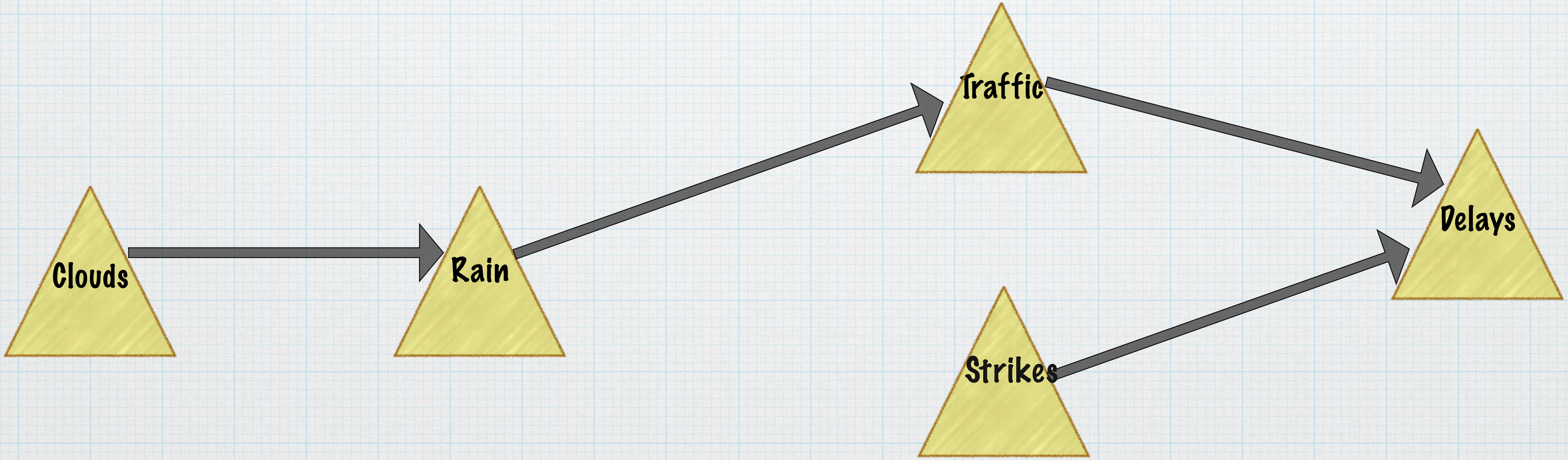
What are causal scenarios (DAGs) ?

Generalised way to represent cause and effect relations among observed events .

Events modelled as random variables.

No directed cycles -> Directed Acyclic Graphs (DAGs)

Edges indicate direct causation.



Causal Markov condition for DAGs

If a probability distribution P over the variables in a DAG G can be factorised as:

$$P(x_1, \ldots x_n) = \prod_i P(x_i \,|\, PA_G(x_i))$$

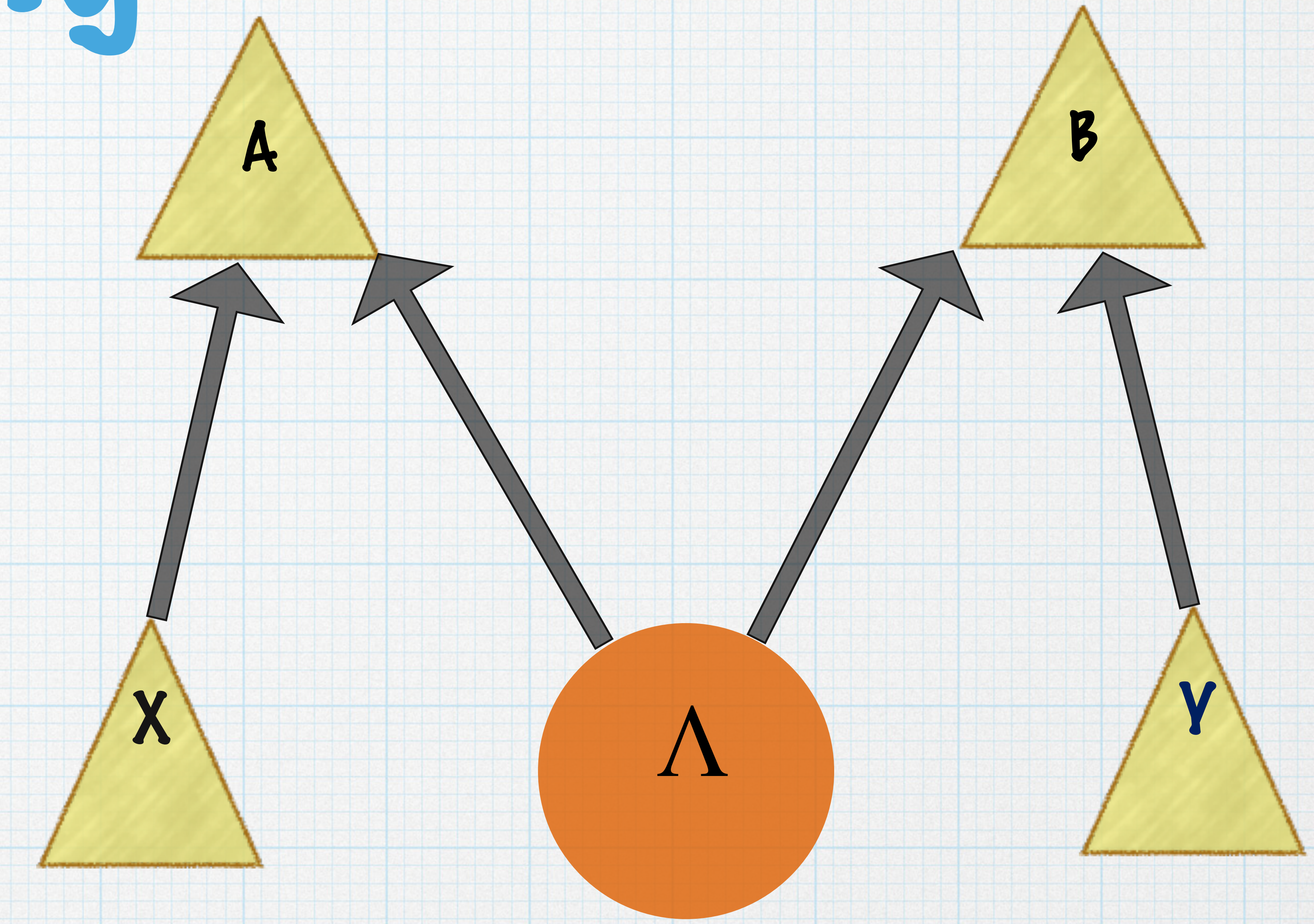
$PA_G(x_i)$ -> parents of x_i in G,

then P is Markov with respect to G

and G is a classically causal explanation of P.

Bell's Theorem recast using DAGs

The causal Markov condition for the Bell DAG encodes the notion of Local Causality.

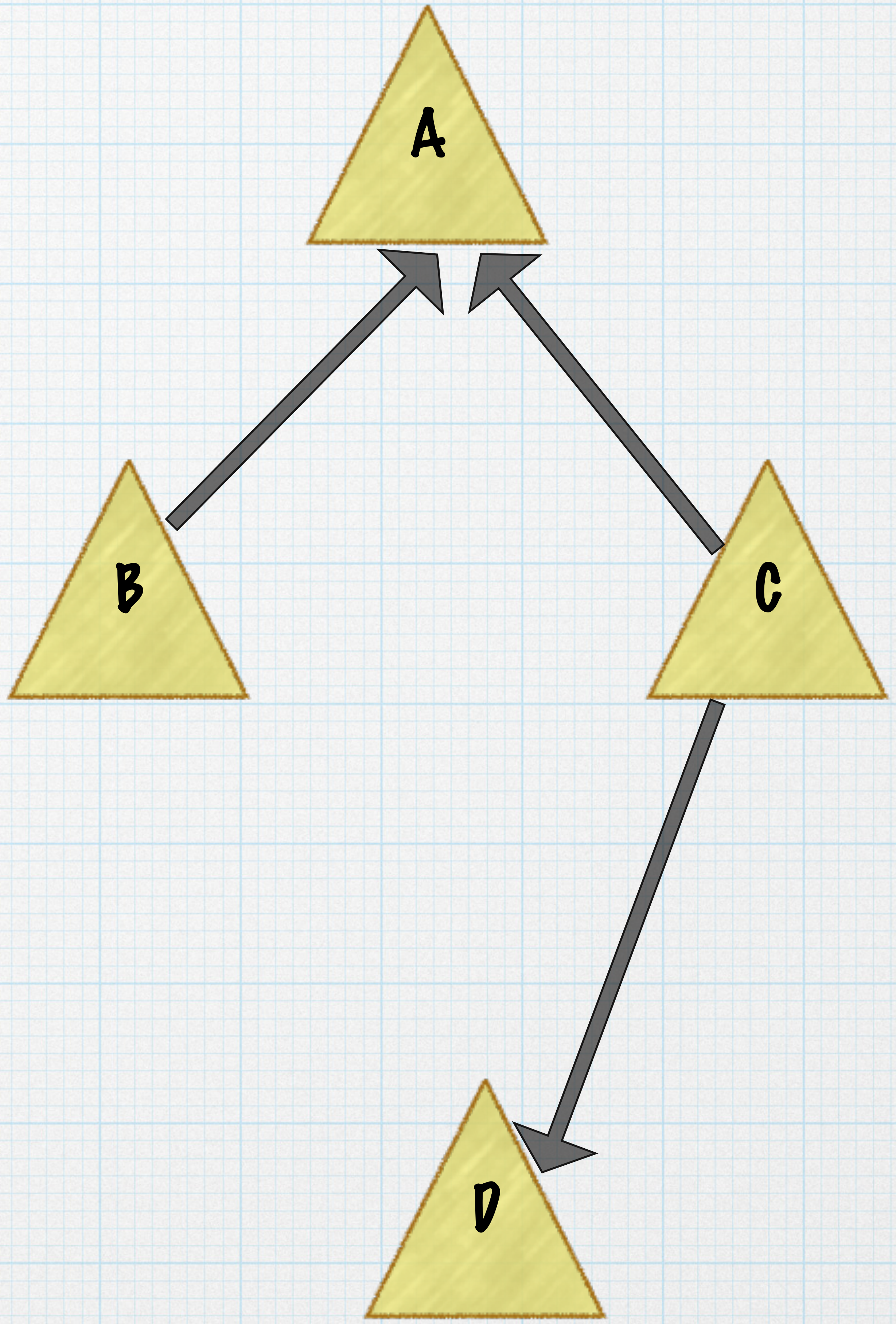


$$P(A, B, X, Y) = \sum_{\Lambda} P(A \mid X, \Lambda) P(B \mid Y, \Lambda) P(X) P(Y) P(\Lambda) \implies \text{Bell's Inequalities}$$

Observed Variables

Latent Variables

Notion of d-separation in DAGs



d-separation -> a graphical condition to read off conditional independences.

$$A \perp_d D \,|\, C \implies P(A \,|\, D, C) = P(A \,|\, C)$$

$$B \perp_d C \implies P(B \,|\, C) = P(B)$$

How to check d-separation ?

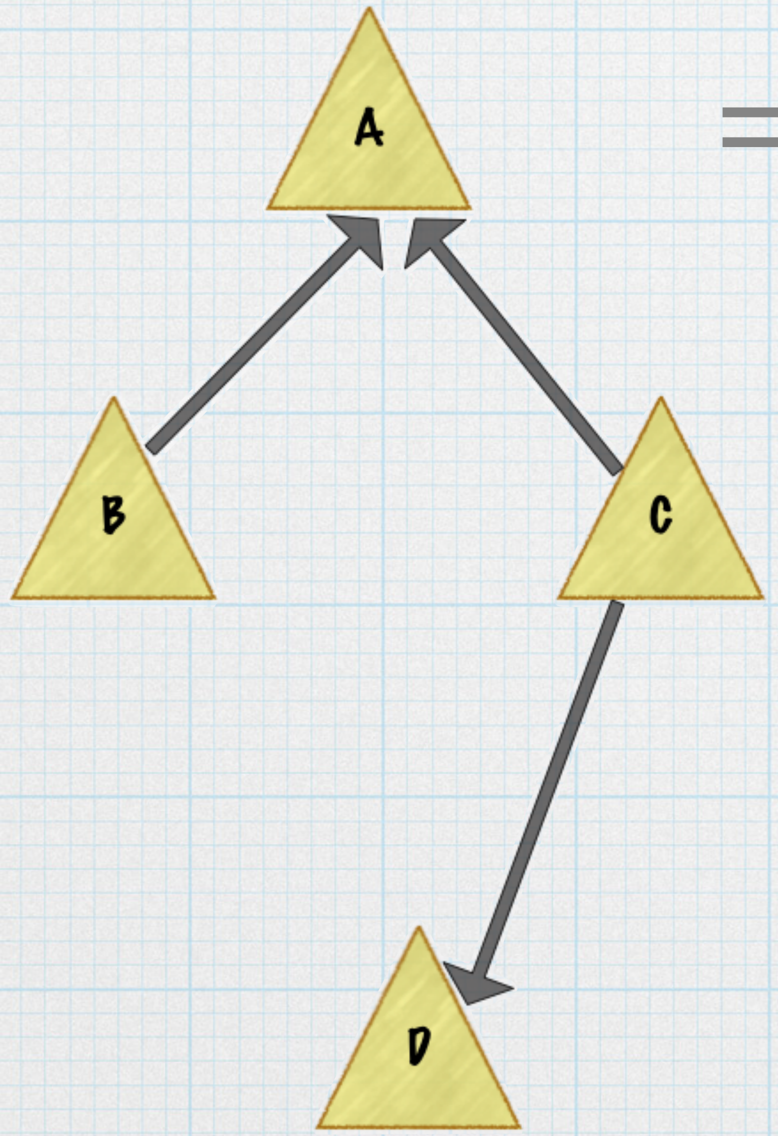
Check if all paths (directed or undirected) between the concerned nodes are blocked.

For eg: $A \perp_d B \mid Z$ if any of the following is true:

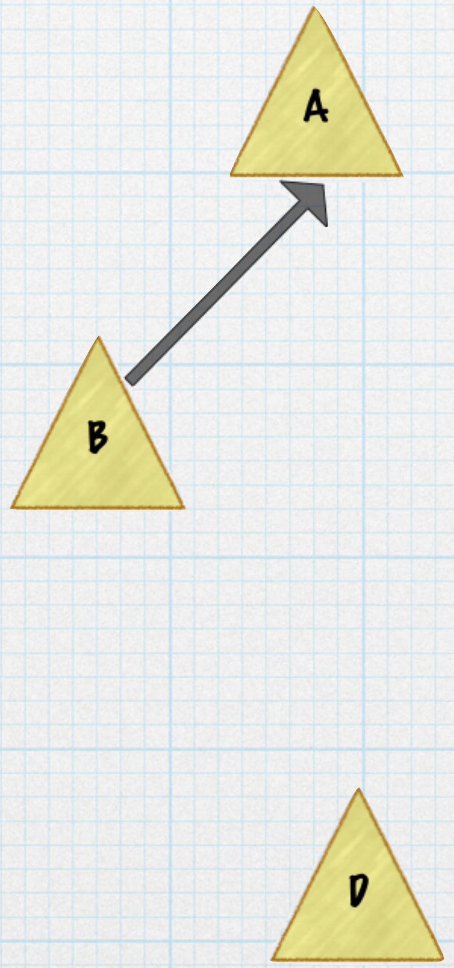
1. Chain of nodes along the path $i \rightarrow m \rightarrow j, m \in Z$
2. Fork along the path $i \leftarrow m \rightarrow j, \quad m \in Z$
3. Collider along the path $i \rightarrow m \leftarrow j, m \notin Z, \; d \notin Z \;\; \forall \;\; d \;\in\; Des(m)$

Towards e-separation ?

If two sets A, B are d-separated by Z after deletion of a set of nodes W in the graph then A and B are e-separated by Z.
For eg:



$\implies A \perp_d D$ is false but after deletion of C we have



$\implies A \perp_d D$ is true

Thus $A \perp_d D$ after deletion of C.

We say $A \perp_e C$.

Different theories allow different types of distributions !

$$C = \{P(x_1 \ldots x_n) : P \; follows \; Causal \;\; Markov \; condition\}$$

$$Q = \{P(x_1 \ldots x_n) : P \, can \; be \; obtained \;\; from \; Quantum \; theory \; by \; Born \; rule\}$$

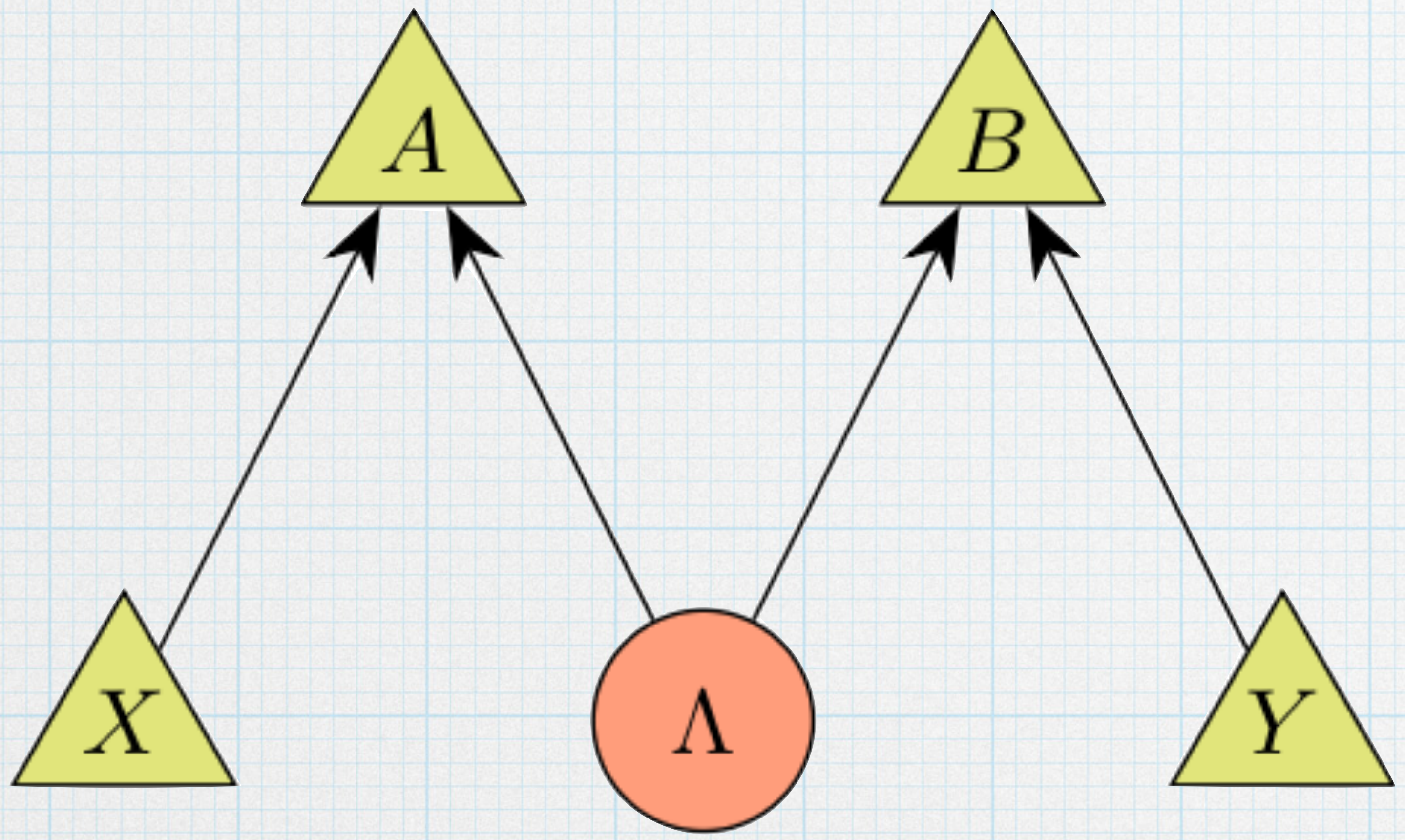
$$G = \{P(x_1 \ldots x_n) : P \, can \; be \; obtained \;\; from \; Generalized \;\; Probabilistic \; Theories\}$$

$$I = \{P(x_1, \ldots . x_n) : P \, respects \; all \; observed \; conditional \;\; independences\}$$

$$C \subseteq Q \subseteq G \subseteq I$$

Refer to: Henson, Lal, Pusey (2014)

Quantum vs Classical: Allowed Probabilities



For Bell DAG:

$$C = \{P(A,B,X,Y) : P(A,B,X,Y) = \sum_{\Lambda} P(A\,|\,X,\Lambda)P(B\,|\,Y,\Lambda)P(X)P(Y)P(\Lambda)\}$$

$$Q = \{P(A,B,X,Y) : P(A,B,X,Y) = tr[(E_X^A \otimes E_Y^B)\ \rho_{\Lambda_{AB}}]\ P(X)P(Y)\}$$

What happens when there are no latent variables in the DAG ?

For a DAG, G, without latent variables, a probability distribution P is Markov with respect to G **if and only if** P satisfies all the observed conditional independences or equivalently all observed d-separation relations.

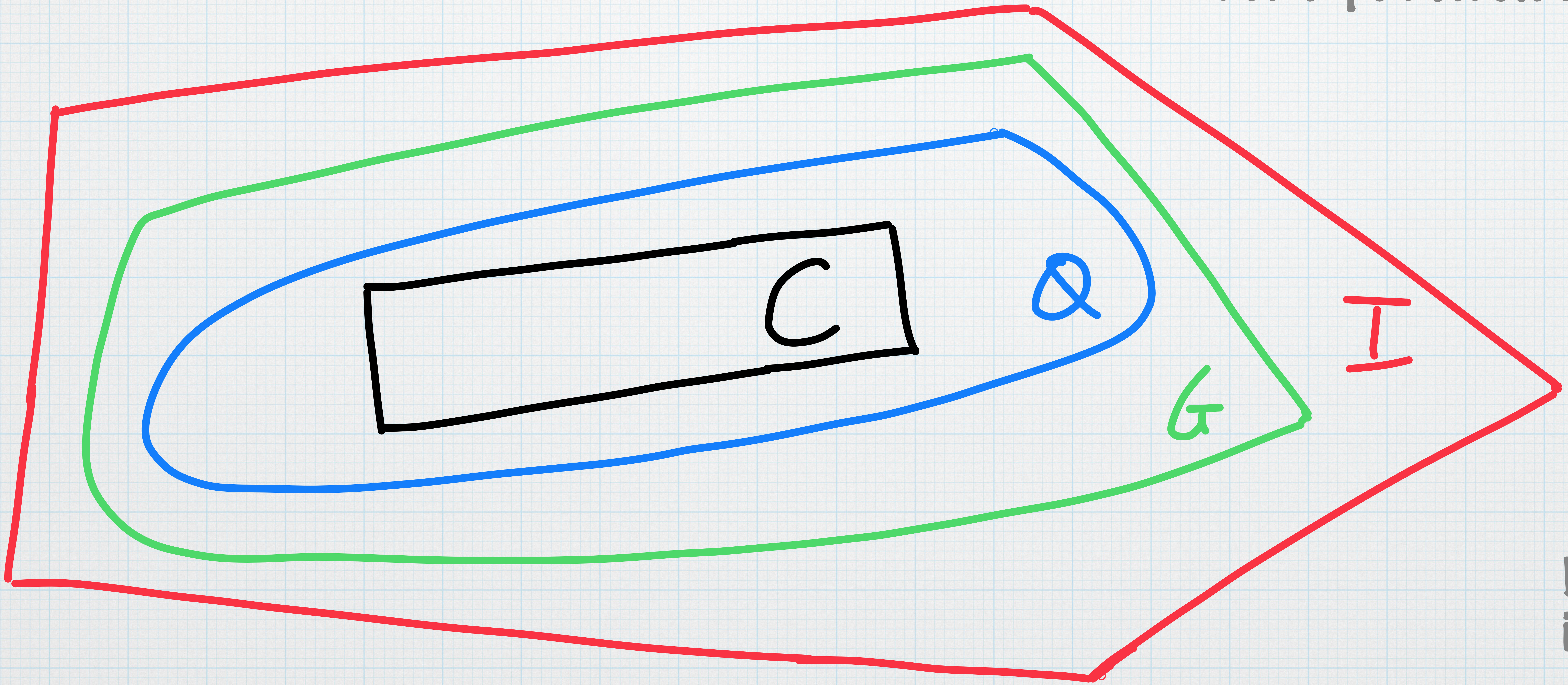
This is because the observed conditional independences are the only possible conditional independences in this case.

Hence for a latent free DAG,

$$C_{LF} = Q_{LF} = G_{LF} = I_{LF}$$

All theories might not allow the same probability distributions !

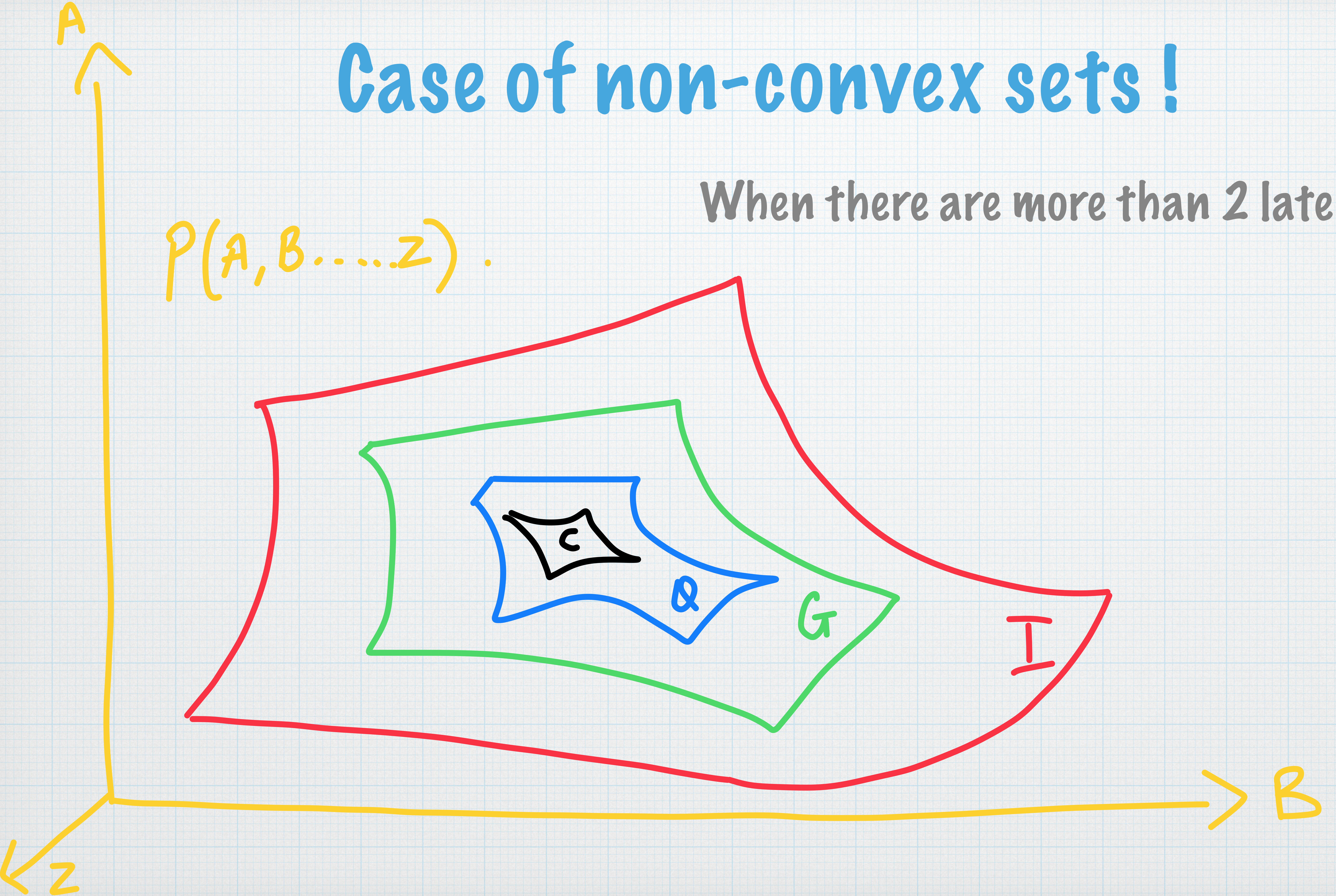
For a particular scenario



Not convex sets
in general !

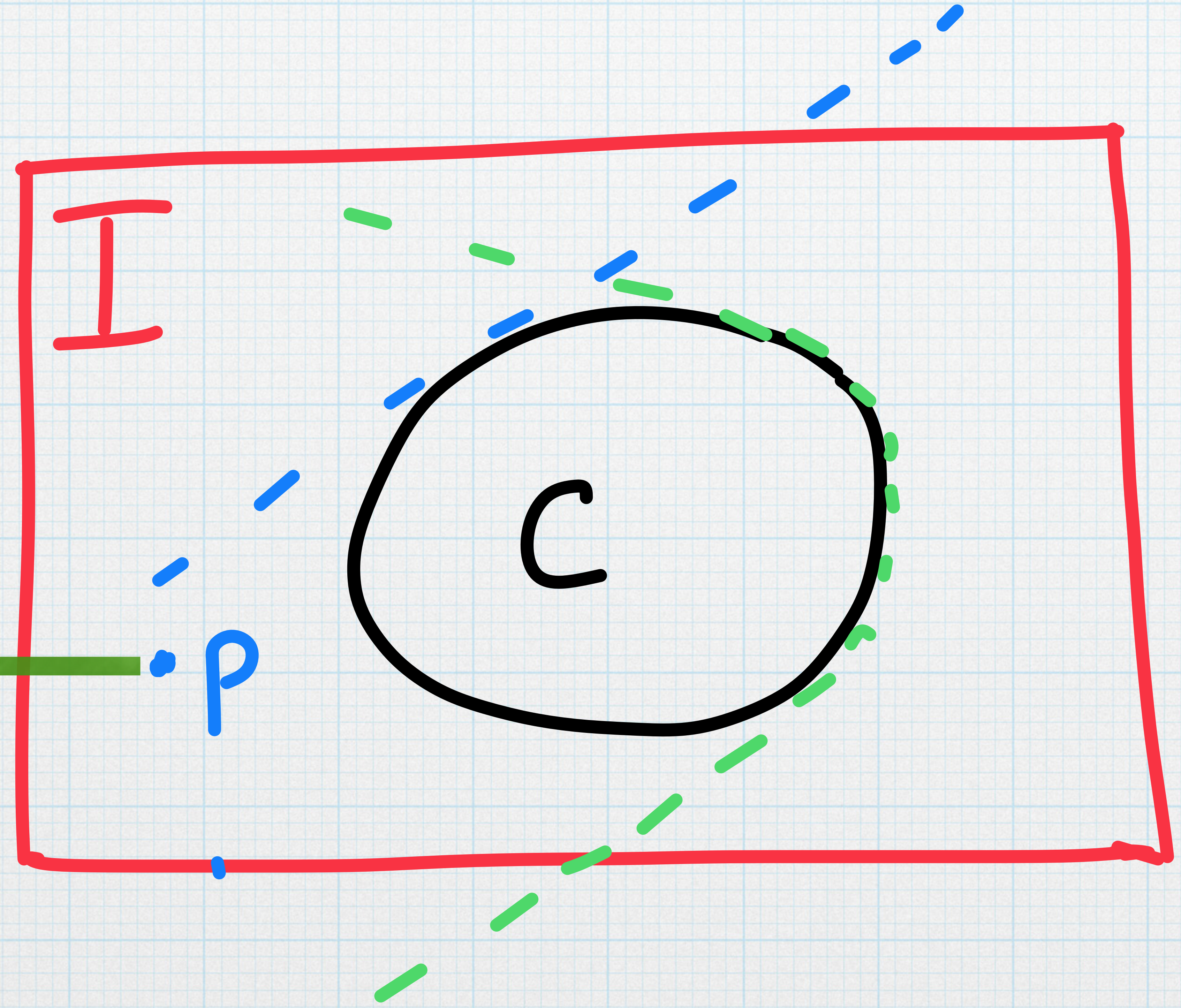
Case of non-convex sets !

When there are more than 2 latent variables !



"Interesting DAGs"

Only those DAGs which have $C \subset I$ can possibly support "Non-Classical" correlations and are termed "Interesting".



Non-Classical Probability Distribution

Very difficult to check if $C \subset I$

We need to attack the problem in a different way !

DAGs can help. Graphical criteria might simplify the difficult algebraic problem.

Henson, Lal and Pusey (HLP): Sufficient condition for "non-interestingness"

* Provided a series of graphical transformations which when met were proof of "non-interestingness".
* When not met the DAG could be "interesting" or not.
* Characterized all DAGs up to 6 nodes as "interesting" or not.
* Couldn't characterise DAGs of 7, 8.. nodes
* Could their condition be necessary as well ?

Introduction to mDAGs

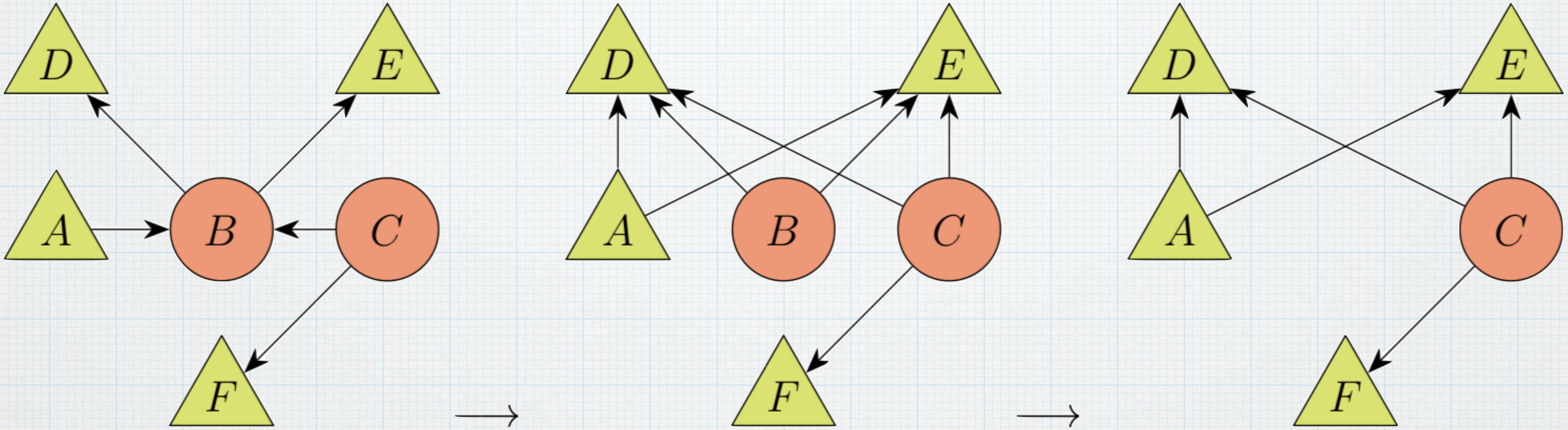
1. Exogenization: In a DAG G, with set of latent nodes $\{\lambda_i\}$, $\forall \lambda_i$ add edges $m \to n \ \forall \ m \in PA_G(\lambda_i)$ to every $n \in CH_G(\lambda_i)$ and delete the edges $m \to \lambda_i$ $\forall \ m \in PA_G(\lambda_i)$

2. Redundancy Removal: Delete all latent variables λ_i for which $CH_G(\lambda_i) \subseteq CH_G(\lambda_j)$ where λ_j is another latent variable s.t $\lambda_i \neq \lambda_j$ and $PA_G(\lambda_i) = PA_G(\lambda_j) = \phi$

These lead to another DAG G' s.t $C_G = C_{G'}$

G' will be called an mDAG.

Example of an mDAG transformation



Exogenization

Redundancy Removal

$$C_G = C_{G'}$$

HLP condition for mDAGs

If G and H are mDAGs s.t H can be obtained from G by applying one of the following transformations:

1. Removal of an edge

2. Addition of an edge $X \to Y$ where previously $PA(X) \subseteq PA(Y)$ and $PA(X)$ contained at least one latent node

Then $C_H \subseteq C_G$

How to apply the HLP conditions ?

1. From the mDAG G, using the transformations defined previously try to obtain an mDAG H , s.t H has no latent variables.

2. And that $I_G = I_H$.

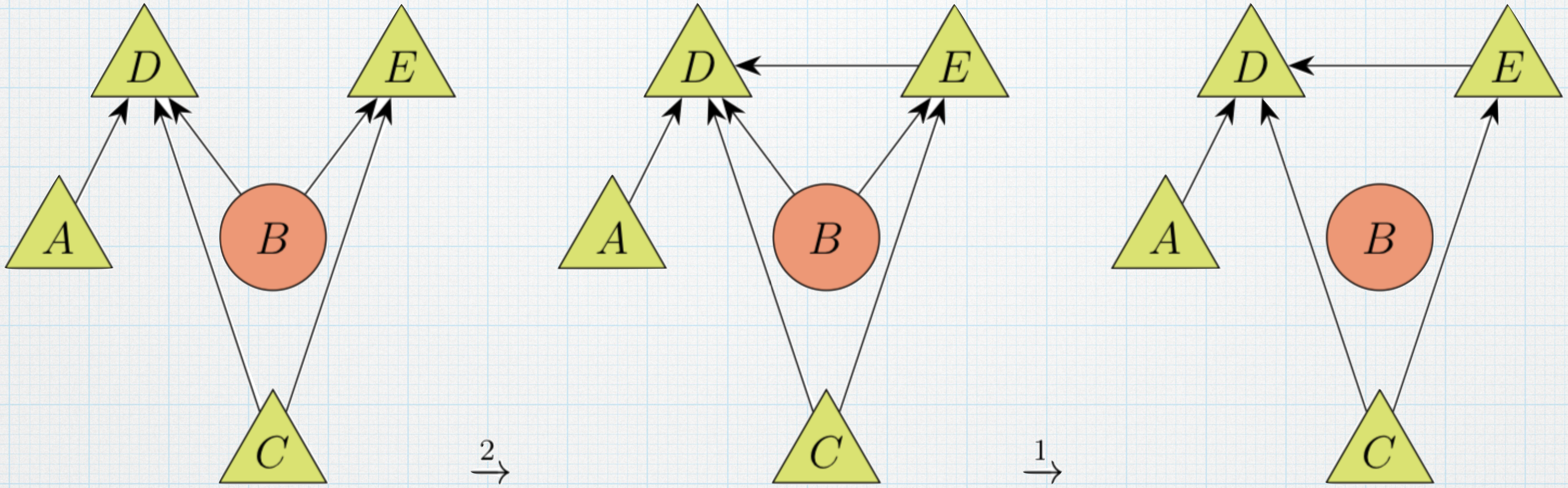
Since, $C_H \subseteq C_G$

$$\implies C_G = I_G$$

And $I_H = C_H = I_G$

And thus G is "non-interesting".

HLP condition example



Addition of an edge **Removal of edges**

$$C_G \supseteq C_H$$

The last mDAG has a latent variable which has no children, so we can simply delete it to get a latent free mDAG which by definition is "non-interesting".

HLP Conjecture !

That these transformations so introduced are both sufficient and necessary to certify “non-interestingness”.

That is,

If using these transformations and nothing more one can get an mDAG that is “non-interesting”, then the original mDAG is “non-interesting” as well, otherwise it is “interesting”.

Evan's result on mDAGs

Any mDAG, G is "non-interesting" if and only if $\exists$ another mDAG H that does not have any latent variables and for which $C_G = C_H$.

Because for the if part we have,

where $C_G = C_H \implies I_G = I_H$

$C_G \subseteq I_G$ *and* $C_G = C_H = I_H$

and thus, $C_G = I_G$

For the only if part refer: Evans(2023)

But HLP's condition is proven to be only a sufficient one.

So how do we test the remaining mDAGs that HLP's condition could certify as "non-interesting" ?

Need other methods that can certify "interestingness"…….and act them on these remaining mDAGs

Can we find other graphical conditions ?

Yes, we can !

Maximality,

d-separation,

e-separation,

Infeasible supports of probability distributions

$\Longrightarrow$ They show "Interestingness"

Refer ->

arXiv:2308.02380

Using d-separation to certify "interestingness"

If an mDAG G has a set of observed d-separation relations that cannot be produced by ANY latent free DAG, then G is "interesting".

Proof: $C_G = C_H \implies I_G = I_H$, the contrapositive leading to

$$I_G \neq I_H \implies C_G \neq C_H \quad \forall \text{ possible latent free } H$$

Hence by Evan's result G is "interesting".

Using e-separation to certify "interestingness"

Firstly, if for any 2 mDAGs, G and H, $C_G = C_H$ then their sets of observed e-separation relations must be identically the same (just like for d-separation).

If the observed e-separation relations in a mDAG, G cannot be reproduced by ANY latent free mDAG H, then G is "interesting".

Refer -> arXiv:2308.02380

Supports of a probability distribution

Given a probability distribution $P(X_1, \ldots. X_n)$ its support is defined as:

$$S(P(X_1, \ldots X_n)) = \{\{x_1, \ldots x_n\} \mid P(X_1 = x_1, \ldots. X_n = x_n) \geq 0\}$$

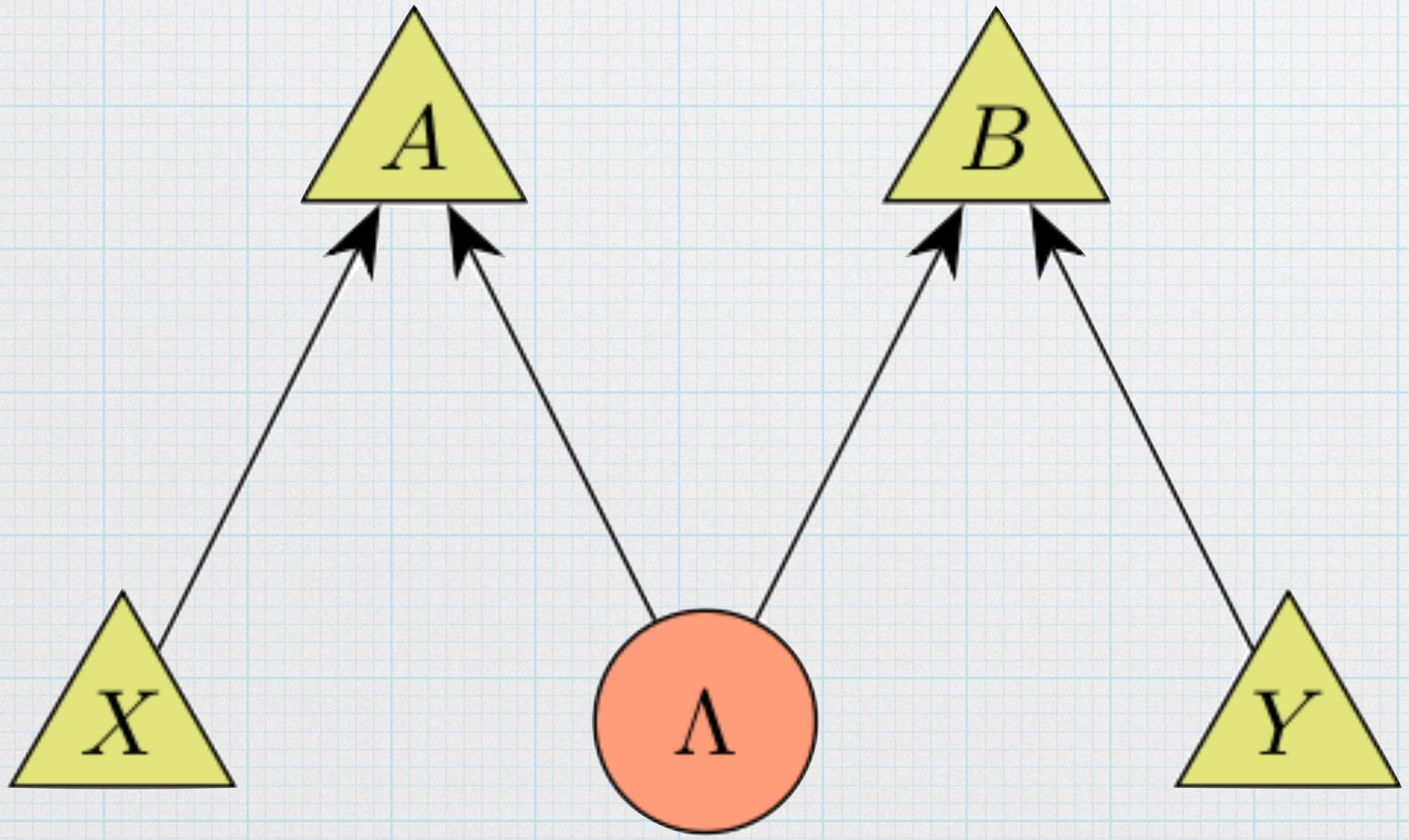
If there exists a $P \in C_G$ s.t $S(P) = S$, where S is a set of events, then we say that S is **classical** w.r.t G.

If there exists a $P \in I_G$ s.t $S(P) = S$, where S is a set of events, then we say that S is **classical-up-to-observed conditional independences** w.r.t G.

Fraser's Important Algorithm

Give a DAG to the algorithm, it finds supports of probability distributions that are not classically feasible.

For the Bell DAG its spits out PR Box and Hardy's supports



$$\mathcal{S}_{\text{Bell}} = \left\{ \begin{array}{l} \{X=0, Y=0, A=0, B=0\} \\ \{X=0, Y=0, A=1, B=1\} \\ \{X=0, Y=1, A=0, B=0\} \\ \{X=0, Y=1, A=1, B=1\} \\ \{X=1, Y=0, A=0, B=0\} \\ \{X=1, Y=0, A=1, B=1\} \\ \{X=1, Y=1, A=1, B=0\} \\ \{X=1, Y=1, A=0, B=1\} \end{array} \right\}$$

PR-Box supports

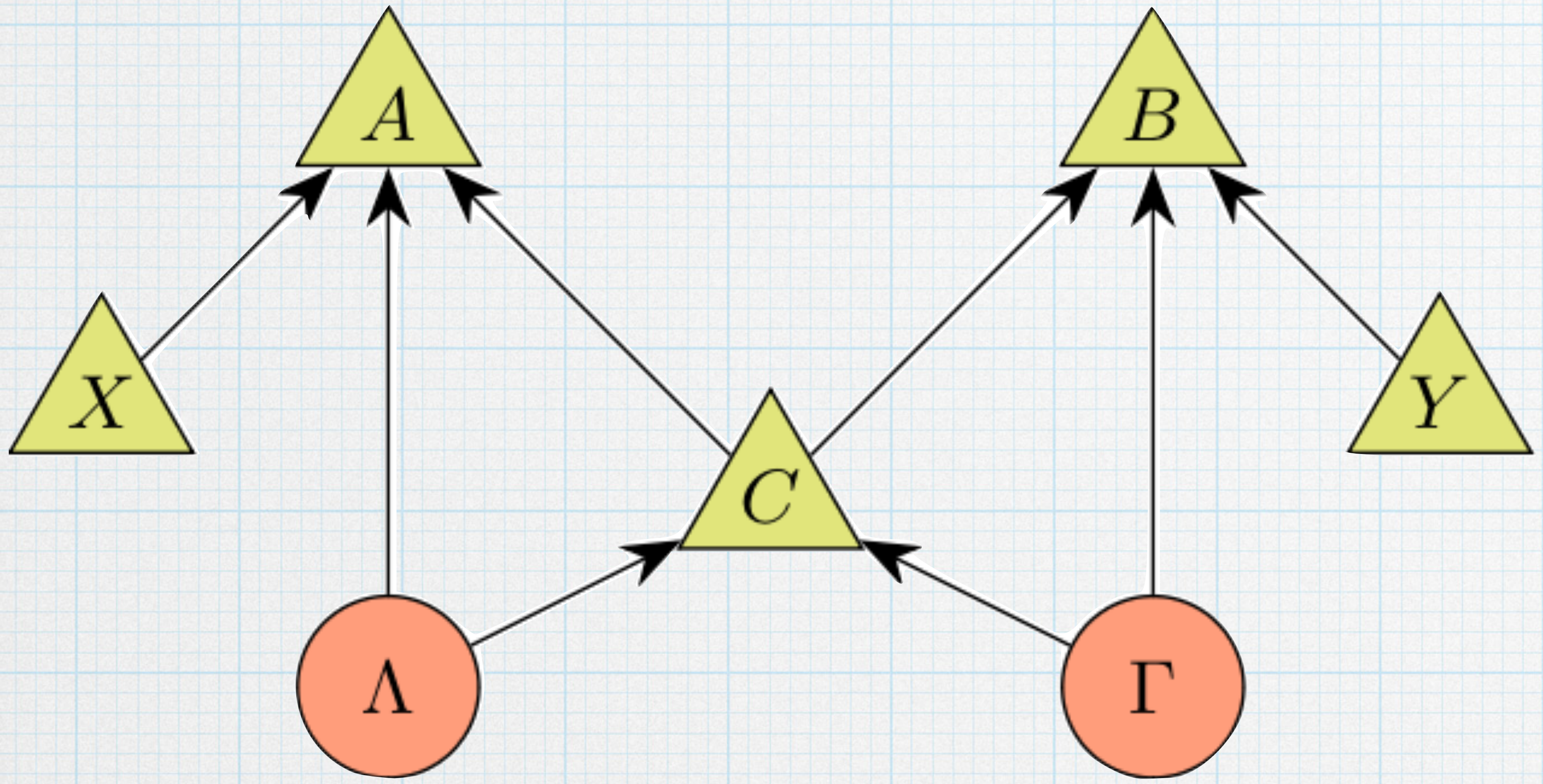
Classically infeasible supports for “interestingness”

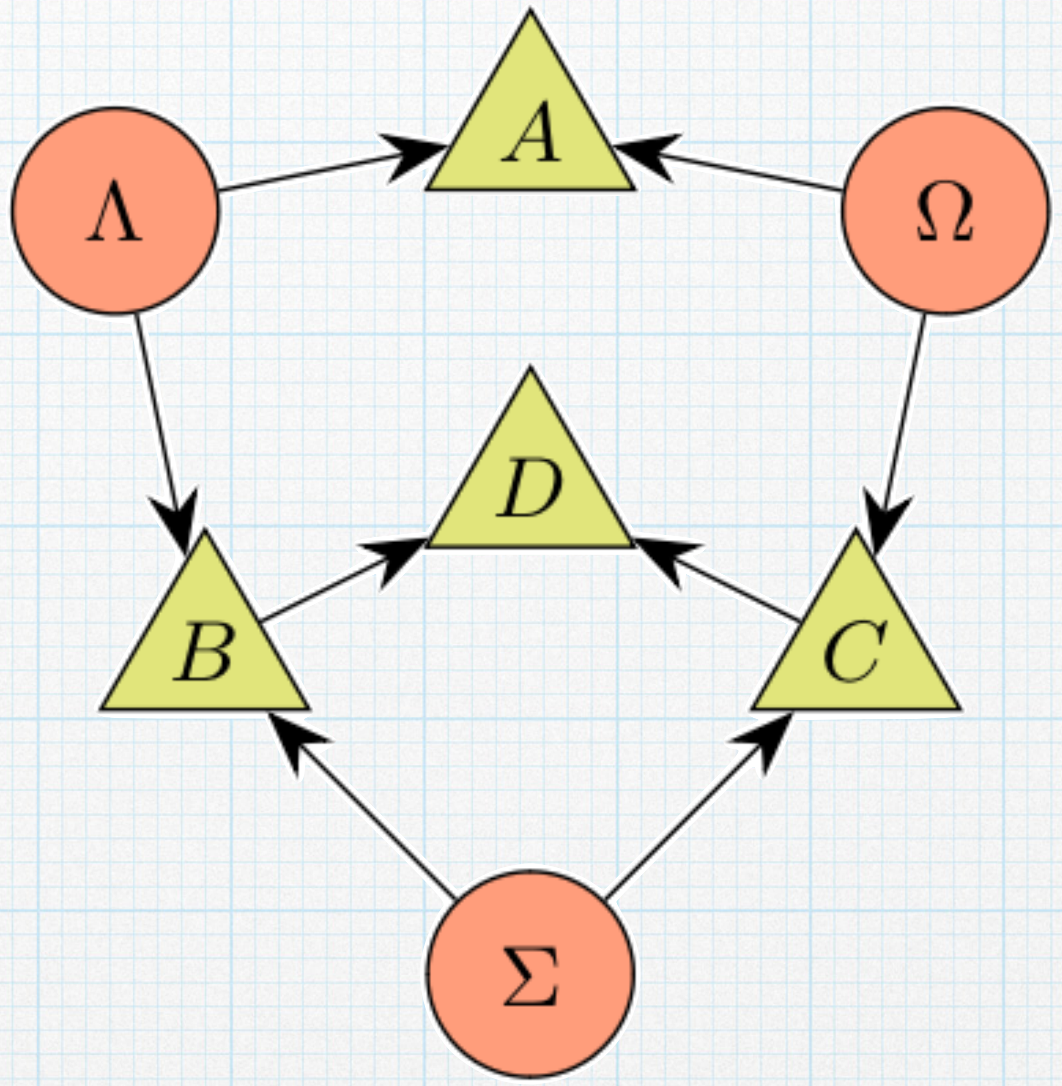
If two mDAGs G and H s.t $C_G = C_H$ then their sets of classical supports must be identical (unknown if this could be only-if as well).

If an mDAG, G has a set of classical supports that cannot be reproduced in ANY latent free mDAG, then G is “interesting”.

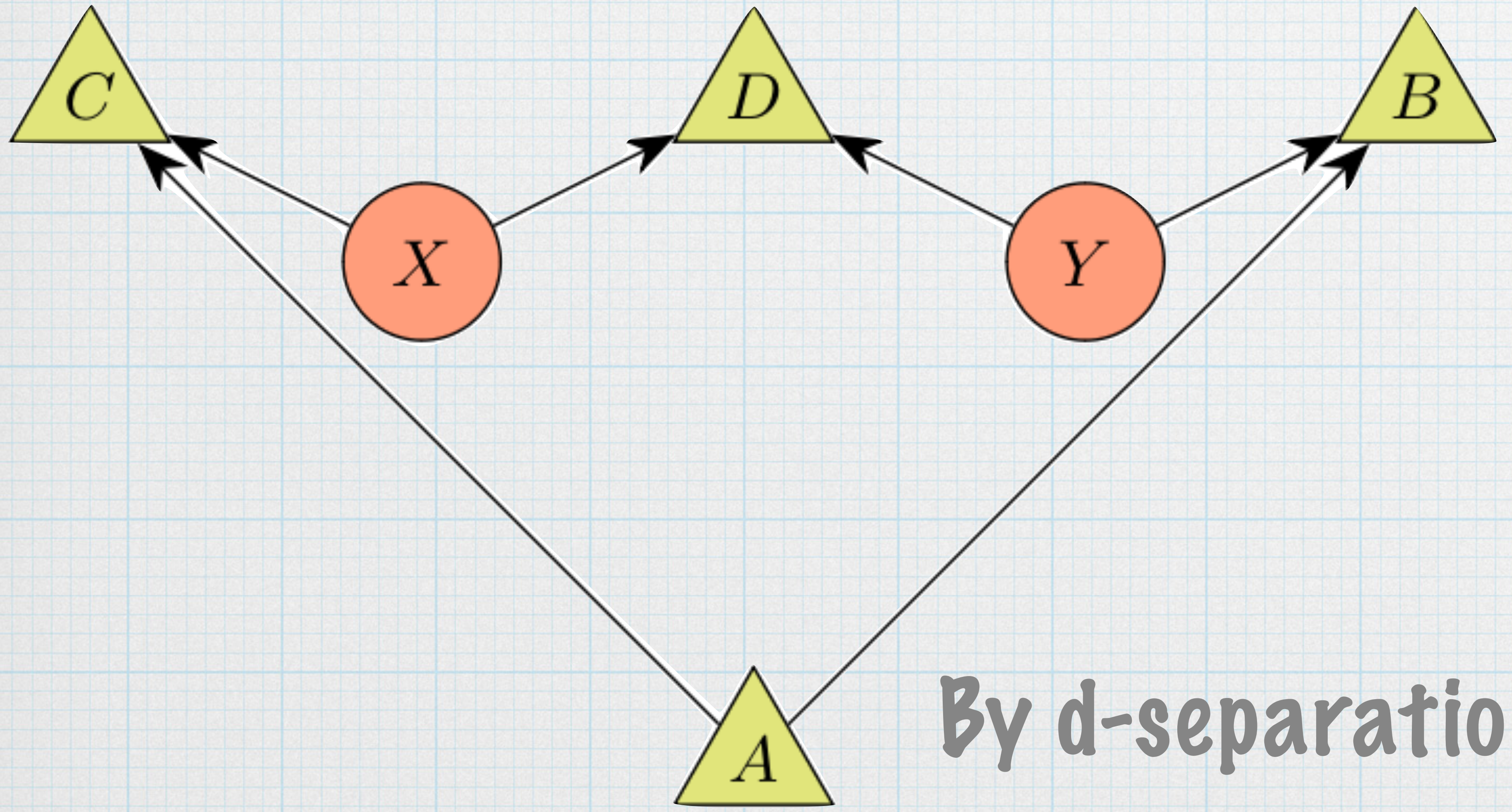
Refer -> arXiv:2308.02380

Some "interesting" DAGs we found

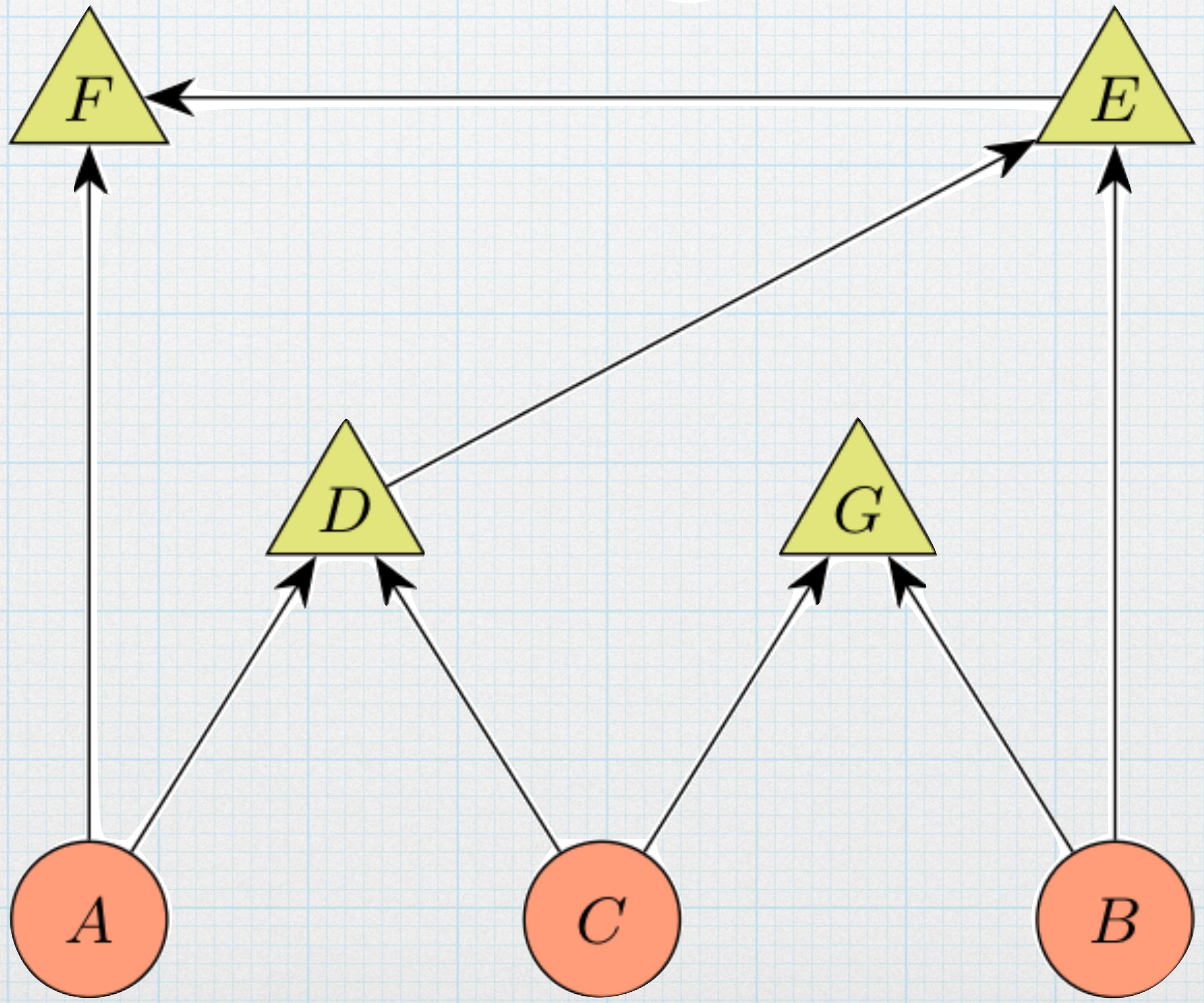




We not only find that they are "interesting" but find the exact probability distributions that are non-classical !

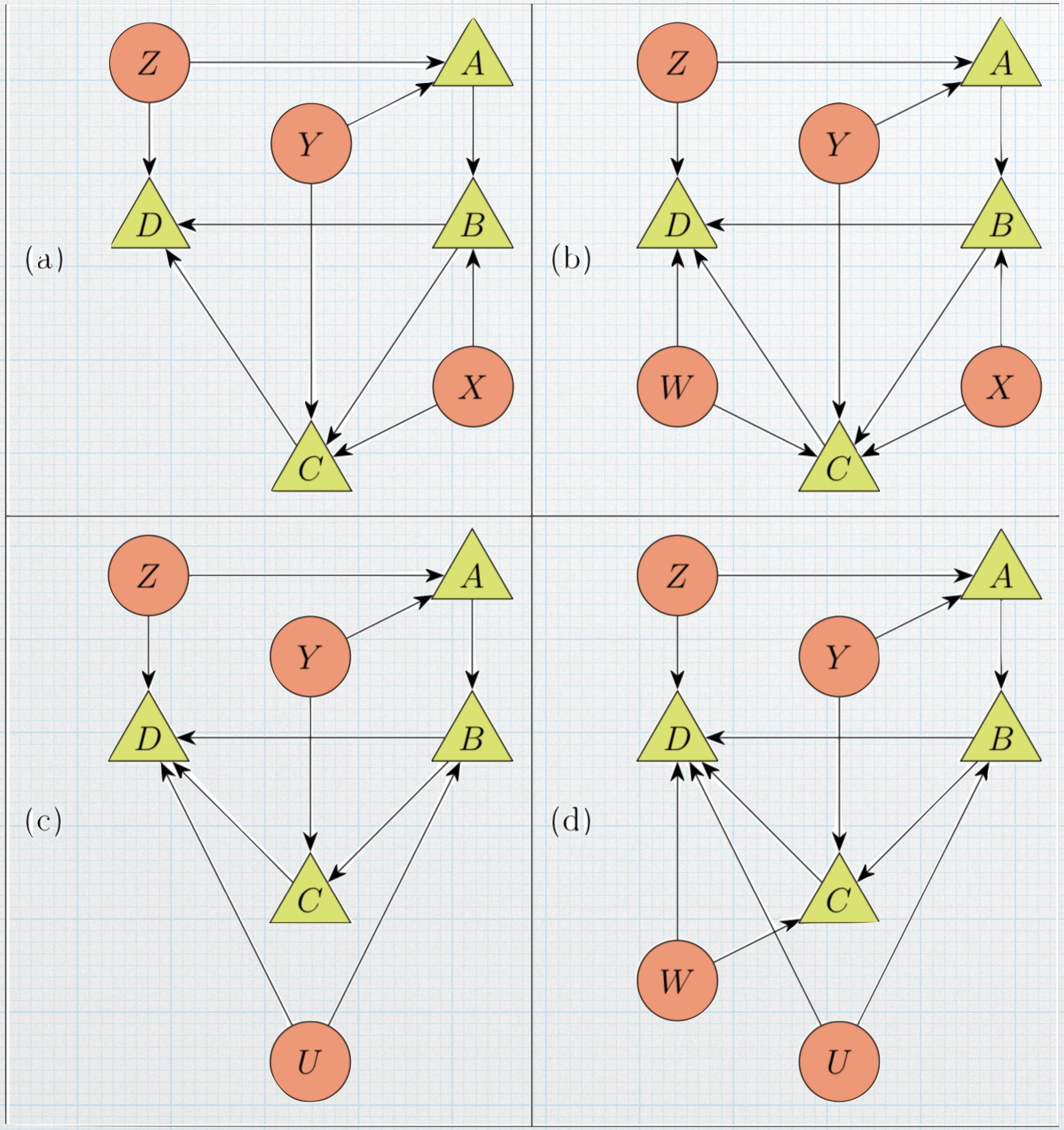


By d-separation



By e-separation

Examples for Supports method



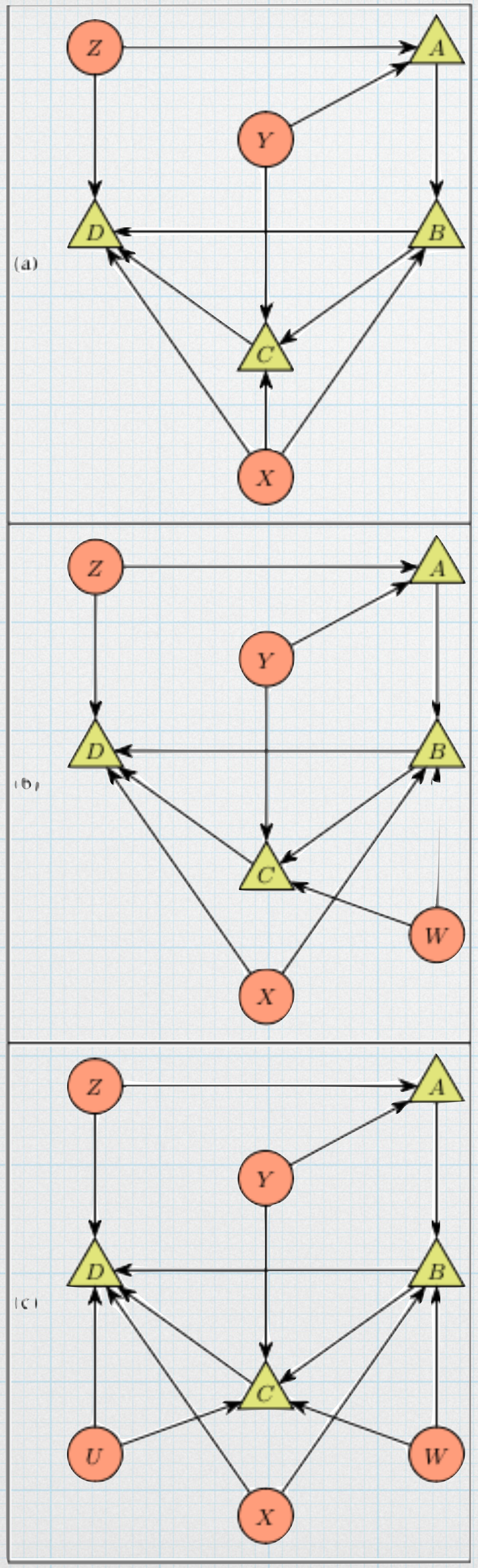
$$\mathcal{S}_{\text{for Table 2}} = \left\{ \begin{array}{c} \{A=0, B=0, C=0, D=0\} \\ \{A=0, B=0, C=1, D=0\} \\ \{A=0, B=1, C=0, D=0\} \\ \{A=1, B=0, C=0, D=0\} \\ \{A=1, B=1, C=0, D=0\} \\ \{A=2, B=0, C=0, D=1\} \\ \{A=2, B=1, C=1, D=0\} \end{array} \right\}$$

Infeasible Support

Computational Results

Category	DAGs with 3 observed nodes	DAGs with 4 observed nodes	DAGs with 5 observed nodes
Total Count of DAGs	46	2809	1,718,596
DAGs remaining after HLP condition (since it is only a sufficient condition)	5	996	1,009,961
DAGs remaining after various graphical criteria, like Maximality, d-separation, e-separation, Infeasible supports of Probability distributions.....	0	3	< 12,834

$\approx$ 99% reduction of uncharacterised DAGs $\Longrightarrow$ HLP condition looks to be necessary as well !



3 unclassified mDAGs

Shannon cones corresponding to sets C and I are the same for these 3 mDAGs, so no difference can be found at the level of Shannon entropic inequalities.

What to do- Explore Non Shannon type inequalities or accelerate Fraser's algorithm to solve these 3.

Summary and Future work

* Evidence towards HLP condition being necessary as well.
* Several graphical criteria to check "interestingness".
* Explicit construction of "Non-Classical" distributions.
* These scenarios can exhibit classical-quantum or post quantum gap.
* Potential candidates for exhibiting quantum or post quantum advantage.
* Importance for classical causal inference (in ML, AI)
* Attacking specific scenarios to confirm classical-quantum advantage.

Thank You !