Title: Causal Inference Lecture - 230403

Speakers: Robert Spekkens

Collection: Causal Inference: Classical and Quantum

Date: April 03, 2023 - 10:00 AM

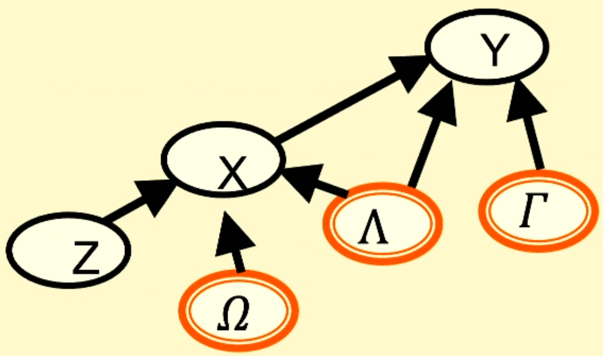
URL: https://pirsa.org/23040000

Abstract: zoom link: https://pitp.zoom.us/j/94143784665?pwd=VFJpajVIMEtvYmRabFYzYnNRSVAvZz09

Final comments on causal compatibility for classical models

Up to now we have considered

The problem of **causal compatibility:**
Which causal structures are compatible with the data?
(nontrivial if some or all of the data is purely observational)

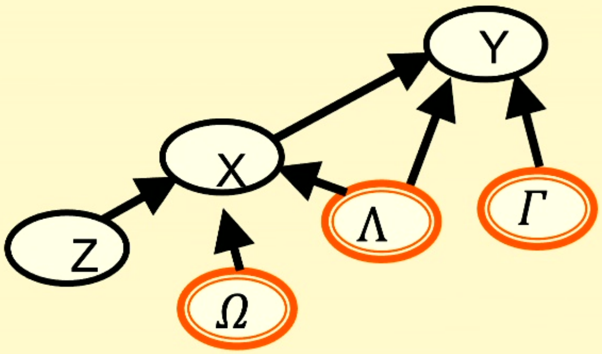


$$X = f(\Lambda, \Gamma, Z)$$
$$Y = g(\Lambda, \Omega, X)$$
$$P_\Lambda$$
$$P_\Gamma$$
$$P_\Omega$$

$$P_{XY|Z} = \sum_{\Lambda,\Gamma,\Omega} \delta_{X,f(\Lambda,\Omega,Z)} \delta_{Y,g(\Lambda,\Gamma,X)} P_\Lambda$$

Restricted functional dependences

- Linear dependences
- Monotonic functions
- Symplectic functions
- Local noise is additive
  - symmetries



$$X = f(\Lambda, \Gamma, Z)$$
$$Y = g(\Lambda, \Omega, X)$$
$$P_\Lambda$$
$$P_\Gamma$$
$$P_\Omega$$

$$P_{XY|Z} = \sum_{\Lambda,\Gamma,\Omega} \delta_{X,f(\Lambda,\Omega,Z)} \delta_{Y,g(\Lambda,\Gamma,X)} P_\Lambda$$

Restricted functional dependences

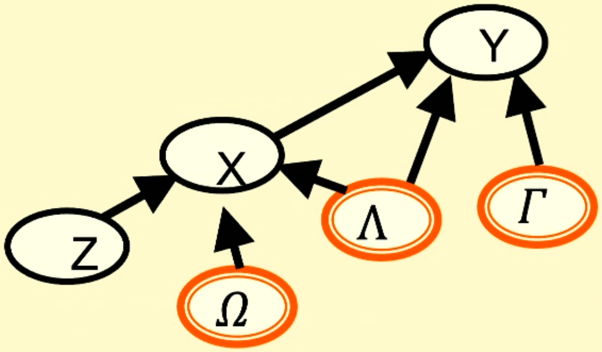
- Linear dependences
- Monotonic functions
- Symplectic functions
- Local noise is additive
- symmetries

Restricted distributions of latents

- Only Gaussian distributions
- Restricted cardinalities
- symmetries

Strength of causal conclusions $\propto$ Strength of causal assumptions

Scope for new techniques all along the spectrum



$X = f(\Lambda, \Gamma, Z)$

$Y = g(\Lambda, \Omega, X)$

P_Λ

P_Γ

P_Ω

$f \in \mathcal{F}$

$g \in \mathcal{G}$

$P_\Lambda \in \mathcal{P}_\Lambda$

$P_\Gamma \in \mathcal{P}_\Gamma$

$P_\Omega \in \mathcal{P}_\Omega$

$$P_{XY|Z} = \sum_{\Lambda,\Gamma,\Omega} \delta_{X,f(\Lambda,\Omega,Z)} \delta_{Y,g(\Lambda,\Gamma,X)} P_\Lambda \implies P_{XY|Z} \in \mathcal{P}_{XY|Z}$$

Adjudicating between causal models

The problem of **adjudicating between causal models**

Given a set of causal structures that are all compatible with the data, which is most likely to be the correct explanation?

Contrast:

Causal explanations of the infinite-run statistics predicted by an operational theory

vs.

Causal explanations of the finite-run statistics accumulated in a real-world experiment or observation

Example of contrast:

No-go theorem establishing that the idealized statistics predicted by operational quantum theory are incompatible with a classical causal model having the causal structure of the Bell DAG.

vs.

An analysis technique for finite-run experimental data that can rule out with high confidence the possibility of a classical causal model having the causal structure of the Bell DAG

Example of contrast:

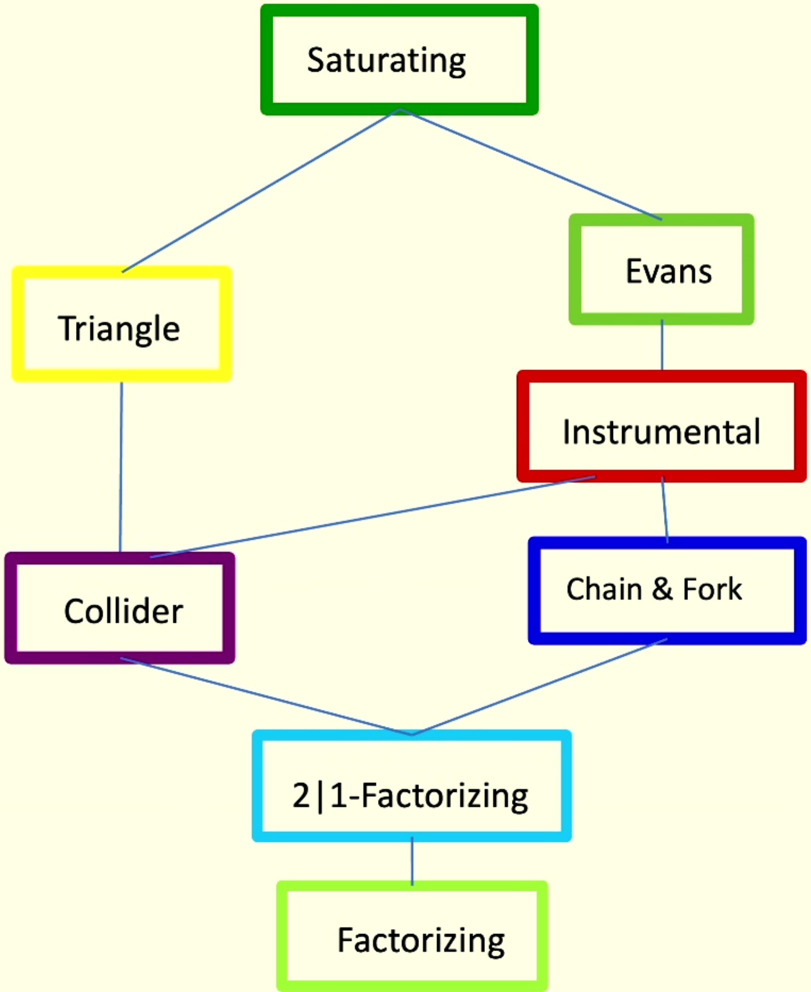
No-go theorem establishing that the idealized statistics predicted by operational quantum theory are incompatible with a classical causal model having the causal structure of the Bell DAG.
E.g., Bell's 1964 argument which appealed to perfect correlations
Hardy's 1993 argument which appealed to events with probability 0

vs.

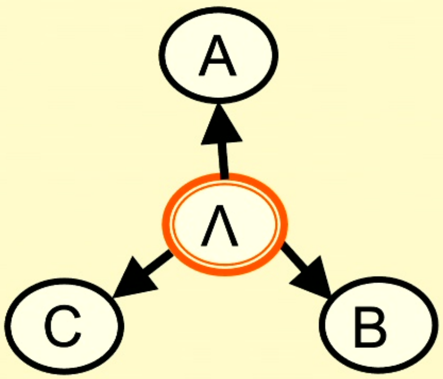
An analysis technique for finite-run experimental data that can rule out with high confidence the possibility of a classical causal model having the causal structure of the Bell DAG
E.g., The noise-robust inequalities proposed by CHSH

Model preference in the idealized scenario

Given the observational dominance order, it makes sense to prefer the models that are lowest in the order, since these are the most falsifiable

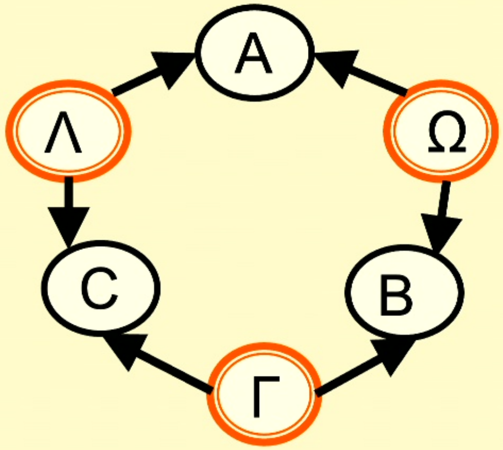

Saturating
Triangle
Evans
Instrumental
Collider
Chain & Fork
2|1-Factorizing
Factorizing

Saturating class



7 parameters

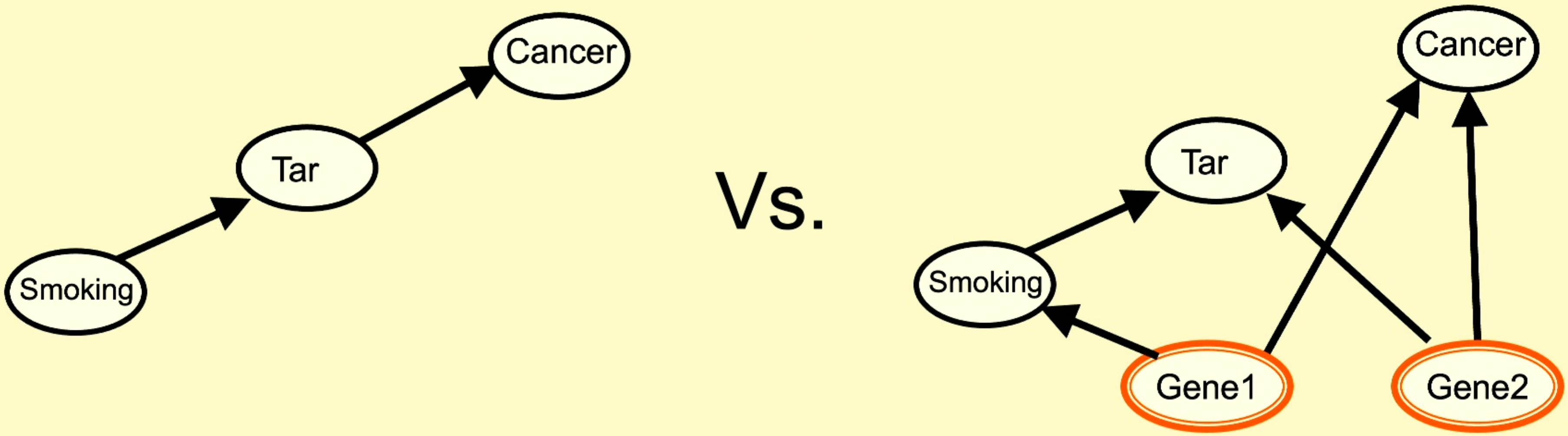
Triangle class



>9 parameters

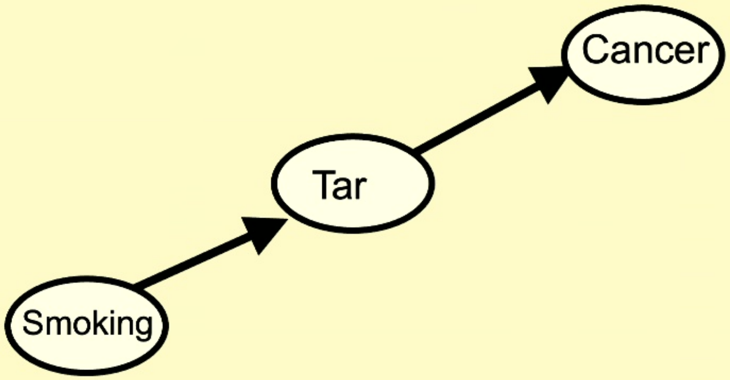
Parameter counting does not capture falsifiability

See also the notion of VC dimension

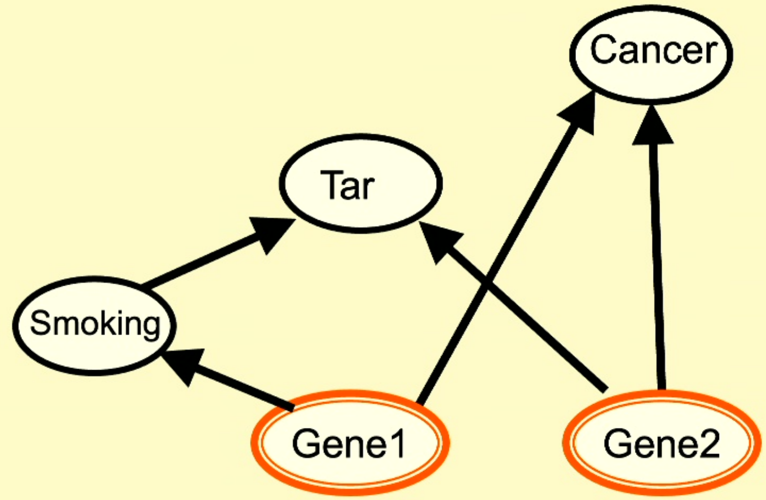


Observe P_{STC} such that

$$S \perp C | T$$



Vs.



Requires fine-tuning

Observe P_{STC} such that

$$S \perp C | T$$

Principle of faithfulness/no fine-tuning:

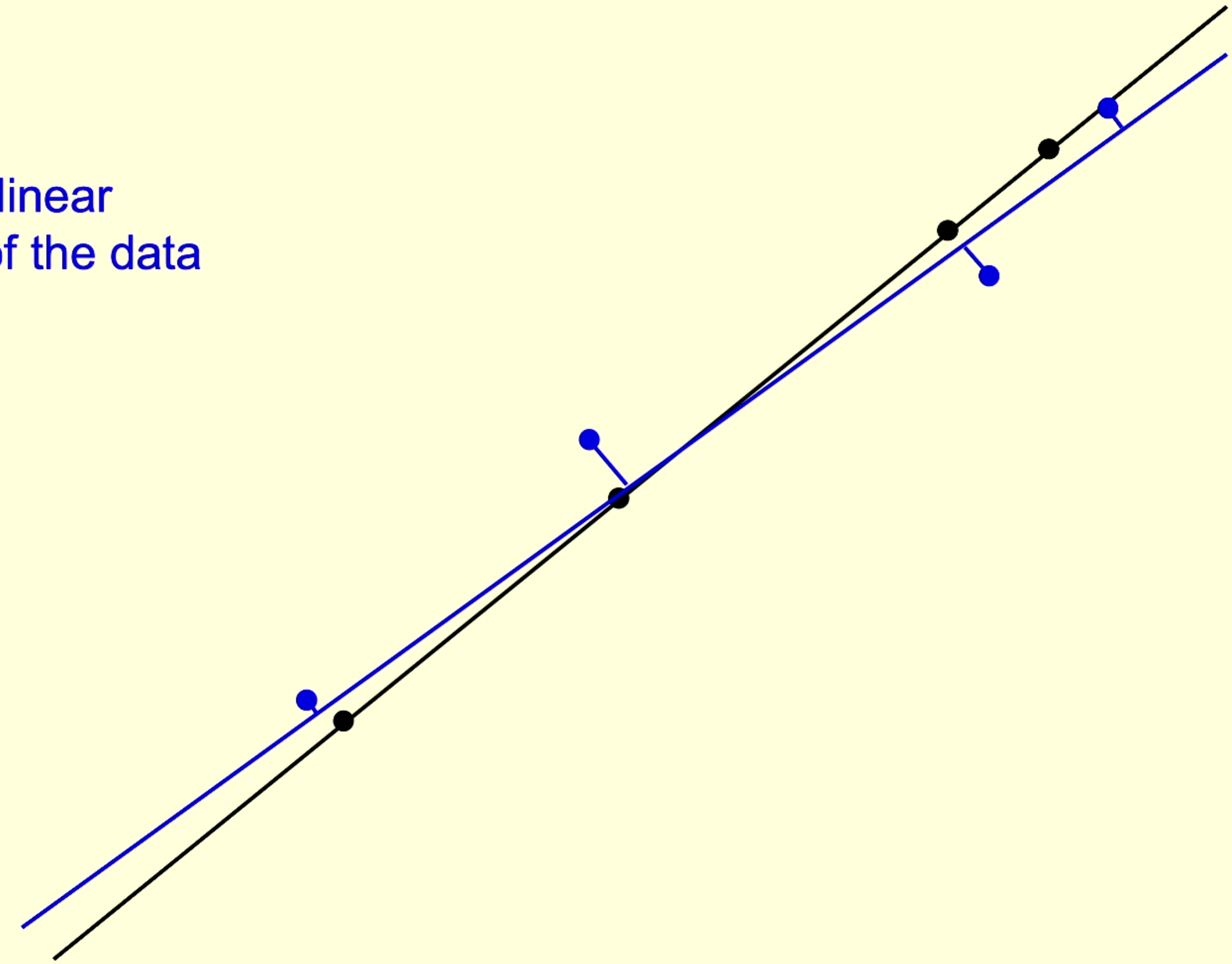
Prefer those causal models for which the conditional independence relations are a consequence of the causal structure rather than the values of the parameters

Finite-run statistics and model-fitting

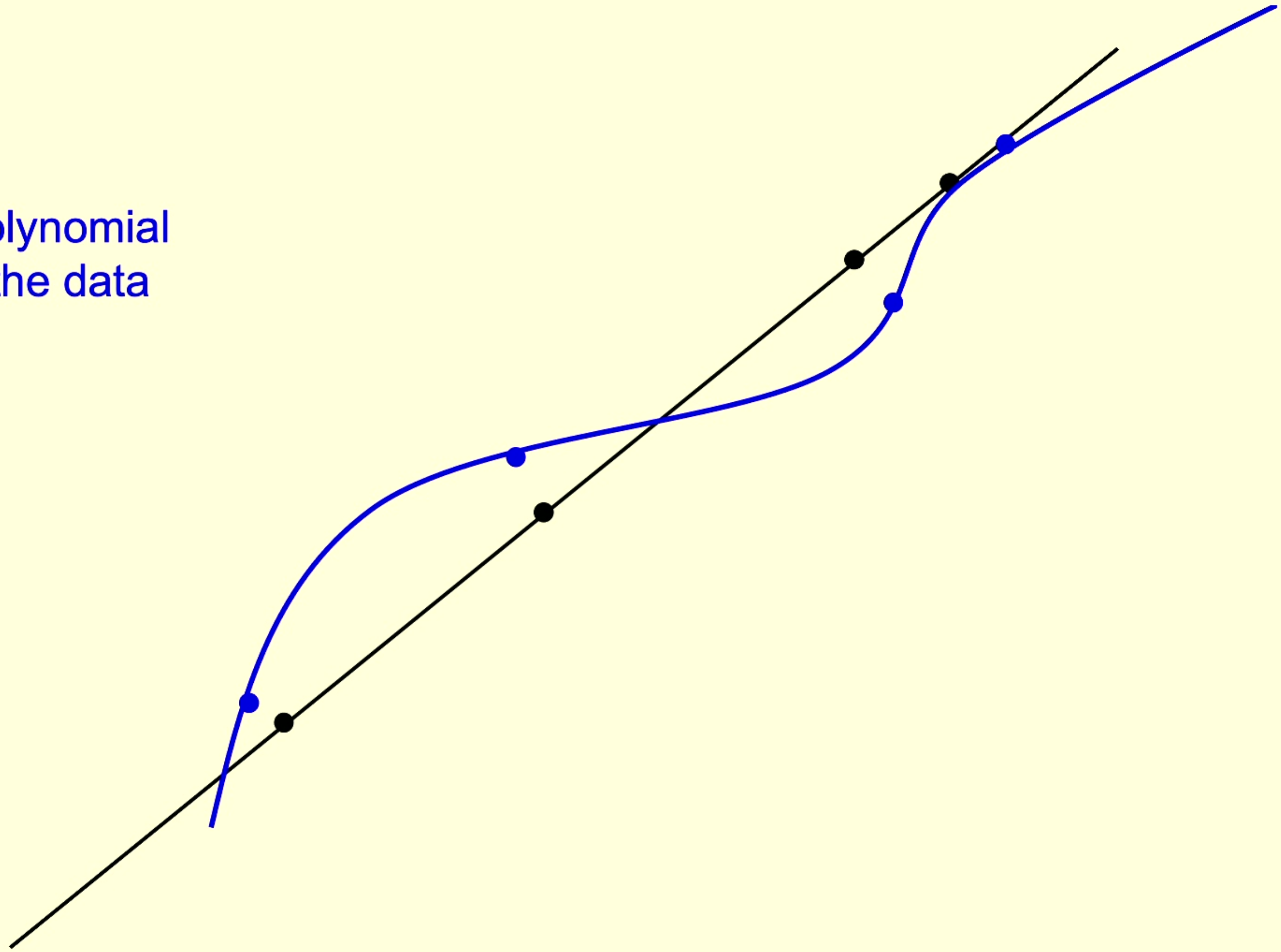
The low bar:

Not underfitting the data

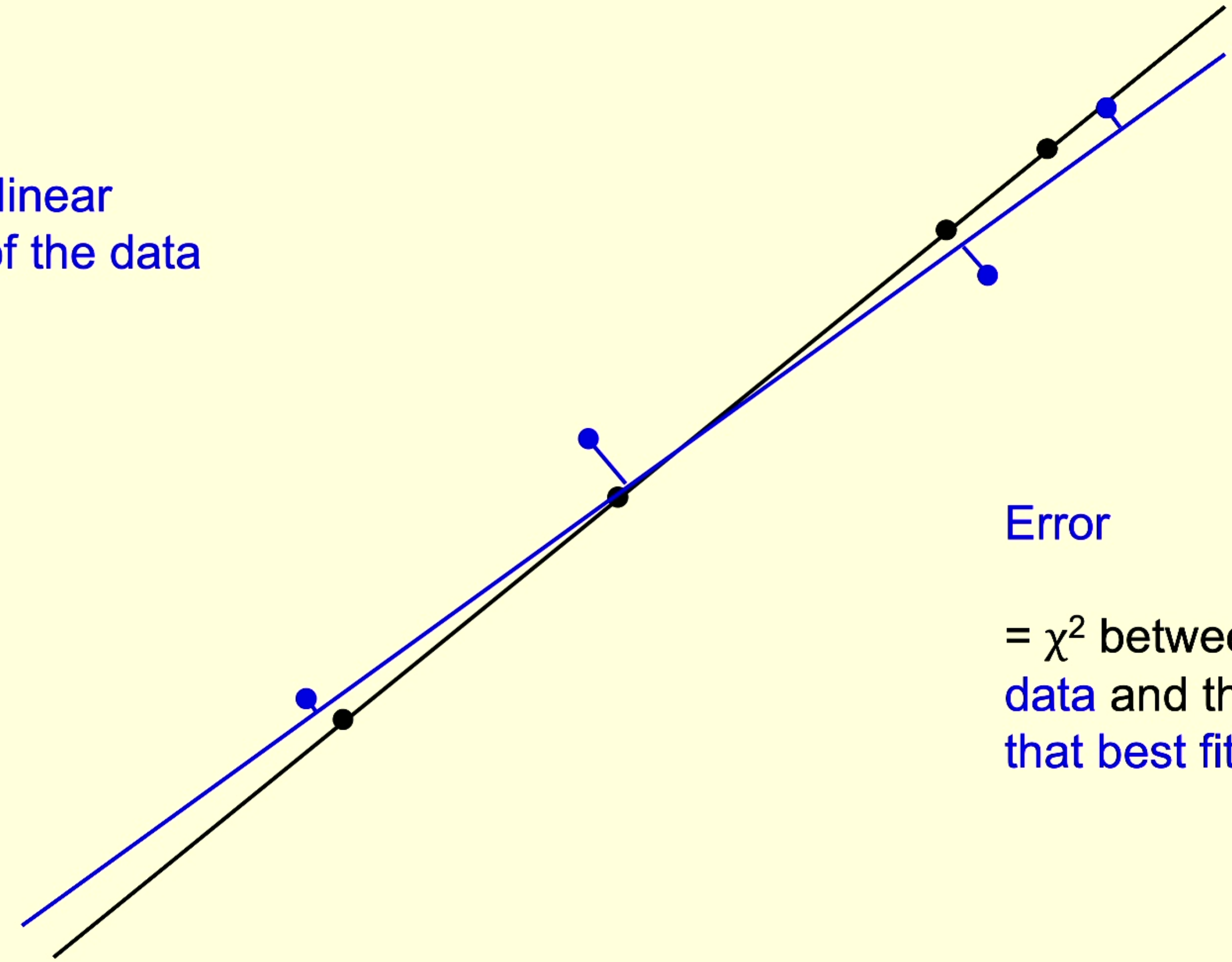
Data
&
Best-fit linear
model of the data



Data
&
Best-fit polynomial
model of the data

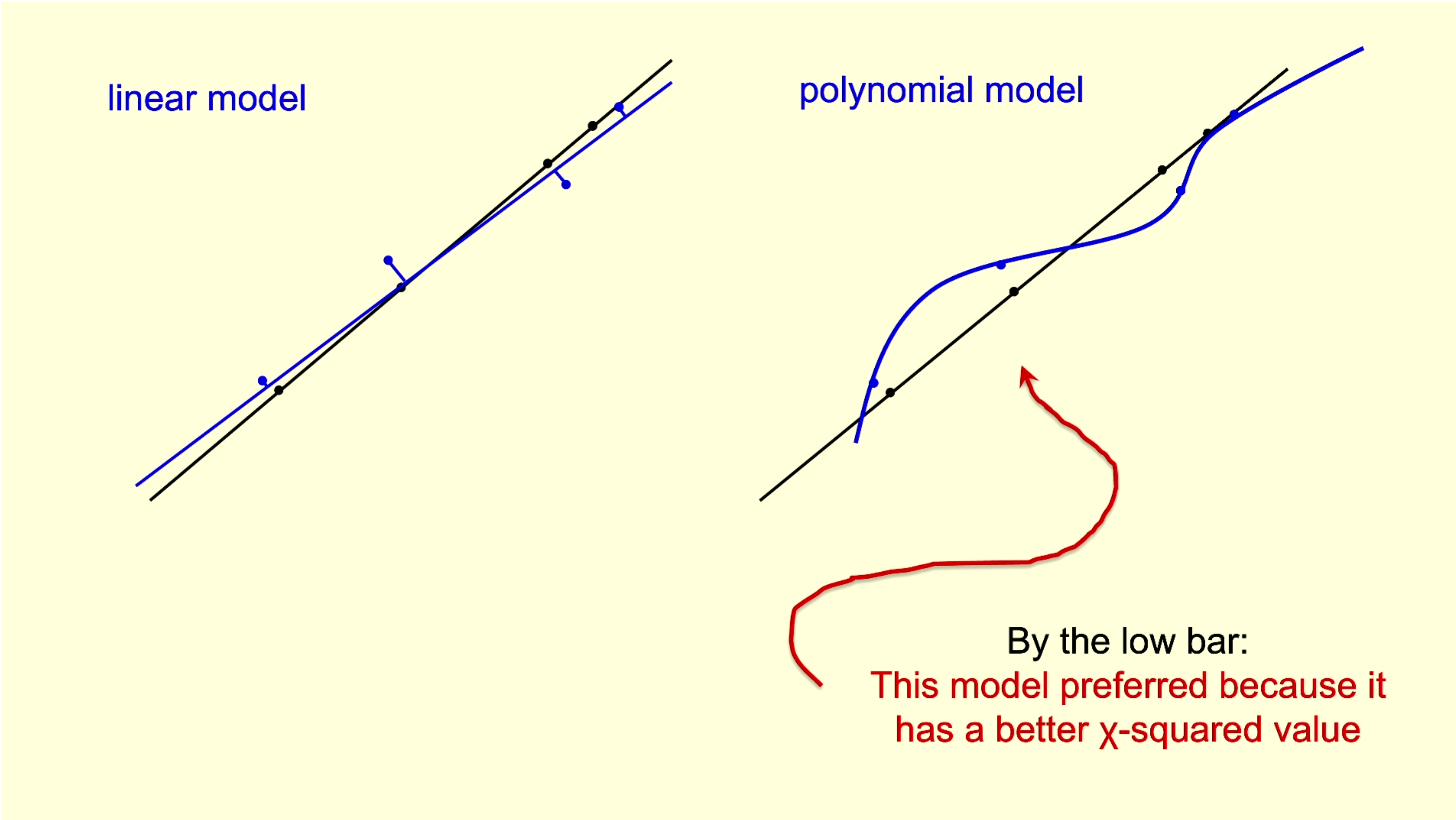


Data
&
Best-fit linear
model of the data



Error

= χ^2 between the
data and the model
that best fits the data

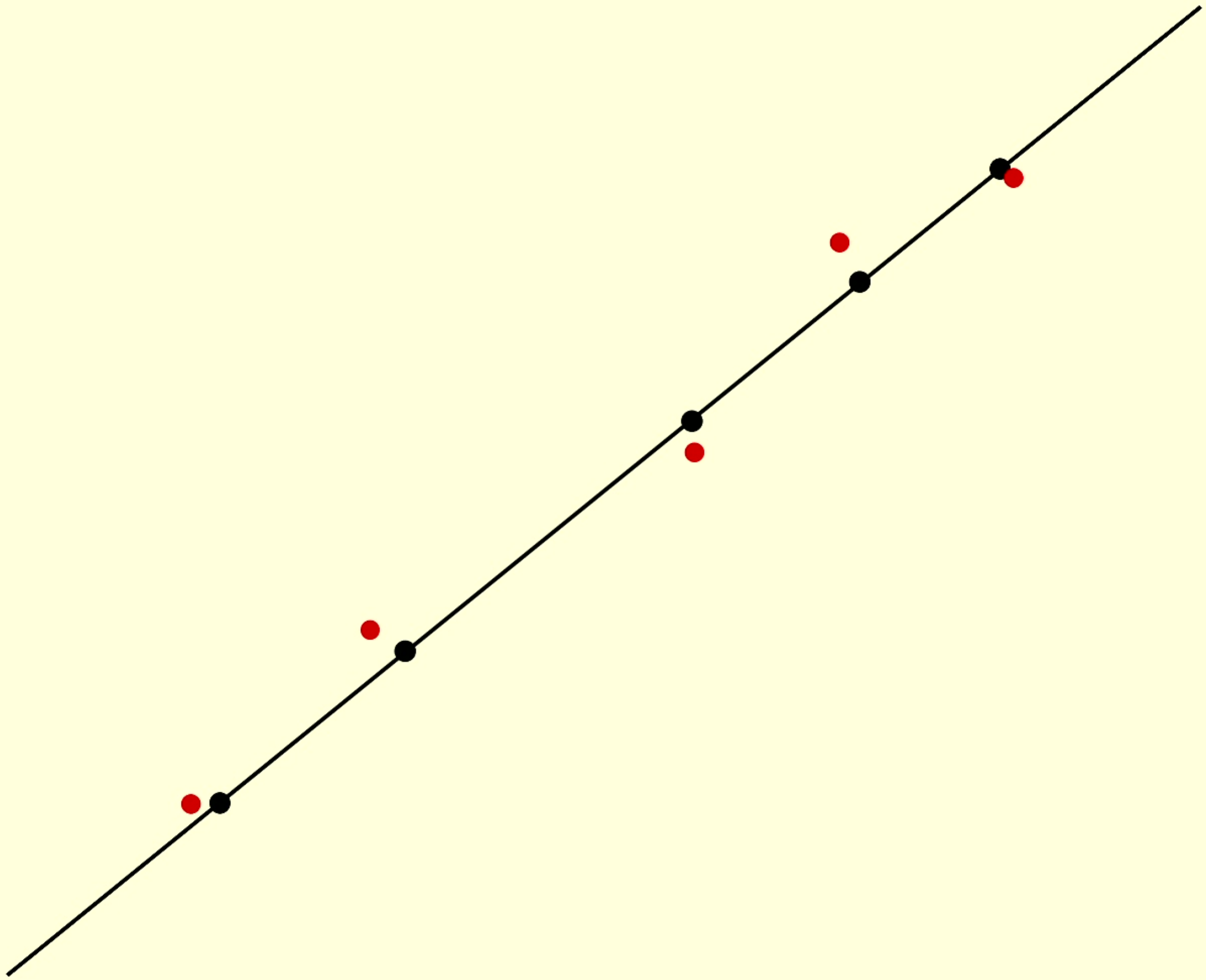

linear model
polynomial model
By the low bar:
This model preferred because it
has a better χ-squared value

The high bar:

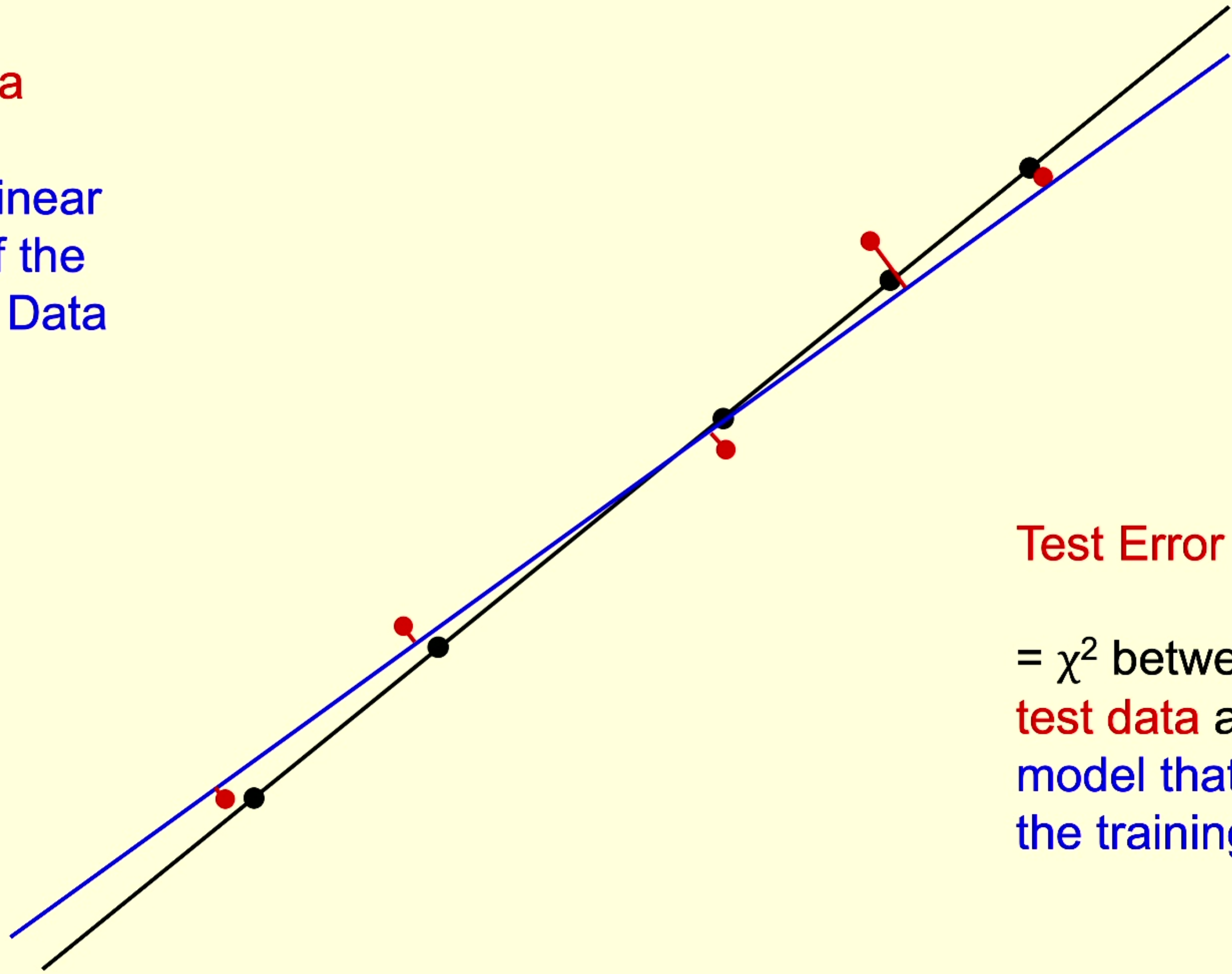
Predictive power

not underfitting the data
and also
not **overfitting** the data

Test Data

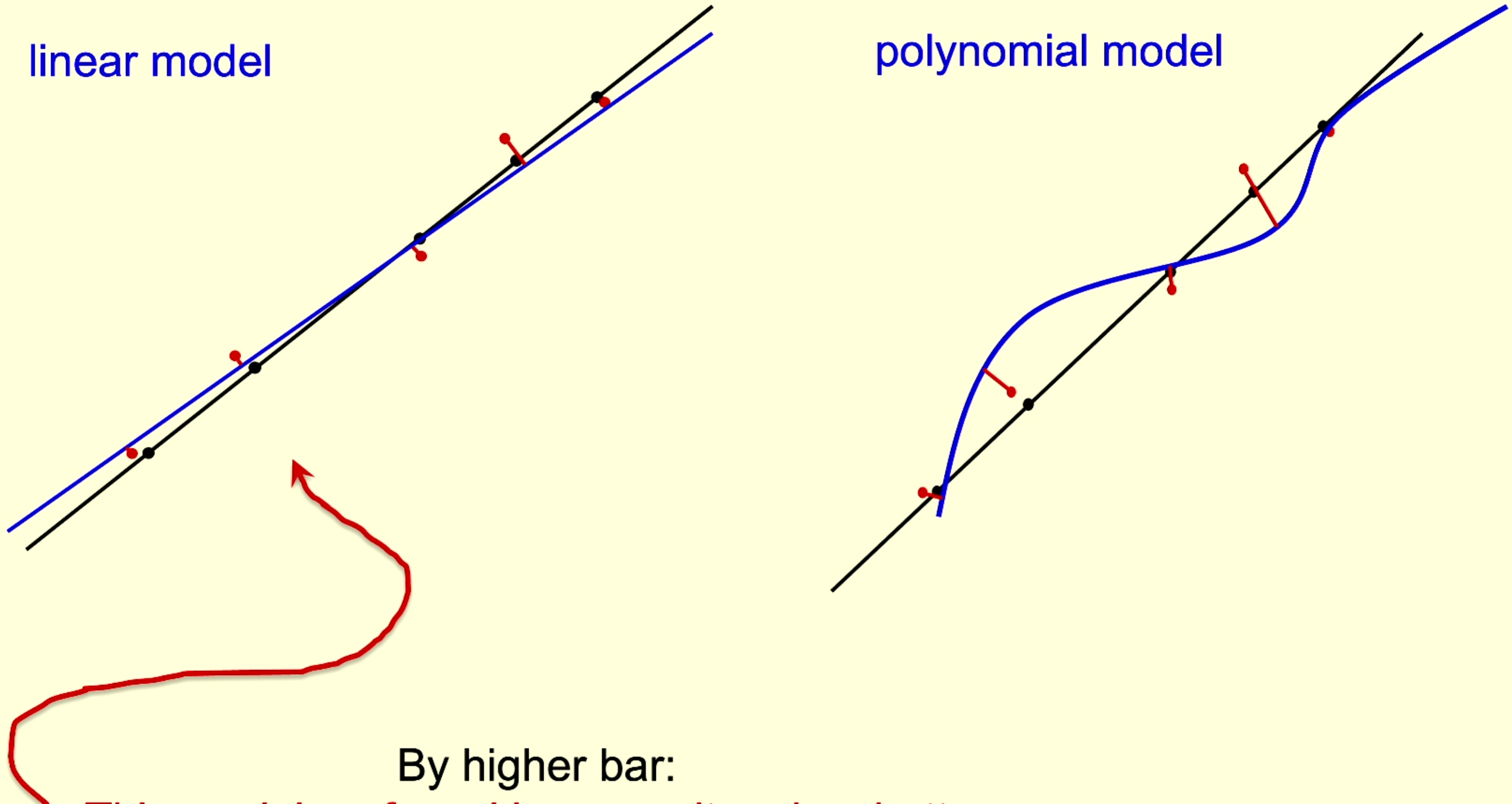


Test Data
&
Best-fit linear
model of the
Training Data



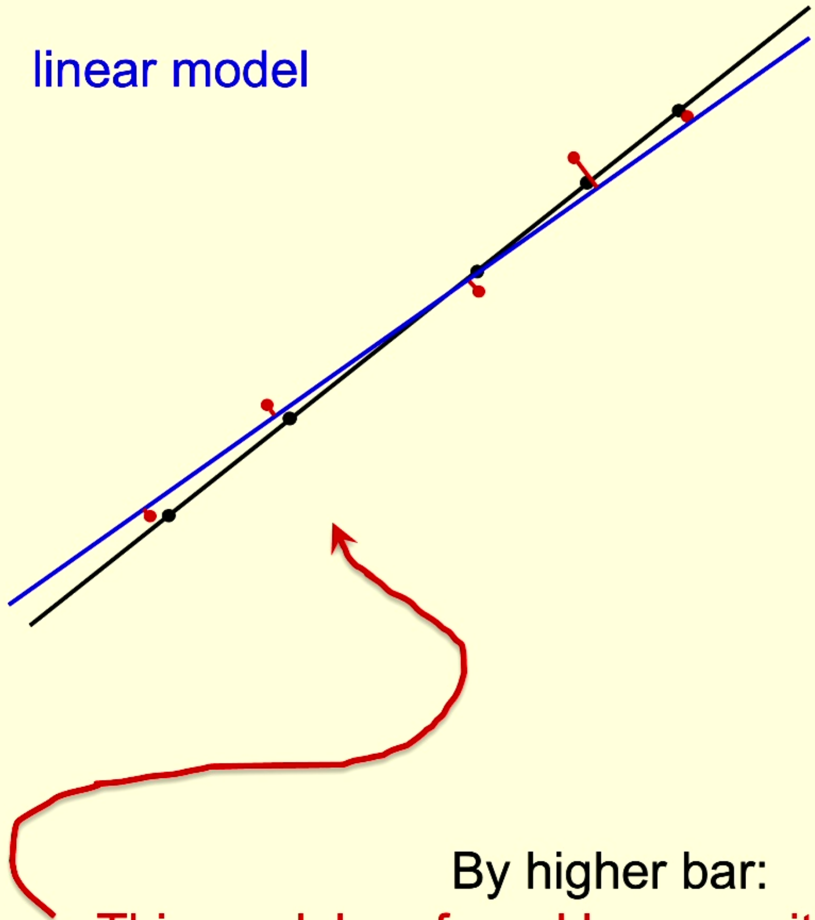
Test Error

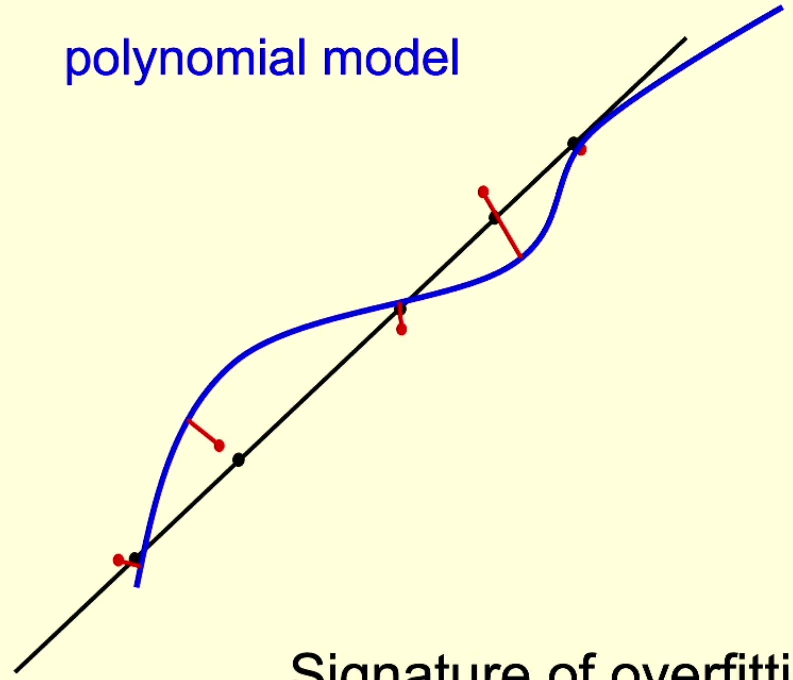
= χ^2 between the
test data and the
model that best fits
the training data



By higher bar:

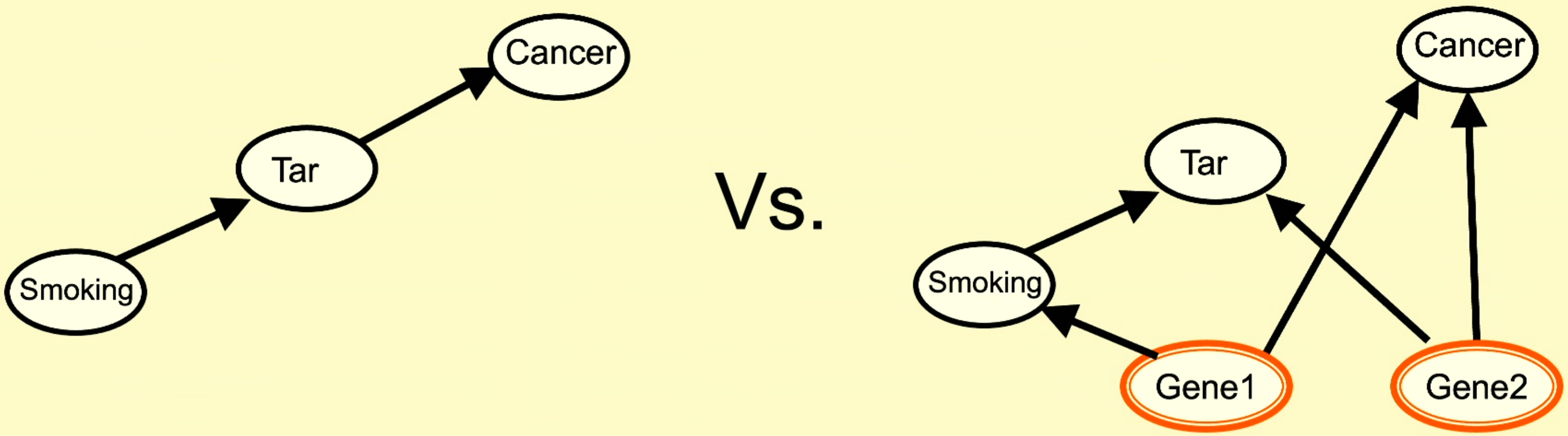
This model preferred because it makes better predictions about unseen data





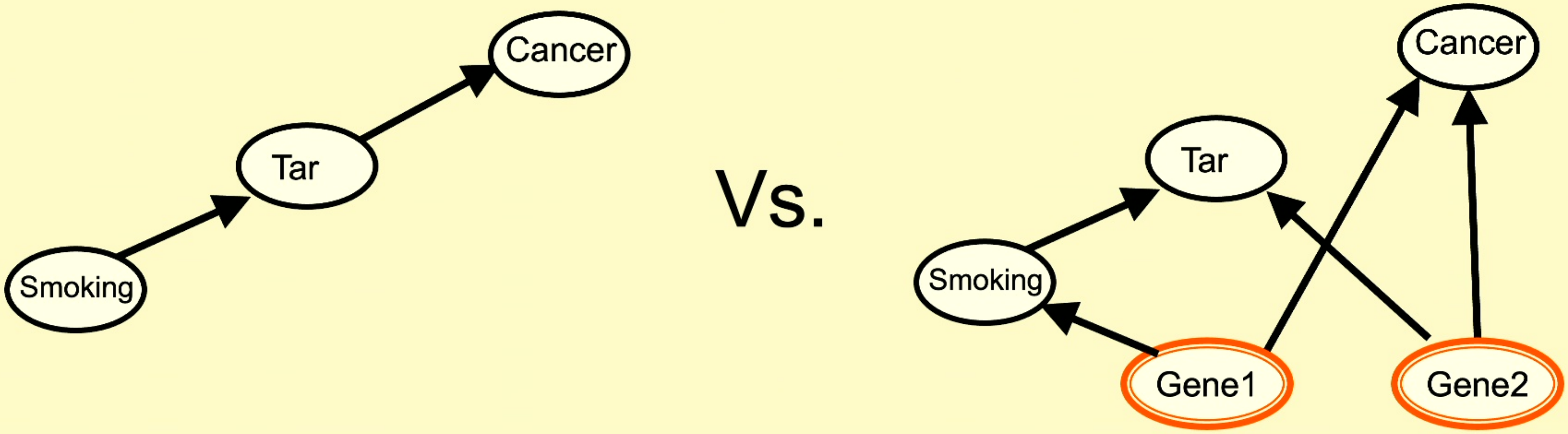
Signature of overfitting:
Lower training error
Higher test error

By higher bar:
This model preferred because it makes better predictions about unseen data



Observe P_{STC} such that **to good approximation**

$$S \perp C | T$$



Observe P_{STC} such that **to good approximation**

$$S \perp C | T$$

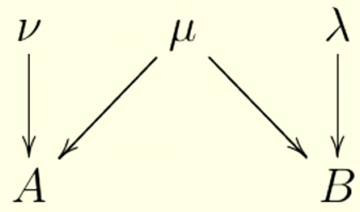
If the deviations are a statistical fluctuations, the model on the right will tend to overfit the data by mistaking these for real features

Estimating strengths of causal mechanisms from observational data

Given a hypothesis about the correct causal structure

How does one estimate the strengths of various parameters of the causal model, in particular the strength of causal mechanisms?

Identifiability of parameters in a structural equation model

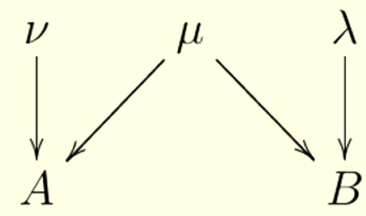


$$A = \mu\nu \qquad B = \mu\lambda$$

μ	ν	λ	$A = \mu\nu$	$B = \mu\lambda$
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	0	0
1	0	0	0	0
1	0	1	0	1
1	1	0	1	0
1	1	1	1	1

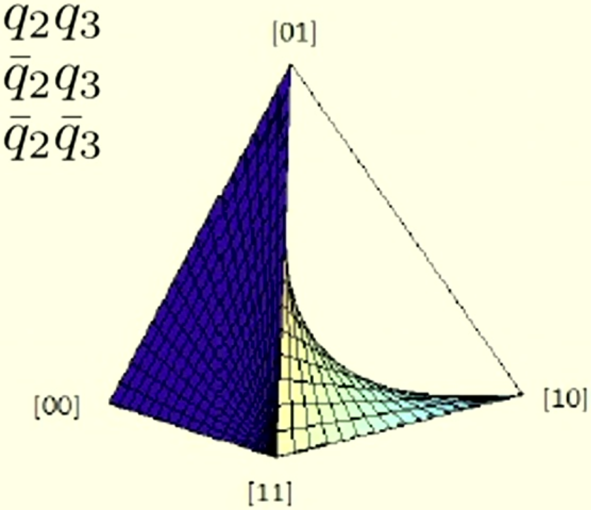
$$q_1 \equiv p(\mu = 0)$$
$$q_2 \equiv p(\nu = 0)$$
$$q_3 \equiv p(\lambda = 0)$$

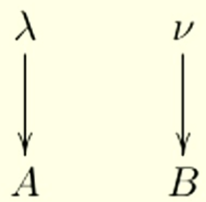
$$p_{00} = q_1 + \bar{q}_1 q_2 q_3$$
$$p_{01} = \bar{q}_1 q_2 \bar{q}_3$$
$$p_{10} = \bar{q}_1 \bar{q}_2 q_3$$
$$p_{10} = \bar{q}_1 \bar{q}_2 \bar{q}_3$$



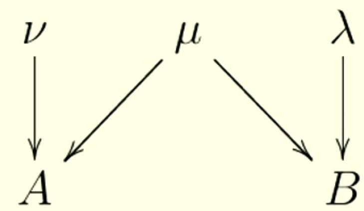
$$A = \mu\nu \qquad B = \mu\lambda$$

$$\begin{aligned} p_{00} &= q_1 + \bar{q}_1 q_2 q_3 \\ p_{01} &= \bar{q}_1 q_2 \bar{q}_3 \\ p_{10} &= \bar{q}_1 \bar{q}_2 q_3 \\ p_{11} &= \bar{q}_1 \bar{q}_2 \bar{q}_3 \end{aligned}$$



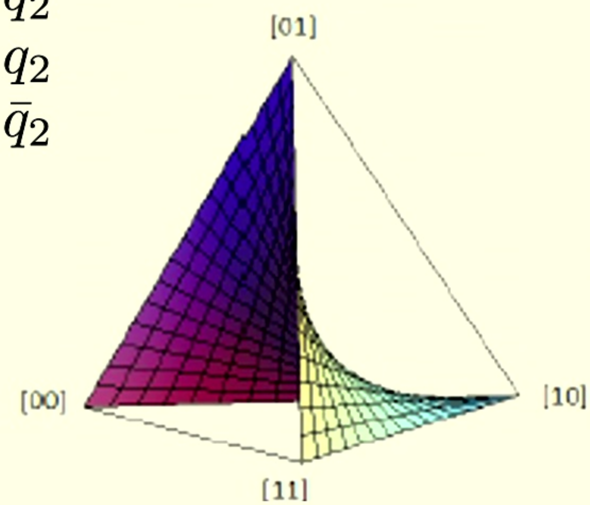


$$A = \lambda \qquad B = \nu$$

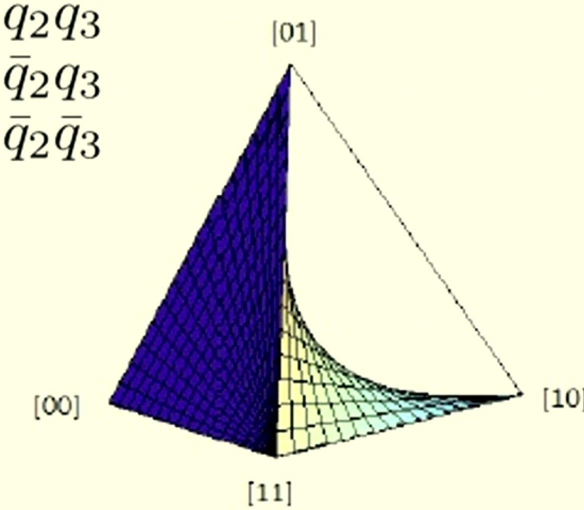


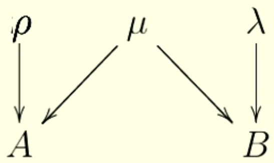
$$A = \mu\nu \qquad B = \mu\lambda$$

$$p_{00} = q_1 q_2$$
$$p_{01} = q_1 \bar{q}_2$$
$$p_{10} = \bar{q}_1 q_2$$
$$p_{10} = \bar{q}_1 \bar{q}_2$$



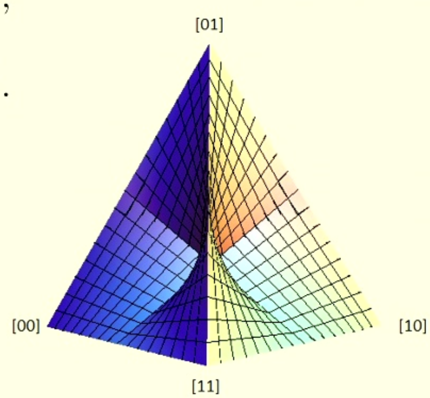
$$p_{00} = q_1 + \bar{q}_1 q_2 q_3$$
$$p_{01} = \bar{q}_1 q_2 \bar{q}_3$$
$$p_{10} = \bar{q}_1 \bar{q}_2 q_3$$
$$p_{11} = \bar{q}_1 \bar{q}_2 \bar{q}_3$$

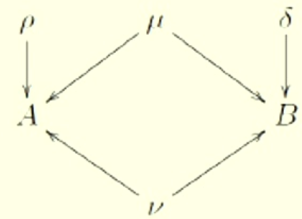




$$A = \rho \oplus \mu \qquad B = \lambda \oplus \mu$$

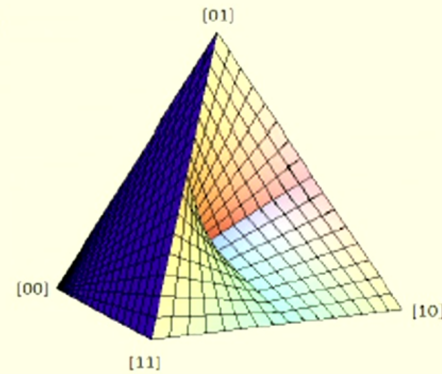
$$\frac{1}{4} > \frac{p_{10}p_{11} - p_{00}p_{01}}{2p_{10} + p_{11} - 1},$$
$$\frac{1}{4} > \frac{p_{01}p_{11} - p_{00}p_{10}}{2p_{01} + p_{11} - 1},$$
$$\frac{1}{4} > \frac{p_{10}p_{01} - p_{00}p_{11}}{2p_{10} + p_{01} - 1}.$$

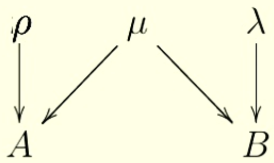




$$A = \nu(\rho \oplus \mu) \qquad B = \nu(\lambda \oplus \mu)$$

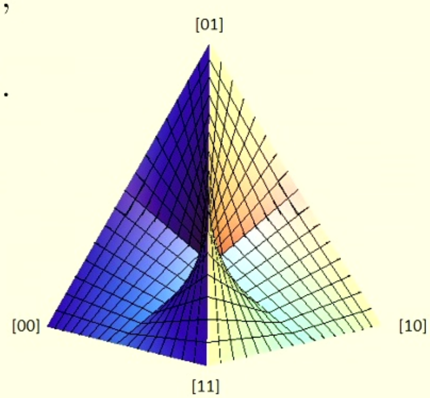
$$p_{00}p_{11} > p_{01}p_{10}$$
$$p_{11}p_{10} > p_{00}p_{01}$$
$$p_{11}p_{01} > p_{00}p_{10}$$

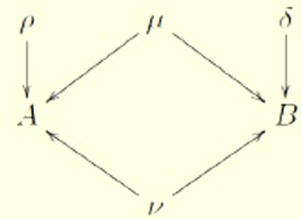




$$A = \rho \oplus \mu \qquad B = \lambda \oplus \mu$$

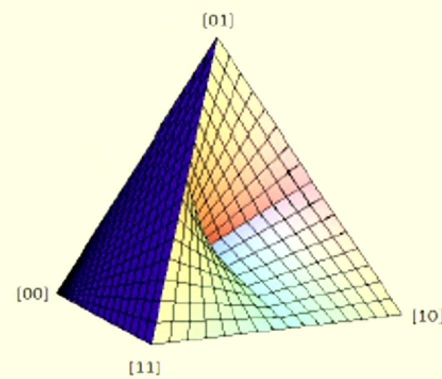
$$\frac{1}{4} > \frac{p_{10}p_{11} - p_{00}p_{01}}{2p_{10} + p_{11} - 1},$$
$$\frac{1}{4} > \frac{p_{01}p_{11} - p_{00}p_{10}}{2p_{01} + p_{11} - 1},$$
$$\frac{1}{4} > \frac{p_{10}p_{01} - p_{00}p_{11}}{2p_{10} + p_{01} - 1}.$$





$$A = \nu(\rho \oplus \mu) \qquad B = \nu(\lambda \oplus \mu)$$

$$p_{00}p_{11} > p_{01}p_{10}$$
$$p_{11}p_{10} > p_{00}p_{01}$$
$$p_{11}p_{01} > p_{00}p_{10}$$



Some paramaters are not identifiable

In the general case, the observational data only implies a **range of possibilities** for the tuple of parameters in the causal model

$$f_{\text{id}}, f_{\text{flip}}, f_{\text{reset}-0}, f_{\text{reset}-1}$$

$$P_{B|A} = \sum_f \delta_{B,f(A)} P_F(f)$$

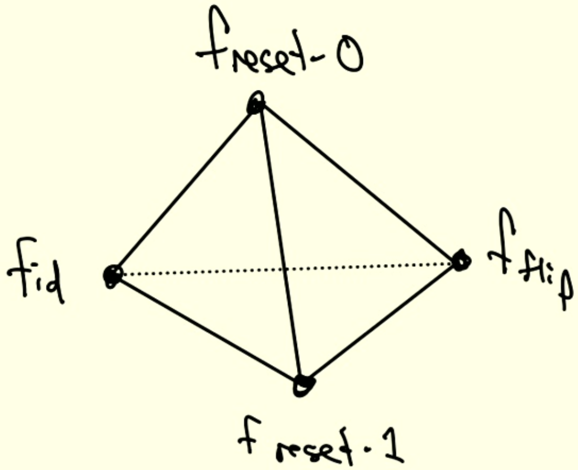
$$P_F = \tfrac{1}{2}[f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}]$$

$$P_{B|A} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

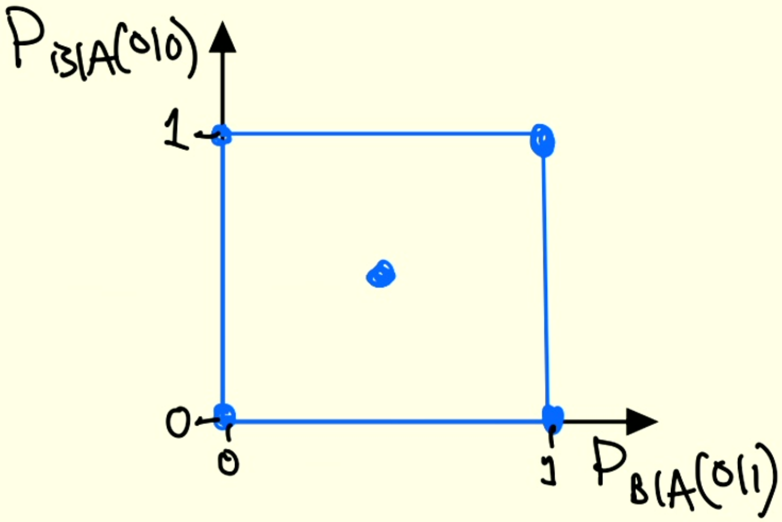
$$P'_F = \tfrac{1}{2}[f_{\text{reset}-0}] + \tfrac{1}{2}[f_{\text{reset}-1}]$$

$$P_{B|A} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

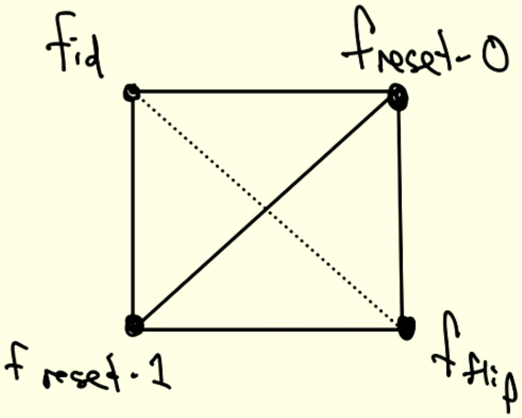
$$P_F = q\left(\tfrac{1}{2}[f_{\text{id}}] + \tfrac{1}{2}[f_{\text{flip}}]\right) + (1-q)\left(\tfrac{1}{2}[f_{\text{reset}-0}] + \tfrac{1}{2}[f_{\text{reset}-1}]\right) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$



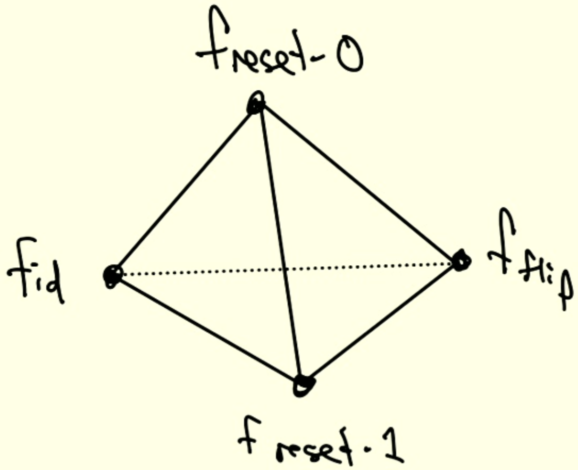
f reset-0
f id
f flip
f reset-1

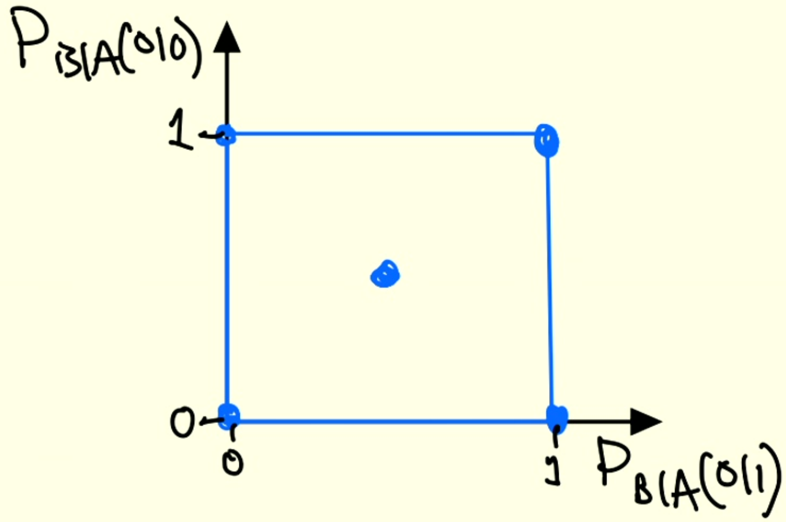


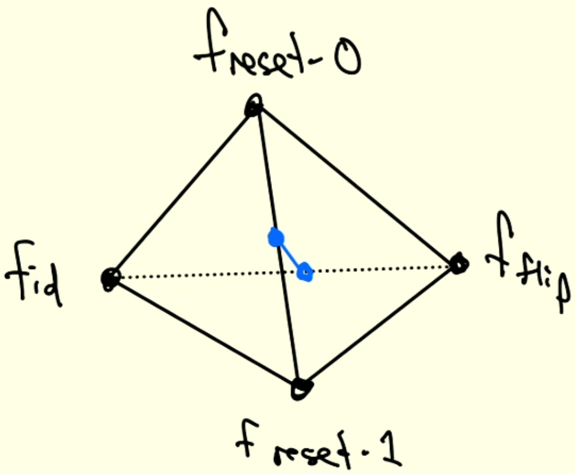
P B|A(0|0)
1
0
0
1
P B|A(0|1)

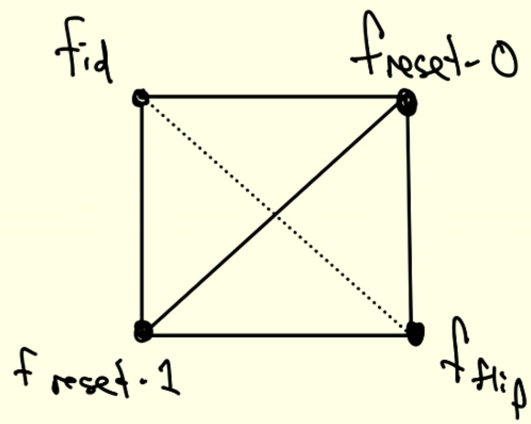


f id
f reset-0
f reset-1
f flip


f reset-0
f id
f flip
f reset-1

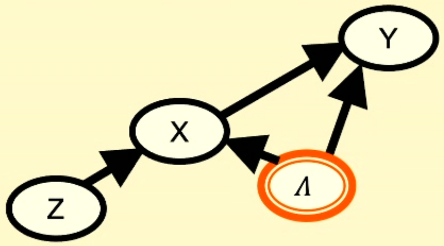

P B|A(0|0)
1
0
0
1
P B|A(0|1)


f reset-0
f id
f flip
f reset-1


f id
f reset-0
f reset-1
f flip

Identifiability concerns particular elements of model classes

Instrumental graph G

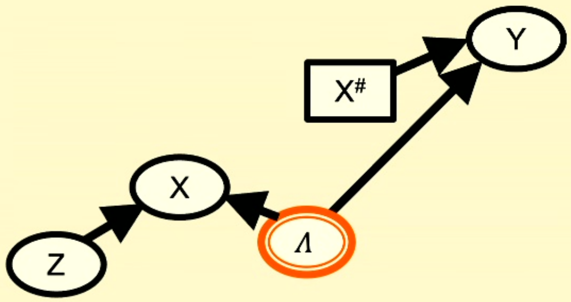


$P_{Y|\Lambda X}$
$P_{X|\Lambda Z}$
P_{Λ}

$$P_{XY|Z} = \sum_{\Lambda} P_{X|\Lambda Z} P_{Y|\Lambda X} P_{\Lambda}$$

The actual world is an element of the model class for the graph G

Interrupted version G'

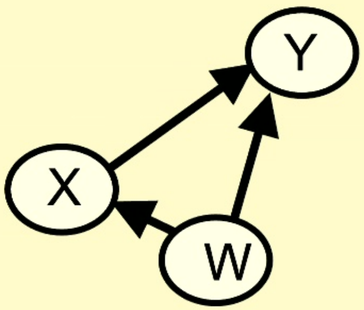


$X^{\#}$ input
$P_{Y|\Lambda X\#}$
$P_{X|Z\Lambda}$
P_{Λ}

$$P_{XY|ZX\#} = \sum_{\Lambda} P_{X|\Lambda Z} P_{Y|\Lambda X\#} P_{\Lambda}$$

The counterfactual world where a split-node intervention is made on X & **every other parameter remains the same**

Actual world G



$P_{X|W}$

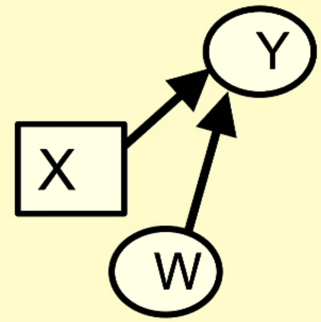
$P_{Y|WX}$

P_W

$$P^G_{Y|X} = \Sigma_W P_{Y|WX} P_{W|X}$$

What X teaches us about Y

Counterfactual world G'



$P_{Y|WX}$

X an input

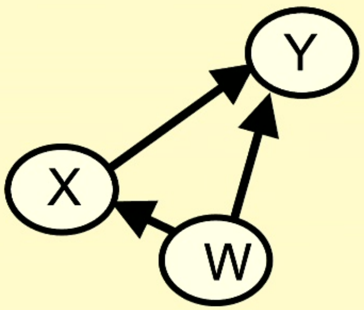
P_W

$$P^{G'}_{Y|X} = \Sigma_W P_{Y|WX} P_W$$

Definition of do-conditional

$$P^G_{Y|\text{do}(X)} := P^{G'}_{Y|X}$$

Actual world G



$P_{X|W}$

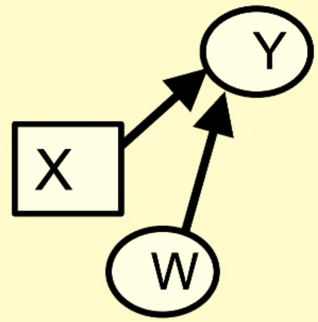
$P_{Y|WX}$

P_W

$$P^G_{Y|X} = \Sigma_W P_{Y|WX} P_{W|X}$$

What X teaches us about Y

Counterfactual world G'



$P_{Y|WX}$

X an input

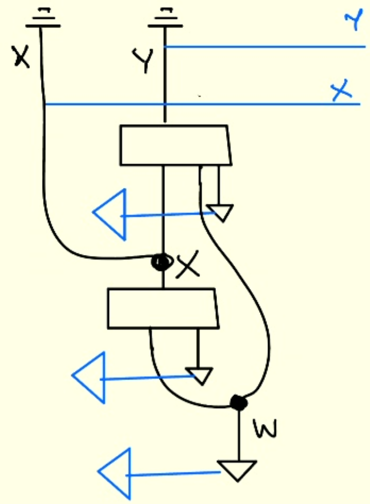
P_W

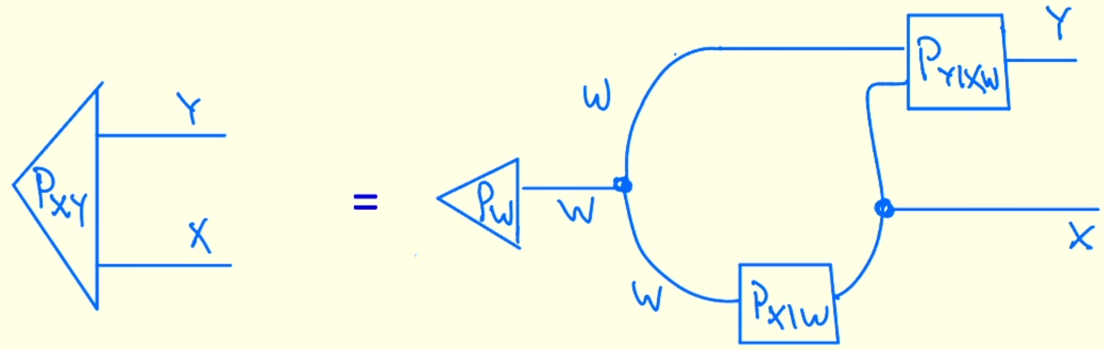
$$P^{G'}_{Y|X} = \Sigma_W P_{Y|WX} P_W$$

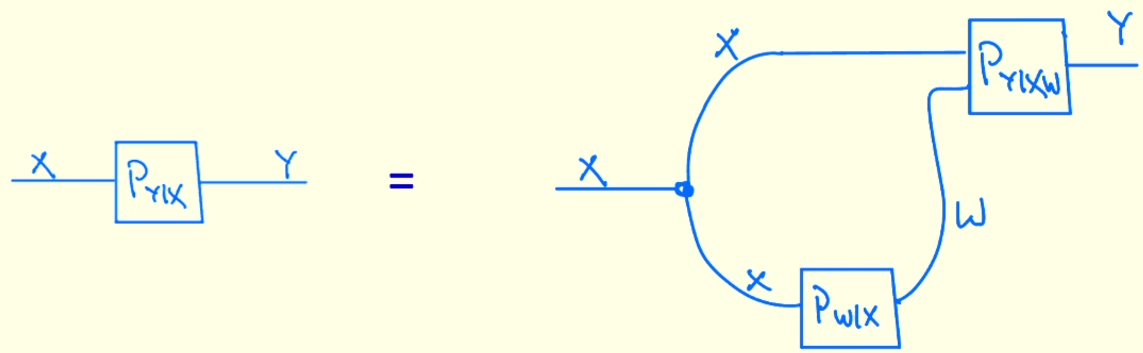
Definition of do-conditional

$$P^G_{Y|\text{do}(X)} := P^{G'}_{Y|X}$$

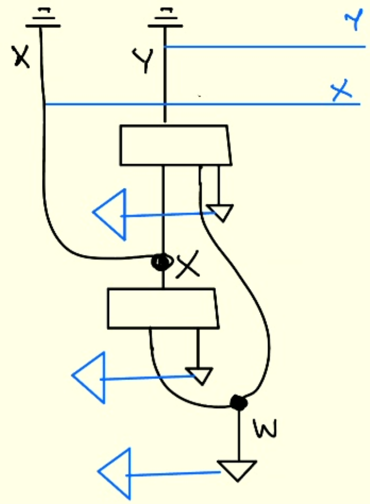
The counterfactual world where an intervention is made on X **& every other parameter remains the same**

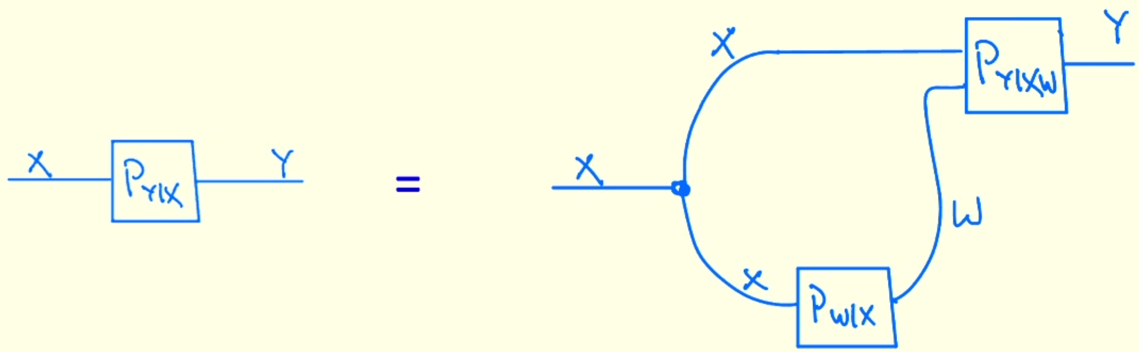




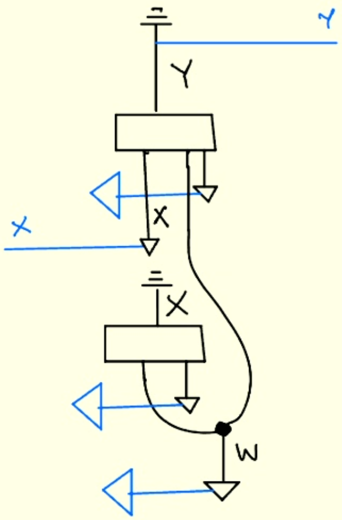


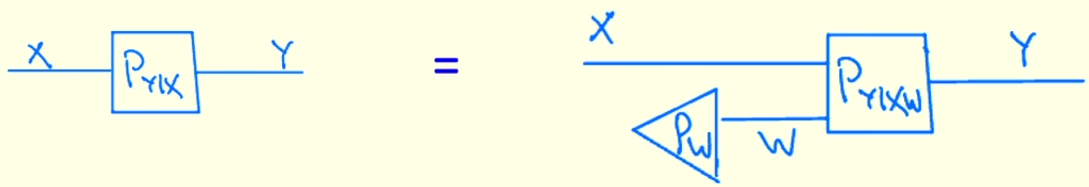
$$P^G_{Y|X} = \Sigma_W \, P_{Y|WX} P_{W|X}$$





$$P^{G}_{Y|X} = \Sigma_W \, P_{Y|WX} P_{W|X}$$





$$P^{G'}_{Y|X} = \Sigma_W \, P_{Y|WX} P_W$$

Recall:

Knowing the do-conditional $P_{Y|doX}$, we can infer the possible dist'ns over functions, i.e., the P_F such that

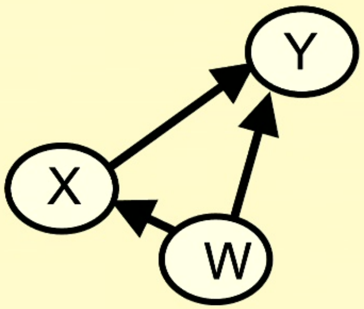
$$P_{Y|\mathrm{do}X} = \sum_f \delta_{Y,f(X)} P_F(f)$$

One can sometimes learn more about P_F by looking beyond $P_{Y|\mathrm{do}X}$ (recall example of Vernam cypher)

But one often settles for just inferring $P_{Y|\mathrm{do}X}$

When X and Y are not only connected by a directed path but also by a common cause, then the regular conditional $P_{Y|X}$ is not equal to the do-conditional $P_{Y|\text{do}X}$ --- this is **confounding**

Actual world G



$P_{X|W}$

$P_{Y|WX}$

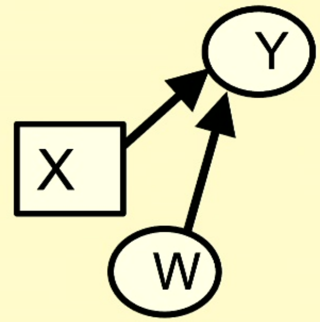
P_W

$$P^{G}_{Y|\text{do}(X)} := P^{G'}_{Y|X}$$

P_W identifiable

$P_{Y|WX}$ identifiable

Counterfactual world G’



$P_{Y|WX}$

X an input

P_W

$$P^{G'}_{Y|X} = \Sigma_W P_{Y|WX} P_W$$

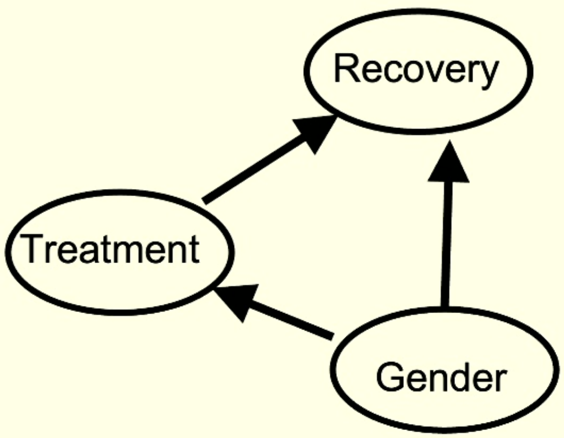
$$P^{G}_{Y|\text{do}(X)} = \Sigma_W P_{Y|WX} P_W$$

identifiable

P(recovery | drug) > P(recovery | no drug)

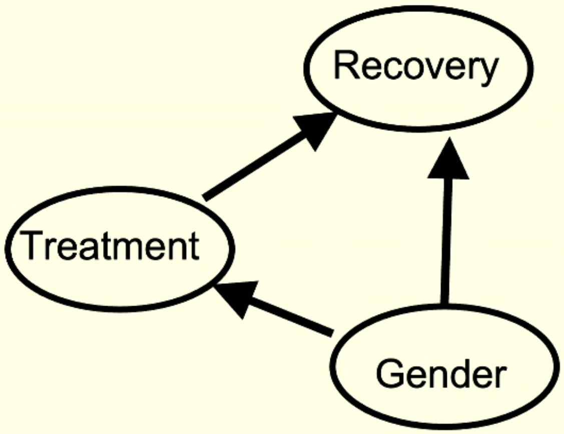
P(recovery | drug, male) < P(recovery | no drug, male) ✓

P(recovery | drug, female) < P(recovery | no drug, female) ✓

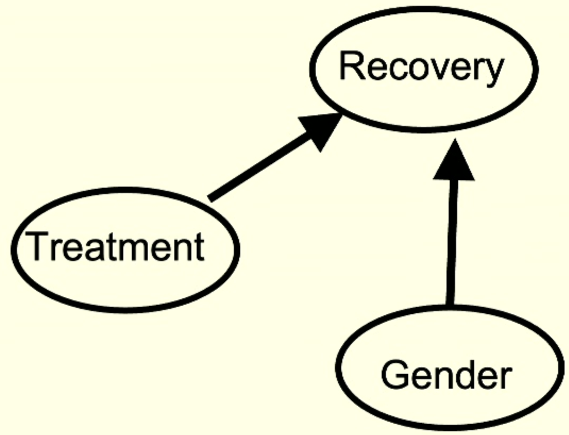


Therefore: stratify the data by the common cause

Actual world



Counterfactual



Standard conditional

$$P_{R|T} = \sum_G P_{R|TG} P_{G|T}$$

Do-conditional

$$P_{R|\mathrm{do}T} = \sum_G P_{R|TG} P_G$$

P(recovery | drug) > P(recovery | no drug)

P(recovery | drug, male) < P(recovery | no drug, male)

P(recovery | drug, female) < P(recovery | no drug, female)

P(recovery | drug) =
P(recovery | drug, male) P(male|drug) +
P(recovery | drug, female) P(female|drug)

>

P(recovery | no drug) =
P(recovery | no drug, male) P(male|drug) +
P(recovery | no drug, female) P(female|drug)

P(recovery | do drug) =
P(recovery | drug, male) P(male) +
P(recovery | drug, female) P(female)

<

P(recovery | do no drug) =
P(recovery | no drug, male) P(male) +
P(recovery | no drug, female) P(female)

Consider DAG G with variable X and set of variables **Z**

$$P_{\mathbf{Z}X} = \left(\prod_{i:Z_i \in \mathbf{Z}} P_{Z_i|\mathsf{Pa}(Z_i)} \right) P_{X|\mathsf{Pa}(X)}$$

By Markov condition in G

$$P_{\mathbf{Z}|\mathsf{do}(X)} = \prod_{i:Z_i \in \mathbf{V}} P_{Z_i|\mathsf{Pa}(Z_i)}$$

By Markov condition in G'

therefore

$$P_{\mathbf{Z}|\mathsf{do}(X)} = \frac{P_{\mathbf{Z}X}}{P_{X|\mathsf{Pa}(X)}}$$

The parental adjustment formula

Suppose $Y \notin X \cup \mathrm{Pa}(X)$

$$P_{Y|\mathrm{do}(X)} = \sum_{\mathrm{Pa}(X)} P_{Y|X\mathrm{Pa}(X)} P_{\mathrm{Pa}(X)}$$

Proof: Recall that

$$P_{\mathbf{Z}|\mathrm{do}(X)} = \frac{P_{\mathbf{Z}X}}{P_{X|\mathrm{Pa}(X)}}$$

Take $\mathbf{Z} := Y \cup \mathbf{W}$

$$P_{Y\mathbf{W}|\mathrm{do}(X)} = \frac{P_{Y\mathbf{W}X}}{P_{X|\mathrm{Pa}(X)}}$$

$$= P_{Y\mathbf{W}X} \frac{P_{\mathrm{Pa}(X)}}{P_{X\mathrm{Pa}(X)}}$$

Defining $\mathbf{W}'$ by $\mathbf{W} := \mathbf{W}' \cup \mathrm{Pa}(X)$

$$P_{Y\mathbf{W}|\mathrm{do}(X)} = P_{Y\mathbf{W}'X\mathrm{Pa}(X)} \frac{P_{\mathrm{Pa}(X)}}{P_{X\mathrm{Pa}(X)}}$$

$$= P_{Y\mathbf{W}'|X\mathrm{Pa}(X)} P_{\mathrm{Pa}(X)}$$

The parental adjustment formula

Suppose $Y \notin X \cup \mathsf{Pa}(X)$

$$P_{Y|\mathsf{do}(X)} = \sum_{\mathsf{Pa}(X)} P_{Y|X\mathsf{Pa}(X)} P_{\mathsf{Pa}(X)}$$

Proof: Recall that

$$P_{\mathbf{Z}|\mathsf{do}(X)} = \frac{P_{\mathbf{Z}X}}{P_{X|\mathsf{Pa}(X)}}$$

Take $\mathbf{Z} := Y \cup \mathbf{W}$

$$P_{Y\mathbf{W}|\mathsf{do}(X)} = \frac{P_{Y\mathbf{W}X}}{P_{X|\mathsf{Pa}(X)}}$$

$$= P_{Y\mathbf{W}X} \frac{P_{\mathsf{Pa}(X)}}{P_{X\mathsf{Pa}(X)}}$$

Defining $\mathbf{W}'$ by $\mathbf{W} := \mathbf{W}' \cup \mathsf{Pa}(X)$

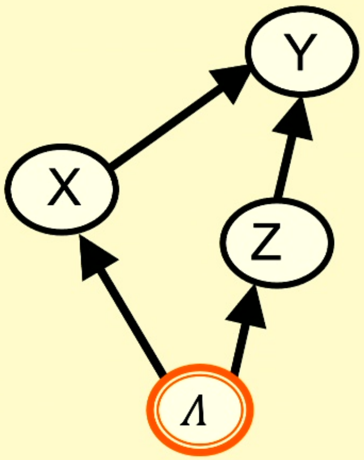
$$P_{Y\mathbf{W}|\mathsf{do}(X)} = P_{Y\mathbf{W}'X\mathsf{Pa}(X)} \frac{P_{\mathsf{Pa}(X)}}{P_{X\mathsf{Pa}(X)}}$$

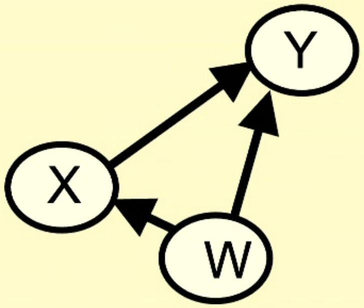
$$= P_{Y\mathbf{W}'|X\mathsf{Pa}(X)} P_{\mathsf{Pa}(X)}$$

Marginalize over $\mathbf{W} := \mathbf{W}' \cup \mathsf{Pa}(X)$

$$P_{Y|\mathsf{do}(X)} = \sum_{\mathsf{Pa}(X)} P_{Y|X\mathsf{Pa}(X)} P_{\mathsf{Pa}(X)}$$ QED

But what about cases where the parents of X are latent?



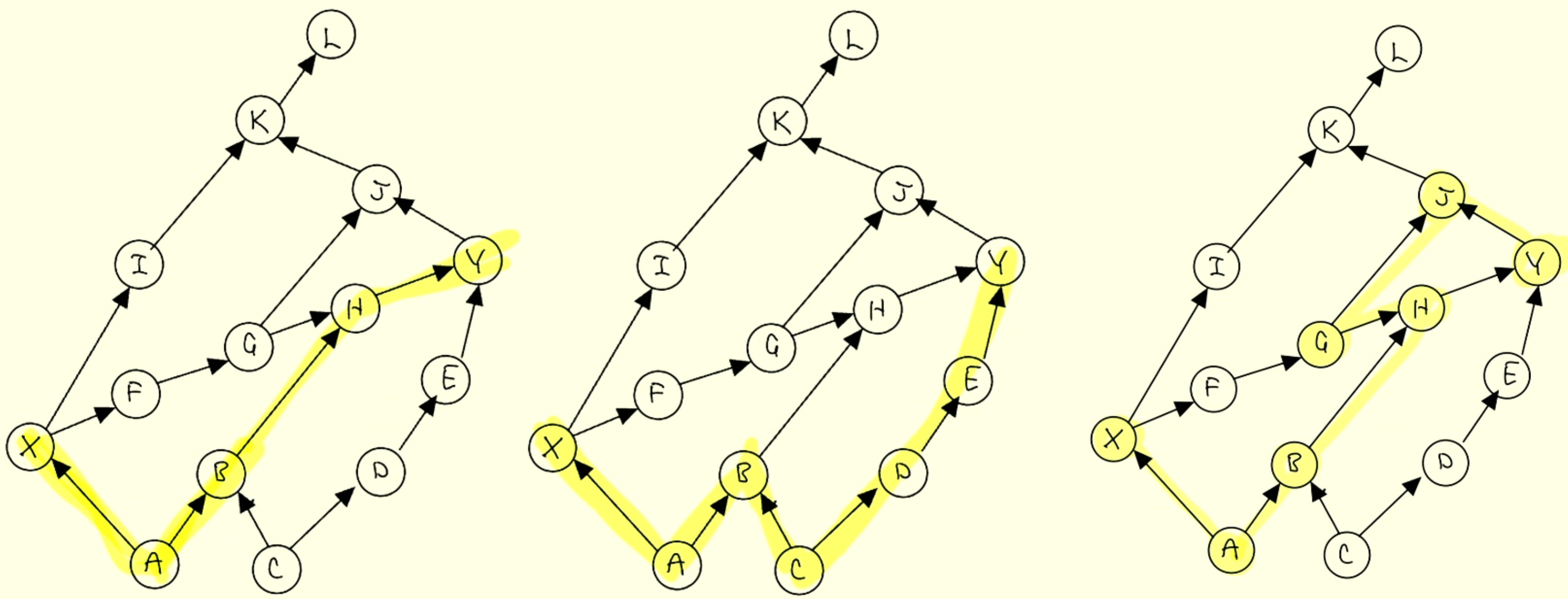


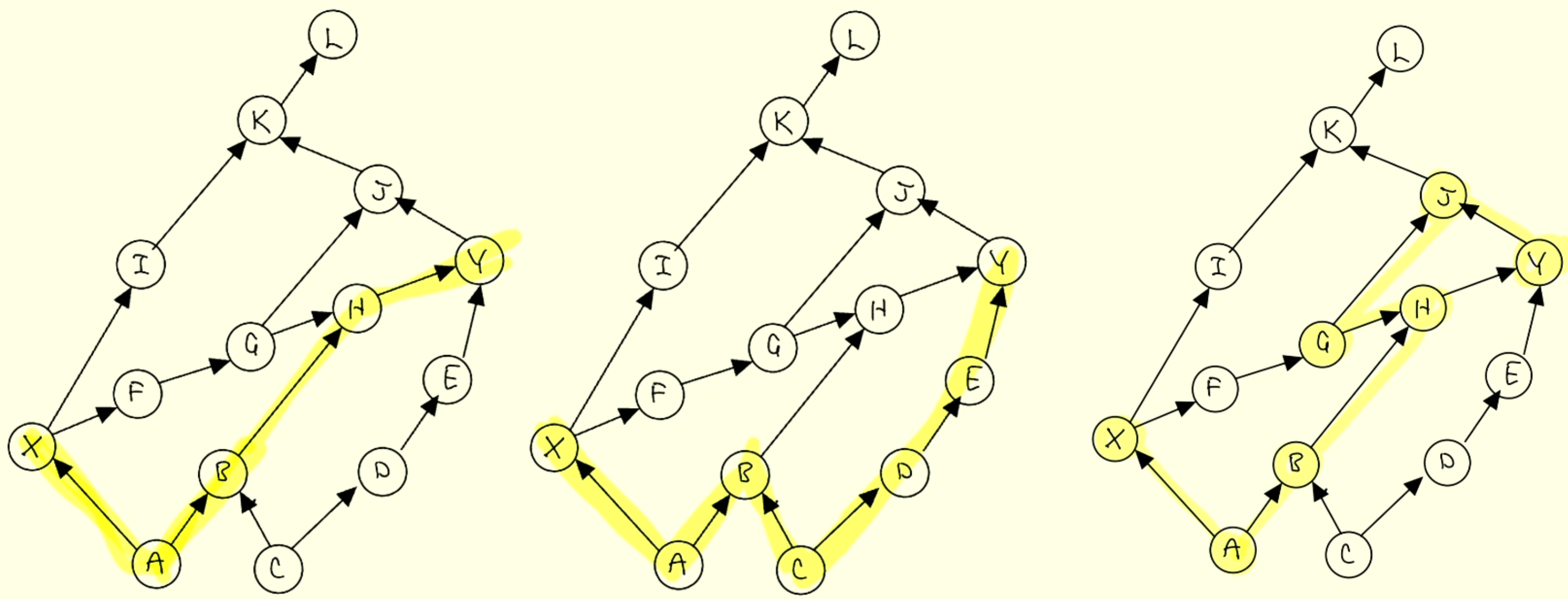
$P_{X|W}$
$P_{Y|WX}$
P_W

$$P_{Y|\mathrm{do}(X)} = \sum_W P_{Y|WX} P_W$$

This is a special case of parental adjustment

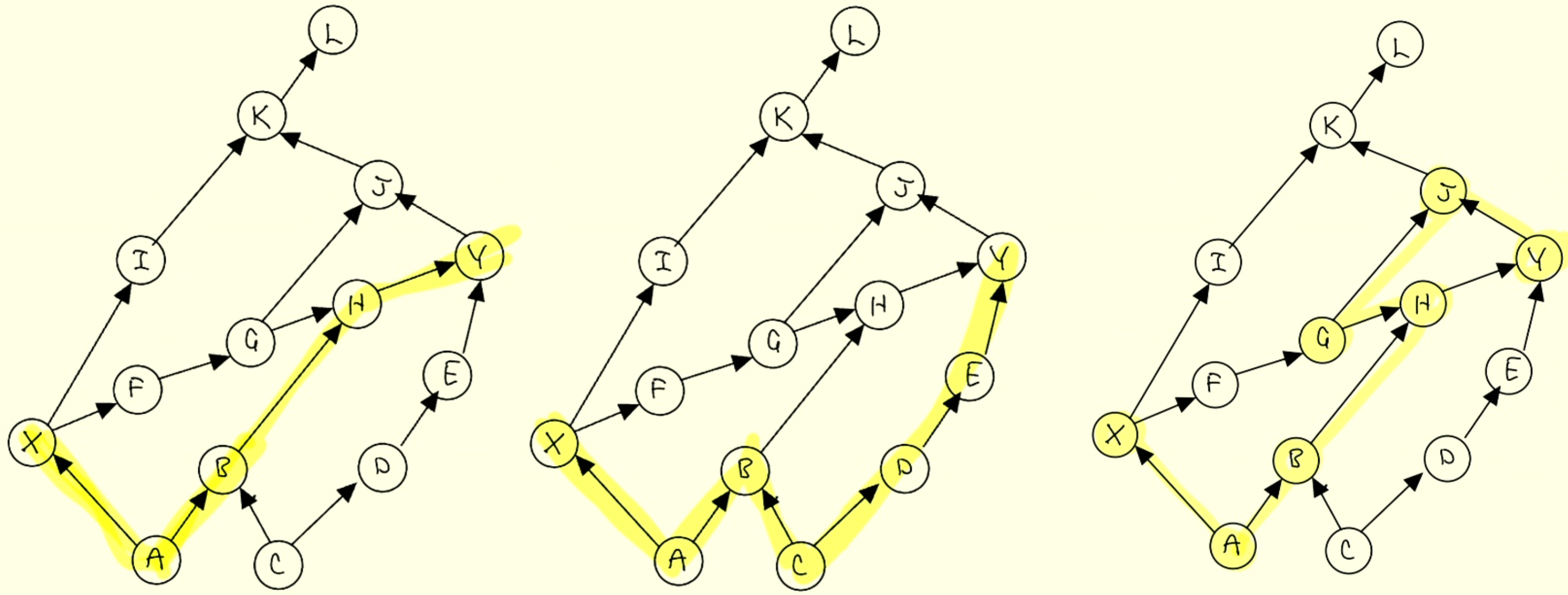
Backdoor paths between X and Y





Sets **Z** that satisfy the backdoor criterion:

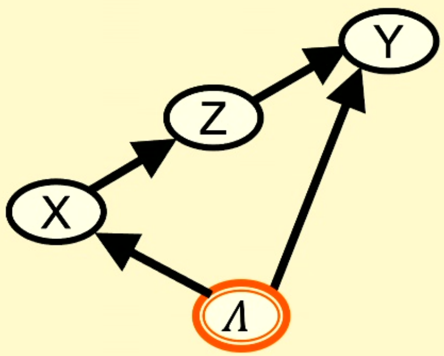
{A}, {A,C}, {A,C,D}, …,
{B,C}, {A,B,C}, …



Sets **Z** that satisfy the backdoor criterion:

{A}, {A,C}, {A,C,D}, …,
{B,C}, {A,B,C}, …

Note: H is a descendent of X so does not satisfy the backdoor criterion



$P_{Y|Z\Lambda}$
$P_{Z|X}$
$P_{X|\Lambda}$
P_{Λ}

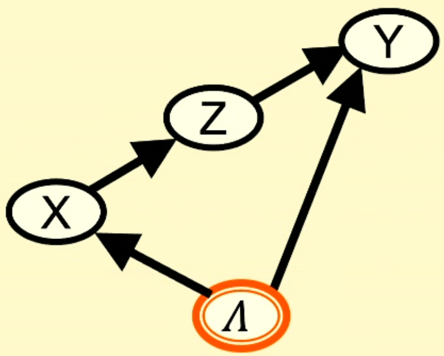
$$P_{Y|\text{do}(X)} = \sum_Z P_{Y|\text{do}(Z)} P_{Z|\text{do}(X)}$$

Consider

$$P_{Z|\text{do}(X)}$$

Noting that the path between X and Z in G is blocked by the empty set because of the collider at Y, backdoor adjustment implies

$$P_{Z|\text{do}(X)} = P_{Z|X}$$

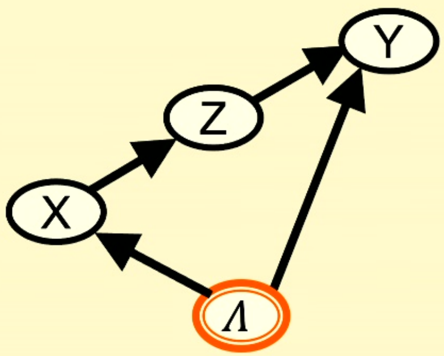


$P_{Y|Z\Lambda}$
$P_{Z|X}$
$P_{X|\Lambda}$
P_{Λ}

$$P_{Y|\mathrm{do}(X)} = \sum_Z P_{Y|\mathrm{do}(Z)} P_{Z|X}$$

Consider

$P_{Y|\mathrm{do}(Z)}$



$P_{Y|Z\Lambda}$
$P_{Z|X}$
$P_{X|\Lambda}$
P_{Λ}

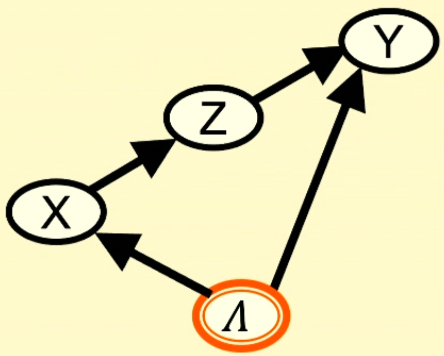
$$P_{Y|\mathrm{do}(X)} = \sum_Z P_{Y|\mathrm{do}(Z)} P_{Z|X}$$

Consider

$$P_{Y|\mathrm{do}(Z)}$$

Noting that the backdoor path between Y and Z in G is blocked by X, backdoor adjustment yields

$$P_{Y|\mathrm{do}(Z)} = \sum_X P_{Y|ZX} P_X$$



$P_{Y|Z\Lambda}$
$P_{Z|X}$
$P_{X|\Lambda}$
P_{Λ}

$$P_{Y|\mathrm{do}(X)} = \sum_Z \left(\sum_{X'} P_{Y|ZX'} P_{X'} \right) P_{Z|X'}$$

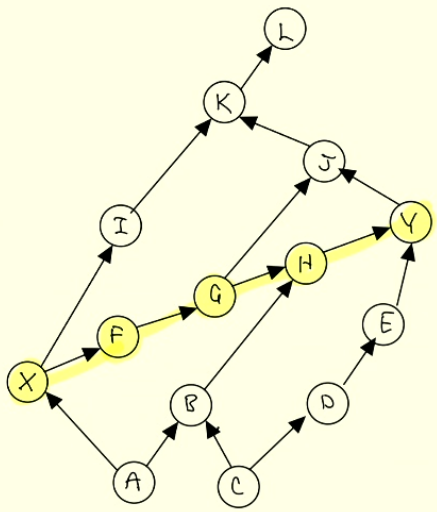
frontdoor adjustment formula

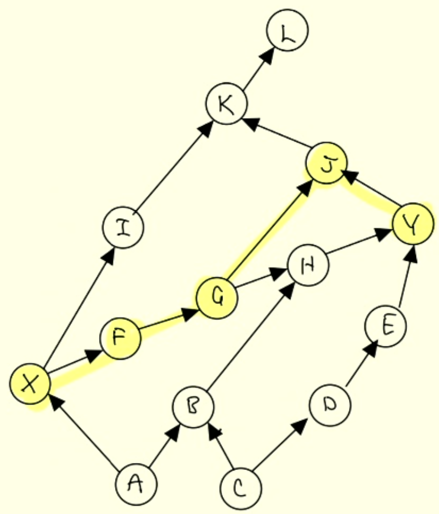
A frontdoor path between X and Y in a DAG G is any path between X and Y that has an arrow out of X

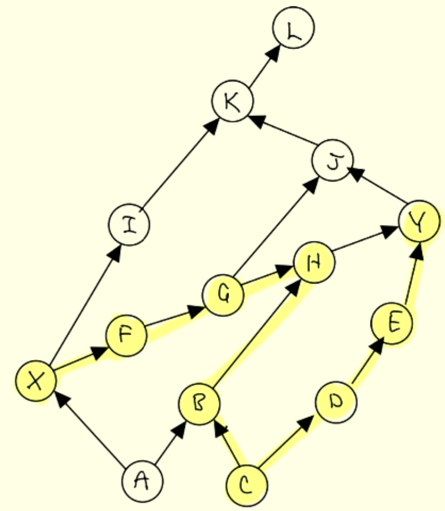
The frontdoor criterion

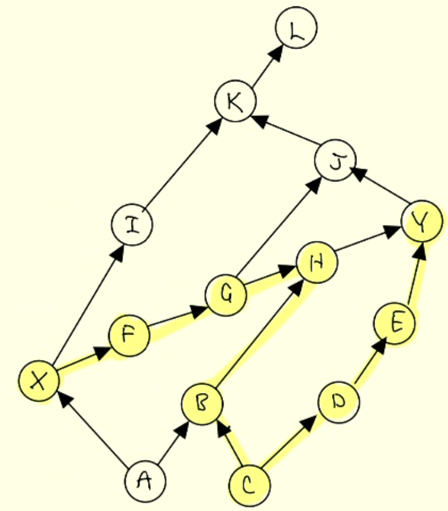
In a DAG G, a set of variables **Z** satisfies the frontdoor criterion relative to (X,Y) if

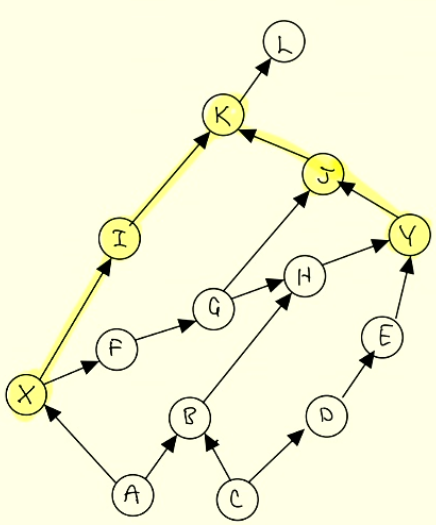
- **Z** intercepts all frontdoor paths from X to Y
- there is no backdoor path between X and **Z**, and
- All backdoor paths between **Z** and Y are blocked by X

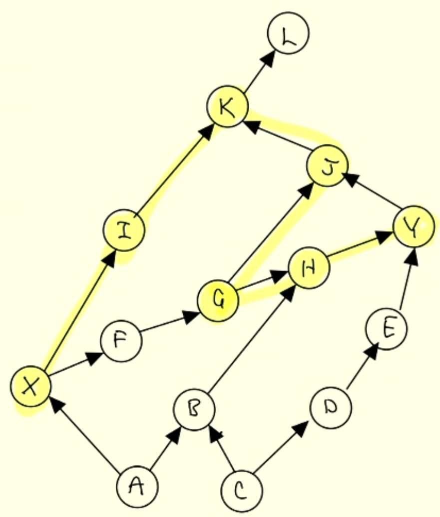








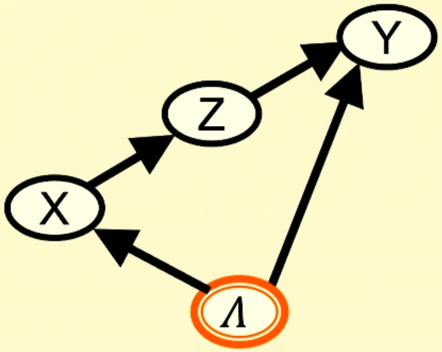




- **Z** intercepts all frontdoor paths from X to Y
- there is no backdoor path between X and **Z**, and
- All backdoor paths between **Z** and Y are blocked by X

Sets **Z** that satisfy the frontdoor criterion:

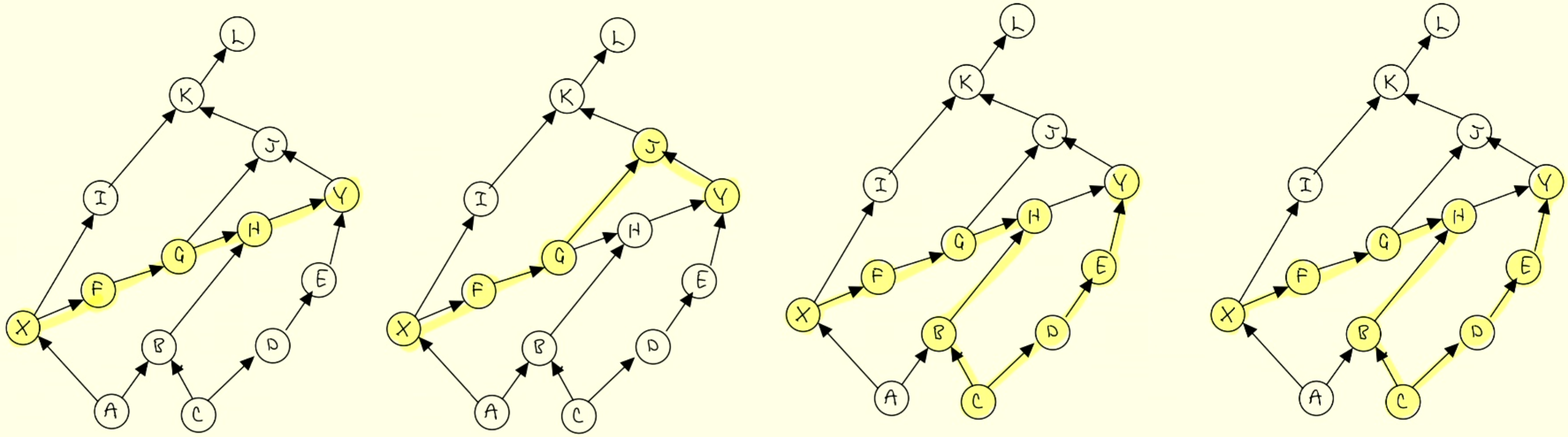
{F,I}, {G,I}, …

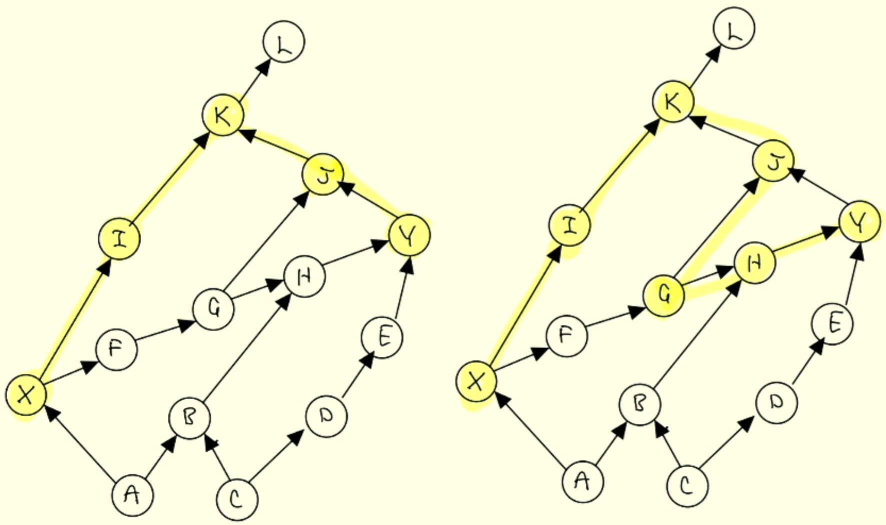


$P_{Y|Z\Lambda}$
$P_{Z|X}$
$P_{X|\Lambda}$
P_{Λ}

$$P_{Y|\text{do}(X)} = \sum_Z \left(\sum_{X'} P_{Y|ZX'} P_{X'} \right) P_{Z|X'}$$

frontdoor adjustment formula

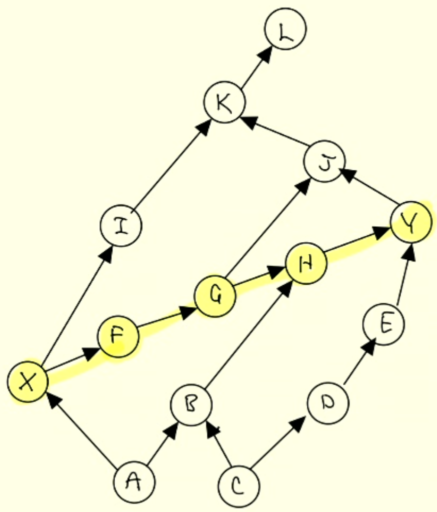


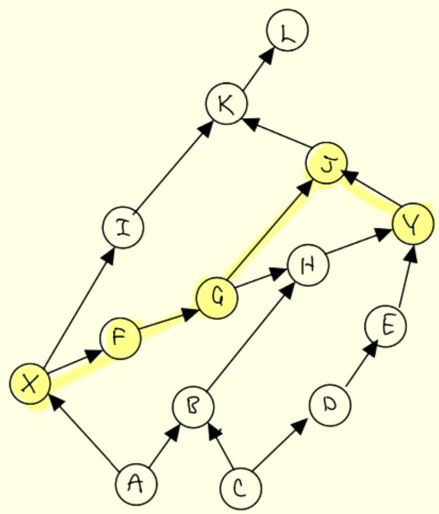


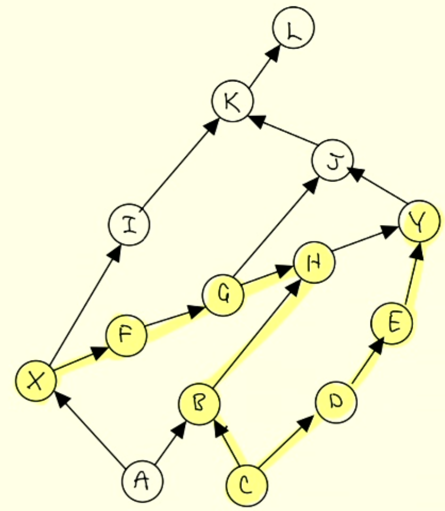
- **Z** intercepts all frontdoor paths from X to Y
- there is no backdoor path between X and **Z**, and
- All backdoor paths between **Z** and Y are blocked by X

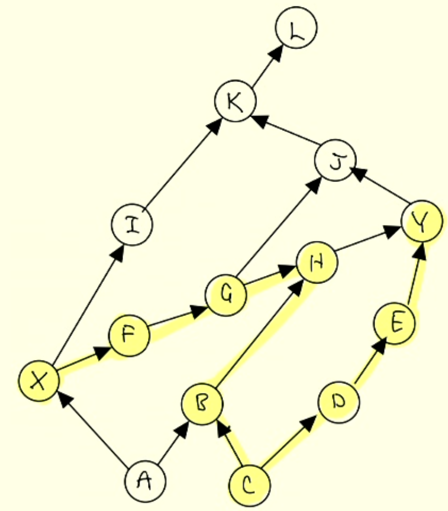
Sets **Z** that satisfy the frontdoor criterion:

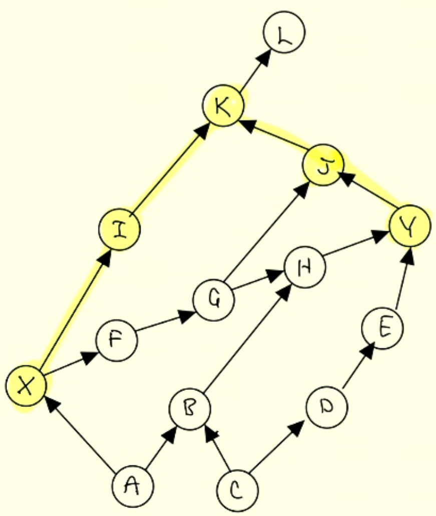
{F,I}, {G,I}, …

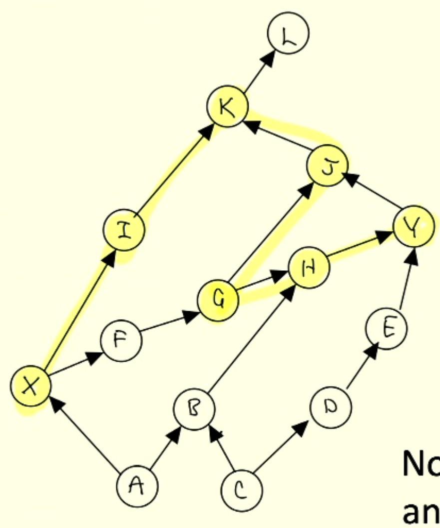










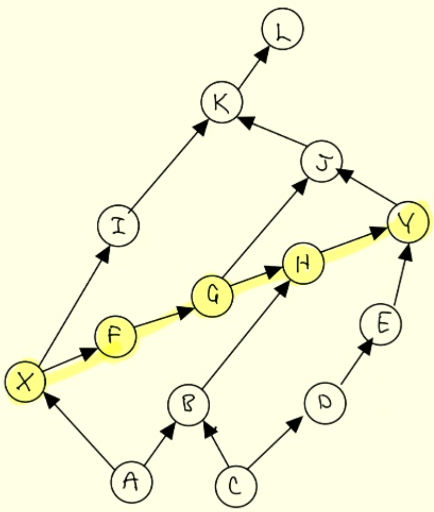


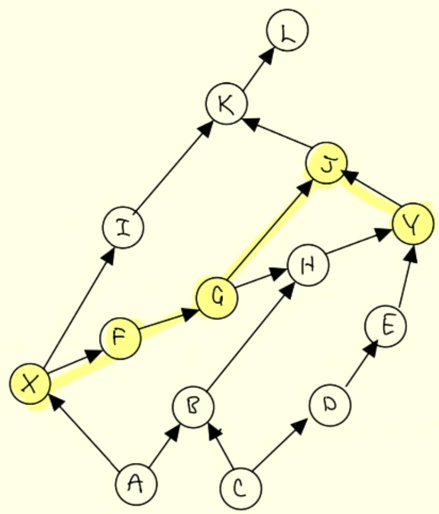
- **Z** intercepts all frontdoor paths from X to Y
- there is no backdoor path between X and **Z**, and
- All backdoor paths between **Z** and Y are blocked by X

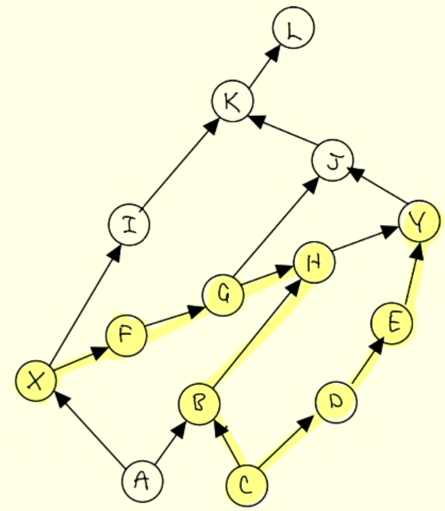
Sets **Z** that satisfy the frontdoor criterion:

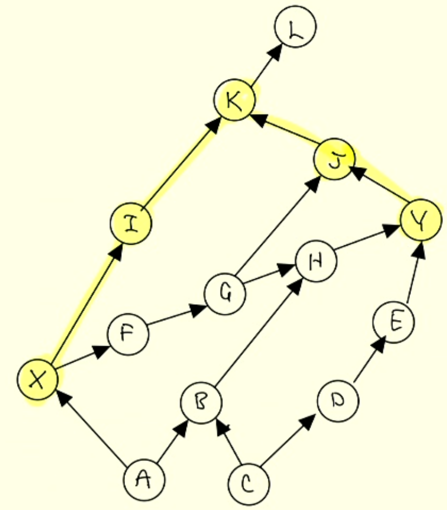
{F,I}, {G,I}, …

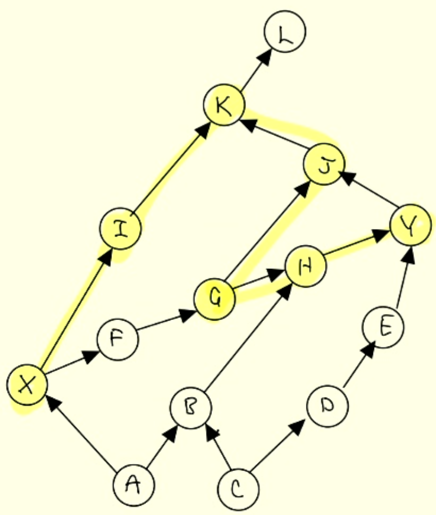
Note: H cannot be in **Z** both because there is a backdoor path with X and because there is a backdoor path with Y that is not blocked by X

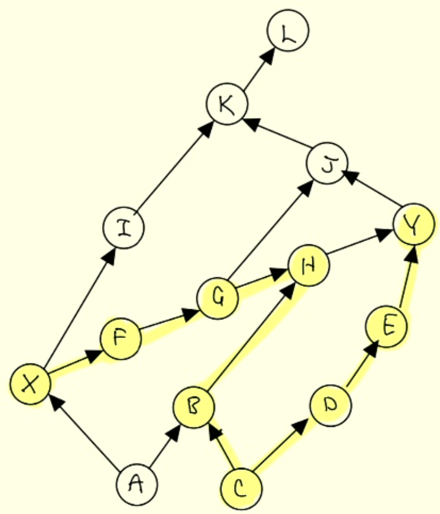












Sets **Z** that satisfy the frontdoor criterion:

{F,I}, {G,I}, …

$$P_{Y|\mathrm{do}(X)} = \sum_{FI} P_{FI|X} \sum_{X'} P_{Y|X'FI} P_{X'}$$

$$P_{Y|\mathrm{do}(X)} = \sum_{GI} P_{GI|X} \sum_{X'} P_{Y|X'GI} P_{X'}$$

…

Rule 1 (Insertion/deletion of observations)

$P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are *d*-separated by $\mathbf{X} \cup \mathbf{W}$ in $\mathbf{G}^*$, the graph obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$.

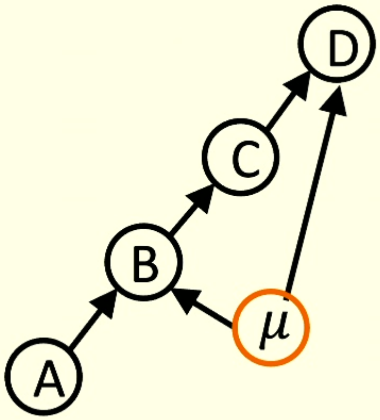
Rule 2 (Action/observation exchange)

$P(\mathbf{Y} \mid do(\mathbf{X}), do(\mathbf{Z}), \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are *d*-separated by $\mathbf{X} \cup \mathbf{W}$ in $\mathbf{G}^\dagger$, the graph obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$ and all arrows pointing out of variables in $\mathbf{Z}$.

Rule 3 (Insertion/deletion of actions)

$P(\mathbf{Y} \mid do(\mathbf{X}), do(\mathbf{Z}), \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are *d*-separated by $\mathbf{X} \cup \mathbf{W}$ in $\mathbf{G}^\ddagger$, the graph obtained from $\mathbf{G}$ by first removing all the arrows pointing into variables in $\mathbf{X}$ (thus creating $\mathbf{G}^*$) and then removing all of the arrows pointing into variables in $\mathbf{Z}$ that are not ancestors of any variable in $\mathbf{W}$ in $\mathbf{G}^*$.

Verma model M



$\mathrm{P}_{D|\mu C}$

$\mathrm{P}_{C|B}$

$\mathrm{P}_{B|A\mu}$

P_{A}

P_{μ}

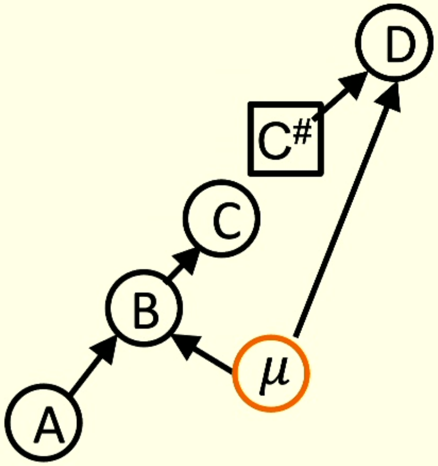
$$P_{ABCD} = \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right)\mathrm{P}_{C|B}\mathrm{P}_{A}$$

$$= Q_{BD|AC}$$

$$Q_{BD|AC} = \frac{\mathrm{P}_{ABCD}}{\mathrm{P}_{C|B}\mathrm{P}_{A}}$$

P_{ABCD}
is compatible with M

Interrupted version M’



$\mathrm{P}_{D|\mu C\#}$

$C^{\#} = c$

$\mathrm{P}_{C|B}$

$\mathrm{P}_{B|A\mu}$

P_{A}

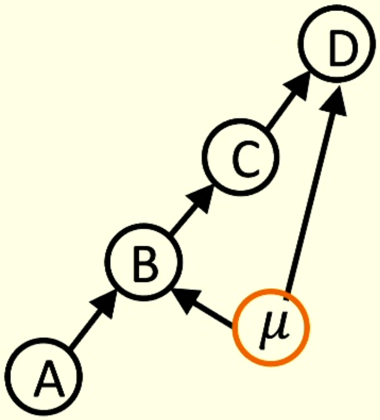
P_{μ}

$$\mathrm{P}_{BD|AC\#} = \sum_{\mu} \mathrm{P}_{D|\mu C\#}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}$$

$$\mathrm{P}_{BD|AC\#} = \frac{\mathrm{P}_{ABCD}}{\mathrm{P}_{C|B}\mathrm{P}_{A}}$$

is compatible with M'

Verma model M



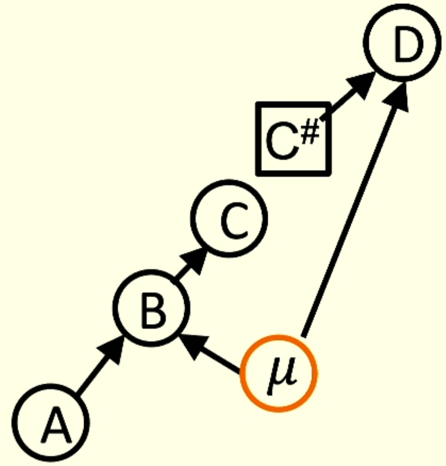
$$\mathrm{P}_{D|\mu C}$$
$$\mathrm{P}_{C|B}$$
$$\mathrm{P}_{B|A\mu}$$
$$\mathrm{P}_{A}$$
$$\mathrm{P}_{\mu}$$

$$P_{ABCD} = \left(\sum_\mu \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right)\mathrm{P}_{C|B}\mathrm{P}_{A}$$
$$= Q_{BD|AC}$$

$$Q_{BD|AC} = \frac{\mathrm{P}_{ABCD}}{\mathrm{P}_{C|B}\mathrm{P}_{A}}$$

P_{ABCD}
is compatible with M

Interrupted version M'



$$\mathrm{P}_{D|\mu C\#}$$
$$C^{\#} = c$$
$$\mathrm{P}_{C|B}$$
$$\mathrm{P}_{B|A\mu}$$
$$\mathrm{P}_{A}$$
$$\mathrm{P}_{\mu}$$

$$\mathrm{P}_{BD|AC\#} = \sum_\mu \mathrm{P}_{D|\mu C\#}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}$$

$$\mathrm{P}_{BD|AC\#} = \frac{\mathrm{P}_{ABCD}}{\mathrm{P}_{C|B}\mathrm{P}_{A}}$$

is compatible with M'

Consider DAG G with set of variables **X** and set of variables **Z**

$$P_{\mathbf{ZX}} = \left(\prod_{i:Z_i \in \mathbf{Z}} P_{Z_i|\mathsf{Pa}(Z_i)} \right) \left(\prod_{j:X_j \in \mathbf{X}} P_{X_j|\mathsf{Pa}(X_j)} \right)$$

By Markov condition in G

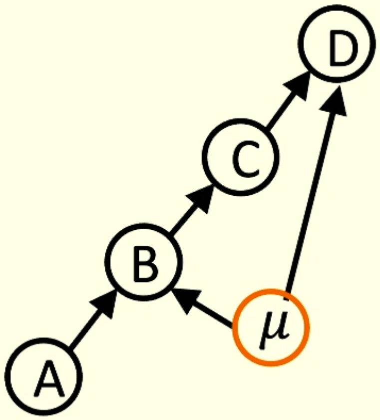
$$P_{\mathbf{Z}|\mathsf{do}(\mathbf{X})} = \prod_{i:Z_i \in \mathbf{V}} P_{Z_i|\mathsf{Pa}(Z_i)}$$

By Markov condition in G'

therefore

$$P_{\mathbf{Z}|\mathsf{do}(\mathbf{X})} = \frac{P_{\mathbf{ZX}}}{\prod_{j:X_j \in \mathbf{X}} P_{X_j|\mathsf{Pa}(X_j)}}$$

Verma graph G



$\mathrm{P}_{D|\mu C}$

$\mathrm{P}_{C|B}$

$\mathrm{P}_{B|A\mu}$

P_{A}

P_{μ}

Suppose we wish to determine

$$P_{BD|\mathrm{do}(AC)}$$

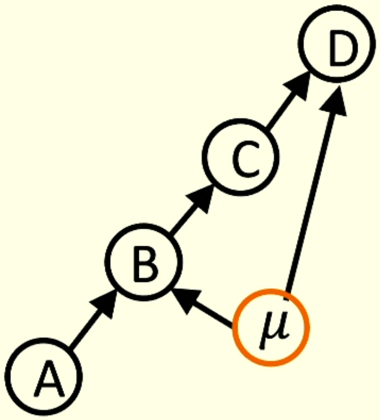
Using

$$P_{\mathbf{Z}|\mathsf{do}(\mathbf{X})} = \frac{P_{\mathbf{ZX}}}{\prod_{j:X_j \in \mathbf{X}} P_{X_j|\mathsf{Pa}(X_j)}}$$

We obtain

$$P_{BD|\mathrm{do}(AC)} = \frac{P_{ABCD}}{P_{C|B}P_{A}}$$

Verma graph G



$\mathrm{P}_{D|\mu C}$

$\mathrm{P}_{C|B}$

$\mathrm{P}_{B|A\mu}$

P_{A}

P_{μ}

Suppose we wish to determine

$$P_{BD|\mathrm{do}(AC)}$$

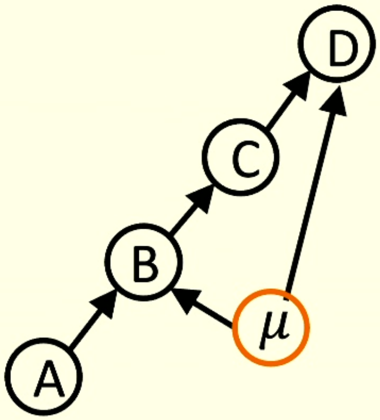
Using

$$P_{\mathbf{Z}|\mathsf{do}(\mathbf{X})} = \frac{P_{\mathbf{ZX}}}{\prod_{j:X_j\in\mathbf{X}} P_{X_j|\mathsf{Pa}(X_j)}}$$

We obtain

$$P_{BD|\mathrm{do}(AC)} = \frac{P_{ABCD}}{P_{C|B}P_{A}}$$

Verma model M



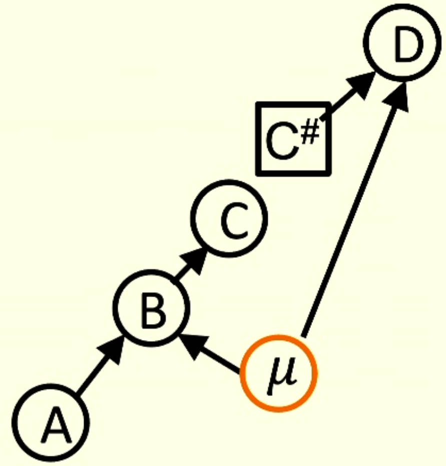
$$\mathrm{P}_{D|\mu C}$$
$$\mathrm{P}_{C|B}$$
$$\mathrm{P}_{B|A\mu}$$
$$\mathrm{P}_{A}$$
$$\mathrm{P}_{\mu}$$

$$P_{ABCD} = \left(\sum_{\mu} \mathrm{P}_{D|\mu C}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}\right)\mathrm{P}_{C|B}\mathrm{P}_{A}$$

$$= Q_{BD|AC}$$

$$Q_{BD|AC} = \frac{\mathrm{P}_{ABCD}}{\mathrm{P}_{C|B}\mathrm{P}_{A}}$$

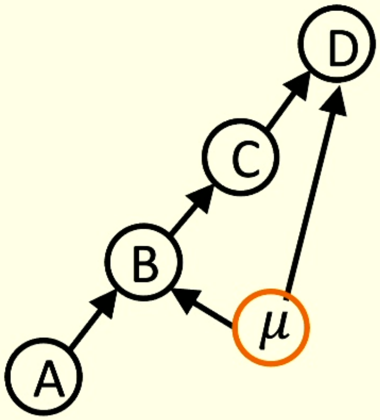
Interrupted version M'



$$\mathrm{P}_{D|\mu C\#}$$
$$C^{\#} = c$$
$$\mathrm{P}_{C|B}$$
$$\mathrm{P}_{B|A\mu}$$
$$\mathrm{P}_{A}$$
$$\mathrm{P}_{\mu}$$

$$\mathrm{P}_{BD|AC\#} = \sum_{\mu} \mathrm{P}_{D|\mu C\#}\mathrm{P}_{B|A\mu}\mathrm{P}_{\mu}$$

Verma graph G



$\mathrm{P}_{D|\mu C}$

$\mathrm{P}_{C|B}$

$\mathrm{P}_{B|A\mu}$

P_{A}

P_{μ}

Suppose we wish to determine

$$P_{BD|\mathrm{do}(AC)}$$

Using

$$P_{\mathbf{Z}|\mathsf{do}(\mathbf{X})} = \frac{P_{\mathbf{ZX}}}{\prod_{j:X_j\in\mathbf{X}} P_{X_j|\mathsf{Pa}(X_j)}}$$

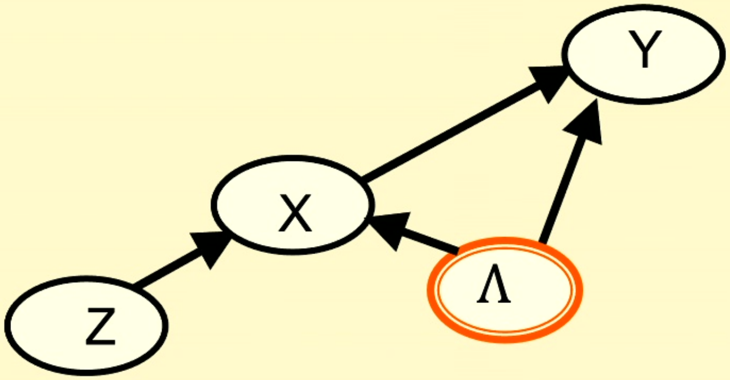
We obtain

$$P_{BD|\mathrm{do}(AC)} = \frac{P_{ABCD}}{P_{C|B}P_{A}}$$

Any identifiability result can be put into the service of deriving causal compatibility constraints

Putting upper and lower bounds on do-conditionals

Instrumental graph



$P_{X|\Lambda Z}$
$P_{Y|\Lambda X}$
P_{Λ}

$$P_{Y|\mathrm{do}(X)}$$

Not identifiable!

But one can find **bounds** in terms of the observed conditional

$$P_{XY|Z}$$

To find bounds on $P_{Y|\mathrm{do}(X)}(1|0)$ in terms of $P_{XY|Z}$

Minimize/Maximize $P_{Y|\mathrm{do}(X)}(1|0) = \sum_g (q_{\mathrm{r}_1,g} + q_{\mathrm{fp},g})$

Subject to $0 \leq q_{fg} \leq 1 \; \forall f, g$

$$p_{00|0} = q_{\mathrm{r}_0,\mathrm{r}_0} + q_{\mathrm{r}_0,\mathrm{id}} + q_{\mathrm{id},\mathrm{r}_0} + q_{\mathrm{id},\mathrm{id}}$$

$$p_{01|0} = q_{\mathrm{r}_0,\mathrm{r}_1} + q_{\mathrm{r}_0,\mathrm{fp}} + q_{\mathrm{id},\mathrm{r}_1} + q_{\mathrm{id},\mathrm{fp}}$$

$$\vdots$$

where $p_{xy|z} := P_{XY|Z}(xy|z)$ are given

One obtains

$$\min\left\{\begin{array}{c} p_{01.0}+p_{10.0}+p_{10.1}+p_{11.1} \\ 1-p_{00.1} \\ 1-p_{00.0} \\ p_{10.0}+p_{11.0}+p_{01.1}+p_{10.1} \end{array}\right\} \leq P_{Y|\mathrm{do}(X)}(1|0) \leq \max\left\{\begin{array}{c} p_{10.0}+p_{11.0}-p_{00.1}-p_{11.1} \\ p_{10.1} \\ p_{10.0} \\ p_{01.0}+p_{10.0}-p_{00.1}-p_{01.1} \end{array}\right\}$$

$$\min\left\{\begin{array}{c} 1-p_{01.1} \\ 1-p_{01.0} \\ p_{00.0}+p_{11.0}+p_{10.1}+p_{11.1} \\ p_{10.0}+p_{11.0}+p_{00.1}+p_{11.1} \end{array}\right\} \leq P_{Y|\mathrm{do}(X)}(1|1) \leq \max\left\{\begin{array}{c} p_{11.0} \\ p_{11.1} \\ -p_{00.0}-p_{01.0}+p_{00.1}+p_{11.1} \\ -p_{01.0}-p_{10.0}+p_{10.1}+p_{11.1} \end{array}\right\}$$

Next Lecture: Quantum causal models