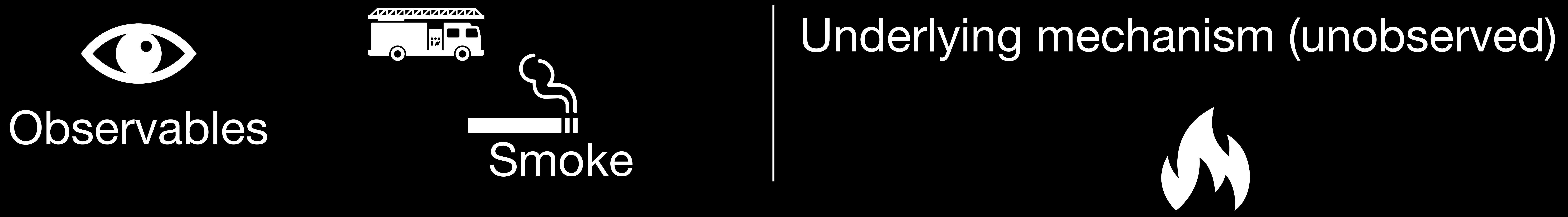
Causal Inference Presentation

For Prof. Elias Barenboim
Student: Nihaar Shah
Date: May 11th, 2020

ns3413@columbia.edu

Overview

- Causal inference stems from the recognition that there exists data generating mechanisms under the phenomenon of interest
- We only have partial knowledge of the model of such mechanism
- In fact, we only observe some features and data that are emitted by this mechanism - an analogy is we see smoke, we know that comes from fire but we don't see the fire or how it was made
-



Different types of data have different value

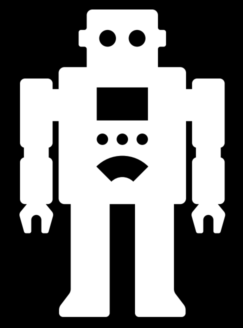
Let's take a very physical viewpoint on this cost

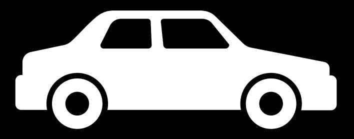
In increasing order

- Observational $P(X)$ - prima facie - done using sensors (camera etc.)
- Joint observations $P(X,Y)$ — done using sensors AND human labelers- more expensive
- Conditional $P(Y|X)$ computed from Joint and then marginalisation
- Interventional (cost of experiments that include changing physical setting etc.)
- Counterfactual - imagining different observations and their outcomes (can be infinite cost if these are not realisable) ….

Kalman Filter (ubiquitous in autonomous cars)

Algorithm to estimate under uncertainty



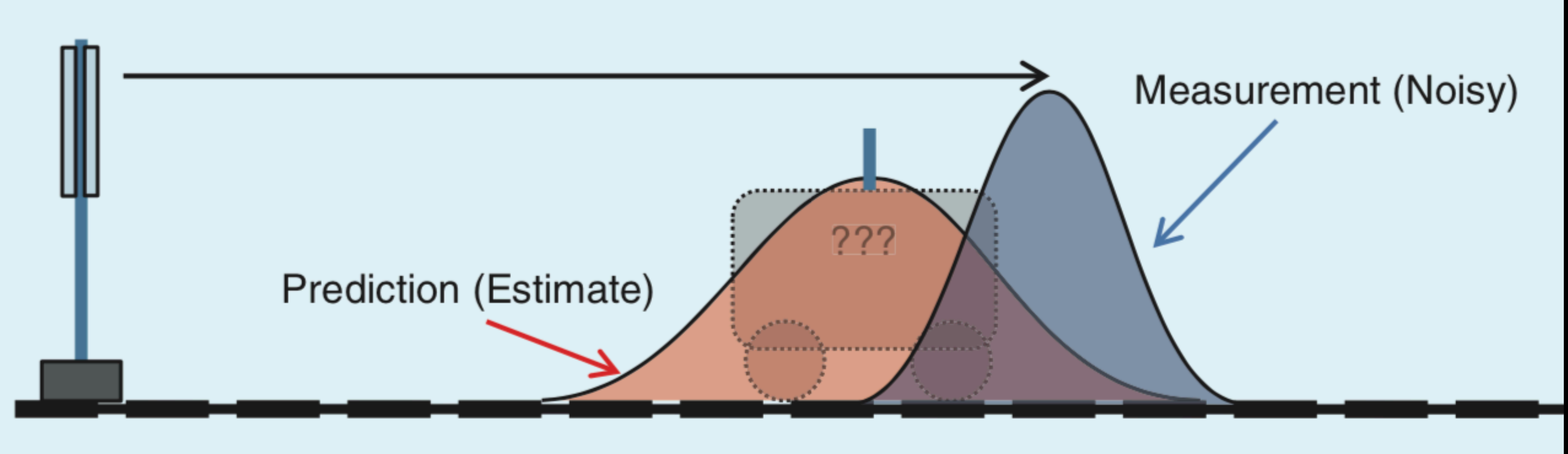


- The most common way to deal with this uncertainty is to use an algorithm called Kalman Filter (KF). A KF is based on a predictor-corrector structure
- The algorithm models physics of the robot such as displacement, velocity, acceleration etc.
- Importantly the algorithm keeps room for uncertainty in the form of a Multivariate Gaussian with zero mean and flexible variance
- The algorithm updates its predictions and uncertainty estimates using measurements (again it allows for uncertainty in measurements)

A motivating example from Robotics

Before going into theory, let's ground our intuition in a real world problem

- Probabilistic robotics deals with uncertainty due to :
- 1. Unmodeled physics 2. Measurement noise
- This image shows the prediction estimate based on physics and measurement noise from a beacon to determine the robot's position



Robot Kalman filter example

Let’s model an autonomous boat that must navigate in the sea

- Predict Equation: $x(k\,|\,k-1) = F\,.\,x(k-1\,|\,k-1) + Bu(k)$
- Where: x(k|k-1) - state prediction at time step k conditioned on previous time step
- F - Plant model (Pre-determined system model e.g. a constant velocity model
- B - Process noise model (noise due to unmodelled uncertainties e.g. wind/air-resistance/unexpected bumps in the road etc.)
- u(k) - Control vector which indicated action taken by robot in units of acceleration

Robot Kalman filter example

Predict equation (closer look)

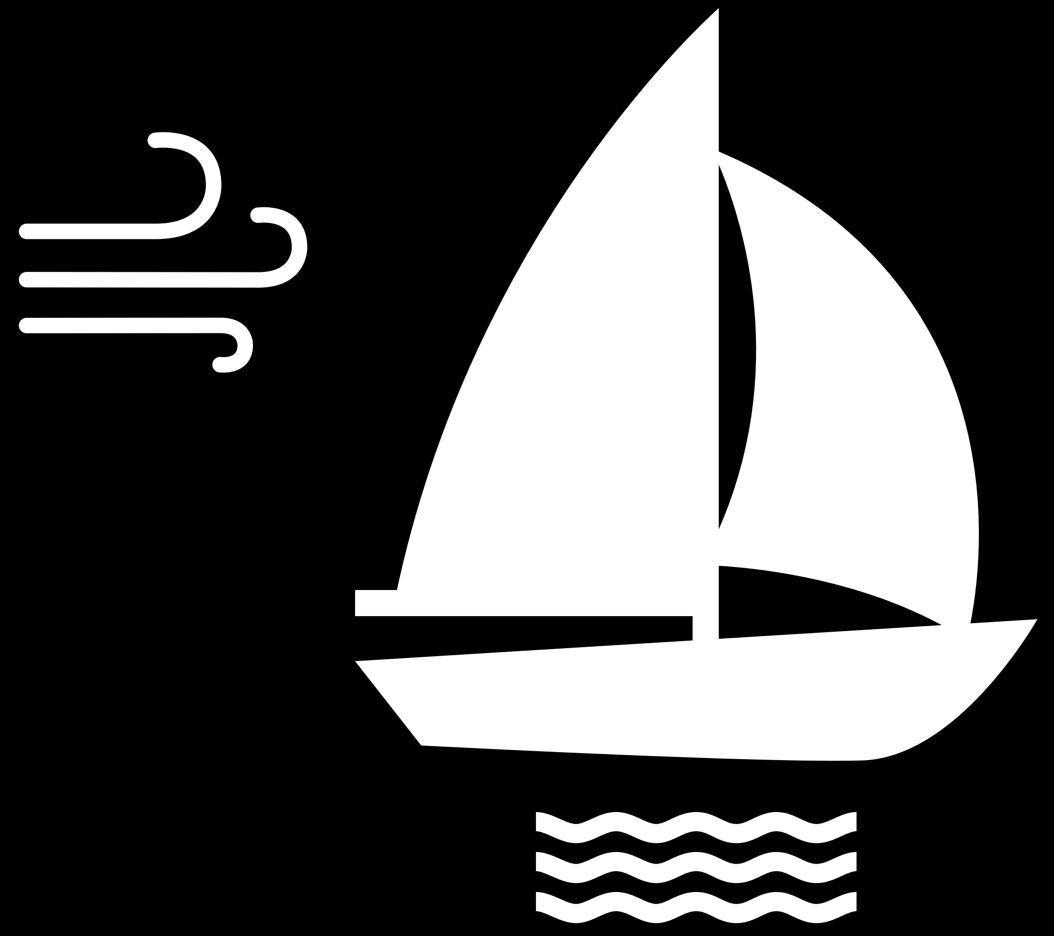
- $x(k|k-1) = F\,.\,x(k-1|k-1) + Bu(k)$
- $\implies \begin{bmatrix} x \\ \dot{x} \end{bmatrix}_t = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix}_{t-1} + \mathbf{Q} \begin{bmatrix} 0.5\Delta t^2 \\ \Delta t \end{bmatrix} \frac{f_t}{m}$
- Here we see the Plant model is chosen to be a constant velocity model
- Any acceleration due to unmodeled factors are considered as noise
- The typical simplifying assumption is that Q, the process noise covariance is that it is drawn iid from $Q \sim N(\mathbf{0}, \mathbf{I})$

Any ways to improve on the prediction model?

- Typically feedback comes from measurement in the form of the difference between the predicted and observed state
- Let’s consider modelling the uncertainty in the plant model in a causal setting
- Let’s say that there are two external forces in our system namely wind and waves of the sea
- Can we utilise this knowledge to formulate a causal problem?

- The relations between wind and waves on the displacement can be captured in a joint distribution
- Both wind and waves will have an effect on the boat’s direction and displacement
- The displacement of the boat is clearly correlated with the velocity of the boat
- Can we make a causal diagram from this information?

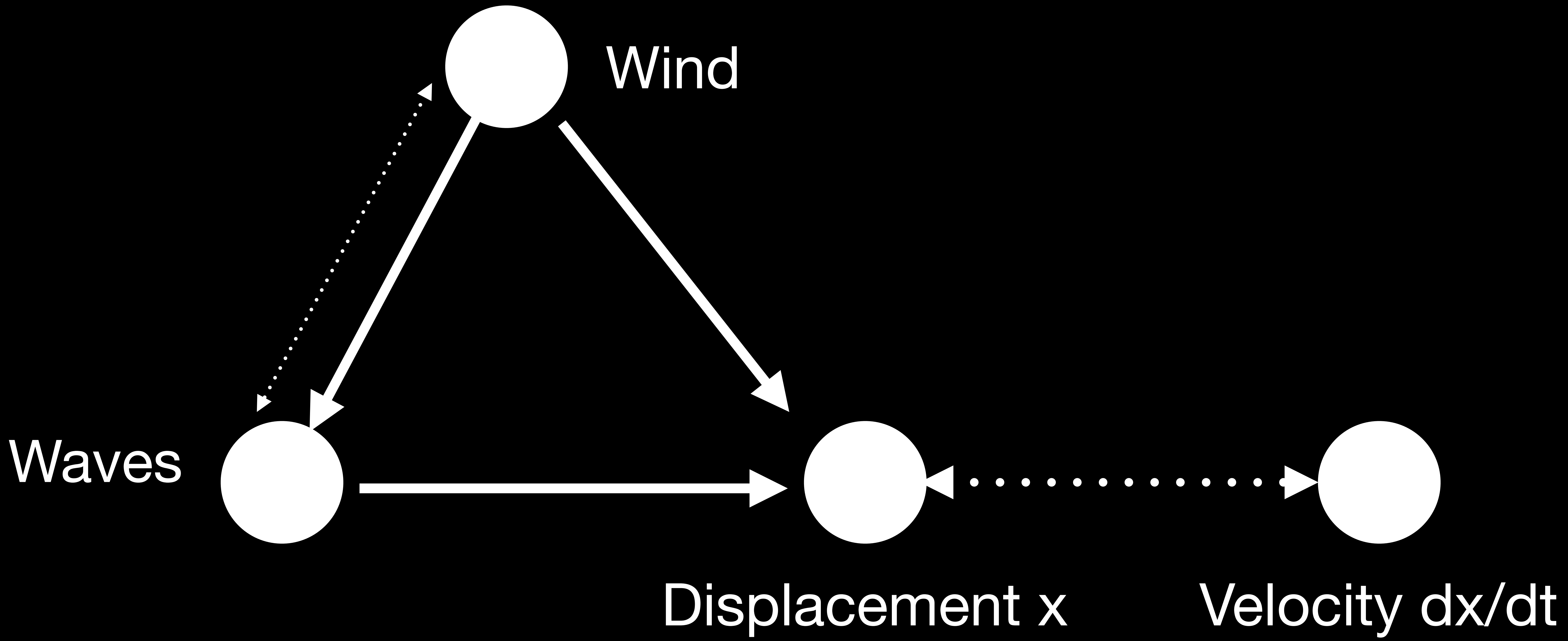
- In the interest of making the joint smaller for illustration only, assume wind and waves can blow along only 1 dimension and the direction in indicated by +1 and -1



Wind	**Waves**	**x**	**dx/dt**
1	1	2	1
1	-1	0	0
-1	1	0	0
-1	-1	-2	-1

Causal diagram of our autonomous boat system

Utilizing causal relations of the surroundings



Can we leverage this knowledge of the physical surrounding ?

Can we improve the Kalman filter navigation algorithm?

- Let’s say we wish to use this causal information to improve:
- 1. The accuracy (low error) of the predicted state vector
- 2. The uncertainty (low covariance) in the predicted state vector
- 3. The time taken for the Kalman filter to converge to the true state
- So that the boat doesn’t hit an obstacle, doesn’t steer off course, is not very bumpy and is safe.
- How can we achieve these aims?

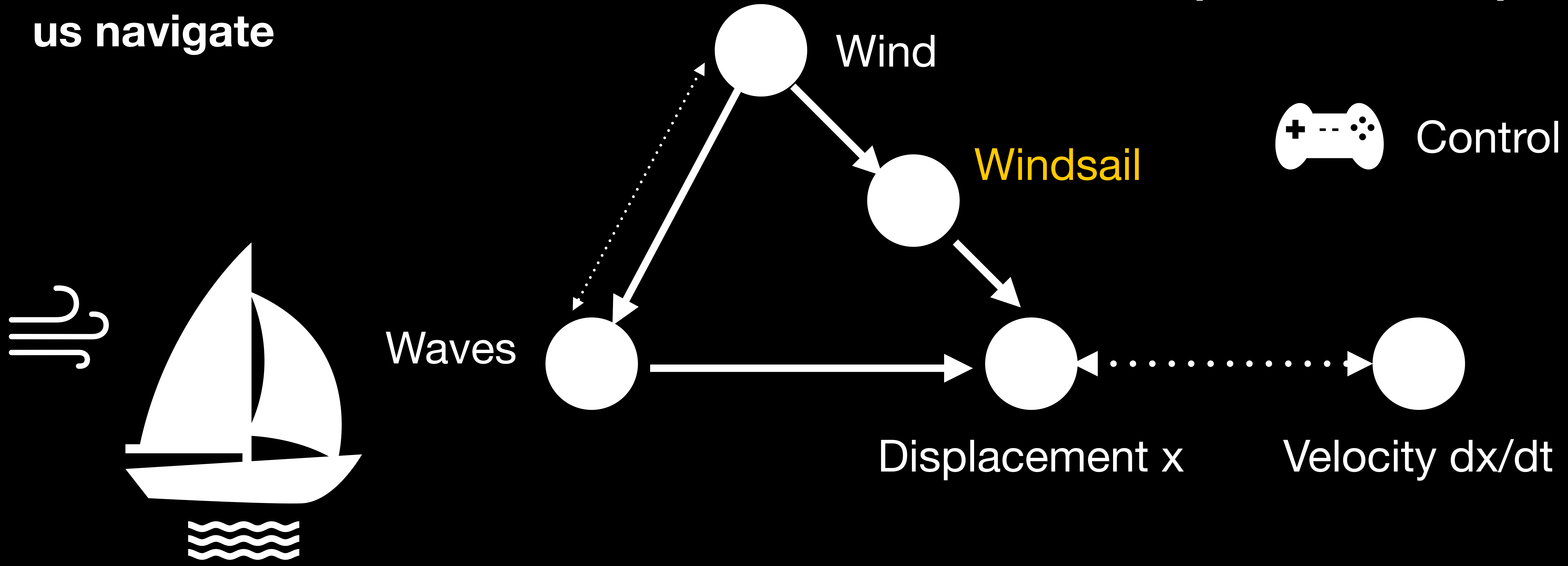
Some ideas:
Observation

- Use wind as an early indicator?
- Wind variable is a parent of the waves so it could be a warning that the boat will be pushed in a certain direction if we measure wind speed.
-

A modification

Now let’s suppose the autonomous boat has a wind sail

- and we wish to use the Kalman filter to control the sail position to help us navigate



Possible joint due to the interaction of wind sail

- SAIL mirrors wind in this case (sail is programmed to follow the direction of wind)

Wind	Windsail	Waves	x	dx/dt
1	1	1	3	2
1	1	-1	-1	0
-1	-1	1	1	0
-1	-1	-1	-3	-2

Intervention

This lets us ask new types of questions now

- What would be the effect on the velocity of the boat if the windsail was set to value SAIL = 1 (a certain direction) & we observed wind =1, waves=-1 (note we are detaching effects of wind on SAIL)

- This intervention has not actually been realised but the autonomous system would like to know the probable effects before actually performing this control action

- Would the system be able to infer this just from the observations of the wind and waves?

Counterfactual

Yet another type of question

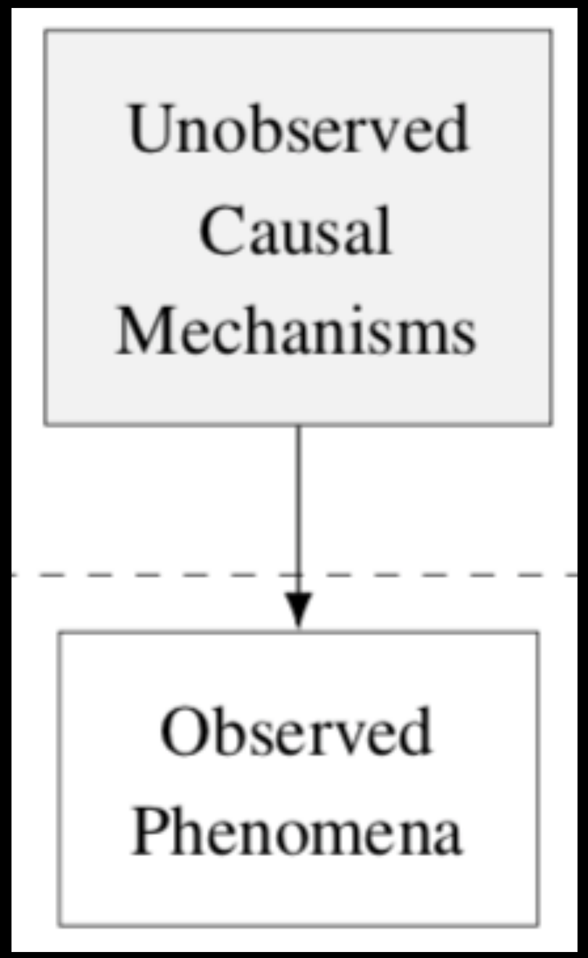
- Given that we have observed for wind = 1 and waves = 1 sail=1 that x = 3
- what position ‘x’ would the boat be at if it were the case that wind=1 waves =-1 sail=1
-

Formalism of the above concepts

Let's conclude the motivating example and make the ideas formal

- It is clear that the underlying mechanisms behind the observational data must be accounted for
- There are categories of questions to be posed
- Aim: use observations to answer questions about

  unobserved events



- We are inspired to breakdown a phenomenon into modular components and describe how these components interact to produce an emergent behaviour
- Doing so allows surgical intervention, decision making & autonomy

Structural Causal Models (What?)

- An SCM denoted by $\mathbb{M}$ is a way to formalise the underlying process that generated the data which we can observe. $\mathbb{M} = \langle \mathbf{U}, \mathbf{V}, \mathbb{F}, P(\mathbf{U}) \rangle$
- $\mathbf{U}$ is a set of background variables i.e. things we do not wish to explicitly model in our system but contribute to the uncertainty of variables we care for
- $\mathbb{V}$ is the set of variables that we care to make causal queries on and are determined by other variables in the model $\mathbb{V} \cup \mathbb{U}$
- $\mathbb{F}$ is a set of functions that relates observable variables to its parents and unobserved variables $v_i \leftarrow f_i(pa_i, u_i)$

How SCMs relate to the autonomous boat?

- U = factor affecting wind and waves e.g. temperature & pressure of an unknown place (because of complex spatial interactions)
- V = wind, waves, sail direction, displacement, velocity
- P(U) = from meteorological data we can determine distributions over wind & waves but they are too complex to have a deterministic function
- $$F = \begin{cases} wind \leftarrow f_{wi}(U_T, U_P) \\ waves \leftarrow f_{wa}(U_T \wedge \text{wind}) \\ sail \leftarrow \text{wind} \\ x \leftarrow \text{wind} \wedge \text{waves} \wedge \text{sail} \end{cases}$$

Seeing (Layer 1)

Formalized as a mapping of events that occur with some uncertainty

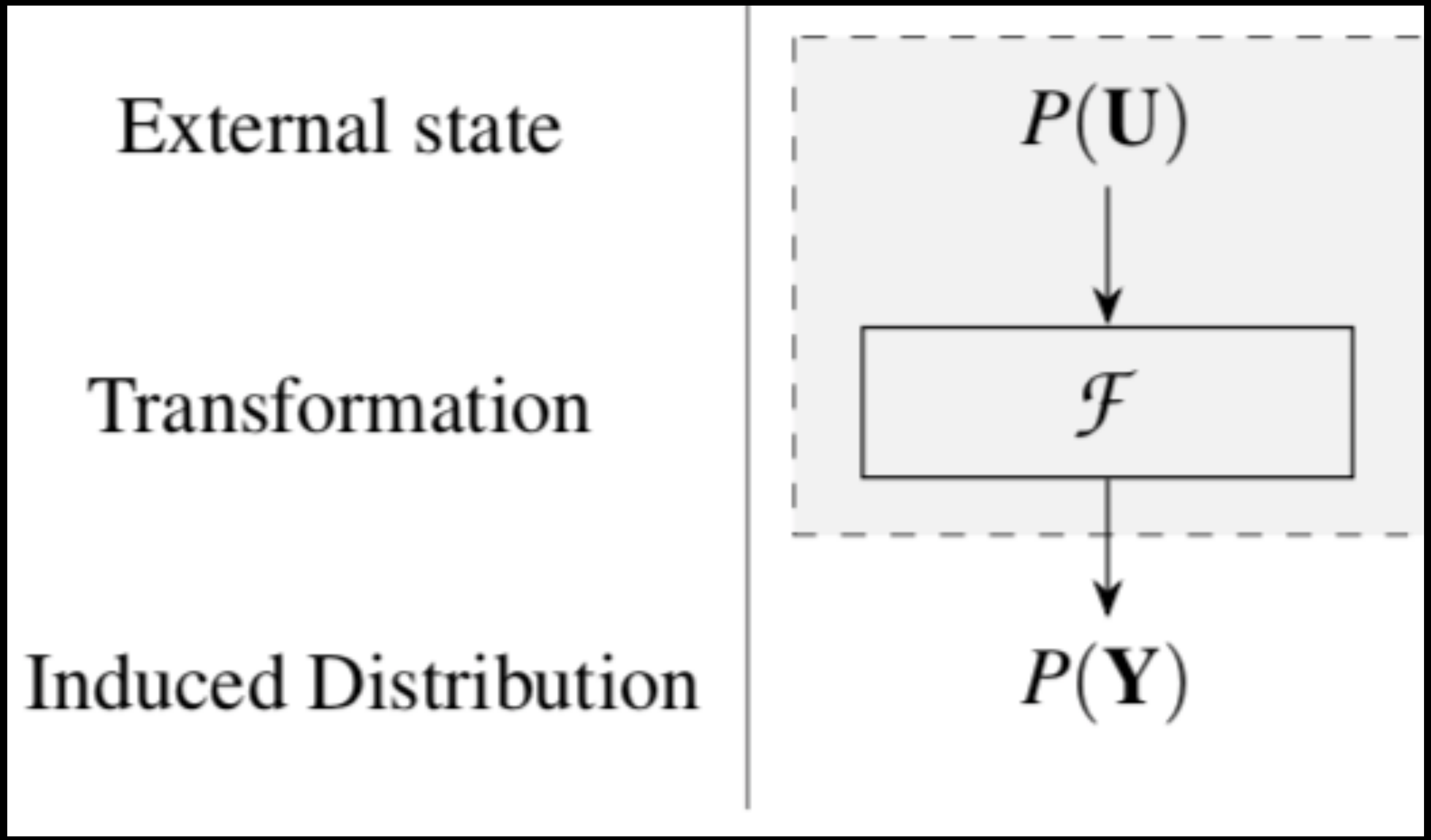
- A joint distribution and conditional $P(X, Y)$ and $P(X \mid Y)$
- Nature evaluates F i.e. maps an external unobserved state which is distributed as P(U) into an observed state distributed as P(V)
- $\mathbf{Y} \subseteq \mathbf{V}$

$$P^M(y) = \Sigma_{\{u \mid Y(u)\}} P(u)$$

- $U = u$ is simply each instantiation of unobserved variables according to their distributions

Illustrating the Layer 1 (Seeing)

- This figure illustrates what we saw in the previous slide in terms of equations.
- Think of F as a mapping of probabilistic events
- The grey shaded region is the Unobserved while white is observed



A seeing type of query

Computation example

$$P(Y=1|X=1) = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{\Sigma_{\{u|Y((u)=1, X(u)=1\}} P(u)}{\Sigma_{\{u|X(u)=1\}} P(u)}$$

- This is easily computed from the joint distribution of U,X,Y

Doing (Layer 2)

Same as Layer 1 except some variables are fixed to constants

- A modification of the SCM gives natural valuations for quantities of this kind
- X = x is fixed while the remaining function mappings are intact :

$$F_x = \{f_i : V_i \notin X\} \cup \{X \leftarrow x\}$$

- Def: Potential response Y_x is the solution for Y (a subset of the endogenous variables V) of the set of equations F_x
- Formally, for a unit $\mathbf{U} = \mathbf{u}$ from unobserved variables, nature evaluates F_x
- This is the same as Layer 1 except a set of variables have been fixed to a constant

Doing (Layer 2)

Taking charge & acting on the situation to bring a certain change

- Evaluation (the only difference from Layer 1 is Y_x instead of Y:

$$P^M(y) = \Sigma_{\{u|Y_x(u)\}} P(u)$$

- Y_x denotes the random variable induced by averaging the potential response $Y_x(u)$ over all u according to $P(U)$
- This procedure disconnects X from any other source of “natural” variation from the original function
- This means changes is Y would be due to changes in X that occurred outside the modelled system in turn guaranteeing a causal effect

Intervention

Computation

- $Y_x := Y | do(x)$ then from Bayes' rule:
- $P(Y | do(x), z) = P(Y_x | z_x) = \frac{P(Y_x, z_x)}{P(z_x)} = \frac{P(Y, z | do(x))}{P(z | do(x))}$

- In our boat example, if we

Intervene by setting sail = 1

Regardless of the wind

$$F = \begin{cases} wind \leftarrow f_{wi}(U_T, U_P) \\ waves \leftarrow f_{wa}(U_T \wedge \text{wind}) \\ sail \leftarrow 1 \\ x \leftarrow \text{wind} \wedge \text{waves} \wedge \text{sail} \end{cases}$$

Calculating in the Autonomous Sail boat example

$$P(x = 1 | do(\text{sail} = 1) = \Sigma_{u|Y_{X=1}(u)=1} P(u)$$

$$= \Sigma_{Y_{x=1}(U_T, U_P)=1} P(U_T, U_P)$$

- Thus we could compute the probability of an outcome after intervening
- In this case, it turned out to be calculable from unobserved factors over which we have a P(U) distribution

Imagining (Layer 3)

Alternative observations AND outcomes than those manifested in reality

- An SCM induces a family of joint distributions over counterfactual events $\mathbf{Y_x}, \ldots, \mathbf{Z_w}$ for any $\mathbf{Y}, \mathbf{Z}, \ldots, \mathbf{X}, \mathbf{W} \subseteq \mathbf{V}$
- $P(Y_{X=1} = 1 \,|\, X = 1, Y = 1)$ is read as the probability of event Y=1 happening in an alternative reality had event X=1 taken place conditioned on the present reality
- P(alternative reality of X & Y | present reality of X & Y)

Evaluating Counterfactual

An example with some particular SCM

$$P^M(y_x, \ldots z_w) = \Sigma_{\{u|Y_x(u)=y,\ldots,Z_w(u)=z\}} P(u)$$

$$P(Y_{X=1} = 1 \,|\, X = 0, Y = 0) = \frac{P(Y_{X=1} = 1, X = 0, Y = 0)}{P(X = 0, Y = 0)}$$

$$= \frac{\Sigma_{\{u|Y_{x=1}(u)=1, X(u)=0, Y(u)=0\}} P(u)}{\Sigma_{\{u|X(u)=0, Y(u)=0\}} P(u)}$$

Counterfactual question for the robot example

- Given that we have observed for wind = 1 and waves = 1 sail=1 that x = 3
- What position 'X' would the boat be at if it were the case that wind (W) =1 waves (A) =-1 sail (S)=1
- Limiting the equation for just X and W $P(X_{W=1} = 1 \,|\, X = -1, W = -1)$
- Note that the case X=1 W=1 has not been realised so it is an alternative world
- The case on which we condition has been realised X=-1 W=-1
- This type of question can be useful for the robot if it wants to learn a navigation policy using reinforcement learning and wishes to calculate 'cumulative regret'

A logical perspective

Questions that have been posed and answered by a PCM

- L1: How likely is Y given that we observe X?
- L2: How likely would Y be if one were to make it the case that X = x?
- L3: Given I have observed X and Y, how likely would Y have been if X’ had been true instead of X?

Symbolic Languages

Definition

- Let variables V be given and X,Y,Z ⊆ V. Each language L_i i=1,2,3 consists of Boolean combinations of inequalities between polynomials over terms $P(\alpha)$ where $P(\alpha)$ is an L_i term
- L_1 terms are of the form $P(Y = y)$encoding the probability that Y takes on values y
- L_2terms additionally include probabilities of conditional expressions $P(Y_x = y)$ giving probability that variable Y would take on values y were X=x
- L_3 encodes probabilities over conjunctions of conditional (L_2) symbolising the joint that all conditional statements hold simultaneously $P(Y_x = y, \ldots, Z_w = z)$

Examples of this syntax

- $L_1 : P(X = 1 \mid Y = 1) = P(X = 1)P(Y = 1)$
- $L_2 : P(Y_{X=1} = 1) = 3/4$ is the probability of Y taking on value 1 were X to take on value 1 is 3/4
- $L_3 : P(y_x, y'_{x'}) \geq P(y \mid x) - P(y \mid x')$ is a statement expressing a lower bound on the probability of necessity and sufficiency
- Def: Pearl Causal Hierarchy (PCH) is the collection of distributions induced by languages L_1, L_2, L_3 (syntax) . If M^* is a fully specified SCM then its PCH is fully specified

Q: Is there an increasing logical expressiveness for L_1, L_2, L_3?

If not, then the hierarchy between these languages collapses

- Consider an example where $M^* = \langle \mathbf{U} = \{U\}, \mathbf{V} = \{X, Y\}, \mathbb{F}, P(U) \rangle$
- U is distributed as a fair coin flip $F = \begin{cases} X \leftarrow U \\ Y \leftarrow X \end{cases}$
- We must have $f_Y(x, u) = x$ for any unit u or else L_2-probability $P(y_x)$ differs between M and M^*. f_X is also determined by L_1 requirement that $P(x)$ match between M, M^*. This is enough to determine all L_3 quantities.
- A 'collapse' would mean we can draw all possible causal conclusions merely from correlations

Collapse relative to $M*$

- Let Ω be the set of all possible SCMs
- Layer j of the causal hierarchy collapses to Layer 'i' with $i < j$ relative to $M* \in \Omega$ if $M* \sim_i M$ implies that $M* \sim_j M$ for all $M \in \Omega$
- Theorem: Causal Hierarchy Theorem (CHT) almost never collapses i.e. for almost any SCM the layers of the hierarchy remain distinct

A graphical perspective

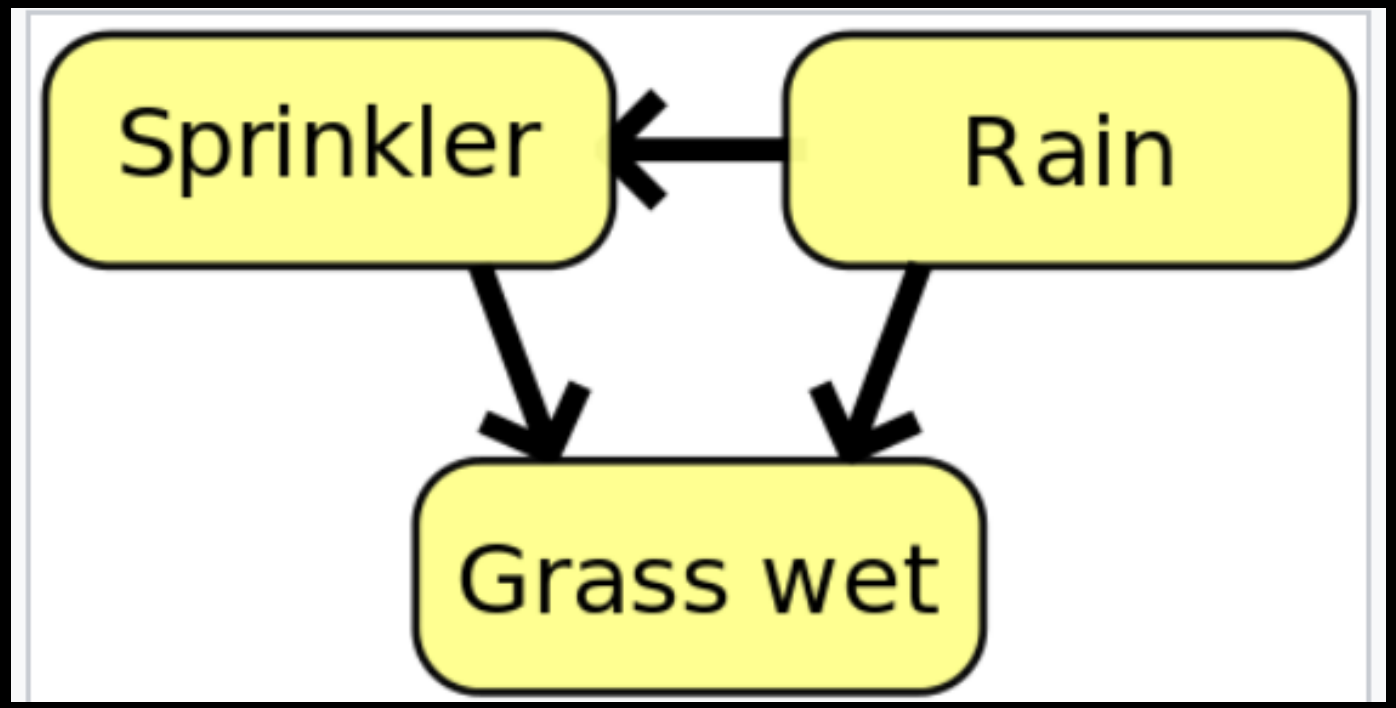
A fundamental way to study causal inference

- So far we have seen a semantic and logical approach to the need for SCMs
- We have also seen the presence of three layers of hierarchy which are in increasing order of expressivity
- In the motivating Kalman Filter example we tried to draw relations between factors in the model as a graph
- Now we formalise the notion of graphs and note the differences between the types of graphical approaches

Bayes Nets

- represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG)
- ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor
- This classic Bayes net shows conditional

Probability relations between the variables



A graphical perspective

Why are Bayes Nets not enough?

- What type of causal knowledge would allow us to make cross-layer inferences i.e. for example, go from $P(Y|X)$ to $P(Y|do(x))$
- Bayes nets although sometimes erroneously believed to encode causal knowledge unfortunately falls short on bridging the gap between layers 1 and 2
- Why? Because it fails to distinguish between two mechanisms that have the same observational data, same conditional independence relations (both of which can be found from a bayes net) but react differently to interventions which the bayes net can't inform us of.

Illustrating the mapping of SCMs and interventions

Bayes net give an illusion of causality yet their test is interventions

- It is possible for different

Data generating mechanisms

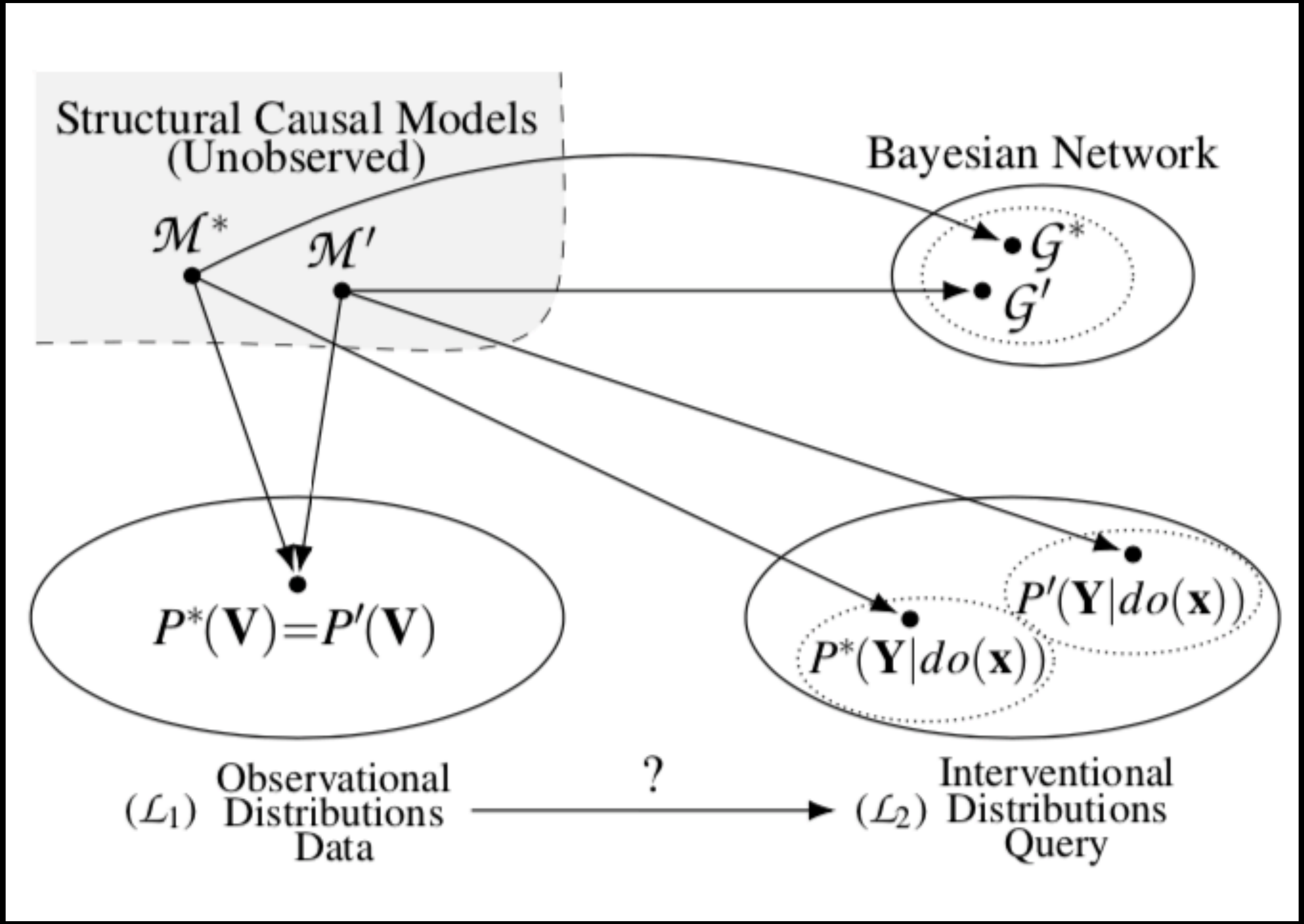
To have the same observations

Data and the same Conditional

independence relations yet clearly

Will have differing interventional

reactions



Example of Bayes Net's inability for L2 inference

$$\mathbb{F}_1 = \begin{cases} X \leftarrow U_x \\ Z \leftarrow X \oplus U_z \\ Y \leftarrow Z \oplus U_y \end{cases}$$

$$\mathbb{F}_2 = \begin{cases} X \leftarrow Z \oplus U_x \\ Z \leftarrow Y \oplus U_z \\ Y \leftarrow U_y \end{cases}$$

	Same	Different	
$P^{1,2}(V)$	**Observational Data P(V)**	SCMs	$\mathbb{M} = \langle \mathbf{U}, \mathbf{V}, \mathbb{F}, P(\mathbf{U}) \rangle$
$X \perp Y \mid Z$	**Conditional Independence relations**	Interventional distributions	$P(Y \mid do(X = x))$

Assymetry between cause and effect

A failing of L1 constraints to make L2 inferences

- Cause (X) may change the effect on a certain variable (Y)
- But that doesn’t imply that changing this effect variable (Y) will alter the cause variable (X)
- Figure shows that observational data is not enough to differentiate different causal mechanisms or make L2 inference
- For example in the Bayes net image seen before, the arrow from Rain -> Sprinkler indicates Rain affects the outcome of sprinkler but not the other way round

Short-coming of Bayes Net (BN)

- Bayes net is thus a tool for formalising L1 data $P(x,y)$ $P(x|y)$ and conditional independences $Y \perp X \mid Z$
- The gap between L1 and L2 is not filled by Bayes net because it can’t differentiate the causal effect of an intervention
- For example in the Sprinkler Bayes net we can not find the effect of altering the sprinkler on the state of wetness
- We need to go one step further to formalise causal effects $P(y|do(x))$

Causal Diagram (Markovian)

Constructed from an SCM

- 1. A vertex for an endogenous variable every the SCM v_i
- 2. A directed edge between each vertex that appears in the function $f_i \in \mathbb{F}$
- Intuitively the arrow represents a master-slave relation ($A \to B$ means B listens to A)
- Functionally, the edge between A and B indicates that $B \leftarrow f(A)$
- X causes Y and both are affected by an exogenous variable

Causal Bayesian Network (stronger than BN)

Properties for a CBN in a Markovian setting

- 1. Markovian $P(v_i | pa_i, pa(pa_i)) = P(v_i | pa_i)$
- 2. Missing Link $P(v_i | do(pa_i), do(x)) = P(v_i | do(pa_i))$ for $V_i \in V, V_i \notin X$
- After intervening on the Parent of a variable, the variable is insensitive to any other intervention in the system
- 3. Parents do/see $P(v_i | do(x), do(pa_i)) = P(v_i | do(x), pa_i)$
- Whether the function takes the value of its arguments by intervention or by observation, the same behaviour for it is observed

Differences between CBN and BN

- Encodes stronger assumptions than BN like constraints 2 and 3 seen in the last slide

- Missing arrows in a BN indicating conditional independence

- While in CBN missing arrow indicates lack of direct effect

A Theorem about CBNs

- Theorem: The causal Diagram induced by the SCM is a CBN for the collection of observational and experimental distributions induced by M
- CBN can act as a basis for causal reasoning when the SCM is not fully known and a collection of interventional distributions is not available.

Another Theorem about CBNs

- Theorem- Truncated Factorization Product (Markovian): Let the graphical model G be a CBN for the set of interventional distributions. For any $\mathbf{X} \subseteq \mathbf{V}$ the interventional L2 distribution $P(V \mid do(X) = x)$ is identifiable through the truncated factorisation product:

$$P(v \mid do(x)) = \Pi_{\{i \mid V_i \notin X\}} P(v_i \mid pa_i) \mid_{X=x}$$

  - The L2 expression on the LHS is written in terms of the L1 observation RHS
  - Note the factors relative to the intervened variables are removed

Back door criterion (Markovian setting)

Markovian Models - those without unobserved confounders

- For *any* treatment X and outcome Y, the interventional distribution

$$P(Y|do(x))) = \Sigma_z P(Y|x,z)P(z)$$

- If the set of covariates Z is constituted by all pre-treatment variables and all relevant sources of variations are measured then adjusting for these variables will lead to the causal effect

Blessings of a Markovian Situation

A strong assumption but has nice properties

- To re-emphasize: A markovian assumption is that every node is conditionally independent of its non-descendants given its parents (has no bearing on nodes which do not descend from it)
- Then the nice properties discussed above follow namely:
- CBN is a perfect surrogate for the Causal diagram induced by the SCM
- For *any* $X \subseteq V$ the $P(v \mid do(x))$ is identifiable through truncated factorization
- L2 quantities (causal effects) are computable from the observational data (L1 data)

Semi-Markovian Causal Bayes Networks

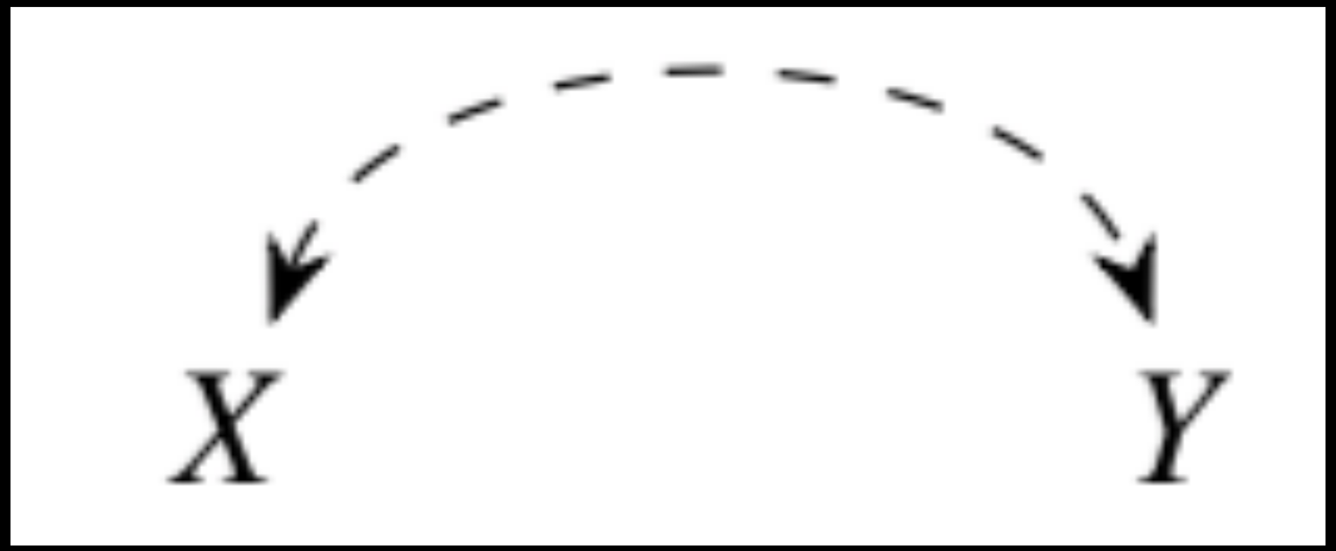
Reality hits us in the form of Unobserved confounders

- All relevant factors about the phenomenon under study are not measured
- Example: Roll two dice and define events X & Y as sum and difference of outcomes.
- If X =2 then the outcomes have to be exactly 1 and 1. So P(Y=0)=1
- So clearly X & Y are not independent. How about put an arrow X ->Y
- That would mean X causes Y and so this should be true: $P(Y|do(X=2)) = P(Y)$ which is not (reporting X=2 doesn't change Y)

Unobserved confounder

- Realize: certain dependencies among endogenous variables cannot be explained by other variables in the model

- Neither can they be ignored because $X \not\perp\!\!\!\perp Y$!



- This dotted arrow is neutral with respect to the interventional invariance i.e.

$$P(Y \mid do(X)) = P(Y)$$

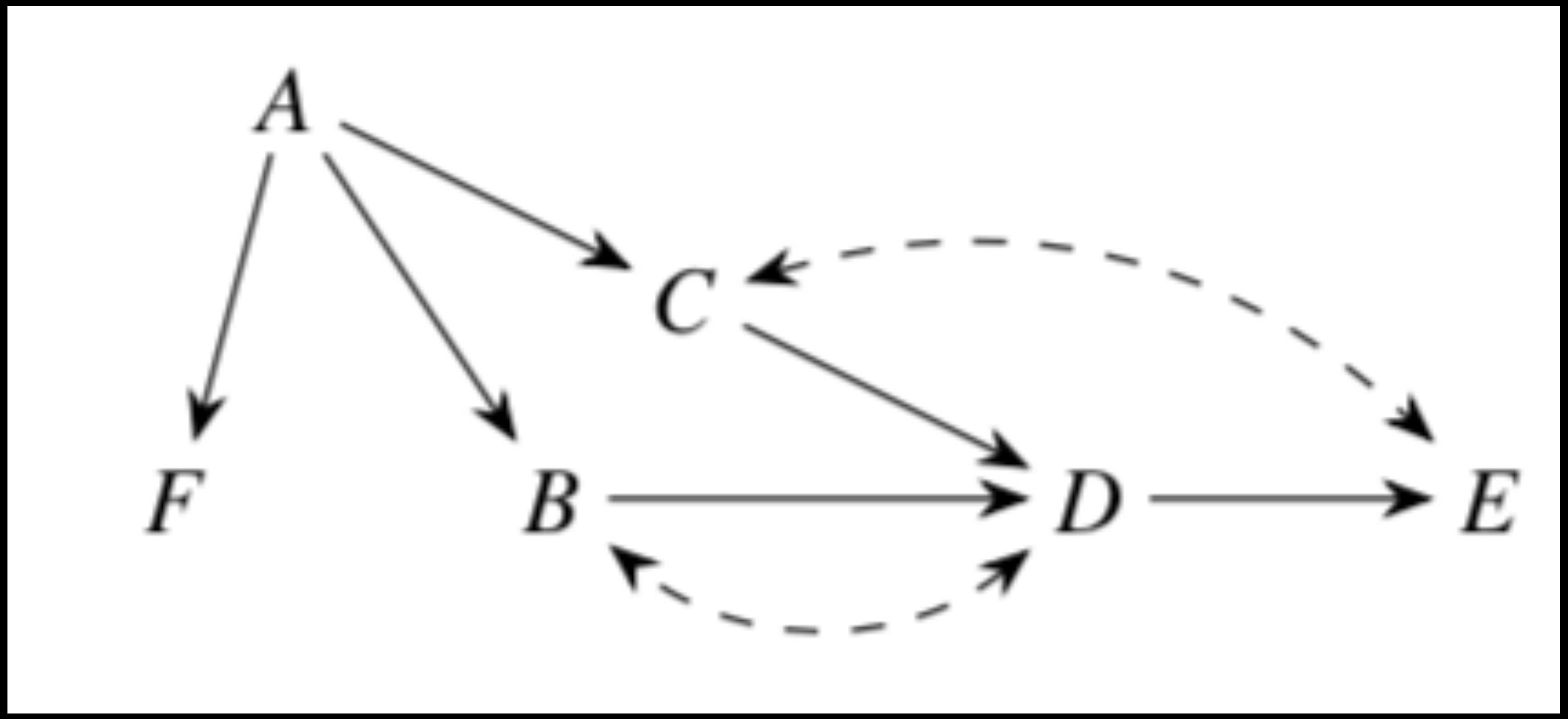
Causal Diagram (Semi-Markovian Models)

- G is a causal diagram of an SCM if it is constructed as:
- 1. A vertex for every endogenous variable in $\mathbf{V}$
- 2. An edge for for every $V_i, V_j \in \mathbf{V}$ if V_j is an argument for $f_i \in \mathbf{F}$
- 3. A bi-directed edge $V_j < - - > V_i$ for every $V_i, V_j \in \mathbf{V}$ if $U_i, U_j \subset \mathbf{U}$ are correlated or the corresponding functions f_i, f_j share some $U \in \mathbf{U}$ as an argument

Family Properties of Causal Diagrams

A familiar Bayes net property does not hold

- Each SCM induces a unique causal diagram (in contrast to Bayes net - an SCM to BN mapping was not 1-1)
- Family relations in Semi-Markovian models are less well-behaved than in Markovian



Notice: The markovian property of a variable being conditionally independent from its non descendants given its parents doesn't hold here. For node D

{Non-desc} $\backslash Pa_d = \{A, F\}$

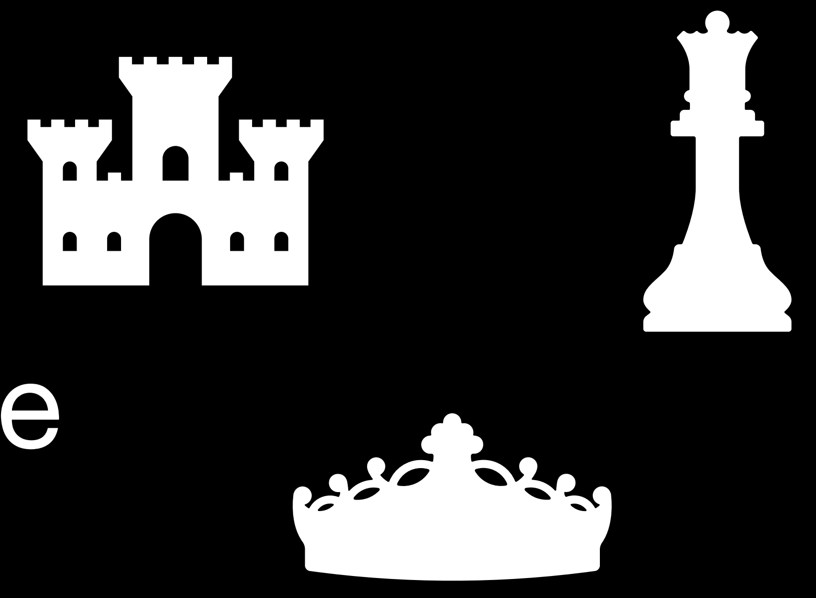
$Pa_d = \{B, C\}$ and $D \not\perp\!\!\!\perp \{A, F\} \mid \{B, C\}$

Because of D <—>B <— A

Boundary of influence of nodes in a Graph

A property about causal graphs that is important to understand

- Think of this property as a region of influence of a node in a graph
- This region can be limited by conditioning on relatives of a node
- An analogy: let nodes represent members of a royal family in Europe
- Let us assume they are fighting for the throne as their objective
- If a member is independent of others then he/she gets the throne
- To condition on other nodes is similar to winning the confidence of these members - once conditioned on certain nodes they help you become independent from other ancestors. For Markovian case conditioning on parents was enough



Confounded Components

A way to determine modularity in semi-markovian models

- Like we saw in the Markovian case - to partition nodes such that we can claim they are conditionally independent we only needed to condition on Parents. In Semi-Markovian we need a new property

- C-Components: Let $\{C_1, C_2, \ldots, C_k\}$ be a partition over the set of variables $\mathbf{V}$ where C_i is a confounded component if for every pair V_i, V_j of nodes in C_i there exists a path made up entirely of bi-directed edges

- Looking at the image on the previous slide note that $\{B, D\}, \{C, E\}, \{A\}, \{F\}$ form C-Components.

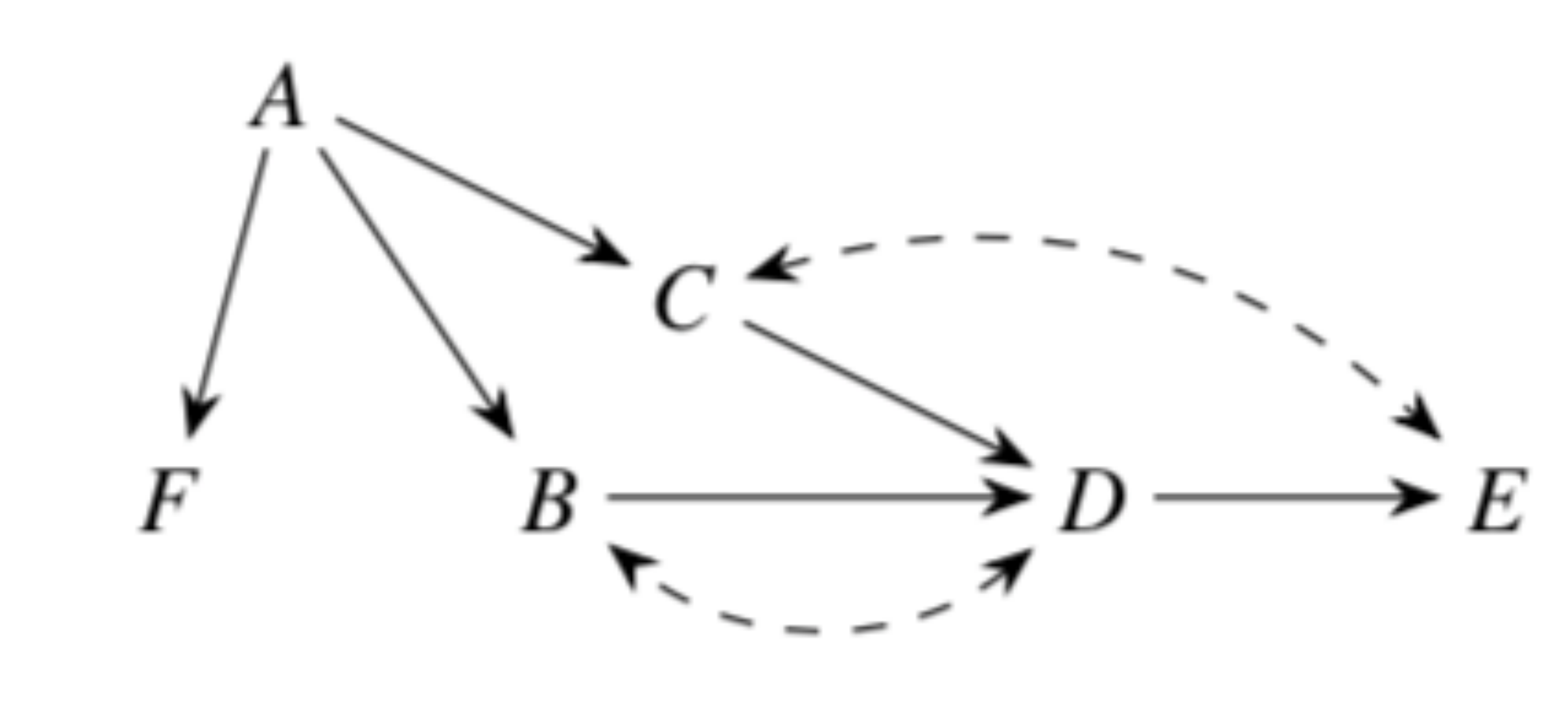
C-Components continued

- For each endogenous variable V_i we need to condition on :
- 1. its parents
- 2. Variables V_j in the same C-component siblings that precedes (topologically) it
- 3. Parents of the V_j
- Then V_i is shielded from other non-descendants in the graph
- A Node should win confidence of its parents and its older sibling's parents to become independent of its non-descendants (easy to remember). Define this set as Pa_i^+

Semi-Markovian Relative

Through an example

- One topological order is A<B<C<D<E<F
- $P(v) = \Pi_{V_i \in V} P_x(v_i | pa_i^{x+})$
- $pa_i^+ = Pa^1(\{V \in C(V_i) : V \leq V_i\}) \setminus \{V_i\}$
- Thus:
- $P(a, b, c, d, e, f) = P(a)P(b | a)P(c | a)P(d | b, c, a)P(e | d, c, a)P(f | a)$



Causal Bayesian Network (Semi-Markovian)

Properties of a CBN in a semi-Markovian setting

- Let P_* be the collection of all interventional distributions $P(V \mid do(X = x))$
- A graphical model with directed and bidirected edges is a CBN if for every intervention :
- 1. $P(V \mid do(x))$ is semi-Markov relative to $G_{\bar{x}}$
- 2. For every $V_i \in V \setminus X, W \subseteq V \setminus (Pa_i^+ \cup X \cup \{V_i\})$
- $P(v_i \mid do(x), pa_i^{x+}, do(w)) = P(v_i \mid do(x), pa_i^{x+})$
- meaning conditioning on the set of augmented parents Pa_i^{x+} renders V_i invariant to an intervention on other variables

- Missing bidirected link: For every node, let us partition into two sets confounded and unconfounded parents Pa_i^c and Pa_i^u in $G_{\bar{x}}$
- Then $P(v_i \mid do(x), pa_i^c, do(pa_i^u)) = P(v_i \mid do(x), pa_i^c, pa_i^u)$
- This means we are relaxing the stringent do/see conditions in a Markovian CBN

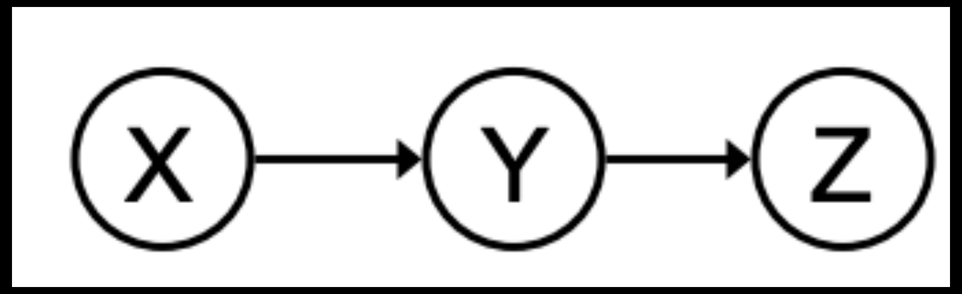
Identifiability

- Effect identifiability: the causal effect of an action is said to be identifiable from P and G if for every two models with the same causal diagram G the observational distributions being equal implies the interventional distributions to also be equal $P^1(v) = P^2(v) \implies P^1(Y|do(x), z) = P^2(Y|do(x), z))$
- This formalises the cross layer queries we wish to answer

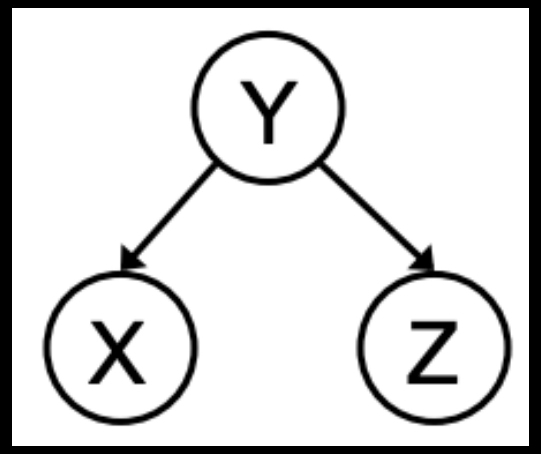
Do -Calculus

Building on top of d-separation for interventional distributions

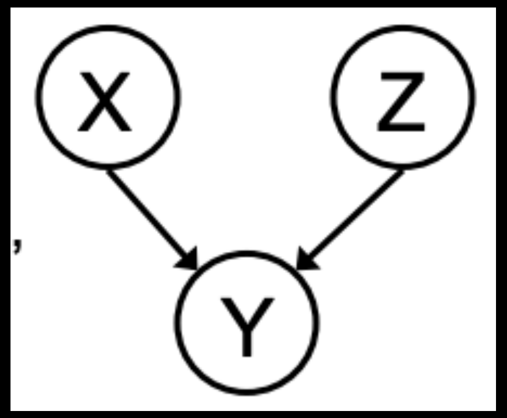
- We need rules that allow us to navigate among interventional distributions and jump across unrealised worlds
- These rules will be licensed by the invariances encoded in the causal graph
- Recall d-Separation



$X \perp Z | Y$



$X \perp Z | Y$

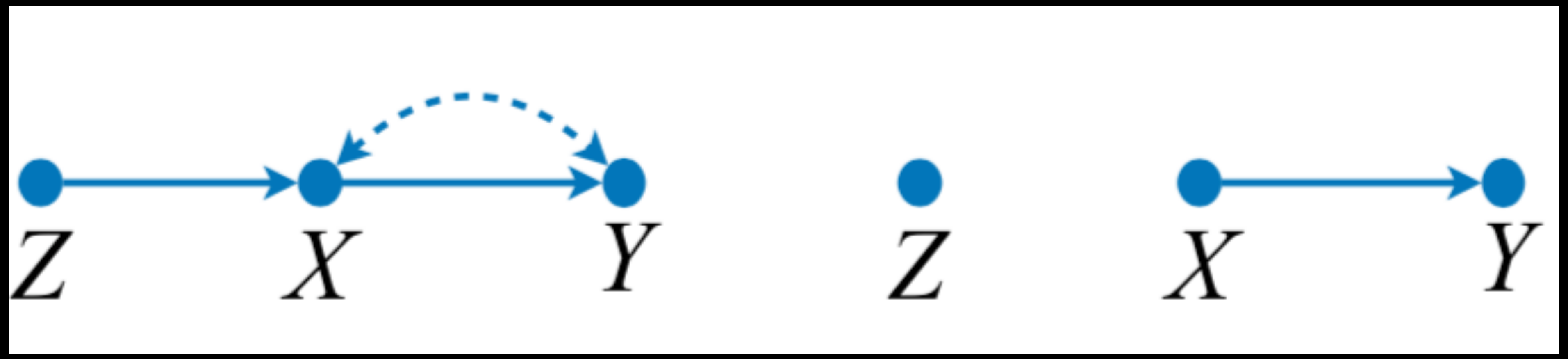


$X \not\perp Z | Y$

- Rule 1 of d-separation is an extension of this but in the causal graph $G_{\bar{X}}$

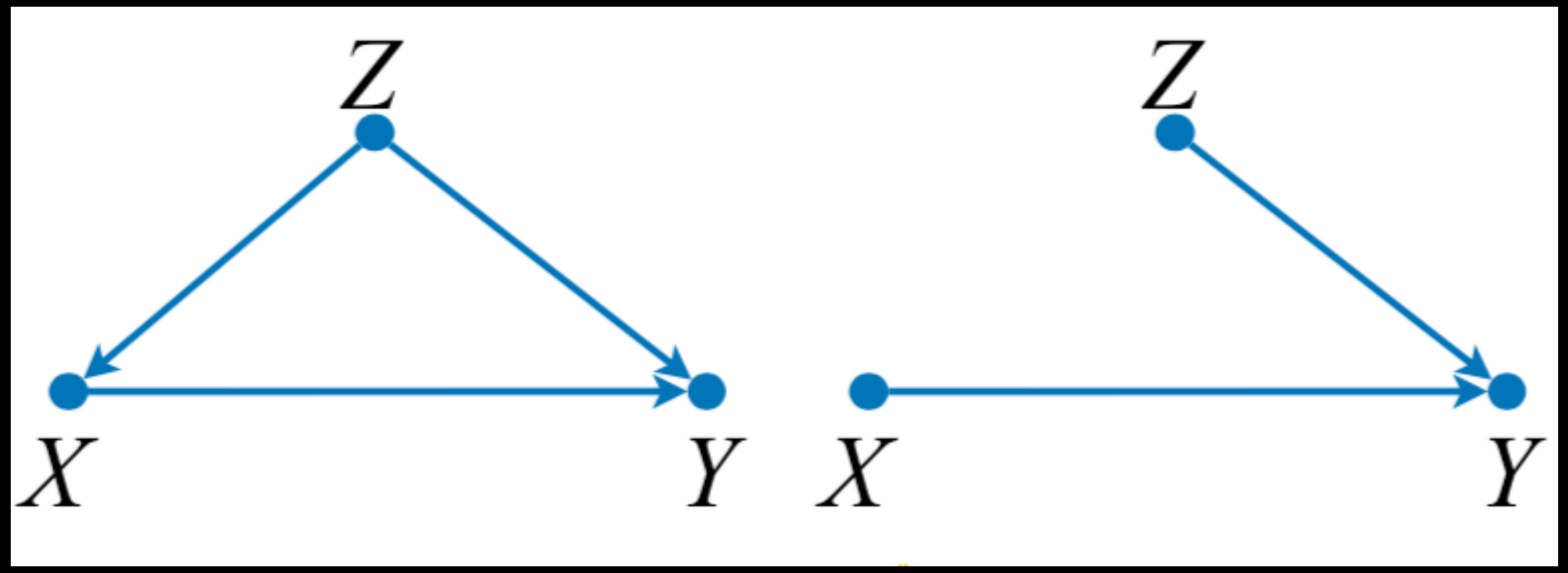
Rule 1

- Let G be the CBN for P_*then for any disjoint sets $X, Y, Z, W \subset V^* \backslash$
- Rule 1: $P(y \mid do(x), z, w) = P(y \mid do(x), w)$ if $(Y \perp Z \mid X, W)$ in $G_{\bar{X}}$
- This is about adding or removing observations

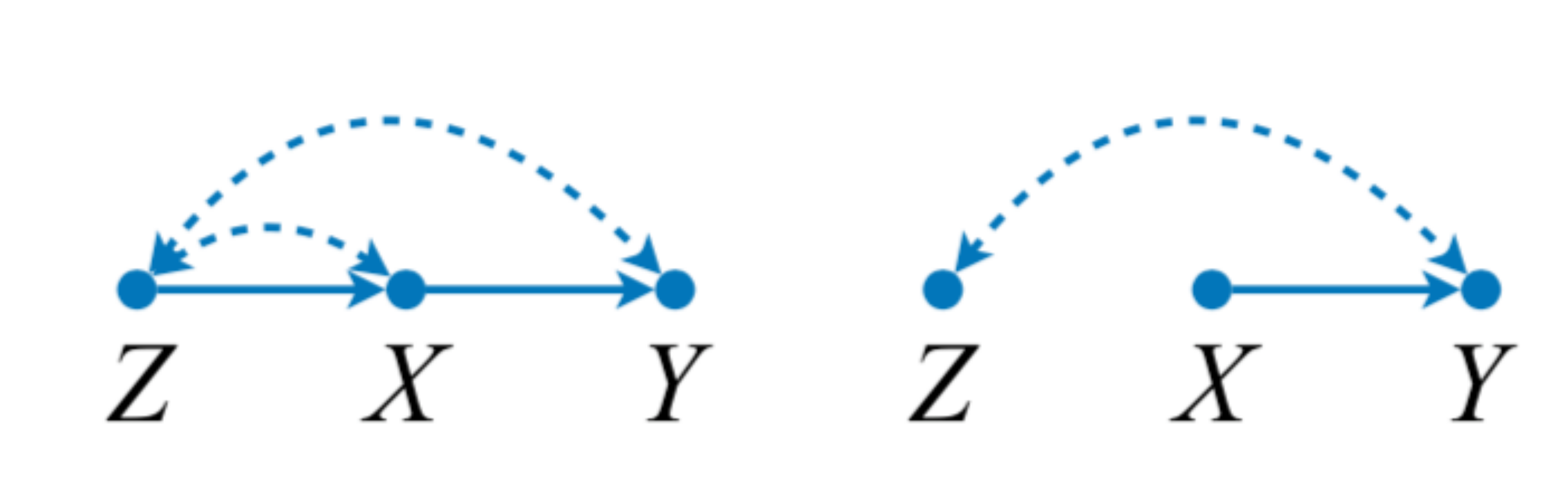


Rule 2

- Rule 2: $P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w)$ if $(Y \perp Z \mid X, W)$ in $G_{\bar{X}\underline{Z}}$
- This is exchanging an action with an observation or vice versa
- After observing Z, Y reacts to X in the same way with or without intervention
-



- Rule 3: $P(y \mid do(x), do(z), w) = P(y \mid do(x), w)$ if $(Y \perp Z \mid X, W)$ in $G_{\bar{X}\bar{Z}(\bar{W})}$
- This is about Adding or removing actions
- If there is no causal path from X to Z then an intervention on X will have no effect on Z



Back door Criterion

- 1. No node in Z should be a descendent of X
- 2. Z should block every path between X and Y that contains an arrow into X
- Then the causal effect is identifiable by $P(Y|do(x)) = \Sigma_z P(Y|x,z)P(z)$
- The most practical way of checking is by removing outgoing arrows from X and confirming whether Z separates X and Y

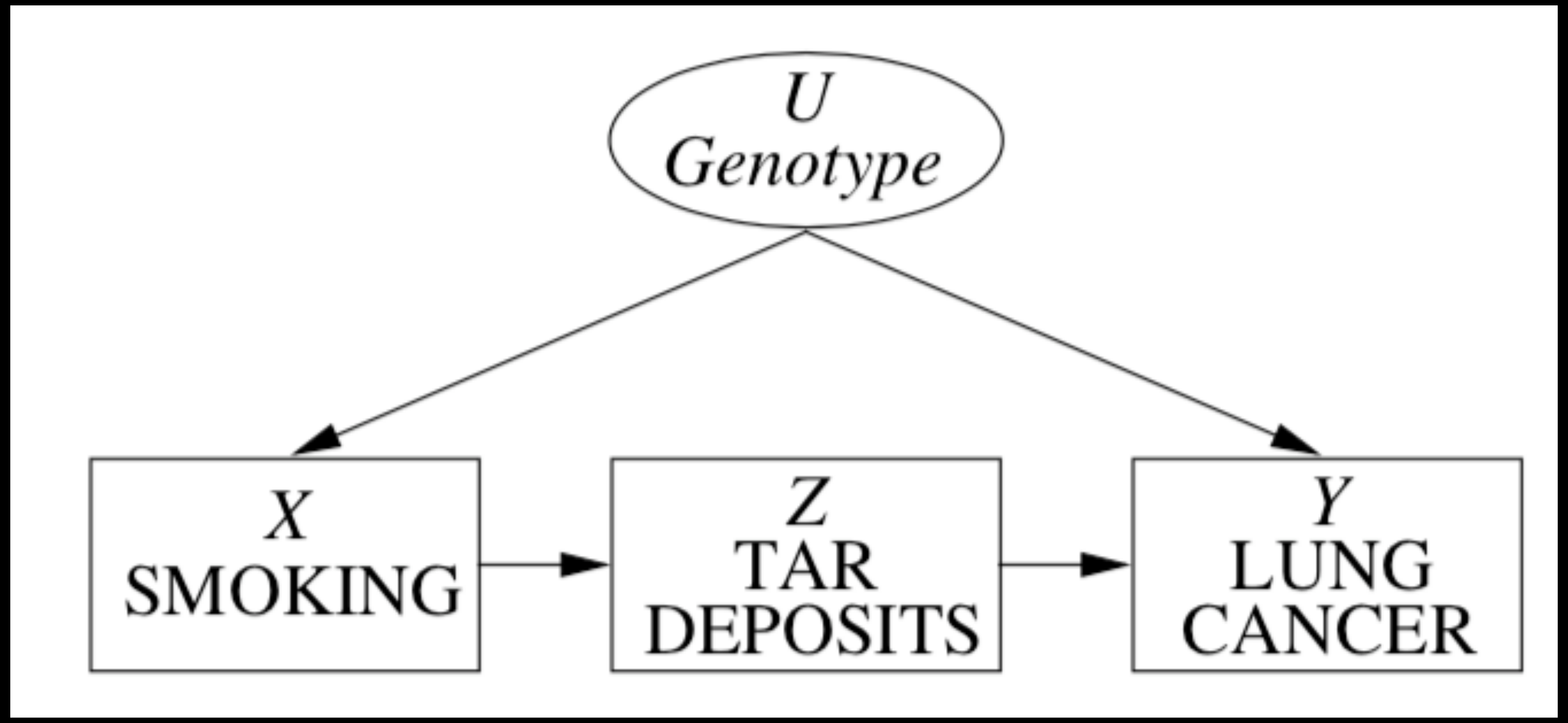
Front door

Another identifiable criterion

- Criteria: Z is said to satisfy the front-door criterion relative to X & Y if
- 1. Z intercepts all directed paths from X to Y .
- 2. There is no unblocked backdoor path from X to Z.
- 3. All backdoor paths from Z to Y are blocked by X.
- The image shows the causal graph for the age old debate

Regarding does smoking cause cancer. This happens to be ID

.



Recent Developments

Stochastic, conditional, and non-atomic interventions

- It may be challenging to assess the effect of new soft intervention from non-experimental data
- Sigma Calculus and Soft interventions introduced by Correa and Bareinboim 2020 is a response to that

Conclusion

- We have started this interesting topic with a motivating example from probabilistic robotics which is very much grounded in reality
- We referred to this time and again as the theories were introduced along with supplemental examples from text
- This slideshow has shown three distinct perspectives : Semantic, Logical and Graphical to tackle causal inference
- Hope the reader is inspired to do research in the same
- Thank you!
- Nihaar Shah (ns3413)