Coming back to the lab

- The set of all paths between two nodes is implemented in the function `all_simple_paths`
- The last question is in some sense still open:
  you can do it brute force (every set of variables in the graph).
  Be careful: a set containing a set that d-separates may not d-separates, due to colliders

  Recent paper on the subject (to get efficient algorithms):
  Finding Minimal d-separators in Linear Time and Applications, Benito van der Zander, Maciej Liśkiewicz, Proceedings of The 35th Uncertainty in Artificial Intelligence Conference, PMLR 115:637-647, 2020

Introduction to structural causal models

Charles K. Assaad, Emilie Devijver, Eric Gaussier

emilie.devijver@univ-grenoble-alpes.fr

Table of contents

Deterministic Structural Causal Models

A set of variables with:

- a causal graph,
- a set of equations describing how each variable depends upon its immediate causal predecessors.

Convention: each equation has one effect variable on the left hand side, and the cause variables on the right hand side. Any variable that makes no difference to the value of the effect variable is excluded from each equation.

Deterministic Structural Causal Models

A set of variables with:

- a causal graph,
- a set of equations describing how each variable depends upon its immediate causal predecessors.

Convention: each equation has one effect variable on the left hand side, and the cause variables on the right hand side.
Any variable that makes no difference to the value of the effect variable is excluded from each equation.

Deterministic Structural Causal Models

A set of variables with:

- a causal graph,
- a set of equations describing how each variable depends upon its immediate causal predecessors.

Convention: each equation has one effect variable on the left hand side, and the cause variables on the right hand side. Any variable that makes no difference to the value of the effect variable is excluded from each equation.

Example

Consider a gas grill, used to cook corn.
Set of variables:

- Gas connected (1 if yes, 0 if no)
- Gas knob (0 for off, 1 for medium, 2 for high)
- Gas level (0 for off, 1 for medium, 2 for high)
- Igniter (1 if pressed, 0 if not)
- Flame (0 for off, 1 for medium, 2 for high)
- Corn on (0 for no, 1 for yes)
- Corn cooked (0 for raw, 1 for medium, 2 for well done)

Set of equations:

- Gas level = Gas connected $\times$ Gas knob
- Flame = Gas level $\times$ Igniter
- Corn cooked = Flame $\times$ Corn on

Example

Consider a gas grill, used to cook corn.
Set of variables:

- Gas connected
- Gas knob
- Gas level
- Igniter
- Flame
- Corn on
- Corn cooked

Set of equations:

- Gas level = Gas connected $\times$ Gas knob
- Flame = Gas level $\times$ Igniter
- Corn cooked = Flame $\times$ Corn on

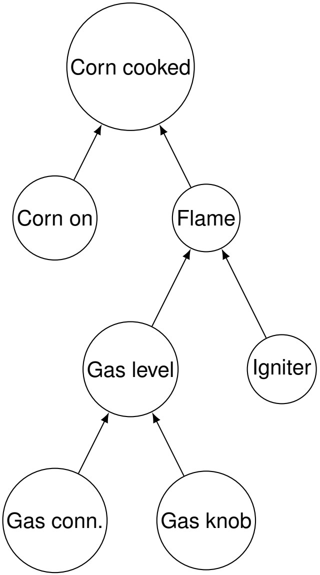
Example

Consider a gas grill, used to cook corn.
Set of variables:

- Gas connected
- Gas knob
- Gas level
- Igniter
- Flame
- Corn on
- Corn cooked

Set of equations:

- Gas level = Gas connected × Gas knob
- Flame = Gas level × Igniter
- Corn cooked = Flame × Corn on



Deterministic Structural Causal Models

Endogenous variables: variables that are determined by other variables in the model.
Exogenous variables: their values are determined outside of the system.

Context: an assignment of values to the exogenous variables[1].
In an acyclic SEM, a context uniquely determines the values of all the variables in the model.
World (or causal setting): a context in an acyclic SEM

[1] Halpern (2016)

Deterministic Structural Causal Models

Endogenous variables: variables that are determined by other variables in the model.
Exogenous variables: their values are determined outside of the system.

Context: an assignment of values to the exogenous variables[1].
In an acyclic SEM, a context uniquely determines the values of all the variables in the model.
World (or causal setting): a context in an acyclic SEM

[1] Halpern (2016)

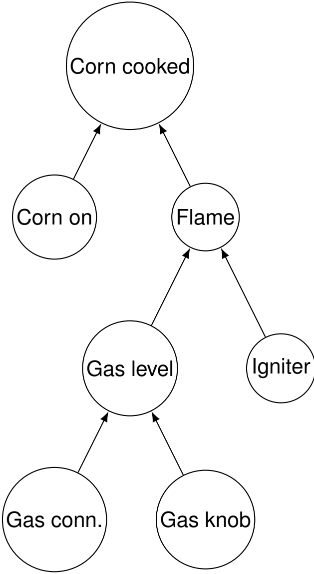
Example (cont'd)

Consider a gas grill, used to cook corn.
Set of exogenous/endogenous variables:

- Gas connected
- Gas knob
- Gas level
- Igniter
- Flame
- Corn on
- Corn cooked

Set of equations:

- Gas level = Gas connected $\times$ Gas knob
- Flame = Gas level $\times$ Igniter
- Corn cooked = Flame $\times$ Corn on



Example (cont'd)

Consider a gas grill, used to cook corn.
If we add the **context**

- Gas connected = 1
- Gas knob = 2
- Igniter = 1
- Corn on = 1

to our three equations,

- Gas level = Gas connected $\times$ Gas knob
- Flame = Gas level $\times$ Igniter
- Corn cooked = Flame $\times$ Corn on

we get Gas level = 2, Flame = 2, and Corn cooked = 2.

Deterministic Structural Causal Models

Independent mechanism principle
The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.

Intervention: set a value for a specified variable by a process that overrides the usual causal structures - without interfering with the causal processes governing the other variables.

To represent an intervention on a variable, we replace the equation for that variable with a new equation stating the value to which the variable is set.

Deterministic Structural Causal Models

Independent mechanism principle
The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.

Intervention: set a value for a specified variable by a process that overrides the usual causal structures - without interfering with the causal processes governing the other variables.

To represent an intervention on a variable, we replace the equation for that variable with a new equation stating the value to which the variable is set.

Deterministic Structural Causal Models

Independent mechanism principle
The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.

Intervention: set a value for a specified variable by a process that overrides the usual causal structures - without interfering with the causal processes governing the other variables.

To represent an intervention on a variable, we replace the equation for that variable with a new equation stating the value to which the variable is set.

Example (cont'd)

If we intervene to set the level of flame at medium, we would represent this by replacing the equation

$$\text{Flame} = \text{Gas level} \times \text{Igniter}$$

with

$$\text{Flame} = 1.$$

One could pour kerosene into the grill and light it with a match.

New causal structure: Flame becomes an exogenous variable.

Example (cont'd)

If we intervene to set the level of flame at medium, we would represent this by replacing the equation

$$\text{Flame} = \text{Gas level} \times \text{Igniter}$$

with

$$\text{Flame} = 1.$$

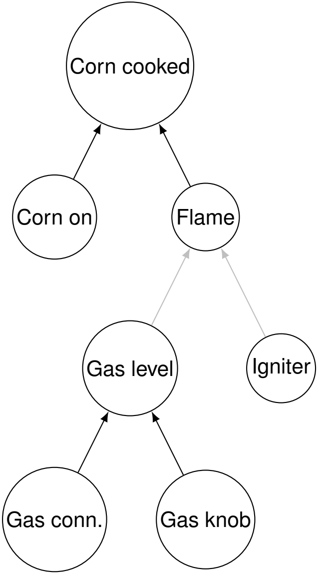
One could pour kerosene into the grill and light it with a match.

New causal structure: Flame becomes an exogenous variable.

Example (cont'd)

Graphically, we can think of the intervention as "breaking the arrows" pointing into Flame.

The new system of equations can then be solved to discover what values the other variables would take as a result of the intervention.



Example (cont'd)

In the world described above, our intervention would produce the following set of equations:

- Gas connected = 1
- Gas knob = 2
- Igniter = 1
- Corn on = 1
- Gas level = Gas connected × Gas knob
- ~~Flame = Gas level × Igniter~~
- Flame = 1
- Corn cooked = Flame × Corn on

Example (cont'd)

The result is a new world with a modified causal structure, with

- Gas level = 2,
- Flame = 1,
- Corn cooked = 1.

Since the equation connecting Flame to its causes is removed, any changes introduced by setting Flame to 1 will only propagate forward through the model to the descendants of Flame.

Example (cont'd)

The result is a new world with a modified causal structure, with

- Gas level = 2,
- Flame = 1,
- Corn cooked = 1.

Since the equation connecting Flame to its causes is removed, any changes introduced by setting Flame to 1 will only propagate forward through the model to the descendants of Flame.

Actual Causation

SCM can help to determine the actual causes from the potential causes.

If the corn is not cooked, is it because the corn has not been put on the grill or because the flame is off?
→ we can directly check the value of the variables.

This is not possible when considering only the graph.

Probabilistic Structural Causal Model

$V = \{X_1, X_2, \ldots, X_n\}$ set of endogenous variables
$U = \{U_1, U_2, \ldots, U_n\}$ corresponding set of exogenous variables.

Suppose that each endogenous variable X_i is a function of its parents in V together with U_i:

$$X_i = f_i(PA(X_i), U_i).$$

Graphical representation is including only the endogenous variables V, and we use $PA(X_i)$ to denote the set of endogenous parents of X_i.

Probabilistic Structural Causal Model

$V = \{X_1, X_2, \ldots, X_n\}$ set of endogenous variables
$U = \{U_1, U_2, \ldots, U_n\}$ corresponding set of exogenous variables.

Suppose that each endogenous variable X_i is a function of its parents in V together with U_i:

$$X_i = f_i(PA(X_i), U_i).$$

Graphical representation is including only the endogenous variables V, and we use $PA(X_i)$ to denote the set of endogenous parents of X_i.

Probabilistic Structural Causal Model

$V = \{X_1, X_2, \ldots, X_n\}$ set of endogenous variables
$U = \{U_1, U_2, \ldots, U_n\}$ corresponding set of exogenous variables.

Suppose that each endogenous variable X_i is a function of its parents in V together with U_i:

$$X_i = f_i(PA(X_i), U_i).$$

Graphical representation is including only the endogenous variables V, and we use $PA(X_i)$ to denote the set of endogenous parents of X_i.

Probabilistic Structural Causal Model

U_i is sometimes called an error variable for X_i : it is responsible for any difference between the actual value of X_i and the value predicted on the basis of $PA(X_i)$ alone. We may think of U_i as encapsulating all of the causes of X_i that are not included in V.

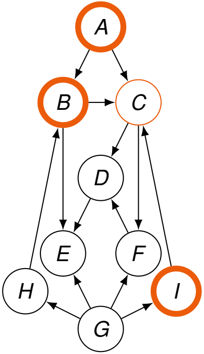
Remark The assumption that each endogenous variable has exactly one error variable is innocuous. If necessary, U_i can be a vector of variables. Moreover, the error variables need not be distinct or independent from one another.

Probabilistic Structural Causal Model

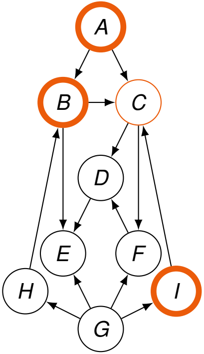
U_i is sometimes called an error variable for X_i : it is responsible for any difference between the actual value of X_i and the value predicted on the basis of $PA(X_i)$ alone. We may think of U_i as encapsulating all of the causes of X_i that are not included in V.

Remark The assumption that each endogenous variable has exactly one error variable is innocuous. If necessary, U_i can be a vector of variables. Moreover, the error variables need not be distinct or independent from one another.

Example

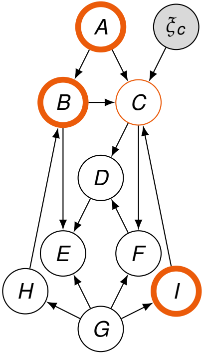


Example



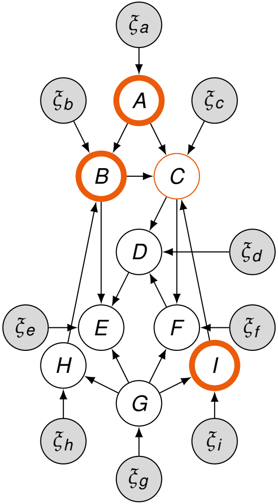
$$C = f(A, B, I)$$

Example



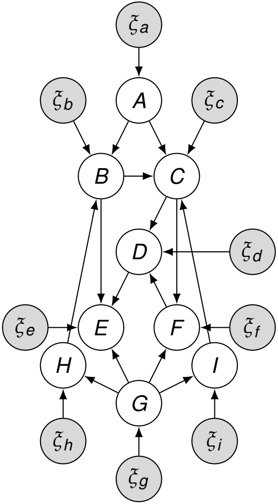
$$C := f_c(A, B, I, \xi_c)$$

Example

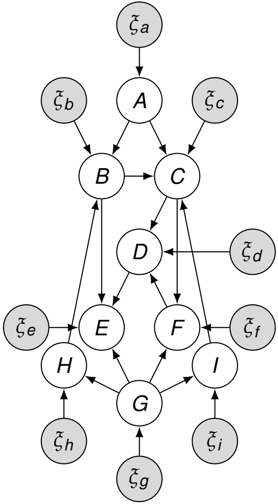


$$C := f_c(A, B, I, \xi_c)$$

Example (cont'd)

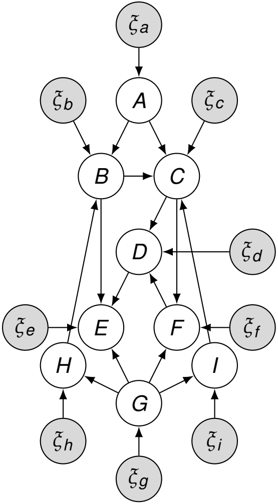


Example (cont'd)



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

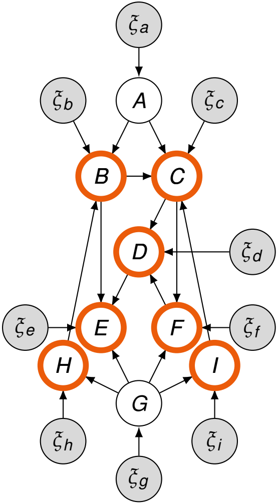
Example (cont'd)



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:

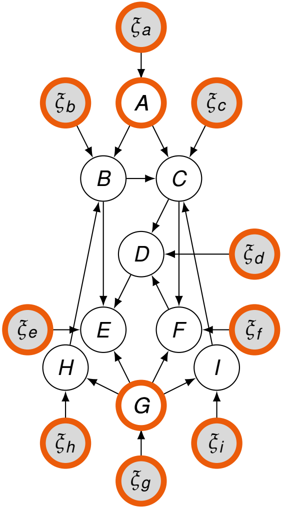
Example (cont'd)



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:
Endogenous variables

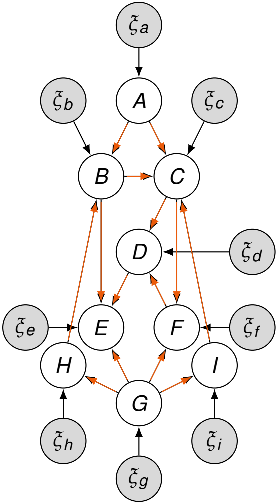
Example (cont'd)



$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

A structural causal model (SCM) is a tuple that contains:
Exogenous variables

Example (cont'd)



$$
M : \begin{cases}
A := f_a(\xi_a) \\
B := f_b(A, H, \xi_b) \\
C := f_c(A, B, I, \xi_c) \\
D := f_d(C, F, \xi_d) \\
E := f_e(B, G, \xi_e) \\
F := f_f(C, G, \xi_f) \\
G := f_g(\xi_g) \\
H := f_h(G, \xi_h) \\
I := f_i(G, \xi_i)
\end{cases}
$$

A structural causal model (SCM) is a tuple that contains:
Causal mechanisms for generating each endogenous variable

Probabilistic Structural Causal Model

Independent Mechanism Principle
In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other conditional distributions.

- Independence of noises, conditional independence of structures
- Independence of information contained in mechanisms
- Intervenability, autonomy, modularity, invariance, transfer

If the system of equations is acyclic, an assignment of values to the exogenous variables $U_1, U_2, \ldots, U_n$ uniquely determines the values of all the variables in the model. Then, if we have a probability distribution P' over the values of variables in U, this will induce a unique probability distribution P on V.

Probabilistic Structural Causal Model

Independent Mechanism Principle
In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other conditional distributions.

- Independence of noises, conditional independence of structures
- Independence of information contained in mechanisms
- Intervenability, autonomy, modularity, invariance, transfer

If the system of equations is acyclic, an assignment of values to the exogenous variables $U_1, U_2, \ldots, U_n$ uniquely determines the values of all the variables in the model. Then, if we have a probability distribution P' over the values of variables in U, this will induce a unique probability distribution P on V.

Probabilistic Structural Causal Model

Modularity assumption:
If we intervene on a subset $S \subset \mathbf{V}$, then $\forall B \in \mathbf{V}$, we have the following:

- If $B \notin S$, then $\Pr(B \mid Pa(B))$ does not change
- If $B \in S$, then $\Pr(B = b \mid Pa(B)) = 1$ if b if the value of B fixed by the intervention; else, $\Pr(B = b \mid Pa(B)) = 0$

If b is the value of B fixed by the intervention, then we say that the value b is *consistant* with the intervention.

Probabilistic Structural Causal Model

Modularity assumption:
If we intervene on a subset $S \subset \mathbf{V}$, then $\forall B \in \mathbf{V}$, we have the following:

- If $B \notin S$, then $\Pr(B \mid Pa(B))$ does not change
- If $B \in S$, then $\Pr(B = b \mid Pa(B)) = 1$ if b if the value of B fixed by the intervention; else, $\Pr(B = b \mid Pa(B)) = 0$

If b is the value of B fixed by the intervention, then we say that the value b is *consistant* with the intervention.

Example (cont'd)

SCM

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

Example (cont'd)

SCM

$$M : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := f_c(A, B, I, \xi_c) \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

Interventional SCM

$$M_c : \begin{cases} A := f_a(\xi_a) \\ B := f_b(A, H, \xi_b) \\ C := c \\ D := f_d(C, F, \xi_d) \\ E := f_e(B, G, \xi_e) \\ F := f_f(C, G, \xi_f) \\ G := f_g(\xi_g) \\ H := f_h(G, \xi_h) \\ I := f_i(G, \xi_i) \end{cases}$$

Probabilistic Structural Causal Model

Violation of the modularity assumption
If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example

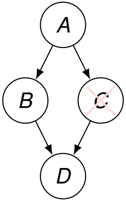
$\Pr(C \mid A)$ changes
$\Pr(B \mid A)$ changes

In other words, without the modularity assumption, the intervention is not necessarily local.

Violation of the modularity assumption

If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example



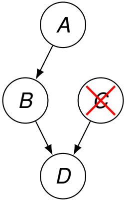
$\Pr(C \mid A)$ changes
$\Pr(B \mid A)$ changes

In other words, without the modularity assumption, the intervention is not necessarily local.

Violation of the modularity assumption

If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example



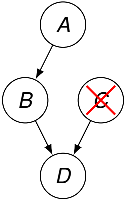
$\Pr(C \mid A)$ changes
$\Pr(B \mid A)$ changes

In other words, without the modularity assumption, the intervention is not necessarily local.

Violation of the modularity assumption

If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example



$\Pr(C \mid A)$ changes

$\Pr(B \mid A)$ changes

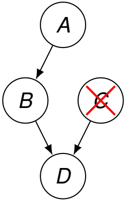
In other words, without the modularity assumption, the intervention is not necessarily local.

Probabilistic Structural Causal Model

Violation of the modularity assumption

If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example



$\Pr(C \mid A)$ changes
$\Pr(B \mid A)$ changes

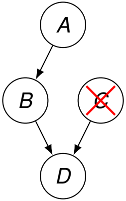
In other words, without the modularity assumption, the intervention is not necessarily local.

Probabilistic Structural Causal Model

Violation of the modularity assumption

If the modularity assumption is not satisfied the the intervention on $S \subset \mathbf{V}$ can change $\Pr(B \mid Pa(B))$ even if $B \notin S$

Example



$\Pr(C \mid A)$ changes
$\Pr(B \mid A)$ changes

In other words, without the modularity assumption, the intervention is not necessarily local.

Probabilistic Structural Causal Model

Reminder: bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

Probabilistic Structural Causal Model

Reminder: bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

$$\Pr(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_i \Pr(\mathbf{V}_i = \mathbf{v}_i \mid Pa(\mathbf{V}_i))$$

Probabilistic Structural Causal Model

Reminder: bayesian network factorization

$$\Pr(\mathbf{V}_1, \cdots, \mathbf{V}_d) = \prod_i \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

$$\Pr(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_i \Pr(\mathbf{V}_i = \mathbf{v}_i \mid Pa(\mathbf{V}_i))$$

Truncated factorization

If we intervene on a subset $S \subset \mathbf{V}$, then

$$\Pr_{\{S=s\}}(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = \prod_{i \notin S} \Pr(\mathbf{V}_i \mid Pa(\mathbf{V}_i))$$

if $\mathbf{v}_1, \cdots, \mathbf{v}_d$ are values consistant with the intervention, else,

$$\Pr_{\{S=s\}}(\mathbf{V}_1 = \mathbf{v}_1, \cdots, \mathbf{V}_d = \mathbf{v}_d) = 0$$

Highlight: differences between regression and causality

- **Need of graph:** in regression, there is no use of graph: the output might be the cause or the effect.
  easier to deal with, but less powerful

- **The noise / the latent variables:**
  In regression, the noise is here to measure the distance between the model and the data.
  In causality, it encodes latent phenomena.

→ the regression is an important tool for causality (causal inference, causal reasoning, ...) but causality goes beyond!

Highlight: differences between regression and causality

- **Need of graph:** in regression, there is no use of graph: the output might be the cause or the effect.
  easier to deal with, but less powerful

- **The noise / the latent variables:**
  In regression, the noise is here to measure the distance between the model and the data.
  In causality, it encodes latent phenomena.

$\rightarrow$ the regression is an important tool for causality (causal inference, causal reasoning, . . .) but causality goes beyond!

Highlight: differences between regression and causality

- **Need of graph:** in regression, there is no use of graph: the output might be the cause or the effect.
  easier to deal with, but less powerful

- **The noise / the latent variables:**
  In regression, the noise is here to measure the distance between the model and the data.
  In causality, it encodes latent phenomena.

$\rightarrow$ the regression is an important tool for causality (causal inference, causal reasoning, ...) but causality goes beyond!

Highlight: differences between regression and causality

- **Need of graph:** in regression, there is no use of graph: the output might be the cause or the effect.
  easier to deal with, but less powerful

- **The noise / the latent variables:**
  In regression, the noise is here to measure the distance between the model and the data.
  In causality, it encodes latent phenomena.

$\rightarrow$ the regression is an important tool for causality (causal inference, causal reasoning, . . .) but causality goes beyond!

Quantification - estimation

Once the model is set (the graph and the set of equations), we can try to estimate the parameters from the data.

The considered graph can be considered as a variable selection method: we know independence / conditional independence between variables, that can reduce the set of features.
If one would not have the graph, a multivariate regression problem would have been considered.

If the model is assumed to be linear, we juste have to estimate the linear coefficients.
The causes can be quantified by those linear coefficients.

Quantification - estimation

One may want to determine if a model is valide (but very difficult to find a suitable model among a collection)
Covariance matrix of the population, compared with

- Covariance matrix of exogenous variables
- Covariance matrix of endogenous variables
- Cross covariance matrix

χ^2 test (among others)