Chapter 1

Structural Causal Models

1.1 Signatures, Augmented Graphs, and Latent Projections

When we start defining causal models, we divide our model of a system into two types of variables: variables that are "in" the system, which call endogenous variables, and variables that are "outside" the system, which we call exogenous variables. We typically think of the endogenous variables as variables whose values we can observe, and the exogenous variables as variables whose values we cannot observe.

Definition 1.1. *A* **signature** $\mathcal{S}$ *consists of:*

- *A set* $\mathcal{X}$ *of* **endogenous variables**,
- *A set* $\mathcal{E}$ *of* **exogenous variables**, *and*
- *A* **range function** $\mathcal{R}$ *mapping each variable* $V \in \mathcal{X} \cup \mathcal{E}$ *to its alphabet* $\mathcal{R}(V)$.

For convenience, we will overload the notation on the range function to take a set as input, i.e., for a set $\mathbf{V} \subseteq \mathcal{X} \cup \mathcal{E}$, we define

$$\mathcal{R}(\mathbf{V}) := \times_{V \in \mathbf{S}} \mathcal{R}(V)$$

Example 1.1 (Genetics in Mice: Signature)**.** *As a running example, we will consider a simplified model of genetic inheritance of weight among a family of mice. We will consider a signature with* $\mathcal{X} = \{X_i\}_{i=1}^5$ *and* $\mathcal{E} = \{\varepsilon_j\}_{j=1}^5$.

- *Let* X_1 *be the presence or absence of a genetic modification. This is a binary, so* $\mathcal{R}(X_1) = \{0, 1\}$.
- *Let* X_2 *and* X_3 *represent the weights, in grams, of two mice, Mickey and Minnie.*
- *Let* X_4 *represent the weight of Mickey and Minnie's offspring, Cheddar.*
- *Let* X_5 *represent the weight of Cheddar's offspring, Gouda.*

These are all continuous variables, so $\mathcal{R}(X_2) = \mathcal{R}(X_3) = \mathcal{R}(X_4) = \mathcal{R}(X_5) = \mathbb{R}_+$.

For the exogenous variables, ε_1 *represents factors determining whether a gene is modified (e.g., a random number generator,* $\mathcal{R}(\varepsilon_1) = [0, 1]$*).* ε_2 *represents factors determining the weight of the male mouse, e.g. the weights of the mouse's parents. Similarly,* ε_3 *represents factors determining the weight of the female mouse, and so on for* $\varepsilon_4, \varepsilon_5$.

Often, one considers variables whose alphabets are real-valued (i.e., $\mathcal{R}(V) = \mathbb{R}$), or categorical (i.e., $\mathcal{R}(V) = \{0, 1, \ldots, K\}$). However, one can also use the formalism we will introduce for more general alphabets, e.g.

any special reason we need this superflous latent notation?

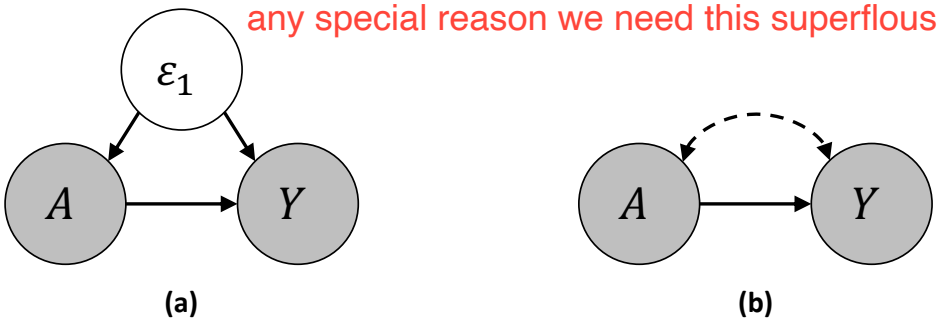

Figure 1.1: **(a)** An augmented graph $\tilde{\mathcal{G}}$, and **(b)** its latent projection $\tilde{\mathcal{G}}(\mathcal{X})$.

vector-valued random variables ($\mathcal{R}(V) = \mathbb{R}^d$) or function-valued random variables (e.g., $\mathcal{R}(V) = L^2(\mathbb{R})$, the set of square-integrable real-valued functions).

One of the main principles behind the theory of structural causal models is to use a graph for summarizing how the variables in a signature relate to one another. The nodes of these graphs are in a one-to-one correspondence with the variables $\mathcal{X} \cup \mathcal{E}$. We begin with a *directed acyclic graph* (DAG), i.e., a directed graph with no cycles. In particular, a directed edge $V_i \to V_j$ in a DAG will have the interpretation that the variable V_i has a direct causal influence on the node V_j.

In graph theory, DAGs have a standard notation that we will use here: given a DAG $\mathcal{G}$ and a node V_i in $\mathcal{G}$, we denote $\text{pa}_{\mathcal{G}}(V_i)$ to denote its **parents**, and $\text{ch}_{\mathcal{G}}(V_i)$ to denote its **children**. We use $\text{an}_{\mathcal{G}}(V_i)$ to denote the **ancestors** of V_i and $\text{de}_{\mathcal{G}}(V_i)$ to denote its **descendants**. We will add a bar to indicate an "inclusive" version of these sets, e.g. $\overline{\text{pa}}_{\mathcal{G}}(V_i) = \text{pa}_{\mathcal{G}}(V_i) \cup \{V_i\}$.

Now, we can relate a DAG to a signature as follows:

Definition 1.2. *We say a DAG $\tilde{\mathcal{G}}$ over the nodes $\mathcal{X} \cup \mathcal{E}$ is* **compatible** *with a signature $\mathcal{S}$ if $\text{pa}_{\tilde{\mathcal{G}}}(\varepsilon_i) = \varnothing$ for all $\varepsilon_i \in \mathcal{E}$. We call the tuple $(\tilde{\mathcal{G}}, \mathcal{S})$ a* **template**.

*tilde means we include exog-vars

It is often convenient to use graphs which only include the endogenous variables. However, to correctly capture the dependencies induced by the exogenous, we must introduce a new edge type to our graphs.

Definition 1.3. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template. The* **latent projection** *of $\tilde{\mathcal{G}}$, denoted $\tilde{\mathcal{G}}(\mathcal{X})$, is an* **acyclic directed mixed graph** *(ADMG) with nodes $\mathcal{X}$ and*

- *A directed edge $X_i \to X_j$ if $X_i \to X_j$ in $\tilde{\mathcal{G}}$, and*
- *A bidirected edge $X_i \leftrightarrow X_j$ if there exists some ε_k such that $\varepsilon_k \to X_i$ and $\varepsilon_k \to X_j$*

Example 1.2. *Let $\mathcal{X} = \{A, Y\}$, where $A \in \{0, 1\}$ is a binary variable indicating whether a treatment is given to a patient, and $Y \in \mathbb{R}$ measures their average blood pressure one month after treatment. Let $\mathcal{E} = \{\varepsilon_1\}$, and let $\tilde{\mathcal{G}}$ be the augmented graph in Figure 1.1(a). Then the latent projection of $\tilde{\mathcal{G}}$ is the graph in Figure 1.1(b).*

As an important special case, we consider graphs where this latent projection operation does *not* introduce any bidirected edges.

Definition 1.4. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template. We call $\tilde{\mathcal{G}}$* **Markovian** *if $|\text{ch}_{\tilde{\mathcal{G}}}(\varepsilon_i)| \leq 1$ for all $\varepsilon_i \in \mathcal{E}$.*

* this means no exo has two children.

It is clear from Definition 1.3 that the latent projection $\tilde{\mathcal{G}}(\mathcal{X})$ of a Markovian graph $\tilde{\mathcal{G}}$ is also a DAG.

Example 1.3 (Genetics in Mice: Augmented Graph and Latent Projection)**.** *In our hypothetical experiment, the presence or absence of the genetic modification is picked prior to measuring the weight of any mice,*

- *The only edge to X_1 is from exogenous noise ε_1, which captures for example the randomness in a computer-generated number.*

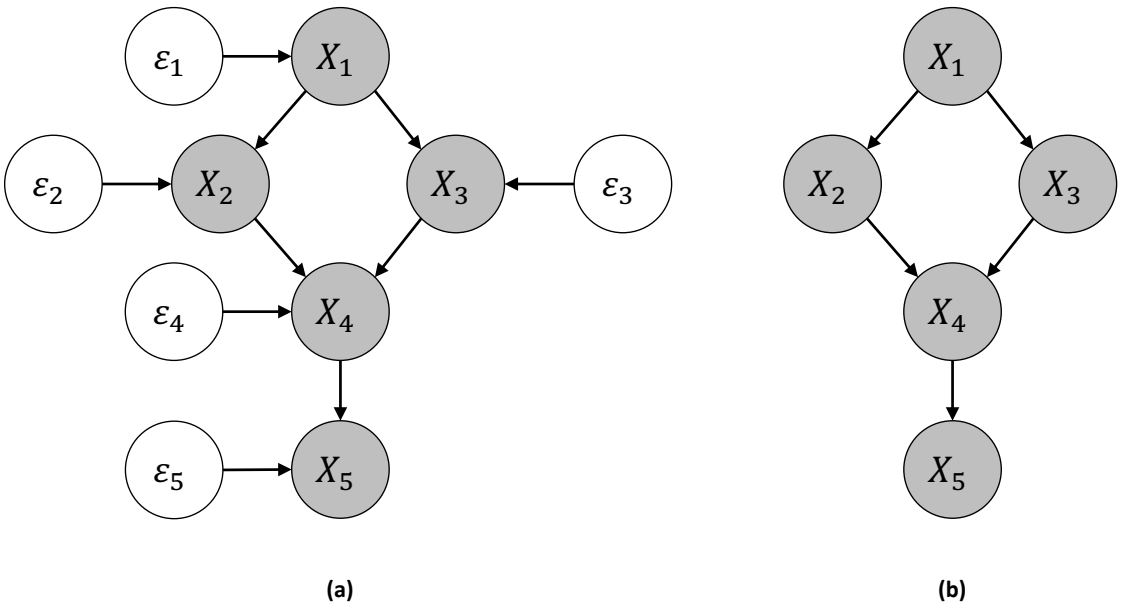

Figure 1.2: **(a)** The augmented graph $\tilde{\mathcal{G}}$ from Example 1.3. **(b)** The latent projection of $\tilde{\mathcal{G}}$.

- *The genetic modification is performed on Mickey, so there is an edge $X_1 \to X_2$. Mickey's weight also depends on some random factors of variation in his DNA, so there is an edge $\varepsilon_2 \to X_2$.*
- *Similarly, the genetic modification is performed on Minnie, so there is an edge $X_1 \to X_3$ and an edge $\varepsilon_3 \to X_3$.*
- *The weight of Cheddar depends on Mickey and Minnie's weights, so there is are edges $X_2 \to X_4$ and $X_3 \to X_4$, as well as an edge $\varepsilon_4 \to X_4$.*
- *Finally, the weight of the Gouda depends on the weight of Cheddar, so there is an edge $X_4 \to X_5$, as well as an edge $\varepsilon_5 \to X_5$.*

The augmented graph $\tilde{\mathcal{G}}$ is shown in Figure 1.2(a). Its latent projection is shown in Figure 1.2(b).

1.2 Causal Mechanisms and Structural Causal Models

tuple of G_tilde and Signature collection of vars)

The definition of compatibility captures the idea that the variables $\varepsilon_i \in \mathcal{E}$ are exogenous, i.e., they are not causally influenced by any other variables. To obtain a structural causal model, we add two elements to a template: causal mechanisms, which say *how* the endogenous variables depend on their parents, and a distribution over the exogenous variables, which introduces stochasticity into the model. First, we define the notion of a causal mechanism:

Definition 1.5. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template. A* **causal mechanism** *for $X_i \in \mathcal{X}$ is a function*

$$f_{X_i} : \mathcal{R}(\text{pa}_{\tilde{\mathcal{G}}}(X_i)) \to \mathcal{R}(X_i).$$

We let $\text{mech}_{\tilde{\mathcal{G}},\mathcal{S}}(X_i)$ denote the set of all such functions.

This leads to the fundamental formal definition:

Definition 1.6. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template. A* **structural causal model** *(SCM) M consistent with this template consists of:*

- *A product distribution $\mathbb{P}_{\mathcal{E}}$ over the exogenous variables, called the* **exogenous distribution**, *and*
- *An indexed set $\{f_{X_i}\}_{X_i \in \mathcal{X}}$ of causal mechanisms, with $f_{X_i} \in \text{mech}_{\tilde{\mathcal{G}},\mathcal{S}}(X_i)$*

We call the latent projection $\tilde{\mathcal{G}}(\mathcal{X})$ the **causal graph** *of M.*

This basic definition is the foundation for a rich field of study. We will begin by thinking about what a structural causal model implies about the distribution over the endogenous variables.

Claim 1.1. *Given a structural causal model, there exists a unique distribution $\mathbb{P}_{\mathcal{X}}(\mathcal{X})$ which is compatible with $\mathbb{P}_{\mathcal{E}}$ and the causal mechanisms. We call this distribution the* **entailed distribution** *over $\mathcal{X}$.*

Example 1.4. *Let $\mathcal{S}$ be the signature from Example 1.1. Let $\tilde{\mathcal{G}}$ be the DAG in Figure 1.2(a). Assume that these variables are related via the following set of assignments:*

$$\begin{aligned} X_1 &= \mathbb{1}_{\varepsilon_1 \leq 0.5} & \varepsilon_1 &\sim \mathsf{Unif}([0,1]) \\ X_2 &= \varepsilon_2 + 2X_1 & \varepsilon_2 &\sim \mathcal{N}(25,1) \\ X_3 &= \varepsilon_3 + 2X_1 & \varepsilon_2 &\sim \mathcal{N}(20,1) \\ X_4 &= {}^1\!/_2\,(X_2 + X_3) + \varepsilon_4 & \varepsilon_4 &\sim \mathcal{N}(0,1) \\ X_5 &= X_4 + \varepsilon_5 & \varepsilon_5 &\sim \mathcal{N}(0,2) \end{aligned}$$

where the distributions of $\varepsilon_1, \ldots, \varepsilon_5$ are mutually independent. The entailed distribution over $\mathcal{X}$ is

$$\begin{aligned} \mathbb{P}_{\mathcal{X}}(\mathcal{X}) = \mathsf{Ber}(X_1; 0.5) &\times \mathcal{N}(X_2; 25 + 2X_1, 1) \times \mathcal{N}(X_3; 20 + 2X_1, 1) \\ &\times \mathcal{N}(X_4; {}^1\!/_2(X_2 + X_3), 1) \times \mathcal{N}(X_5; X_4, 2) \end{aligned}$$

1.3 Interventions

remember (SCM) M consists of Exogs and CauMechs

Now, we will define the notion of an intervention on a causal model, which changes some SCM M into a related SCM M^I.

Definition 1.7. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template. An* **intervention** *I (also called a* **mechanism change***) which is consistent with this template consists of:*

- *A set $\mathcal{X}(I) \subseteq \mathcal{X}$ of* **intervened variables** *(also called* **intervention targets***), and*
- *An indexed set $\{g_{X_i}\}_{X_i \in \mathcal{X}(I)}$ of* **interventional causal mechanisms**, *with $g_{X_i} \in \text{mech}_{\tilde{\mathcal{G}},\mathcal{S}}(X_i)$.*

Definition 1.8. *Given an SCM M with causal mechanisms $\{f_{X_i}\}_{X_i \in \mathcal{X}}$ and an intervention I, both consistent with the template $(\tilde{\mathcal{G}}, \mathcal{S})$, the* **interventional SCM**, *denoted M^I, is a SCM with:*

- *The same exogenous distribution as M, and*
- *Causal mechanisms $f^I_{X_i} = g_{X_i}$ if $X_i \in \mathcal{X}(I)$, and $f^I_{X_i} = f_{X_i}$ otherwise.*

why does intervention change the causal mechnism..?

only change the causal mechanism of intervened

We call the entailed distribution of M^I the **interventional distribution**, *and denote it $\mathbb{P}^I_{\mathcal{X}}$.*

Remark 1.1. *Note that the causal mechanisms for non-intervened variables are* **invariant**, *e.g., they do not change between M and M^I. This is one of the defining properties of causal models: invariance of causal mechanisms unless they are explicitly postulated to change. An important consequence of invariance is* **consistency**: *if a variable X_i is not intervened, then we have*

$$\mathbb{P}^I_{\mathcal{X}}(X_i \mid \text{pa}_{\mathcal{G}}(X_i)) = \mathbb{P}_{\mathcal{X}}(X_i \mid \text{pa}_{\mathcal{G}}(X_i))$$

fr

This definition of an intervention is one of the most general definitions considered in the literature, and is often called a **soft** or **imperfect** intervention to distinguish it from more restricted classes of interventions. Two important restricted classes are "hard" (aka "perfect") interventions and "do" interventions.

Definition 1.9. *A* **hard** *or* **perfect** *intervention is one where each of the intervened variables does not depend on its endogenous parents, i.e., g_{X_i} is a function only of $\mathcal{E}$.*

A **do***-intervention is a perfect intervention where g_{X_i} is a constant. In the case of a do-intervention which sets the variables $\mathcal{A}$ to the values $\mathbf{a}$, we denote the interventional distribution as $\mathbb{P}_{\mathcal{X}}(\mathcal{X} \mid \texttt{do}(\mathbf{A} = \mathbf{a}))$.*

Example 1.5. *Let M be the SCM from Example 1.4.*

(a) *Let I be an intervention with $\mathcal{X}(I) = \{X_1\}$ and $g_{X_1}(\varepsilon_1) = 1$, i.e., the genetic modification is* always *made. Since g_{X_1} is constant, this is a do-intervention. Then the interventional distribution is*

$$\mathbb{P}^I_{\mathcal{X}}(\mathcal{X}) = \delta_1(X_1) \times \mathcal{N}(X_2; 27, 1) \times \mathcal{N}(X_3; 22, 1) \times \mathcal{N}(X_4; 1/2(X_2 + X_3), 1) \times \mathcal{N}(X_5; X_4, 2)$$

isn't 25 not 27? because epsilon_1 = 0

isn't 20 not 22? because epsilon_1 = 0

(b) *If instead we let $g_{X_1}(\varepsilon_1) = \mathbb{1}_{\varepsilon_1 \leq 0.9}$, this would be a perfect intervention.*

(c) *Suppose we switch Mickey to a breed of mice which is naturally heavier and for which the genetic modification has a larger effect, so $g_{X_2}(X_1, \varepsilon_2) = \varepsilon_2 + 4X_1 + 5$. This is a soft intervention.*

It is natural to treat an intervention as a transformation of a structural causal model, as done in Definition 1.7. However, from a technical standpoint, it is often convenient to treat an intervention as an expansion of a causal model into a new model which encapsulates both the original SCM and the intervened SCM, as we now describe. First, we introduce an expansion on the signature:

Definition 1.10. *Let $(\mathcal{S}, \tilde{\mathcal{G}})$ be a template, and let I be an intervention consistent with this template. The* **interventional signature** *$\mathcal{S}^I$ has:*

- *Endogenous variables $\mathcal{X}^I := \mathcal{X} \cup \{\zeta^I\}$, for a new endogenous variable ζ^I called the* **intervention indicator**,
- *Exogenous variables $\mathcal{E}^I := \mathcal{E} \cup \{\varepsilon^I\}$, for a new exogenous variable $\mathcal{E}^I$, and*
- *Range function $\mathcal{R}^I$, which matches $\mathcal{R}$ with the additional criterion $\mathcal{R}^I(\zeta^I) = \mathcal{R}^I(\varepsilon^I) = \{0, 1\}$*

Remark 1.2. *For the sake of intuition, we may consider adding only the endogenous variable ζ^I and treating it as a special type of variables, which indexes different SCMs. The exogenous variable ε^I shows up only out of technical necessity, so that we can condition on different values of ζ^I.*

Next, we introduce an expansion on the augmented graph:

Definition 1.11. *Let $(\tilde{\mathcal{G}}, \mathcal{S})$ be a template and let I be an intervention consistent with this template. Then the* **interventional augmented graph***, denoted $\tilde{\mathcal{G}}^I$, is a DAG with nodes $\mathcal{X}^I \cup \mathcal{E}^I$, and the following edges:*

- *For $V_i, V_j \in \mathcal{X} \cup \mathcal{E}$, let $V_i \to V_j$ if and only if $V_i \to V_j$ in $\tilde{\mathcal{G}}$*
- *$\zeta^I \to X_i$ for all $X_i \in \mathcal{X}(I)$, and*
- *$\varepsilon^I \to \zeta^I$*

From these pieces, we may define a new structural causal model.

Definition 1.12. *Let M be a structural causal model with template $(\mathcal{S}, \tilde{\mathcal{G}})$, exogenous distribution $\mathbb{P}_{\mathcal{E}}$, and causal mechanisms $\{f_{X_i}\}_{X_i \in \mathcal{X}}$. Let I be an intervention consistent with the template and with interventional mechanisms $\{g_{X_i}\}_{X_i \in \mathcal{X}(I)}$. Let $\mathbb{P}_{\varepsilon^I}$ be a Bernoulli distribution over whether or not the intervention is performed.*

The **expanded interventional SCM** *is a structural causal model M^I_+, where M^I_+ has*

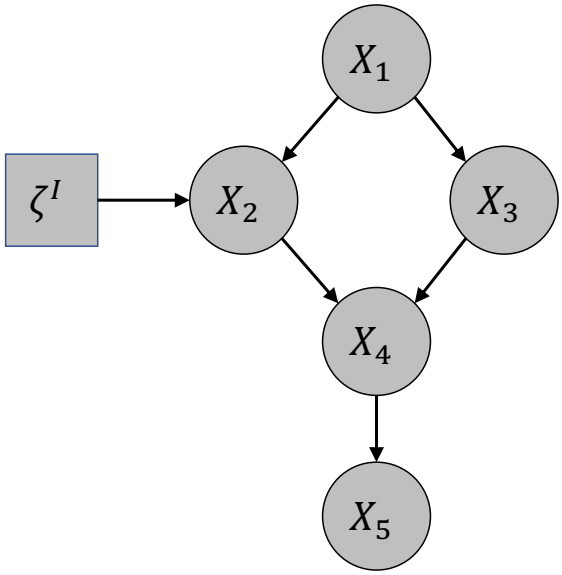

Figure 1.3: The interventional graph $\mathcal{G}^I$ from Example 1.6.

- *Signature* $\mathcal{S}^I$
- *Augmented graph equal to the interventional augmented graph* $\tilde{\mathcal{G}}^I$,
- *Exogenous distribution* $\mathbb{P}_{\mathcal{E}^I} = \mathbb{P}_{\mathcal{E}} \times \mathbb{P}_{\varepsilon^I}$ *and*
- *Causal mechanisms*
  - *For* $X_i \in \mathcal{X}$, *we have*

$$f'_{X_i}(\mathrm{pa}^a_{\mathcal{G}}(X_i), \zeta) = \mathbb{1}_{\zeta^I=0} f_{X_i}(\mathrm{pa}_{\tilde{\mathcal{G}}}(X_i)) + \mathbb{1}_{\zeta^I=1} g_{X_i}(\mathrm{pa}_{\tilde{\mathcal{G}}}(X_i))$$

  - *For* ζ^I, *we have* $f'_{\zeta^I} = \varepsilon^I$

The causal graph of M^I_+ *is called the* **interventional causal graph**. *If* $\mathcal{G}$ *is the causal graph of* M, *the interventional causal graph is denoted* $\mathcal{G}^I$.

Example 1.6. *Let* M *be the SCM in Example 1.4 and* I *be an intervention such that* $\mathcal{X}(I) = \{X_2\}$. *The interventional graph* $\mathcal{G}^I$ *is in Figure 1.3.*

The next claim formalizes the fact that M^I_+ encapsulates the interventional SCM:

Claim 1.2. *Let* $\mathbb{P}^I_{\mathcal{X}}$ *be the entailed distribution of* M^I, *and let* $\mathbb{P}_{\mathcal{X}^I}$ *be the entailed distribution of* M^I_+. *Then*

$$\mathbb{P}^I_{\mathcal{X}}(\mathcal{X}) = \mathbb{P}_{\mathcal{X}^I}(\mathcal{X} \mid \zeta^I = 1)$$

1.4 Counterfactuals

The formalism of structural causal models allows us to address *counterfactual* questions. Such questions are concerned with reasoning about how things would have turned out differently, if a different action had been taken, with "all else being equal". This depends on taking into account the *factual* outcome: the outcome that really did happen.

For example, in healthcare, we might want to answer:

" My patient took aspirin, and their headache went away. Would it have gone away if they had not taken aspirin? "

In epidemiology, we might want to answer:

" My state started COVID quarantines in March 2020. How many deaths would there have been if they had started quarantines in February? "

Formally, we define counterfactuals in terms conditioning the exogenous variables on the factual outcome, as follows.

Definition 1.13. *Let M be a structural causal model consistent with template $(\tilde{\mathcal{G}}, \mathcal{S})$. Let $\mathbf{S} \subseteq \mathcal{X}$ be a subset of endogenous variables and $\mathbf{s} \in \mathcal{R}(\mathbf{S})$ be a realization of $\mathcal{S}$.*

The **counterfactual SCM**, *denoted $M_{\mathbf{S}=\mathbf{s}}$, is a SCM with:*

- *The exogenous distribution $\mathbb{P}_{\mathcal{E}|\mathbf{S}=\mathbf{s}}$*
- *The same causal mechanisms as M*

Example 1.7. *Let M be the structural causal model from Example 1.4. Let $\mathbf{S} = \mathcal{X}$ with $x_1 = 0$, $x_2 = 26$, $x_3 = 20$, $x_4 = 23$, and $x_5 = 23$. Then*

$$\mathbb{P}_{\mathcal{E}|\mathbf{S}=\mathbf{s}}(\mathcal{E}) = \mathsf{Unif}(\varepsilon_1; (0.5, 1]) \times \delta_1(\varepsilon_2) \times \delta_0(\varepsilon_3) \times \delta_0(\varepsilon_4) \times \delta_0(\varepsilon_5)$$

We can use $M_{\mathbf{S}=\mathbf{s}}$ to ask about counterfactual statements, e.g. "what if we had set $X_1 = 1$?". This questions corresponds to asking about the distribution $\mathbb{P}^I_{\mathbf{S}=\mathbf{s}}$ for $\mathcal{X}(I) = \{X_1\}$ and $g_{X_1} = 1$, and we find that

$$\mathbb{P}^I_{\mathbf{S}=\mathbf{s}}(X_1, \ldots, X_5) = \delta_1(X_1) \times \delta_{28}(X_2) \times \delta_{22}(X_3) \times \delta_{25}(X_4) \times \delta_{25}(X_5)$$

1.5 Additional Reading

- **Cyclic models**: In this course, we will only consider causal models without cycles in their augmented graph. This definition can be extended to allow for cycles, see for example Bongers et al. (2021).
- **Margins of DAG models**: The latent projection is a "lossy" operation: there might be constraints on $\mathbb{P}_{\mathcal{X}}$ that are implied by the augmented graph $\tilde{\mathcal{G}}$, which cannot be read off of $\tilde{\mathcal{G}}(\mathcal{X})$. Evans (2016) shows how to define graphical models over $\mathcal{X}$ which capture of constraints implied by the augmented graph.
- **Counterfactuals**: We will not use counterfactuals again in this course. There is a wide set of questions that one can ask in terms of counterfactuals, and indeed counterfactuals are a primary object in the "potential outcomes" approach to causality. The graphical modeling and potential outcomes approaches to causality can be unified via *Single World Intervention Graphs* (SWIGs), introduced by Richardson and Robins (2013).

Bibliography

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915.

Evans, R. J. (2016). Graphs for margins of bayesian networks. *Scandinavian Journal of Statistics*, 43(3):625–648.

Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013.