Section 4

Lecture 2

Structural equation model

Definition

A structural equation model (SEM) is a model that describes how values are assigned to each variable in a system

Think about nature (or whoever) assigning values to each variable in the system. This describes a generative story of how the data came to be as follows.

Or think about each equation above represents a physical mechanism that determines the value of the variable on the left (output) from values of the variable on the right (inputs).

We motivate structural equation models (SEMs) with an example

Consider

$$
\begin{aligned}
L &= f_L(U_L) \\
A &= f_A(L, U_A) \\
Y &= f_Y(A, L, U_Y) = Y^{a=A, l=L} = \sum_{a,l} I(A = a, L = l) Y^{a,l}.
\end{aligned} \tag{1}
$$

Here U_L, U_A, U_Y are external unmeasured factors that are mutually independent. A generative story goes as follows:

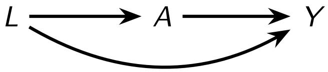
- The value of L is determined as a function of the value of U_L as given by the function f_L.
- The value of A is determined as a function of the value of L, U_A as given by the function f_A.
- The value of Y is determined as a function of the value of L, A, U_Y as given by the function f_Y.

We motivate structural equation models with an example

Consider the SEM $\mathcal{M}$

$$
\begin{aligned}
L &= f_L(U_L) \\
A &= f_A(L, U_A) \\
Y &= f_Y(A, L, U_Y)
\end{aligned}
\tag{2}
$$

and the graph $\mathcal{G}$,



How does $\mathcal{M}$ induce a distribution over the observable law $P(L = l, A = a, Y = y)$ and can this distribution be fully described in some way by simply looking at the graph $\mathcal{G}$?
And how about the distributions under interventions on A, that is, $P(L = l, A = a, Y^a = y)$?

We accompanied the structural equations with a picture

Structural equation models are typically accompanied with a corresponding picture known as a path diagram (formally, this is a Directed Acyclic Graph): that is, a graph which makes explicit the directionality of the underlying process.

For a more compact representation, unmeasured factors that do not determine two or more variables in the system can be left out of the diagram (graph).

We will later see that these graphs can be interpreted as causal DAGs.

Non-parametric structural equation model (NSPEM)

There exist unknown functions $f_1, \ldots, f_m$ such that the observed ("factual") variables $V_1, \ldots, V_m$ satisfy

$$
\begin{aligned}
V_1 &= f_1(U_1) \\
V_2 &= f_2(PA_2, U_2) \\
V_3 &= f_3(PA_3, U_3) \\
&\vdots \\
V_m &= f_m(PA_m, U_m)
\end{aligned} \tag{3}
$$

where:

- $f_0, f_1, \ldots$ are unknown deterministic functions.
- PA_i is the set of random variables that are direct causes ("parents") of V_i.
- $U_0, U_1, \ldots$ are random variables ("disturbances" or "error terms") (not drawn in the graph). Sometimes called exogenous variables.

NPSEM continues

For any treatment regime $g = (g_{j_1}, \dots, g_{j_t})$, the counterfactual variables under g are generated by replacing the functions $(f_{j_1}, \dots, f_{j_t})$ with the functions $(g_{j_1}, \dots, g_{j_t})$, where $t \leq m$. Below is an illustration. This is called performing recursive substitution.

$$\begin{aligned}
V_1^g &= f_1(U_1) \\
V_2^g &= f_2(PA_2^g, U_2) \\
&\vdots \\
V_{j_1}^{g+} &= g_{j_1}(PA_{j_1}^g, U_{j_1}) \\
&\vdots \\
V_m^g &= f_m(PA_m^g, U_m)
\end{aligned} \tag{4}$$

The superscript "g" indicates that V_i^g is a counterfactual variable (in other words, potential outcome variable). The superscript "$g+$" is given to the variables on which we intervene. A NPSEM requires (3) and (4) to hold.

Some remarks

- Structural: f_k not only generates observed (factual variables), but also variables in other counterfactual worlds where we have done interventions.
- Counterfactual: The variable $V_j^g, j \in \{1, \dots, m\}$ are called counterfactual variables under treatment regime g.
- A cause: A variable A is a cause of a variable Y if a change in A can lead to a change in Y.

Example: Point intervention

Let the regime g be defined by the intervention that sets V_2 to a.

$$
\begin{aligned}
V_1^a &= f_1(U_1) \\
V_1^{a+} &= a \\
V_3^a &= f_3(PA_3^a, U_3) \\
&\vdots \\
V_m^a &= f_m(PA_m^a, U_m)
\end{aligned}
\tag{5}
$$

The superscript "a" indicates that V_i^a is a counterfactual variable (or potential outcome variable).

Now comes the causal inference part!

We must say something about the dependencies between the U's to encode causal relations.

Definition (Independent error model)

A NPSEM wrt. a DAG $\mathcal{G}$ such that $U_0, \ldots, U_M$ are mutually independent.

`This is Pearl's NPSEM-IE`[8]`.`
`"IE" stands for independent errors.`
`NB: The independent error assumption is not really needed, and can be relaxed in the more general FFRCISTG model`[9]

The U_ks represents all other variables that are used by nature, the decision maker or anyone else to determine the value of V_k.

---

[8] Judea Pearl. *Causality: Models, Reasoning and Inference 2nd Edition*. Cambridge University Press, 2000.

[9] James M Robins. "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". In: *Mathematical modelling* 7.9-12 (1986), pp. 1393–1512.

Section 5

Causal graphs

Graphs

Some things you need to know about graphs

- Graphs encode conditional independendcies
- Graphs allow us to *represent* and *organize* assumptions and prior knowledge.
- Graphs make the assumptions transparent and explicit.

What is the role of causal graphs?

- Graphs help us to reason about independencies; that is, they help us reason about whether certain exchangeability assumptions (conditional independencies) hold.
- This agrees with the mantra: "draw your assumptions before your conclusions."[10]
- Graphs help us to conceptualize problems and have intuitive appeal, also for researchers who are illiterate in math.
- However, the intuitive graphical representations have a **mathematical justification**. Therefore you – as mathematicians – can translate the intuitive subject-matter expertise (from doctors, economists, social scientists) to precise mathematical statements.
- Graphs allow us to encode *causation* and *association*.

[10] Miguel A Hernan and James M Robins. *Causal inference: What if?* CRC Boca Raton, FL:, 2018.

Section 6

DAGs

Basic definitions

In a DAG $\mathcal{G}$ we define the following sets (parents, children, ancestors and descendants):

- $\mathbf{pa}_G(V_i) \equiv \{V_t : V_t \rightarrow V_i \text{ exists in } \mathcal{G}\}$.
- $\mathbf{ch}_G(V_i) \equiv \{V_t : V_i \rightarrow V_t \text{ exists in } \mathcal{G}\}$.
- $\mathbf{an}_G(V_i) \equiv \{V_t : V_t \rightarrow V_a \rightarrow \cdots \rightarrow V_j \rightarrow V_i \text{ exists in } \mathcal{G}\} \cup V_i$.
- $\mathbf{de}_G(V_i) \equiv \{V_t : V_i \rightarrow V_a \rightarrow \cdots \rightarrow V_j \rightarrow V_t \text{ exists in } \mathcal{G}\}$.

Further terminology:

- A path where $V_a \rightarrow V_i \leftarrow V_b$ is called a collider path, and here V_i is a collider.
- A path where $V_a \leftarrow V_i \rightarrow V_b$ is called a fork.
- A path is *blocked* if it contains a collider. Otherwise it is *open*.
- A DAG is complete if there is an arrow between every pair of nodes.
  A complete DAG imposes no restrictions on the observed data distributions.

Topological order with respect to a graph

Definition (Topological order of a DAG)

The nodes $V_1, V_2, \dots$ follow a topological order relative to a DAG $\mathcal{G}$, if V_j is not ancestor of V_i whenever $j > i$.

Note that topological orders are not necessarily unique, but in the DAG in Figure 65 the only possible topological order is $\langle L, A, Y \rangle$.

Some preliminaries (for your reference)

- Consider a study population Ω.
- Let ω be an element (i.e. unit or individual) in Ω.
- Note that we used subscript i to denote an individual in the first lecture, but now the subscript just indicates a particular random variable, and we write $V_i(\omega)$ when we consider the value for individual ω.
- Consider a discrete random variable V_j.
- Let $V_j(\omega)$ be the value of V_j in ω.
- Let $\mathcal{G}$ be a DAG with nodes $V = \{V_1, V_2, \cdots, V_m\}$.
- We use overlines to denote histories of variables, e.g. $\overline{v}_j \equiv (v_1, v_2, \ldots, v_j) \in \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_m$.
- Let $PA_k = \{V_j : V_j \in \mathbf{pa}_G(V_k)\ \}$. A random vector
- Let $pa_k = \{v_j : V_j \in \mathbf{pa}_G(V_k)\ \}$ for a $\overline{v} \equiv (v_1, v_2, \ldots, v_m) \in \mathcal{V}_1 \times \mathcal{V}_2 \times \cdots \times \mathcal{V}_m$ A realisation of PA_k.
- From now on I will use $p(v_i \mid v_j)$ to denote conditional densities $P(V_i = v_i \mid V_j = v_j)$.

What is a graph?

Definition (Graph)

A graph $\mathcal{G}$ is a collection of

- Nodes (vertices), $V = \{V_1, V_2, \cdots, V_m\}$.
- Edges $(V_i V_j)$ connecting some of the vertices.

We write $(V_i V_j)$ to denote an edge that connects V_i and V_j.
A **path** is a sequence of edges of the form
$\langle (V_1, V_2), (V_2, V_3), \cdots, (V_{k-1}, V_k) \rangle$,

What is a directed graph?

Definition (Directed Graph)

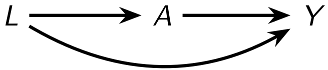
A directed graph is a graph with a set of nodes and *arrows* connecting some of the nodes. A graph $\mathcal{G}$ is a collection of

- Nodes (vertices) $V = \{V_1, V_2, \cdots, V_k\}$.
- **Directed** edges connecting some of the nodes.

We write $(V_i V_j)_{\rightarrow}$ to denote a directed edge from V_i to V_j.
It is directed, because the graphs A **directed path** is a sequence of edges of the form
$\langle (V_1, V_2)_{\rightarrow}, (V_2, V_3)_{\rightarrow}, \cdots, (V_{k-1}, V_k)_{\rightarrow} \rangle$,
A directed graph has a **cycle** if there exists a path
$\langle (V_1, V_2)_{\rightarrow}, (V_2, V_3)_{\rightarrow}, \cdots, (V_{k-1}, V_k)_{\rightarrow}, (V_k, V_1)_{\rightarrow} \rangle$.
A Directed Acyclic Graph is a directed graph with no cycles.
PS: Now the subscript does not longer indicate an individual. V_1 is now a random variable. From now on, I will use $V_1(\omega)$ when I talk about the value for a particular individual.

Example

We can now define the graph below as a causal DAG that describes the conditional randomised trial,



where $V_1 = L, V_2 = A, V_3 = Y$.
Here $\mathbf{pa}_G(Y) = (L, A)$.
The graph is complete because there is an arrow between every pair of nodes.

What is a model

Definition (Statistical model)

A statistical model $\mathcal{P}$ is a collection of laws, $\mathcal{P} = \{P_\eta : \eta \in \Gamma\}$.

Here Γ could be an infinite dimensional space. We will typically only restrict ourselves to the space of models with finite variance.

Bayesian network

Definition (Bayesian network)

A Bayesian Network with respect to a DAG $\mathcal{G}$ with nodes $V = (V_1, \ldots, V_m)$ is a statistical model for the random vector V specifying that V belongs to the collection of laws $\mathcal{B}$ satisfying the Markovian factorisation

$$p(v) = \prod_{j=1}^{m} p(v_j \mid pa_j)$$

Here, $p(x \mid y) \equiv P(X = x \mid Y = y)$.
We say that the DAG $\mathcal{G}$ represents the Bayesian Network $\mathcal{B}$.
For any law p in $\mathcal{B}$, we say that p factors according to $\mathcal{G}$, or that p is represented by $\mathcal{B}$.

Causal DAG

Definition (Robins EPI 207)

A causal model associated with a DAG satisfies:

1. The lack of an arrow from node V_i to V_j can be interpreted as the absence of a direct causal effect of V_i on V_j (relative to the other variables on the graph).
2. Any variable is a cause of all its descendants. Equivalently, any variable is caused by all its ancestors.
3. All common causes, even if unmeasured, of any pair of variables on the graph, are themselves on the graph.
4. The Causal Markov Assumption (CMA): The causal DAG is a statistical DAG, i.e., the distribution of V factors.
5. Because of the causal meaning of parents and descendants on a causal DAG, the Causal Markov Assumption is equivalent to the statement:
   - Conditional on its direct causes (i.e., parents), a variable V_i is independent of any variable it does not cause (i.e., any nondescendant).