Outline

1. Causal DAGs and intervention
2. Linear structural equation models
3. Estimation of causal effect
4. Semi-Markov causal models
5. Potential outcome approach
Causal DAGs and intervention

(Reference: Pearl (2000) §3.1 and §3.2; Pearl (1995))

Definition: A causal model among \( X_1, \ldots, X_p \) is defined by a DAG \( G \) and a distribution \( P(\epsilon) = P(\epsilon_1, \ldots, \epsilon_p) \).

- Each child-parent relationship in \( G \), \((X_j, PA_j)\), represents a functional relationship (structural equation model, SEM):
  \[
  X_j = f_j(PA_j, \epsilon_j), \quad j = 1, \ldots, p.
  \] (1)

- The background (error) variables are jointly independent:
  \[
  P(\epsilon_1, \ldots, \epsilon_p) = \prod_j P(\epsilon_j). \] (2)

- (18) and (2) imply that \( P(X_1, \ldots, X_p) \) is Markovian with respect to the DAG \( G \):
  \[
  P(X_1, \ldots, X_p) = \prod_{j=1}^{p} P(X_j | PA_j). \] (3)
Causal DAGs and intervention

Causal effect defined via external intervention:

- Consider an atomic intervention that forces $X_i$ to some fixed value $x_i$, which we denote by $do(X_i = x_i)$ or $do(x_i)$ for short.
- Effect of $do(x_i)$: to replace the SEM for $X_i$ by $X_i = x_i$ and substitute $X_i = x_i$ in the other SEMs.
- For two distinct sets of variables $X$ and $Y$, the causal effect of $X$ on $Y$ is determined by the mapping

$$x \mapsto \mathbb{P}[Y \mid do(X = x)] \equiv \mathbb{P}(Y \mid do(x)).$$

Examples of causal effects.

1. **linear SEM**: Causal effect $\frac{\partial \mathbb{E}(Y|do(x))}{\partial x}$.
2. **Treatment ($X = 1$) vs control ($X = 0$)**: Causal effect $\mathbb{E}(Y \mid do(X = 1)) - \mathbb{E}(Y \mid do(X = 0))$. 
Causal DAGs and intervention

Model interventions as variables:

- Treat intervention as additional variable in the DAG: \( F_j \) for intervention on \( X_j \).
- SEM for \( X_j \) change to

\[
X_j = h_j(PA_j, F_j, \varepsilon_j) = \begin{cases} 
  f_j(PA_j, \varepsilon_j), & \text{if } F_j = \text{idle} \\
  x, & \text{if } F_j = \text{do}(x). 
\end{cases}
\]

(4)

- Augment the parents of \( X_j \) to \( PA_j \cup \{F_j\} \):

\[
\mathbb{P}(X_j = x_j \mid PA_j, F_j) = \begin{cases} 
  \mathbb{P}(X_j = x_j \mid PA_j), & \text{if } F_j = \text{idle} \\
  I(x_j = x), & \text{if } F_j = \text{do}(x), 
\end{cases}
\]

assuming all \( X_j \) are discrete for convenience.
Computing causal effect (of interventions): To simplify notation, consider discrete \( X_j \) and write \( \mathbb{P}(X = x) = P(x) \).

- **Truncated factorization** of \( P(x_1, \ldots, x_p) \) given \( do(X_i = x_i^*) \):

\[
P(x_1, \ldots, x_p \mid do(x_i^*)) = I(x_i = x_i^*) \prod_{j \neq i} P(x_j \mid pa_j), \tag{5}
\]

where \( pa_j = (x_k : k \in PA_j) \).

- Multiple interventions \( do(X_S = x^*), S \subset \{1, \ldots, p\} \):

\[
P(x_1, \ldots, x_p \mid do(x^*)) = I(x_S = x^*) \prod_{j \notin S} P(x_j \mid pa_j). \tag{6}
\]

- Graph structure change when \( do(X_i = x_i^*) \): delete edges \( X_j \rightarrow X_i \) for all \( j \in PA_i \), i.e. change \( G \) to \( G_{\overline{X_i}} \).
Causal DAGs and intervention

Difference between $P(y \mid do(x))$ and $P(y \mid x)$.

- Two DAGs $G_1$ and $G_2$ on $X_1, X_2$:

  - $G_1$:
    
    $P(x_1 \mid do(x_2)) = P(x_1)$,

  - $G_2$:
    
    $P(x_1 \mid do(x_2)) = P(x_1 \mid x_2)$.

Find $P(x_1 \mid do(x_2))$ with respect to $G_1$ and $G_2$. 
From (5), putting $x_i = x_i^*$:

$$P(x_{-i} \mid do(x_i^*)) = \prod_{j \neq i} P(x_j \mid pa_j) \cdot \frac{P(x_i^* \mid pa_i)}{P(x_i^* \mid pa_i)}$$

$$= \frac{P(x_1, \ldots, x_p)}{P(x_i^* \mid pa_i)}$$

$$= P(x_j, j \in B \mid x_i^*, pa_i) P(pa_i), \quad (7)$$

where $B = [p] \setminus \{i, PA_i\}$ and $[p] := \{1, \ldots, p\}$.

- Intervention event (do-operator) not on the right-hand side.
- Compute causal effect (intervention probability) by conditional probabilities (pre-intervention probabilities) that can be estimated from observational data.
Theorem 1 (Adjustment for direct causes)

Let $PA_i$ be the parents of $X_i$ and $Y$ be any set of other variables in a causal DAG $\mathcal{G}$. Then the causal effect of $do(X_i = x_i)$ on $Y$ is given by

$$P(y \mid do(x_i)) = \sum_{pa_i} P(y \mid x_i, pa_i)P(pa_i),$$

where $P(y \mid x_i, pa_i)$ and $P(pa_i)$ are pre-intervention probabilities.

Proof.

Marginalize out $X_j \notin Y \cup \{X_i\}$ on both sides of (7).
A simple implication of Theorem 1:
If \( Y \) is a set of non-descendants of \( X_i \), then
\[
Y \perp X_i \mid PA_i.
\]

By Theorem 1
\[
P(y \mid do(x_i)) = \sum_{pa_i} P(y \mid x_i, pa_i)P(pa_i)
\]
\[
= \sum_{pa_i} P(y \mid pa_i)P(pa_i) = P(y),
\]
which is independent of the intervention on \( X_i \). Thus, \( X_i \) has no causal effect on \( Y \).
A causal model \((\mathcal{G}, \mathbb{P}_\varepsilon)\) with linear SEMs:

- A linear model for each child-parent relationship:
  \[
  X_j = \sum_{i \in PA_j} \beta_{ij} X_i + \varepsilon_j, \quad j = 1, \ldots, p. \tag{9}
  \]

- \(\varepsilon_j\)'s are independent and \(\mathbb{E}(\varepsilon_j) = 0\);

- Usually assume \(\varepsilon_j \sim \mathcal{N}(0, \omega_j^2)\). In this case, the DAG is called a Gaussian DAG and the graphical model is called a Gaussian Bayesian network.
Linear structural equation models

Causal effect:

- The causal effect of $X_k$ on $X_j$

$$
\gamma_{kj} := \frac{\partial \mathbb{E}(X_j \mid do(X_k = x))}{\partial x} = \mathbb{E}(X_j \mid do(X_k = c + 1)) - \mathbb{E}(X_j \mid do(X_k = c)), \quad (10)
$$

for any $c \in \mathbb{R}$, due to the linear model assumption.

- Using modified DAG $\mathcal{G}_{\tilde{X}_k}$ after intervention,

$$
\mathbb{E}(X_j \mid X_k = x; \mathcal{G}_{\tilde{X}_k}) = \gamma_{kj} x,
$$

where $\mathbb{E}(\bullet; \mathcal{G}_{\tilde{X}_k})$ takes expectation with respect to $\mathcal{G}_{\tilde{X}_k}$. 
Linear structural equation models

Apply Theorem 1 to find $\gamma_{kj}$:

- Let $Z = PA_k$ and $z$ denote the value of $PA_k$,

$$p(x_j \mid do(X_k = x_k)) = \int_Z p(x_j \mid x_k, z)p(z)dz,$$

where the $p$ on the right side is given by the pre-intervention distribution (that of $G$).

- Let $(\beta, \alpha)$ be the regression coefficient of $X_j$ on $(X_k, PA_k)$, that is, $\mathbb{E}(X_j \mid X_k, Z) = \beta X_k + \alpha^T Z$, which can be estimated from observational data.

- Then the causal effect

$$\gamma_{kj} = \frac{\partial}{\partial x_k} \mathbb{E}(X_j \mid do(X_k = x_k))$$

$$= \frac{\partial}{\partial x_k} \int_Z \left\{ \beta x_k + \alpha^T z \right\} p(z)dz = \beta.$$
Estimation of causal effect

Reference: Pearl (2000) §3.3.

Problem setup:

- Given a causal DAG $G$, if $P(y \mid do(x))$ can be uniquely computed from the (pre-intervention) distributions of observed variables in $G$, then we say the causal effect of $X$ on $Y$ is identifiable.

- Note that we allow unobserved nodes in $G$.

- Only observational data are collected.
Example: Observed nodes $X \rightarrow Z \rightarrow Y$; hidden node $U$, a common parent of $X$ and $Y$ (sometimes called a confounder).

Can we estimate the causal effect of $X$ on $Y$ or of $Z$ on $Y$ from observational data collected for $(X, Y, Z)$?
Back-door adjustment:

- Theorem 1 implies: If $X, PA_X, Y$ are observed, then $P(y \mid do(x))$ is identifiable by (8).
- Theorem 1 is a special case of back-door adjustment: $PA_X$ satisfies the back-door criterion relative to $X$ and $Y$.
- *Back-door criterion:* A set of variables $Z$ satisfies the back-door criterion relative to an ordered pair of variables $(X, Y)$ in a DAG $G$ if
  1. no nodes in $Z$ is a descendant of $X$;
  2. $Z$ blocks every chain between $X$ and $Y$ that contains an arrow into $X$ (backdoor path).
Estimation of causal effect

Theorem 2 (Back-door adjustment)

If $Z$ satisfies the back-door criterion relative to $(X, Y)$. Then the causal effect of $X$ on $Y$ is given by

$$P(y \mid do(x)) = \sum_{z} P(y \mid x, z)P(z). \quad (11)$$

Proof.

Add intervention variable $F_X \rightarrow X$ to $G$:

$$P(y \mid do(x)) = \sum_{z} P(y \mid do(x), z)P(z \mid do(x))$$

$$= \sum_{z} P(y \mid F_X = do(x), x, z)P(z).$$

Invoke that $(X, Z)$ d-separates $F_X$ and $Y$. 

Zhou, Q Graphical Models 16/32
Estimation of causal effect

Linear SEM: By (11), the causal effect can be identified by regressing $Y$ on $(X, Z)$:

$$\gamma_{X \rightarrow Y} := \frac{\partial}{\partial x} \mathbb{E}(Y \mid do(x)) = \beta_X(Y \sim X + Z).$$

Suppose we have data observed for the three random variables $X, Y, Z$. Then to estimate the causal effect $X$ on $Y$:

1. Discrete data: estimate $P(y \mid x, z)$ and $P(z)$ from data. Then plug into (11).
2. Linear SEM: least-squares regression $Y$ on $(X, Z)$, then

$$\hat{\gamma}_{X \rightarrow Y} = \hat{\beta}_X(Y \sim X + Z).$$
Estimation of causal effect

Example:

By Theorem 2,

\[ P(y \mid do(z)) = \sum_x P(y \mid x, z)P(x), \quad P(z \mid do(x)) = P(z \mid x), \]

(12)

without observing \( U \).
Estimation of causal effect

Is $P(y \mid do(x))$ identifiable? Yes, because:

$$P(y \mid do(x)) = \sum_z P(y, z \mid do(x))$$

$$= \sum_z P(z \mid do(x))P(y \mid z, do(x))$$

$$= \sum_z P(z \mid do(x))P(y \mid do(z)). \quad (13)$$

Last step uses $Y \perp F_Z \mid \{Z, do(x)\}$:

$$P(y \mid z, do(x)) = P(y \mid do(z), do(x)) = P(y \mid do(z)).$$

Then, plug (12) into (13) to get

$$P(y \mid do(x)) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x'). \quad (14)$$
Eq. (14) is an example of \textit{front-door adjustment}.

- Front-door criterion: \( Z \) satisfies the front-door criterion relative to \((X, Y)\) if
  1. \( Z \) intercepts all directed paths from \( X \) to \( Y \);
  2. there is no back-door path from \( X \) to \( Z \); and
  3. all back-door paths from \( Z \) to \( Y \) are blocked by \( X \).

\textbf{Theorem 3 (Front-door adjustment)}

If \( Z \) satisfies the front-door criterion relative to \((X, Y)\), then

\[ P(y \mid do(x)) = \sum_{z} P(z \mid x) \sum_{x'} P(y \mid x', z) P(x'). \tag{15} \]

- Linear SEMs:

\[ \gamma_{X \rightarrow Y} = \gamma_{X \rightarrow Z} \times \gamma_{Z \rightarrow Y} = \beta_X(Z \sim X) \times \beta_Z(Y \sim Z + X). \]
Proof of Theorem 3.

(i) Condition 1 implies

\[ P(y \mid do(x)) = \sum_z P(z \mid do(x))P(y \mid do(z)). \]

(ii) Backdoor adjustment with Condition 2 shows that

\[ P(z \mid do(x)) = P(z \mid x). \]

(iii) Backdoor adjustment with Condition 3 shows that

\[ P(y \mid do(z)) = \sum_{x'} P(y \mid x', z)P(x'). \]

Rules of do-calculus (Pearl (2000) §3.4): a set of inference rules for transforming intervention and observational probabilities, say to translate causal effect to conditional probabilities.
Estimation of causal effect

Instrumental variable formula (Bowden and Day 1984) (assume linear SEMs)

Observed nodes $Z \rightarrow X \rightarrow Y$, and $U$ is hidden common parent of $X$ and $Y$. Is $\gamma_{X \rightarrow Y} = \alpha_2$ identifiable?
Estimation of causal effect

1. Z has no parents, thus \( \alpha_1 \) is identifiable by regressing \( X \) on \( Z \): \( \alpha_1 = \beta_Z(X \sim Z) \).

2. Similarly, the causal effect of \( Z \) on \( Y \), \( \alpha_1 \alpha_2 \), is also identifiable: \( \alpha_1 \alpha_2 = \beta_Z(Y \sim Z) \).

3. Combined we have the \textit{instrumental variable formula}:

\[
\alpha_2 = \frac{\beta_Z(Y \sim Z)}{\beta_Z(X \sim Z)} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}.
\]
Estimation of causal effect

Two-stage least-squares:

1. Regress $X$ on $Z$ so $\alpha_1 = \beta_Z(X \sim Z)$ and let $\hat{X} = \alpha_1 Z$.
2. Regress $Y$ on $\hat{X}$ and then $\alpha_2 = \beta_{\hat{X}}(Y \sim \hat{X})$:

$$
\beta_{\hat{X}}(Y \sim \hat{X}) = \frac{\text{Cov}(Y, \alpha_1 Z)}{\text{Var}(\alpha_1 Z)} = \frac{\text{Cov}(Y, Z)}{\alpha_1 \text{Var}(Z)} = \alpha_2.
$$

Note: To estimate $\alpha_2$ from samples of $(X, Y, Z)$, $\beta \to \text{LSE} \hat{\beta}$. 
Conditional instrumental variable (Brito and Pearl 2002): 

$Z$ is said to be a *conditional instrumental variable* given $S$ relative to $(X, Y)$ if

1. $S$ contains no descendants of $X$ or $Y$;
2. $S$ $d$-separates $Z$ from $Y$ but not from $X$ in the graph obtained after deleting all edges emerging from $X$.

Then, the causal effect of $X$ on $Y$

$$
\gamma_{X \rightarrow Y} = \frac{\text{Cov}(Y, Z \mid S)}{\text{Cov}(X, Z \mid S)} = \frac{\beta_Z(Y \sim Z + S)}{\beta_Z(X \sim Z + S)}.
$$

(17)
If two nodes $X_i$ and $X_j$ share a common hidden parent $U$, remove $U$ from the DAG and add a bidirected edge $X_i \leftrightarrow X_j$: acyclic directed mixed graph (ADMG).

- $X_i \leftrightarrow X_j$: their background variables $\varepsilon_i$ and $\varepsilon_j$ are dependent.

- A causal model with dependent background variables is called a semi-Markov causal model (SMCM).
Semi-Markov causal models

DAG with hidden variables

\[ U_1 \]

\[ X_1 \rightarrow X_2 \rightarrow X_3 \]

ADMG

\[ X_1 \rightarrow X_2 \rightarrow X_3 \]

- SEM for SMCM over \( X = \{X_1, \ldots, X_p\} \):

\[ X_j = f_j(PA_j, \varepsilon_j), \quad j = 1, \ldots, p. \quad (18) \]

\( \varepsilon_i \perp \varepsilon_j \) if no bidirected edge between \( i \) and \( j \).

- The joint distribution \( \mathbb{P}(X) \) is obtained by marginalization of \( \mathbb{P}(X, U) \) defined by a DAG on \( X \cup U \):

\[
P(x_1, \ldots, x_p) = \sum_{u_1, \ldots, u_d} P(x_1, \ldots, x_p \mid u_1, \ldots, u_d) \prod_i P(u_i).
\]
Semi-Markov causal models

Let \( Y(x) \equiv [Y \mid do(X = x)] \). Restrictions encoded by SMCM:

1. **Exclusion**: For any \( S \subset V \setminus (PA_Y \cup \{Y\}) \) (no directed edge from \( S \) to \( Y \)),

   \[
   Y(pa_Y) = Y(pa_Y, s). \tag{19}
   \]

   \[\because\] both = \( f_Y(pa_Y, \varepsilon_Y) \).

2. **Independence**: For any \( Z \in V \) not connected to \( Y \) via bidirected edges,

   \[
   Y(pa_Y) \perp Z(pa_Z). \tag{20}
   \]

   \[\because\] \( Y(pa_Y) = f_Y(pa_Y, \varepsilon_Y) \), \( Z(pa_Z) = f_Z(pa_Z, \varepsilon_Z) \) and \( \varepsilon_Y \perp \varepsilon_Z \).
Semi-Markov causal models

- Exclusion restrictions: \( Y(x) = Y(x, z) \) and \( X = X(y, z) \).
- Independence restrictions: \( X \perp \{ Y(x), Z(y) \} \), but \( Y(x) \not\perp Z(y) \).
Potential outcome approach

Under potential outcome framework (Rubin 1990):

- $Y(x)$ is a counterfactual entity representing the potential outcome of $Y$ had $X$ been $x$.

- Suppose $X \in \{0, 1\}$ (treatment vs control). Want to estimate causal effect $\mathbb{E}[Y(1) - Y(0)]$ or $\mathbb{E}[Y(1) - Y(0) | X = 1]$.

- $\mathbb{P}^*[Y(x)]$ corresponds to $\mathbb{P}(Y | do(x))$. Making assumptions to calculate when $Y(x)$ is missing.

\[
\begin{array}{cccc}
X & Y(1) & Y(0) & Z \text{ (covariates)} \\
1 & ✓ & ? & ✓,\cdots,✓ \\
1 & ✓ & ? & ✓,\cdots,✓ \\
1 & ✓ & ? & ✓,\cdots,✓ \\
0 & ? & ✓ & ✓,\cdots,✓ \\
0 & ? & ✓ & ✓,\cdots,✓ \\
\end{array}
\]
Assume conditional ignorability (Rosenbaum and Rubin 1983): 
\( Y(x) \perp X \mid Z. \)

\[
P^*(Y(x) = y) = \sum_z P^*(Y(x) = y \mid z)P(z)
\]

\[
= \sum_z P^*(Y(x) = y \mid x, z)P(z)
\]

\[
= \sum_z P(Y = y \mid x, z)P(z) \quad \text{backdoor adjustment.}
\]

\[
Z \quad X \quad Y
\]

\[
Y(x) = f_Y(x, Z, \varepsilon_Y) = h(Z, \varepsilon_Y)
\]

\[
X = f_X(Z, \varepsilon_X)
\]

\[
Y(x) \perp X \mid Z \iff \varepsilon_Y \perp \varepsilon_X \mid Z.
\]


