Reading

10. *Causal inference in statistics : a primer*
Judea Pearl, Madelyn Glymour, Nicholas Jewell.
Pearl, Judea, author
Chichester, West Sussex, UK : John Wiley & Sons Ltd, 2016. / Book / Online

- Primer book - Chapter 3

**Acknowledgement (big thanks!):**
Many slides are by Martina Contisciani that have been modified here.
Recap: d-Separation

If a set of nodes Z blocks every path between two nodes X and Y, then X and Y are d-separated conditioned on Z.

- H and D are d-separated by \{XG\}
- G and X are d-separated by \{H\}
- G and X are NOT d-separated by \{HD\}
Recap: Structural Causal Model & Graphical Causal Model

- \( M = \langle U, V, F \rangle \)
- Endogenous (observable) variables \( V = \{G, X, D, O\} \)
- Exogenous (noise) variables \( U = \{U_G, U_X, U_D, U_O\} \)
- Structural equations \( F \):
  \[
  \begin{align*}
  G &= F_G(U_G), \\
  X &= F_X(U_X, G), \\
  D &= F_D(U_D, G, X), \\
  O &= F_O(U_O, X, D)
  \end{align*}
  \]
  Can be linear, exp, ...

Quantitative

Corresponding Graphical Causal Model

Qualitative
Recap: Structural/Graphical Causal Model to Probabilistic Model

• \(<M, \Pr>\)
• \(M\) is a Structural Causal Model
• \(\Pr\) is the Probability distribution
  • Satisfies Causal Markov Condition
  • Conditional independence in directed graphical models
• \(\Pr(X_1, X_2, \ldots) = \prod_i \Pr(X_i \mid Pa(X_i))\)

If we knew the values of the exogenous variables and the structural equations in \(F\), we exactly know the values of endogenous \(V\)
But not in practice – so assume a probability distribution \(\Pr(U = u)\), which gives a \(\Pr\) distribution on \(V\)
Assumption: \(U\)’s are independent of each other
Correlation ≠ Causation

Recall: Randomized controlled experiment

All factors that influence the outcome variable are either static, or vary at random, except for one.

So, any change in the outcome variable must be due to that one input variable.
Intervention ≠ conditioning

Effect of an intervention on ice cream sales (X) on crime (Y)

\[ Z = \text{weather temperature} \]

**Fix** the value of a variable. **Remove** the tendency to vary according to other variables, by severing all arrows that enter the manipulated \( X \).

**Change the structure** of the graph by removing the incoming edges (**surgery** on the graphical model).

No causal path from \( X \) to \( Y \) here
Intervention ≠ conditioning

Effect of an intervention on ice cream sales (X) on crime (Y)

\[ Z = \text{weather temperature} \]

Change the structure of the graph by removing the incoming edges (surgery on the graphical model).

Change the perception, not the world. Focus to the subset of cases in which the variable takes the value we are interested in.
Intervention ≠ conditioning

Effect of an intervention on ice cream sales (X) on crime (Y)

Z = weather temperature

Intervention P(Y=y|do(X=x))

Conditioning P(Y=y| X=x)

Change the structure of the graph by removing the incoming edges (surgery on the graphical model).

Change the perception, not the world. Focus to the subset of cases in which the variable takes the value we are interested in.
Intervention ≠ conditioning

Question:
Given observational data + valid graph
How to use do-expressions + graph surgery
To get causal information from data

Change the structure of the graph by removing the incoming edges (surgery on the graphical model).

Change the perception, not the world.
Focus to the subset of cases in which the variable takes the value we are interested in.
The Adjustment Formula

How to go from do-expressions to conditional probabilities by graph surgery?
X: Drug Usage (treatment)
Y = Recovery rate (outcome)
Z = Gender (confounder)

**Causal effect difference (Average Causal Effect)**
P(Y=1|do(X=1)) - P(Y=1|do(X=0))

Irrespective of the intervention: (1) Proportion of genders remain the same
(2) Conditional probability of how Y responds to X and Z remain the same
Intervention

\[
P(Y = y | \text{do}(X = x)) = P_m(Y = y | X = x)
\]

\[
= \sum_z P_m(Y = y | X = x, Z = z) P_m(Z = z | X = x)
\]

\[
= \sum_z P_m(Y = y | X = x, Z = z) P_m(Z = z)
\]

\[
= \sum_z P(Y = y | X = x, Z = z) P(Z = z)
\]

(\text{use equalities from previous slides})

Do-expressions reduced to conditional probability expressions – can be estimated from data!

Adjustment formula: we are “adjusting for” Z or “controlling for” Z
Why don’t we need adjustment in RCT?

We start with the above graph and $P = P_m$

\[
P(Y = y|\text{do}(X = x)) = P_m(Y = y|X = x)
= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z|X = x)
= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z)
= \sum_z P(Y = y|X = x, Z = z)P(Z = z) \tag{use equalities from previous slides}
\]
Try yourself: Simpson Paradox

Probability of recovery if people take the drug...

Table 1.1  Results of a study into a new drug, with gender being taken into account

<table>
<thead>
<tr>
<th></th>
<th>Drug</th>
<th>No drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>81 out of 87 recovered (93%)</td>
<td>234 out of 270 recovered (87%)</td>
</tr>
<tr>
<td>Women</td>
<td>192 out of 263 recovered (73%)</td>
<td>55 out of 80 recovered (69%)</td>
</tr>
<tr>
<td>Combined data</td>
<td>273 out of 350 recovered (78%)</td>
<td>289 out of 350 recovered (83%)</td>
</tr>
</tbody>
</table>

And compare answer with the Primer book!
Do we need adjustment here?

\[ Z \]

X: Drug Usage (treatment)  
Y = Recovery rate (outcome)  
Z = Blood pressure measured after the treatment
Do we need adjustment here?

X: Drug Usage (treatment)
Y = Recovery rate (outcome)
Z = Blood pressure measured after the treatment

No arrow entering X
⇒ No surgery needed
⇒ Treatment is as if randomized

Pr (Y = y | do (X = x)) = Pr (Y = y | X = x)

No adjustment needed!
The Causal Effect Rule:

Given a graph $G$ in which a set of variables $PA$ are designed as the parents of $X$, the causal effect of $X$ on $Y$ is given by

$$P(Y = y | do(X = x)) = \sum_z P(Y = y | X = x, PA = z)P(PA = z)$$

where $z$ ranges over all the combinations of values that the variables in $PA$ can take.
We can always condition on Treatment’s parents – great!

But ----- what if the treatments have "unmeasured" parents?

e.g., inaccessible for measurements -- Intellect, quality, efforts, ....

Under what conditions is the structure of the causal graph sufficient for computing a causal effect from a given dataset?

Generalized adjustment formula is given by the backdoor criterion
Backdoor Criterion

Given an ordered pair of variables (X,Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X,Y) if

- no node in Z is a descendant of X, AND
- Z blocks every path between X and Y that contains an arrow into X.
Given an ordered pair of variables \((X,Y)\) in a directed acyclic graph \(G\), a set of variables \(Z\) satisfies the backdoor criterion relative to \((X,Y)\) if

- no node in \(Z\) is a descendant of \(X\), AND
- \(Z\) blocks every path between \(X\) and \(Y\) that contains an arrow into \(X\).

\[
P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)
\]

with \(Z\) satisfying the backdoor criterion between \(X\) and \(Y\)

Note: \(PA(X)\) always satisfies the backdoor criterion.
Backdoor Criterion - Intuition

Given an ordered pair of variables \((X,Y)\) in a directed acyclic graph \(G\), a set of variables \(Z\) satisfies the backdoor criterion relative to \((X,Y)\) if

- no node in \(Z\) is a descendant of \(X\), AND
- \(Z\) blocks every path between \(X\) and \(Y\) that contains an arrow into \(X\).

1. Block all spurious paths between \(X\) and \(Y\).
2. We leave all directed paths from \(X\) to \(Y\) unperturbed.
3. We create no spurious paths.
Backdoor Criterion – Intuition #1

Given an ordered pair of variables \((X,Y)\) in a directed acyclic graph \(G\), a set of variables \(Z\) satisfies the backdoor criterion relative to \((X,Y)\) if

- no node in \(Z\) is a descendant of \(X\), AND
- \(Z\) blocks every path between \(X\) and \(Y\) that contains an arrow into \(X\).

1. Block all spurious paths between \(X\) and \(Y\).
2. We leave all directed paths from \(X\) to \(Y\) unperturbed.
3. We create no spurious paths.

To estimate causal effect from \(X\) to \(Y\), block all “backdoor paths” that have an arrow into \(X\): These paths make \(X\) and \(Y\) dependent but do not transmit causal influence from \(X\) to \(Y\)
Backdoor Criterion – Intuition #2

Given an ordered pair of variables (X,Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X,Y) if

• no node in Z is a descendant of X, AND
• Z blocks every path between X and Y that contains an arrow into X.

1. Block all spurious paths between X and Y.
2. We leave all directed paths from X to Y unperturbed.
3. We create no spurious paths.

Descendants of X are affected by intervention and may affect Y: We would block those paths if we condition on them.
Given an ordered pair of variables \((X,Y)\) in a directed acyclic graph \(G\), a set of variables \(Z\) satisfies the backdoor criterion relative to \((X,Y)\) if

- no node in \(Z\) is a descendant of \(X\), AND
- \(Z\) blocks every path between \(X\) and \(Y\) that contains an arrow into \(X\).

1. Block all spurious paths between \(X\) and \(Y\).
2. We leave all directed paths from \(X\) to \(Y\) unperturbed.
3. We create no spurious paths.

We should not condition on colliders.
Backdoor Criterion – Example 1

Effect of $X$ on $Y$

$X$: Drug

$Y$: Recovery

$Z$: socioeconomic status (unmeasured)

$W$: Body weight
The backdoor path is blocked by $W$, which is no descendant of $X$, and doesn’t create new spurious paths. In addition, $Z$-$W$-$Y$ is a chain structure and conditioning on the middle node will make $Z$ and $Y$ conditionally independent, and the path would be blocked.
Backdoor Criterion – Example 2

Effect of X on Y?
No unblocked backdoor paths from $X$ to $Y$

Effect of $X$ on $Y$?

$$P(Y = y | do(X = x)) = P(Y = y | X = x).$$
Backdoor Criterion – Example 2

What if we condition on W?
Backdoor Criterion – Example 2

We open the spurious backdoor path From $X$ to $Y$ – not correct

What if we condition on $W$?

Effect modification or modetation

What if we want to compute the causal effect for a specific value of $W$?

$$P(Y = y|do(X = x), W = w)$$
Backdoor Criterion – Example 2

What if we want to compute the causal effect for a specific value of W?

\[ P(Y = y | do(X = x), W = w) \]

We adjust for another variable like T to block this path

\[ P(Y = y | do(X = x), W = w) = \sum_t P(Y = y | X = x, W = w, T = t)P(T = t | W = w) \]
Backdoor Criterion – Example 3

To adjust or not to adjust a collider

Which variables to condition on?

Causal effect of X on Y
Backdoor Criterion – Example 3

To adjust or not to adjust a collider

Which variables to condition on?

Need to condition on \{E, Z\}, \{Z, A\}, or \{E, Z, A\} – must include collider Z
Goal:
Reduce causal relationship as “do-operators” to observed conditional probabilities

\[
\Pr(D = \text{yes} \mid \text{do}(Q) = \text{MS})
\]

- Find the right variables to condition on
  - “d-separation” (from graphical models in AI)
  - “Back-door condition”

\[
\sum_g \Pr(D = \text{yes} \mid Q = \text{MS}, G = g) \Pr(G = g)
\]

Can be estimated from data:
says that to understand the causal effect
Of having an MS on PhD admission decision, condition on gender
Backdoor Criterion – Other Benefits

1. Any set that conforms the backdoor criterion must return the same result for $P(Y=y \mid \text{do}(X=x))$. We have a choice! Once set of variables may be less expensive or easier to measure than the others.

2. When all adjustment variables are observed, we get a testable constraint on the data – if the model we are trying to fit does not satisfy these equalities, we can discard the model!
Summary:

• Treatment X, Outcome Y
• Goal is to estimate causal effect \( \Pr(Y = y \mid \text{do}(X = x)) \)

• A set \( Z \) of variables is called admissible covariates for estimating the above causal effect if

\[
\Pr(Y = y \mid \text{do}(X = x)) = \sum_z \Pr(Y = y \mid X = x, Z = z) \Pr(Z = z)
\]

In short

\[
\Pr(y \mid \text{do}(X = x)) = \sum_z \Pr(y \mid x, z) \Pr(z)
\]

We are adjusting for \( Z \) here

• How to find \( Z \)? Use backdoor criterion if you have a graphical causal model and can find such \( Z \) (backdoor is sufficient not necessary)