Causality in statistical inference
Goals (and Agenda)

- Why is causality interesting?
- There is some vocabulary to learn
- We need causality in statistical inference
- And we can infer about causality, too
Introduction, motivation
What an econometric model really is?

- \( Y_i = \alpha_0 + \alpha_1 \cdot X_{1,i} + \alpha_2 \cdot X_{2,i} + ... \)

- What does this equation mean?
  - Simplified data set summary?
  - Equation used to create predictions?
  - Relationship useful in planning results of taken decisions?

- Inspiration – how does it differ in machine learning (e.g., decision trees)?
Why causality?

Democritus: *I would* rather discover one cause than gain the kingdom of Persia

- Knowledge
- Consequences of decisions
  - (e.g. what will happen if minimum wage rises?)
- World controlling methods
  - (e.g. how to improve demographic situation?)

- Questions
  - What influences what
  - Influence strength
Causality – definitions

- Correlation & time sequence
- Time & space processes
- Bradford-Hill criteria (medicine)
- Human agency
- Manipulating
- Counterfactuals
- Invariance
- Probabilistic approach
- Eclecticism
- ... (e.g. Aristotle, Kant, ..., Cartwright, ...)
### Confounding

<table>
<thead>
<tr>
<th>Treatment</th>
<th># patients</th>
<th># live</th>
<th># die</th>
</tr>
</thead>
<tbody>
<tr>
<td>expensive</td>
<td>100</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td>none</td>
<td>100</td>
<td>80</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient status</th>
<th># patients</th>
<th># live</th>
<th># die</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expensive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>90</td>
<td>36</td>
<td>54</td>
</tr>
<tr>
<td>Initial</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>20</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Initial</td>
<td>80</td>
<td>80</td>
<td>0</td>
</tr>
</tbody>
</table>
Applications (University of California, Berkeley)

Are women being discriminated?

<table>
<thead>
<tr>
<th></th>
<th># applications</th>
<th>% hired</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>8442</td>
<td>44%</td>
</tr>
<tr>
<td>female</td>
<td>4321</td>
<td>35%</td>
</tr>
</tbody>
</table>

But...

<table>
<thead>
<tr>
<th>Department</th>
<th># male applications</th>
<th>% hired men</th>
<th># female applications</th>
<th>% hired women</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>825</td>
<td>62%</td>
<td>108</td>
<td>82%</td>
</tr>
<tr>
<td>B</td>
<td>560</td>
<td>63%</td>
<td>25</td>
<td>68%</td>
</tr>
<tr>
<td>C</td>
<td>325</td>
<td>37%</td>
<td>593</td>
<td>34%</td>
</tr>
<tr>
<td>D</td>
<td>417</td>
<td>33%</td>
<td>375</td>
<td>35%</td>
</tr>
<tr>
<td>E</td>
<td>191</td>
<td>28%</td>
<td>393</td>
<td>24%</td>
</tr>
<tr>
<td>F</td>
<td>272</td>
<td>6%</td>
<td>341</td>
<td>7%</td>
</tr>
</tbody>
</table>

Simpson paradox
Simpson Paradox

- Two batting (baseball) averages

<table>
<thead>
<tr>
<th>Batter \ Year</th>
<th>1995</th>
<th>1996</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derek Jeter</td>
<td>12/48</td>
<td>.250</td>
<td>183/582</td>
</tr>
<tr>
<td>David Justice</td>
<td>104/411</td>
<td>.253</td>
<td>45/140</td>
</tr>
</tbody>
</table>

- Justice, despite having better season averages, had worse career one!
Our problem

- We have observational data
- We estimate model
  - frequencies, econometrics, DM, ...
- Yet we want to do interventions!
  - We believe, that causal relationships are stronger

Solutions

- Stratified analysis (e.g. CMH test over $\chi^2$)
- Additional variables in econometric model (DM)
- Models of *propensity score* type
- ...

This lecture – how causal thinking can improve process of analysis development
Graphical representation
Used notation

- We observe three binary variables: X, Y, Z
- Simplifications
  - We ignore estimation uncertainty while using frequency analysis
  - We assume >0 probability of any combination

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Pr(X=x,Y=y,Z=z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5,6%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>22,4%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8,4%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>33,6%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2,4%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9,6%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3,6%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>14,4%</td>
</tr>
</tbody>
</table>
Previous table - shortened

- $P(X=1)=30\%, P(Y=1)=60\%, P(Z=1)=80\%$; independent and conditionally independent variables

For example:

- $X \perp \perp Y$
  - $P(X=1 \mid Y=1) = P(X=1)$, ...

- $X \perp \perp Y, Z$
  - $P(X=1 \mid Y=1, Z=1) = P(X=1)$, ...

- And $X \perp \perp Y \mid Z$
  - $P(X=1 \mid Y=1, Z=1) = P(X=1 \mid Z=1)$, ...

(let’s check if those hold for previous table)

Such reduction may not be possible!
Another table

- Is $X \perp\!\!\!\!\!\!\perp Y$?
- Is $X \perp\!\!\!\!\!\!\perp Z$?
- Or is $X \perp\!\!\!\!\!\!\perp Y \mid Z$?

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>$\Pr(X=x,Y=y,Z=z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>36%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>24%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8%</td>
</tr>
</tbody>
</table>
Chain rule in probability calculus

- \( P(X=x, Y=y, Z=z) = \)
  \[ = P(X=x) \cdot P(Y=y \mid X=x) \cdot P(Z=z \mid X=x, Y=y) \]

- If any independencies (conditional) apply, equation may be simplified

  - E.g we have \( X \perp \perp Y \), hence
  
  \[ P(X=x, Y=y, Z=z) = \]
  \[ = P(X=x) \cdot P(Y=y \mid X=x) \cdot P(Z=z \mid X=x, Y=y) \]
  \[ = P(X=x) \cdot P(Y=y) \cdot P(Z=z \mid X=x, Y=y) \]

- Further reduction is impossible, since below don’t hold:
  
  - \( Z \perp \perp X,Y \)
  - \( Z \perp \perp X \mid Y \)
  - \( Z \perp \perp Y \mid X \)
Final table

- **Data generating process (DGP):**
  - X – cut while shaving (10%)
  - Y – coffee burn during breakfast (20%)
  - Z – bad mood during day
    - 90%, if an accident in the morning;
    - 5% otherwise

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Pr(X=x,Y=y,Z=z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>68.4%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.9%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>17.1%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.4%</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7.6%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.1%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.9%</td>
</tr>
</tbody>
</table>
Chain rule application

- \( P(X, Y, Z) = P(X) \cdot P(Y) \cdot P(Z \mid X,Y) \)
- Graphically equivalent to:

\[
\begin{align*}
X \rightarrow Z & \leftarrow Y
\end{align*}
\]

- Yet we can analyze variables in different order, e.g. \( P(Z, Y, X) \)
- Then we obtain

\[
P(Z, Y, X) = P(Z) \cdot P(Y \mid Z) \cdot P(X \mid Y,Z)
\]

- And graphically

\[
Z \rightarrow Y \rightarrow X
\]
Which image is proper?

- How to present interventions on such graphs?
  - What an intervention really is?
  - Typical interventions?
- Now we assume, that graph structure is known
  - During next lecture we will try to retrieve it from data
Graphical notation (Bayes Networks)

- **Notation**
  - Vertices are variables (some of them latent)
  - Edges are mechanisms (lack of arrow is also important)
  - **Intervention removes arrows coming in**
  - Chain of arrows (not necessarily $\rightarrow \rightarrow \rightarrow$) is path
  - Some paths are not causal (e.g. $\rightarrow \leftarrow \rightarrow \rightarrow$)
  - Some paths transport information (end features dependant)

- **We have**
  - $X$ and $Y$ are dependant
    
    $X \leftarrow$ Third cause $\rightarrow Y$

    - Intervention on $X$ does not affect $Y$

    Intervention $\rightarrow X \leftarrow$ Third cause $\rightarrow Y$
Confounding graphically
Our problems in terms of graphs

- Divorces $\leftarrow$ time $\rightarrow$ Margarine consumption

- Patients state $\rightarrow$ Treatment $\rightarrow$ Result

- Smoking $\leftarrow$ Genes $\rightarrow$ Lung cancer
  - (according to industry, yellow arrow does not exist)

- Gender $\rightarrow$ Department $\rightarrow$ Result
  - (Discrimination is yellow arrow)
  - (How to consider abilities on the graph?)
What we want to be able to do?

- Assuming known graph structure
- Estimate causal impact strength
- (strength of relationships between variables along causal paths →...→)
- Taking under consideration, that dependencies in observational data may result from non-causal paths as well (e.g ←...←, ←......→)
Graphical solution

- During analysis of observational data we have to block non-causal paths (e.g. stratified analysis, etc.)
  - In order to get causal interpretation of analysis results

- In our examples:
  - Zero in regression model
  - Expensive drug \( \rightarrow \) % survival

- Why does randomization in an experiment helps?

![Graphical representation showing Advanced illness leading to Drug and Death, with Drug affecting Death.]

Advanced illness

Drug \rightarrow Death
Attention (Berkson paradox)

- Sometimes introduction of a variable may open path!

Cut while shaving $\longrightarrow$ Bad day $\longleftarrow$ Coffee burned

- Another example

<table>
<thead>
<tr>
<th></th>
<th>handsome</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>intelligent</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>no</td>
<td>25%</td>
<td>25%</td>
</tr>
</tbody>
</table>

$25\% = 50\% \times 50\%$

<table>
<thead>
<tr>
<th></th>
<th>handsome</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>intelligent</td>
<td>33,(3)%</td>
<td>33,(3)%</td>
</tr>
<tr>
<td>no</td>
<td>33,(3)%</td>
<td>(no recall)</td>
</tr>
</tbody>
</table>

$1/3 < 2/3 \times 2/3$

- For proofs and assumptions, see Pearl
Too many controlled variables

- We block one of causal paths
  - (we get direct causal effect)
  - (if exists)

- In terms of correlation this example is identical to the previous one!
  - It is not possible to define confounding in the field of statistics

Characteristics (e.g. cholesterol)

Drugs (e.g. statins)  →  Effect (e.g. CHD)
Demographic features controlling

When YES

Drug → Features → Death

When NO

Drug → Feature → Death
Medical documentation controlling

When YES

When NO
Difficult situations (and smoking)

\[ \text{confounder} \]

\[ X \rightarrow M \rightarrow Y \]
Paths vs instrumental variables

```
IV 🔵 X 🔵 Y
```

- **confounder**
Propensity score method
Appendix
Independence in samples

- We assumed that probability distribution is known
- We have samples and have to deal with random errors
- Tests exist
  - Independence, e.g. $\chi^2$, LIS-test
  - Conditional independence, e.g. CMH, KCI-test
Counterfactuals

- Three levels of inference:
  - observational
    - „patient takes drug A, will he recover?”
  - interventional
    - „what will happen if I will prescribe patient drug A?”
  - counterfactual
    - „patient took drug A and recovered, what would happen if took drug B?”
    - Recovery resulting from A / independently of A / despite A

- Are counterfactuals necessary?
  - Repeatable decisions
  - What would happen without treatment? Question linked to another one regarding % of patients who will die when treated
Counterfactual inference dilemma. Neyman-Rubin model

- Patients have been randomly assigned to treatment ($X=1$) or not ($X=0$)
- Recovery has been observed ($Y=1$) or not ($Y=0$)

<table>
<thead>
<tr>
<th></th>
<th>$Y=0$</th>
<th>$Y=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X=0$</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>$X=1$</td>
<td>30%</td>
<td>45%</td>
</tr>
</tbody>
</table>

- For patient $X=0$, $Y=1$, what would happen if $X=1$?
- Extreme cases

<table>
<thead>
<tr>
<th>$Y$ for $X=0$</th>
<th>$Y$ for $X=1$</th>
<th>% patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>40%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>40%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Y$ for $X=0$</th>
<th>$Y$ for $X=1$</th>
<th>% patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>40%</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>60%</td>
</tr>
</tbody>
</table>
Summary

- Observational data (non-experimental) ≠ Interventional data
  - e.g. forgotten OLS assumption (not random X)!
- Statistics alone is unable to conduct causal inference
  - How to say 'puddle does not cause rain’?
- Black-box approach is not sufficient
  (not always, the more variables the better!)
- Causal inference is important!
References

Please think if below is true?

- When a country's debt rises above 90% of GDP, growth slows.
- Therefore, high debt causes slow growth.
Please think if below is true?

- As ice cream sales increase, the rate of drowning deaths increases sharply.
- Therefore, ice cream consumption causes drowning.
Thank you!