STAT/BIOST 572
Update Student Presentation

Wen Wei Loh

April 26, 2012
Recap

- Transparent Parametrizations of Models for Potential Outcomes
  - Model used to estimate causal effects
  - Imperfect compliance
  - Separate the identifiable parameters from the unidentified
General Potential Outcomes model

- Observed
  - $Z = \text{Assignment to treatment (Instrument)}$
  - $X = \text{Receipt/Exposure to treatment}$
  - $Y = \text{Response}$

- Unobserved (describing potential outcomes)
  - $t_X = \text{Underlying compliance "type"}$
  - $t_Y = \text{Underlying response "type"}$
Compliance Types ($t_X$)

- $Z = $ Assigned treatment
- $X_z = $ Potential receipt/exposure to treatment given assignment $Z = z$
- $X_z = $ observed $X$ under the consistency axiom

$X_{z=0}$ $X_{z=1}$ Compliance Type $t_X$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>NT Never Taker</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>DE Defier</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>CO Complier</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>AT Always Taker</td>
<td></td>
</tr>
</tbody>
</table>

Table: Compliance types ($t_X$) based on potential outcomes
Potential Outcomes

- $Y_{xz} =$ Potential response under receipt/exposure to treatment $x$, and treatment assignment $z$
- In a *general* model, $Z$ may have an effect on $Y$, and $t_Y$ would take $16 = 2^{2^2}$ possible states.
- Under the Instrumental Variable model [Angrist et al., 1996], the effect of $Z$ on $Y$ is only through $X$. 
Response Types ($t_Y$)

- This is represented by the Exclusion Restriction: $Y_{xz} = Y_{xz'} = Y_x$. for $z, z' \in \{0, 1\}$.
- $Y_x$. = Potential response under receipt/exposure to treatment $x$, regardless of assigned treatment

<table>
<thead>
<tr>
<th>$Y_0$.</th>
<th>$Y_1$.</th>
<th>Response Type $t_Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>NR  Never Recover</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>HU  Hurt</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>HE  Helped</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>AR  Always Recover</td>
</tr>
</tbody>
</table>

Table: Response types ($t_Y$) based on potential outcomes under Instrumental Variable model
**Instrumental Variable Model**

- **Exclusion Restriction**
  - Effect of Assignment ($Z$) on Response ($Y$) is only through Receipt/Exposure ($X$)

- Causal estimand of interest: Average Causal Effect (ACE)
  - $ACE(X \rightarrow Y) = P(\text{Helped}) - P(\text{Hurt})$
What’s wrong with this approach?

- Non-identifiability
  - Distribution over potential outcomes $p(t_X, t_Y)$ may only be partially-identified
  - Causal estimands of interest (e.g. ACE) would hence depend on parameters that are not fully-identified.

- How to get meaningful information about causal effects?
  - Pearl [2000] proposed bounds on ACE
  - Find the set of distributions $p(t_X, t_Y)$ compatible with observed data
  - Find the minimum and maximum ACE
Motivating Example

• Data from a double-blind placebo-controlled randomized trial
• Compliance and response were dichotomized
• There are no DEfiers and no Always Takers since \( Z = 0 \Rightarrow X = 0 \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>count</th>
<th></th>
<th></th>
<th></th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>158</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>78</td>
</tr>
</tbody>
</table>

Table: Lipid data; there are two structural zeros
Finding the ACE

- Bounds

\[
\left[ p(y = 1 \mid z = 1) - p(y = 1 \mid z = 0) \right] \\
- \left[ p(y = 1, x = 0 \mid z = 1) + p(y = 0, x = 1 \mid z = 0) \right] \\
\leq \text{ACE}(X \to Y) \leq \\
\left[ p(y = 1 \mid z = 1) - p(y = 1 \mid z = 0) \right] \\
+ \left[ p(y = 0, x = 0 \mid z = 1) + p(y = 1, x = 1 \mid z = 0) \right]
\]

- From the data: (0.39, 0.78)
Bayesian Approach

- Prior distribution over potential outcomes $p(t_X, t_Y)$
  - Reflects proportion of individuals in population that possess characteristics corresponding to $(t_X, t_Y)$
  - Dirichlet prior

- Posterior
  - Gibbs sampling to sample $(t_X, t_Y)$ from resulting posterior
  - Find $ACE(X \rightarrow Y) = p(\text{Helped} \mid \text{data}) - p(\text{Hurt} \mid \text{data})$
Posterior: Sensitive to prior?

- Uniform
  - Dir \((1, \ldots, 1) \rightarrow \text{Dir} (1, \ldots, 1.2, 1, 0.8)\)

- Unit
  - Dir \((\frac{1}{8}, \ldots, \frac{1}{8}) \rightarrow \text{Dir} (1, \ldots, \frac{3}{16}, \frac{1}{8}, \frac{1}{16})\)

- Such perturbation should not have large effect on posterior ACE
Posterior: Sensitive to prior?
Transparent Parametrizations

- Re-parameterize $p(t_X, t_Y)$ into $f(\theta, \zeta)$
  - $\theta =$ identifiable parameter (estimable from observed $(X, Y, Z)$)
  - $\zeta =$ non-identifiable parameter
Figure 4: A graph representing the functional dependencies in the analysis of the simple IV model with no Always Takers or Defiers. Rectangular nodes are observed; oval nodes are unknown parameters. \( p(x = 1|z = 0) = 0 \), so \( p(y|x = 1, z = 0) \) is undefined, hence these nodes are omitted.
Next Steps

- Inference based on posterior distribution
  - Work out the distribution of the observed data $p(y, x \mid z)$ implied by the transparent model
  - Compute the posterior distribution of the parameters
  - Apply inequality restrictions ("truncate") by Monte-Carlo rejection sampling
  - Find ACE ($X \rightarrow Y$) as a function of the unidentified parameter(s)
Next Steps

• Extend to general setting
• Various assumptions or restrictions: 8 different models
• Derive bounds on conditional causal contrasts e.g. Average Controlled Direct Effect
  \[ ACDE_{NT}(x_0) = E\left( Y_{x=0,z=1} - Y_{x=0,z=0} \mid NT \right) \]
• Evaluate bounds conditional Intention-To-Treat effect e.g. \( ITT_{CO} = E\left( Y_{x=1,z=1} - Y_{x=0,z=0} \mid CO \right) \)