STAT/BIOST 572
Update Student Presentation

Wen Wei Loh

April 26, 2012

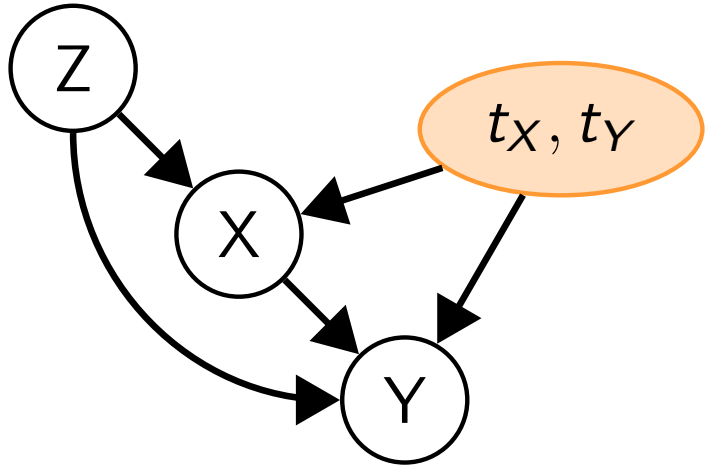
Recap

- 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 = Assignment to treatment (Instrument)
  - X = Receipt/Exposure to treatment
  - Y = Response

- Unobserved (describing *potential outcomes*)
  - t_X = Underlying compliance "type"
  - t_Y = 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	
0	0	NT	Never Taker
1	0	DE	Defier
0	1	CO	Complier
1	1	AT	Always Taker

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: $Y_{xz} = Y_{xz'} = Y_{x\cdot}$ for $z, z' \in \{0, 1\}$.
- $Y_{x\cdot}$ = Potential response under receipt/exposure to treatment x, regardless of assigned treatment

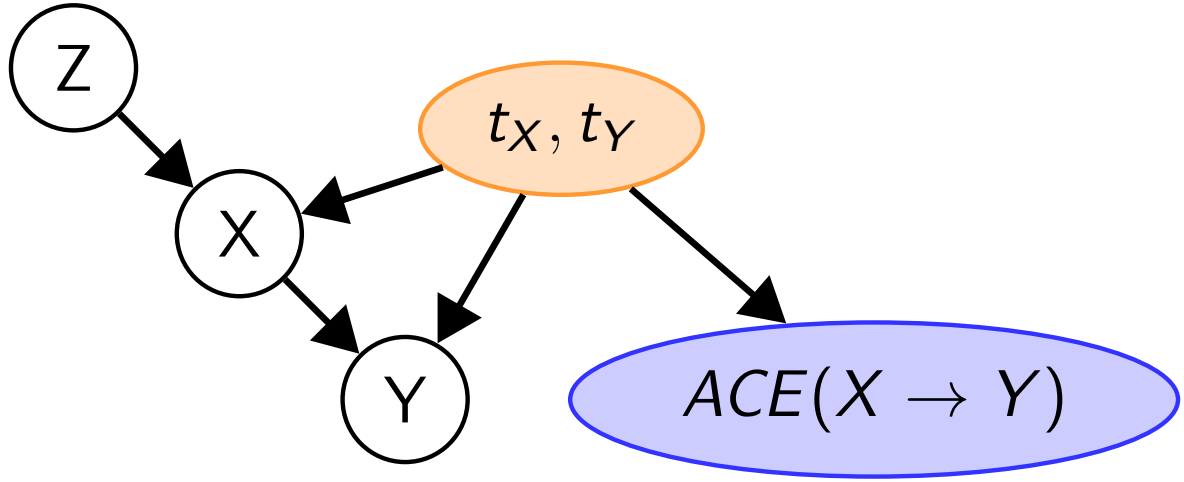
$Y_{0\cdot}$	$Y_{1\cdot}$	Response Type t_Y	
0	0	NR	Never Recover
1	0	HU	Hurt
0	1	HE	Helped
1	1	AR	Always Recover

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)
  - $ACE(X \rightarrow Y) = p(Helped) - p(Hurt)$



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 $Z = 0 \Rightarrow X = 0$

z	x	y	count	z	x	y	count
0	0	0	158	1	0	0	52
0	0	1	14	1	0	1	12
0	1	0	*0*	1	1	0	23
0	1	1	*0*	1	1	1	78

Table: Lipid data; there are two structural zeros

Finding the ACE

- Bounds

$$\Big[p(y=1 \mid z=1) - p(y=1 \mid z=0)\Big]$$
$$- \Big[p(y=1, x=0 \mid z=1) + p(y=0, x=1 \mid z=0)\Big]$$
$$\leq ACE(X \rightarrow Y) \leq$$
$$\Big[p(y=1 \mid z=1) - p(y=1 \mid z=0)\Big]$$
$$+ \Big[p(y=0, x=0 \mid z=1) + p(y=1, x=1 \mid z=0)\Big]$$

- 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

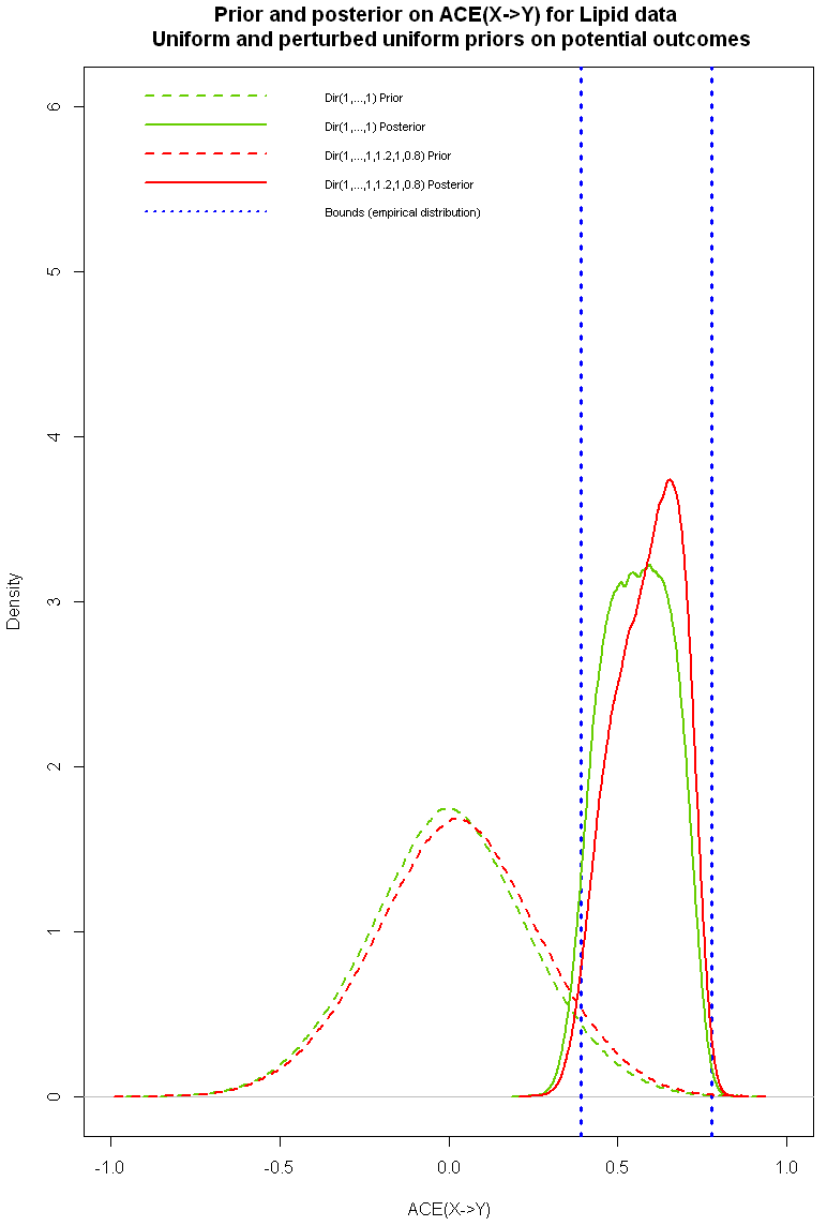
- Posterior
  - Gibbs sampling to sample (t_X, t_Y) from resulting posterior
  - Find $ACE(X \rightarrow Y) = p(Helped \mid data) - p(Hurt \mid data)$

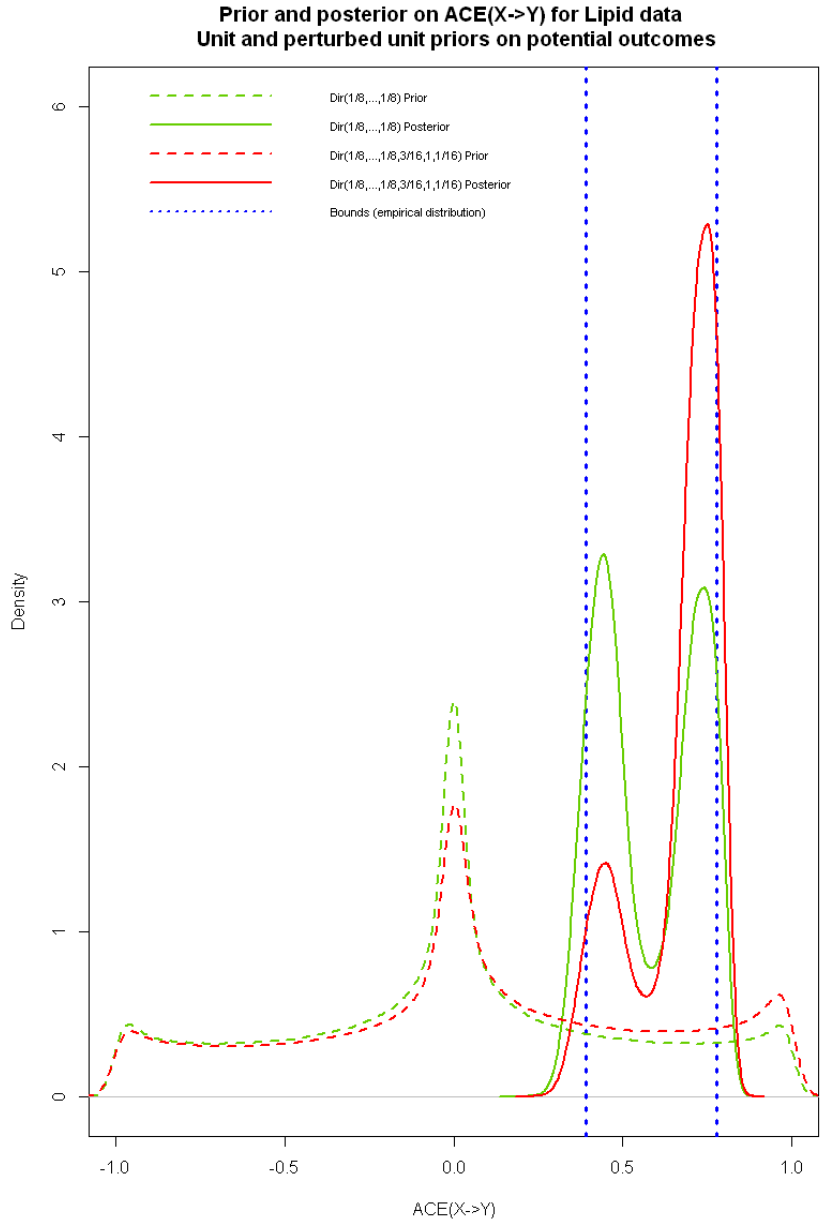
Posterior: Sensitive to prior?

- Uniform
  - Dir $(1, \ldots, 1) \rightarrow$ Dir $(1, \ldots, 1.2, 1, 0.8)$
- Unit
  - Dir $(\frac{1}{8}, \ldots, \frac{1}{8}) \rightarrow$ Dir $(1, \ldots, \frac{3}{16}, \frac{1}{8}, \frac{1}{16})$

- Such perturbation should not have large effect on posterior ACE

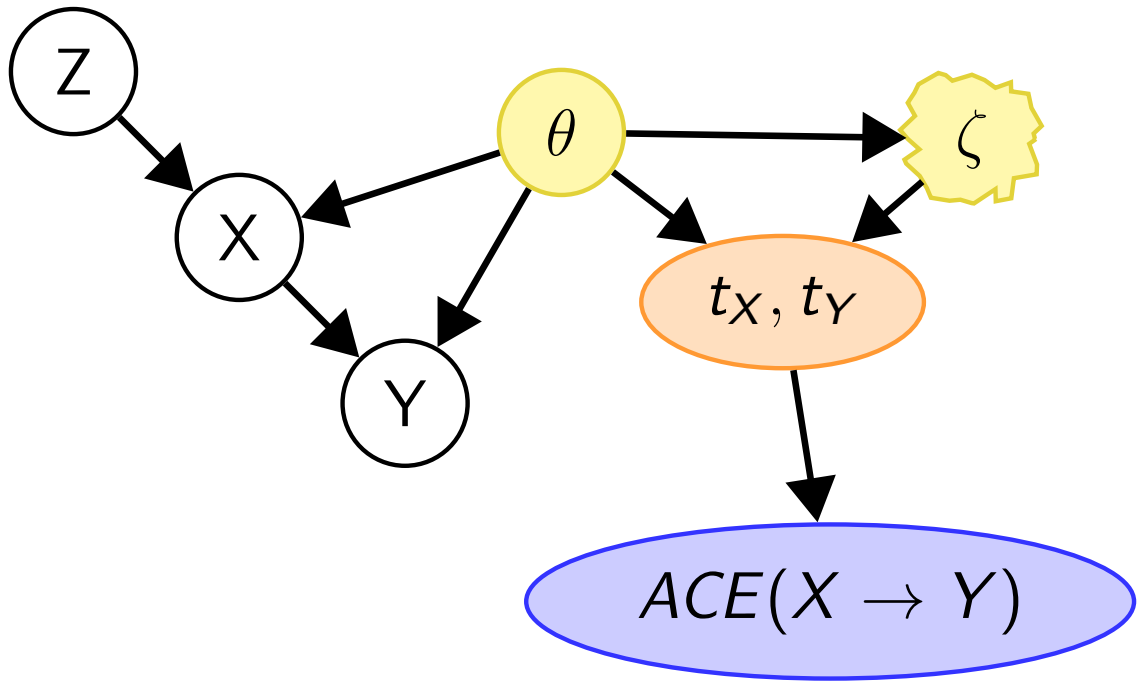
Posterior: Sensitive to prior?





Transparent Parametrizations

- Re-parameterize $p(t_X, t_Y)$ into $f(\theta, \zeta)$
  - θ = identifiable parameter (estimable from observed (X, Y, Z))
  - ζ = non-identifiable parameter



Transparent Parametrizations

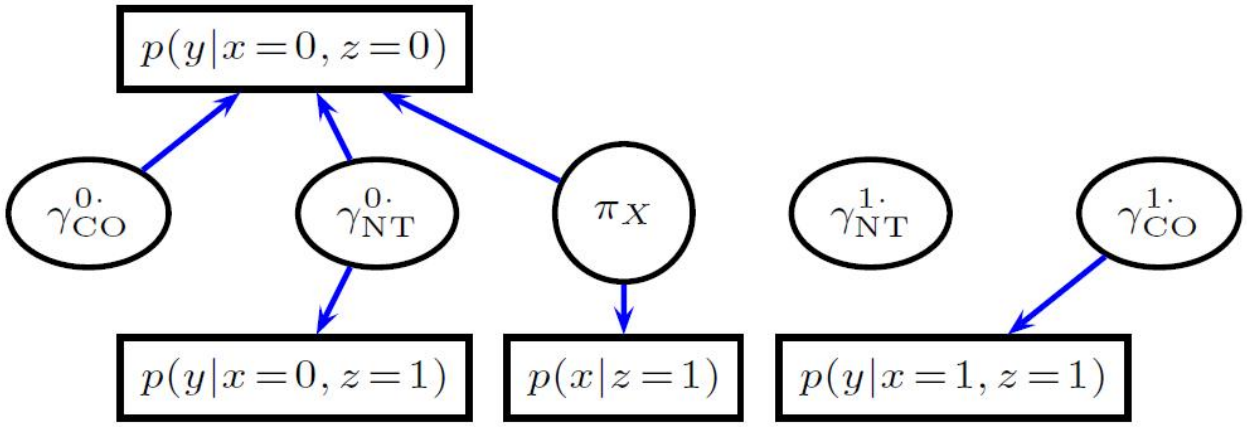


Figure 4: *A graph representing the functional dependencies in the analysis of the simple IV model with no Always Takers or Defiers. Rectangular nodes are observed; oval nodes are unknown parameters.* $p(x=1|z=0)=0$, *so* $p(y|x=1,z=0)$ *is undefined, hence these nodes are omitted.*

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 $(X \to Y)$ 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
    $ACDE_{NT}(x_0) = E\Big(Y_{x=0,z=1} - Y_{x=0,z=0} \mid NT\Big)$
  - Evaluate bounds conditional Intention-To-Treat effect e.g. $ITT_{CO} = E\Big(Y_{x=1,z=1} - Y_{x=0,z=0} \mid CO\Big)$

References

J.D. Angrist, G.W. Imbens, and D.B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

J. Pearl. *Causality: models, reasoning, and inference*, volume 47. Cambridge Univ Press, 2000.