On the relationship between identif
causality and instrumental varia

Perugia - Convegno Causalità

29-30 aprile 2005

Outline

- Nonparametric identification of causal effects
- Parametric notion of identification
- Identification within the Potential Outcome context
- Instrumental variables

Nonparametric identification of causal e

- Spirtes, Glymour and Scheines (1993), Pearl (1995, 20

- Joint distribution of r.v. $(Y_1, \ldots, Y_k)$ factorizing accord with $V = (1, \ldots, k)$ the set of nodes:

$$P_{1,\ldots,k}(Y_1 \ldots Y_k) = P_k(Y_k)\ \Pi_{i=1}^{k-1} P_i(Y_i \mid Y_{par(}$$

- Nonparametric as the functional form of each factor is

► *Causal graph*

- Enhance the graph with a **causal interpretation**:

  (a) all relevant variables are in the graph (*causally suffi*

  (b) represents the system under intervention and
  conditions (*stability*).

► *What is a causal effect?*

- Causal effect: $P(Y_1, \ldots, Y_k \mid \text{do}(Y_j))$ (post-intervention

- If the graph is *causal* it can be derived from the pr
distribution:

- $P(Y_1, \ldots, Y_k \mid \text{do}(Y_j) = v) = \Pi_{i \neq j} P(Y_i \mid Y_{par(i)}) \mid_{Y_j = v}$

- Truncated factorization (Pearl, 2000, 2003)

When all variables are observed and the graph is caus
effects are identifiable.

► *Unobserved variables*

- Partition $V = \{O, L\}$ with O observed and L latent (i.e.
over).

- Let the effect of Y_j on Y_k be of interest:

  $P(Y_k \mid \text{do}(Y_j) = v) = \sum_{par(j)} P(Y_k \mid v, par(j))P(par(j)$

- The effect of Y_j on Y_k is identifiable if $P(Y_k \mid \text{do}(Y_j)$
computed uniquely from the observed variables.

- Simple criterion: it is identifiable whenever nodes $j,$
correspond to variables that are observed.

► *Graphical criteria*

The simple criterion can be sharpened:

- *back-door* criterion
- *front-door* criterion
- Galles and Pearl (1995) criterion
- Tian and Pearl (2002)...

These are all *sufficient criteria*.

Parametric identification

- Rothenberg (1971) Bowden (1973)
- A statistical model $P(Y_1, \ldots, Y_k; \theta)$ is assumed
- θ_0 and θ_1 are observationally equivalent if $P(y_1,$ $P(y_1, \ldots, y_k; \theta_1)$ for all $y_i \in \mathcal{R}^k$.

▶ *Identification at θ_0*

- Global Identification: a parameter point θ_0 is identifiable
  no other $\theta \in \Omega$ such that θ and θ_0 are observationally e

- Local Identification: a parameter point θ_0 is *locally* iden
  exists an open neighborhood of θ_0 containing no other
  and θ_0 are observationally equivalent;

► *Model identification*

- A model il globally identifiable if every parameter po globally identified

- A model is locally identified if every parameter point θ identified

- *Another look at the definition*

Let $m_1(\theta), \ldots, m_r(\theta)$ be the moment characterizing P
sume they exist). Then global identification of a mode
$m_1(\theta), \ldots, m_r(\theta)$ should be invertible w.r. to θ.

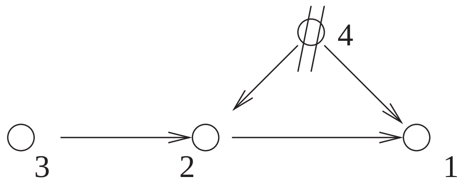
Sometimes θ is partitioned $\theta = \{\theta_a, \theta_b\}$ with θ_a of
Identification is then assessed only for a θ_a

- *When a parametric model is assumed* the graphical c
sharpened

For the Gaussian model (Structural Equation Approach):

- Instrumental Variable;
- Series of papers by Brito and Pearl (2004a,2004b,...);
- Grzebyk et al. (2004), Kuroki and Mayakawa (2004), St
Wermuth (2005);

- *Gaussian case: Instrumental variable* $O = \{1, 2, 3\}$ $L = \{4$

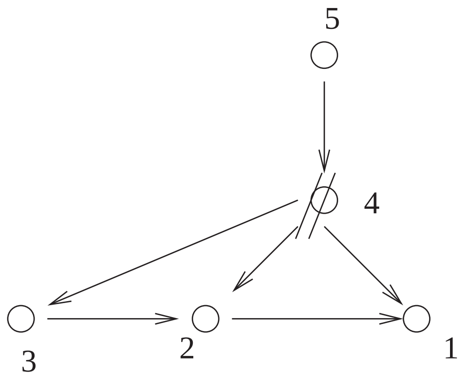


  - β_{kj} is the partial regression coefficient of j on k give
  - α_{jk} is the simple regression coefficient;
  - We want to estimate β_{12};
  - $Y_3 \perp\!\!\!\perp Y_4$ and $Y_3 \perp\!\!\!\perp Y_1 \mid Y_2\, Y_4$;

Y_3 is an instrument for β_{12} as

$$\beta_{12} = \alpha_{13}/\alpha_{23}$$

- *Another example*: $O = \{1, 2, 3, 5\}$ $L = \{4\}$



- *Under the Gaussian assumption*: β_{12} can still be identified
  of elements of the observed covariance matrix.

But are these results really results on cau

- Strongly relying on the assumption of *causal sufficienc*
- Strongly relying on the group-level representation (Ec
  Ch. 8).

One step behind

- The definition of causal effect is at the individual level, comparison between $P(Y_k \mid \text{do}(Y_j) = v)$ and $P(Y_k \mid \text{d}$
each individual in the population;

- The way we are doing this comparison is between *group* defined according to the variables Y.

Now, if (a) the individuals are random variables and (b) th of *causal sufficiency* holds, then the two things are the sa 2000, Pearl, 2003).

Potential Outcome Framework

Y outcome, D cause a binary variable (usually called *trea*

- Causal Effect: comparison between $P(Y \mid \text{do}(D) =$
  $\text{do}(D) = 0)$ at the individual level.

- New notation for the individual level i: comparison betw

$$Y_i(D_i = 1) \text{ and } Y_i(D_i = 0)$$

  such as average treatment effect $E[Y_i(D_i = 1)] - E[Y_i($

- Missing variable problems: we observe (Y_i, D_i).

▶ *Assignment mechanism*

- If the assignment mechanism of D is random the av effect is identified as:

$$E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0)$$

- If strongly ignorable given covariates then average trea identified (Rubin and Rosenbaum papers);

- If it cannot be ignored (as in studies with partial co search for an instrumental variable (Angrist, Imber papers).

▶ *Angrist, Imbens and Rubin papers*

There is an instrument Z (usually called *intention to treat*) t
D such that:

- $D_i = D_i(w)$ is either 0 or 1, an indicator of whether i
$Z_i = w$;

- Potential outcomes: $Y_i(D_i, Z_i)$;

- Missing variable problem: we observe (Y_i, D_i, Z_i)

▶ *The causal effect:*

$$E[Y_i(D_i = 1) - Y_i(D_i = 0) \mid D_i(w) - D_i(v) = 1] =$$

It is the average difference of the potential outcomes in peo
have taken the treatment of $Z_i = w$ and not taken if $Z_i =$
people).

We can identify the effects of Z (intention-to-treat effects)

- $E[Y_i \mid Z_i = w] - E[Y_i \mid Z_i = v]$
- $P(D_i = 1 \mid Z_i = w) - P(D_i = 1 \mid Z_i = z)$.

Without further assumptions the causal effect cannot be id

Further assumptions

- The usual assumption of Rubin causal models holds (S
- $Y_i(D_i = d, z) = Y_i(D_i = d, w)$ (exclusion restriction)
- For each $w > v$, $D_i(w) \geq D_i(z)$ for all i (monotonicity).

$\mu_1 - \mu_0 = \frac{E[Y_i|Z_i=w]-E[Y_i|Z_i=v]}{P(D_i=1|Z_i=w)-P(D_i=1|Z_i=v)}$

Instrumental variable formula.

For this example

- The Potential Outcome assumptions (Rubin Cau
Monotonicity) lead to the same estimator than the c
Gaussian (Pearl's Structural Equation Models);

- D should be binary in Potential Outcome Context anc
(Durbin 1954);

- Again, reliance on *different* but strong assumptions.

Nevertheless the two approaches are different:

- the definition of causal effect is different: $P(Y \mid$ d
  $E[Y_i(D_i = 1) - Y_i(D_i = 1)]$;
- Pearl's causal models can be generalised to *any* s
  Rubin's single cause and single effect (?);
- However, Rubin's does not assume a DAG as depende
- Overall, the use of instrumental variables within
  Outcome Framework seems a more convincing approa

Conclusions

- Within the Structural Equation Approach there are se ization of IV;
- They must have a counter part within the Poter Approach (Principal Strata?);
- They must also have a role in constructing Boundin Causal Effects (Pearl, 2003)
- Testing of Causal Hypothesis? (Robins, Spirtes, 2003).