Causal Inference, Self Selection and Identification

Yitian (Sky) Liang

Department of Statistics

2011/03/24

Outline

- What is causality?
- Objective of causal inference
- Self-selection
- Identification methods

What is Causality? (My Understanding)

- Holding all factors save one at a constant level, the change in the outcome associated with manipulation of the varied factor is called a causal effect of the manipulated factor.
- What is the object that bears the "causal effect"? (In this talk, human being)
- Unit of causal effect: individual vs population
- Key feature:
  - *ceteris paribus*
  - hypothetical

Examples

- (In Ehsan's talk) Does participating in a training program *cause* an increase in earnings?
- (Epidemiological topic?) Does smoking *cause* a deterioration in health?
- (Labor economics) Does education *cause* an increase/decrease in labor supply?
- (Political science topic?) Does migration to Canada *cause* an increase/decease in salary?
- (Business problem) Will a price increment of SMS decrease the usage of SMS?

If I know A cause B, so what?

- (My understanding) The final purpose of causal inference is to evaluate policy intervention.
  - Historical policy intervention
  - Policy intervention in new environment
  - New policy intervention

(Ehsan's) Example

- Two treatments: participation ($X = 1$) vs non-participation ($X = 0$)
- Interested outcome: earnings (Y)
- Objective: Does participation cause a higher earning?

Individual Causal Effect

- Population: Ω. Individual: $\omega \in \Omega$.
- Individual causal effect:

$$Y\left(X = 1, \omega\right) - Y\left(X = 0, \omega\right).$$

- They are called "counterfactuals".
- The ideas of "ceteris paribus" and "hypothetical" are embedded.

Population Causal Effects

- Suppose we could observe Y_1 and Y_0 for everyone.
- We have the population distribution $f_{Y_1,Y_0}(y_1, y_0)$.
- In principle, one could define various causal effects/parameters based on f_{Y_1,Y_0}.
- Example:

$$E_\omega\{Y_1 - Y_0\} \; or \; E_\omega\{Y_1 - Y_0|X = 0\}.$$

Causal Inference

- Why do we need "inference"?
- Main problem: we don't observe Y_1 and Y_0 for the same person!
- Define $Y = XY_1 + (1 - X)\, Y_0$. We observe Y and X.
- The general task of causal inference is to estimate f_{Y_1, Y_0} from $f_{Y,X}$.
- More specifically: estimate $E_\omega \{Y_1 - Y_0\}$ from $\{Y, X\}$.

Self Selection

- A key feature: X is NOT independent of Y.
- This is generally true for many social science and epidemiological problems.
- In contrast with non-random sampling, self selection problem is from the perspective of the individuals, not the researchers.
- Even though we could observe the population data, self selection problem is still there.

Additive Assumption (Only for Simplification)

- Assume

$$\begin{cases} Y_{1\omega} = \mu_1 + U_{1\omega} \\ Y_{0\omega} = \mu_0 + U_{0\omega} \end{cases}, \; E\left(U_{1\omega}\right) = E\left(U_{0\omega}\right) = 0.$$

- What is U_1 and U_2?
- So the causal parameter of interest becomes

$$E_\omega\left\{Y_1 - Y_0\right\} = \mu_1 - \mu_0 + E\left(U_{1\omega}\right) - E\left(U_{0\omega}\right) = \mu_1 - \mu_0.$$

Bias From STAT200 Analysis

- Recall what we observe: $\{Y, X\}$.
- Suppose we observe the whole population.
- Does a linear regression "estimate" give what we want?
- Linear regression "estimate"

$$
\begin{aligned}
& E\left(Y_1|X=1\right)-E\left(Y_0|X=0\right) \\
= \;& \mu_1-\mu_0+\left[E\left(U_1|X=1\right)-E\left(U_0|X=0\right)\right].
\end{aligned}
$$

Essentials of Bias

- The dependence between the "outcome process" and the "selection process".
- Emphasis again: not from any biased-sampling procedure.
- Why randomization in the design context is so important?

Identification

- Without further assumptions, one could not (I think!!) identify the causal parameter from the observed data.
- Difference identification methods essentially require
  - Different identification assumptions
  - Different knowledge about the outcome process, the selection process and the relationship between them.
- Focus on a general survey of different identification methods.

Matching (Propensity Score)

- In addition to $\{Y_\omega, X_\omega\}$, we also observe $\{W_\omega\}$.
- Rosenbaum and Rubin (1983) showed that

$$(Y_1,\ Y_0) \perp X \mid P(W),$$

under assumptions

$$(A1) \quad (Y_1, Y_0) \perp X \mid W$$

and

$$(A2) \quad 0 < Pr(X = 1|W) = P(W) < 1.$$

Matching Cont.

- Under the assumptions, we have

$$
\begin{aligned}
E\left(Y_1 | X = 1, P\left(W\right)\right) &= E\left(Y_1 | P\left(W\right)\right) \\
E\left(Y_0 | X = 0, P\left(W\right)\right) &= E\left(Y_0 | P\left(W\right)\right).
\end{aligned}
$$

- So the causal parameter could be obtained as

$$
\begin{aligned}
E\left(Y_1 - Y_0\right) &= \int \{E\left(Y_1 | X = 1, P\left(w\right)\right) - \\
&\quad E\left(Y_0 | X = 0, P\left(w\right)\right)\} f_W\left(w\right) dw.
\end{aligned}
$$

Why Can Matching Help?

- What do the following mean?
  - $(Y_1, Y_0) \perp X \mid W$
  - $(Y_1,\ Y_0) \perp X \mid P(W)$
- Conditional on the observed information W, the reason that two observationally equivalent individuals make difference choices is purely due to randomization!
- Can we justify it? What does it imply for the requirement of W?

Instrumental Variable

- Re-write the observed Y as

$$\begin{aligned} Y &= XY_1 + (1-X)Y_0 \\ &= \mu_0 + \Delta \cdot X + U_0, \end{aligned}$$

  where $\Delta = \mu_1 - \mu_0 + U_1 - U_0$.
- To illustrate a fundemental idea of IV, assume $U_1 = U_0$.
- In this case, does a linear regression of Y on X give us a consistent estimate of Δ?

IV Cont.

- The problem of linear regression is the dependence between X and U_0.
- If we can find such an instrument Z that satisfies

$$E\left(U_0|Z\right) = E\left(U_0\right),$$

and

$$Pr\left(X = 1|Z\right) \textit{ is a nontrivial function of } Z.$$

- The IV estimator could be defined as

$$\Delta^{IV} = \frac{E\left(Y|Z = z\right) - E\left(Y|Z = z'\right)}{Pr\left(X = 1|Z = z\right) - Pr\left(X = 1|Z = z'\right)}.$$

Other Identification Methods

- Control function
- MLE

Miscellaneous

- The definition of causal effects could be very different in other fields.
- We don't have to focus on the mean!!
- When use different identification methods, I urge one to critically evaluate the invoked assumptions under the specific subject area context.

Reference

James Heckman, "The Scientific Model of Causality," Sociological Methodology, 35: 1–97, 2005.