1. A Brief Review of Counterfactual Causality

Felix Elwert, Ph.D.
elwert@wisc.edu
University of Wisconsin–Madison
Version: May 2013

This workshop focuses on graphical causal models. The graphical approach to causal inference using directed acyclic graphs (DAGs) is equivalent to the potential outcomes approach to causal inference.

⇨ Same concepts, same theorems, different notation.

Since we are a diverse group, we'll first spend about an hour reviewing basic counterfactual notation and concepts and from there motivate the use of DAGs.

We will use this notation at the end of the second day. We will use the concepts throughout.

1. The Counterfactual (Potential Outcomes/Neyman-Rubin) Framework of Causal Inference

Protagonists:

Roots in Neyman (1923)
Statistics: Donald B. Rubin, Paul Holland, Paul Rosenbaum
Economics: James Heckman, Charles Manski

Accomplishments:

1. A precise definition of causal effects
2. A formal model of causality against which we can assess the adequacy of various estimators

Approach:

Causal questions are "what if" questions.

Extend the logic of randomized experiments to observational data.

2. Total Causal Effects in the Potential Outcomes Framework

Example: What is the causal effect of attending catholic school vs. public school on high school graduation?

Consider a binary treatment $T = \{0,1\}$:

T: Attending catholic school (=1) vs. attending public school (=0)

Each individual has two potential outcomes, Y_T, one for each value of the treatment. A potential outcome is the outcome that would be realized if the individual received a specific value of the treatment. (I'm going to suppress "i" subscripts for convenience.)

Y_1: Potential outcome if attending catholic school

Y_0: Potential outcome if attending public school.

For each particular individual, one can generally observe only one, but not both, of the two potential outcomes. The unobserved outcome is called the "counterfactual" outcome.

Let Y be the observed outcome (note: no subscript). Then:

$Y = Y_1$ if $t = 1$

and

$Y = Y_0$ if $t = 0$.

The individual level causal effect (ICE) is defined as the difference between an individual's two potential outcomes,

$ICE = \delta = Y_1 - Y_0$

The ICE answers the question "what would happen if the individual received treatment rather than control?"

$Y_0 = Y_1 = 1 \rightarrow \delta = 0$	Kid would graduate from both catholic and from public school. No effect.
$Y_0 = Y_1 = 0 \rightarrow \delta = 0$	Kid would neither graduate from catholic nor from public school. No effect.
$Y_0 = 1, Y_1 = 0 \rightarrow \delta = -1$	Kid would not graduate from catholic but would graduate from public school. Negative effect.
$Y_0 = 0, Y_1 = 1 \rightarrow \delta = 1$	Kid would graduate from catholic but would not graduate from public school. Positive effect.

We usually cannot rule out that the ICE *differs* across individuals ("effect heterogeneity"). Thus, we define the average causal effect (ACE) as the population average of the individual level causal effects,

ACE = E[δ] = E[Y_1] – E[Y_0].

The ACE is a difference at the population level: it's the high school graduation rate if all kids in a study population had attended catholic school minus the high school graduation rate if all kids had instead attended public school.

3. Fundamental Problem of Causal Inference, Identification, & Assumptions

The so-called "fundamental problem of causal inference" (Holland 1986) is that one can never *directly observe* causal effects (ACE or ICE), because we can never observe both potential outcomes for any individual. We need to compare potential outcomes, but we only have observed outcomes

		Outcome	
		Treatment	Control
Assignment	Treatment	(Y | T=1) = (Y_1 | T=1)	???
	Control	???	(Y | T=0) = (Y_0 | T=0)

→ Causal inference is a missing data problem

So what's an analyst to do? We want to estimate the ACE,

ACE = E[Y_1] – E[Y_0],

where E[.] ranges over the entire population.

But we only have the so-called standard estimator (S*),

$$S^* = E[Y_1|T = 1] - E[Y_0|T = 0],$$

where E[.] ranges over the domain of the treatment and control groups, respectively, not over the entire population.

ACE measures *causation* whereas S* measures *association.*

There's no good reason why the ACE should automatically equal S*.

"Causation ≠ Association"

The "identification problem" refers to the difficulty of separating causation from association.

Assumptions are unavoidable!

Since the fundamental problem of causal inference is a missing data problem, we need to make assumptions to fill in the missing values. *Assumption-free causal inference is impossible!*

Not the existence but the quality of the assumptions is the issue.
Later, we'll use DAGs to get a handle on these assumptions

Specifically, we need to make assumptions (have a "theory") about how the data were generated and collected.

4. Solving the Fundamental Problem with Randomized Experiments

A sufficient condition for the standard estimator (S*) to be an unbiased and consistent estimate of the average causal effect (ACE) is:

(1) $E[Y_1 |T=1] = E[Y_1 |T=0] = E[Y_1]$

The mean potential outcome under treatment for those in the treatment group equals the mean potential outcome under treatment for those in the control group,

and

(2) $E[Y_0 |T=1] = E[Y_0 |T=0] = E[Y_0]$

The mean potential outcome under control for those in the treatment group equals the mean potential outcome under control for those in the control group.

Because then

$$E[Y_1 |T=1] - E[Y_0 |T=0] = E[Y_1] - E[Y_0]$$

association = causation

Note that conditions (1) and (2) are comparability conditions: we must assume that the people in the treatment group on average are identical to the people in the control group with respect to their potential outcomes.

You might call this the "compare apples to apples principle."

This condition can be achieved by random assignment of individuals to the treatment and control group (in a randomized experiment).

5. Causal Inference in Observational Studies

Often random assignment is not possible. Studies where the researcher cannot directly manipulate treatment assignment are called "observational studies."

Then what condition must hold for the standard estimator (S*) to be unbiased and consistent for the average causal effect (ACE)?

"Ignorability" is sufficient (Rubin 1974):

$$(Y_1, Y_0) \perp T$$

In words, the potential outcomes, Y_1 and Y_0, must be jointly independent ("⊥") of treatment assignment.

Ignorability holds in an ideal experiment.

In observational studies, ignorability seldom holds on its own (i.e., without any adjustments). But it may hold within groups defined by some other variables, X.

"Conditional ignorability":

$$(Y_1, Y_0) \perp T | X$$

In words, the potential outcomes, Y_1 and Y_0, must be jointly independent ("⊥") of treatment assignment within groups defined by the value of X (X may be a vector).

If (conditional) ignorability holds, then the standard estimator (S*) will give an unbiased and consistent estimate of the ACE.

But how would we know whether ignorability holds?

⇨ **DAGs can answer this question.**