Counterfactuals and the Potential Outcome Model Summary

Siddhartha Basu

April 21, 2017

0.1 Potential Outcomes

In causal inference, we are generally interested in individual-level causal impacts of the treatment:

$$\delta_i = y_i^1 - y_i^0$$

Let D be the causal exposure variable, which is 1 if you are exposed to the treatment state and 0 if you are exposed to the control state. The observed outcome variable Y is therefore:

$$Y = Y^1 \text{ if } D = 1$$
$$Y = Y^0 \text{ if } D = 0$$
$$Y = DY^1 + (1 - D)Y^0$$

The last equation shows the **fundamental problem of causal inference**, that we cannot observe the potential outcome under the control state for those in the treatment state, and that we cannot observe the potential outcome under the treatment state for those in the control state. This means that it is **impossible to calculate individual-level causal effects**.

0.2 The Average Treatment Effect

Since we cannot calculate individual-level causal effects, we focus our attention on carefully defined aggregate causal effects. The broadest possible average effect is the average treatment effect (ATE) for the population as a whole:

$$E[\delta] = E[Y^1 - Y^0] = E[Y^1] - E[Y^0]$$

0.3 The Stable Unit Treatment Value Assumption

0.3.1 Definition

In most applications, the potential outcome model retains its tractability through the presence of the stable unit treatment value assumption (SUTVA). SUTVA requires that the potential outcomes of individuals be unaffected by changes in the treatment exposures of all other individuals. Mathematically, it requires that, if $\mathbf{d}$ is the vector of treatment assignments for all N individuals, the treatment effect for each individual i:

$$\delta_i(\mathbf{d}) = y_i^1(\mathbf{d}) - y_i^0(\mathbf{d})$$

depends on $\mathbf{d}$, only through d_i, individual i's assignment. SUTVA is what allows us to declare $y_i^1(\mathbf{d}) = y_i^1$ and $y_i^0(\mathbf{d}) = y_i^0$, and as a result, individual-level causal results δ_i exist that are independent of the overall configuration of causal exposure.

0.3.2 Violations

Typical SUTVA violations share two interrelated features:

- **Influence patterns** that result from contact across individuals in social or physical space.

- **Dilution/concentration patterns** that one can assume would result from changes in the prevalence of treatment.

In the worker training example, SUTVA is violated for large programs in small markets but not small programs in large markets. For Catholic schooling, there are both influence patterns between students and potential dilution patterns.

If the violation can be interpreted as a dilution/concentration problem, even when generated in part by an underlying influence pattern, the analyst can proceed by scaling back the asserted relevance of any estimates where the prevalence of treatment is not substantially different. The idea is to state that estimates of average causal effects hold only for what-if movements of relatively small numbers of individuals. This doesn't work when influence patterns are inherent to the causal process of interest, eg. vaccinations. In situations like this you have:

- The indirect effect: difference in outcome for a non-vaccinated person in a community with a vaccination program vs. his outcome in a similar community without a vaccination program.

- Mathematically, Y_i^0(community vaccination program) $- Y_i^0$(no community vaccination program)

- The total effect: Y_i^1(community vaccination program) $- Y_i^0$(no community vaccination program)

Effectively estimating these types of effects requires a **nested randomization structure**, where (1) vaccine programs are randomly assigned to a subset of communities and (2) people within these communities are randomly given vaccinations.

0.4 Treatment Assignment and Observational Studies