Statistical Inference in Experiments I

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- We learned the potential outcome framework last time.
- This framework enables us to define quantities with a causal interpretation.
- Our estimands are usually averages of the individualistic treatment effects on a fixed group.
- These estimands can be identified under certain assumptions.
- We can rely on the scientific solution or the statistical solution to solve the fundamental problem of causal inference.
- The latter is more common in social science.
- It requires 1) a large sample, and 2) random assignment of the treatment.
- Then, we will be able to construct estimators for the estimand.

The Horvitz-Thomspon estimator

- Let's consider a fixed sample with the Bernoulli trial $p_i = p$.
- One estimator we often use to estimate the ATE is the Horvitz-Thompson estimator:

$$\hat{\tau}_{HT} = \frac{1}{N}\sum_{i=1}^{N}\frac{D_i Y_i}{p} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}$$

- We can show that $\hat{\tau}_{HT}$ is unbiased and consistent for τ_{SATE}:

$$\begin{aligned}
E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_i Y_i}{p}\right] &= \frac{1}{N}\sum_{i=1}^{N} E\left[\frac{D_i Y_i}{p}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{p}E[D_i Y_i | D_i = 1]P(D_i = 1) \\
&= \frac{1}{N}\sum_{i=1}^{N} E[Y_i(1) | D_i = 1] = \frac{1}{N}\sum_{i=1}^{N} Y_i(1)
\end{aligned}$$

Variance of the Horvitz-Thomspon estimator (*)

- Similarly, $E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}\right] = \frac{1}{N}\sum_{i=1}^{N} Y_i(0)$.
- Hence, $E[\hat{\tau}_{HT}] = \tau_{SATE}$.
- Note that we treat $\{Y_i(0), Y_i(1)\}_{i=1}^{N}$ as fixed values in the sample.
- Now, variance:

$$
\begin{aligned}
Var\left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_i Y_i}{p}\right] &= \frac{1}{N^2}\sum_{i=1}^{N} Var\left[\frac{D_i Y_i}{p}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{p^2}E[D_i Y_i^2] - \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{p^2}E^2[D_i Y_i] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(1)}{p} - \frac{1}{N^2}\sum_{i=1}^{N} Y_i^2(1)
\end{aligned}
$$

Variance of the Horvitz-Thomspon estimator (*)

- Similarly,

$$Var\left[\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}\right]=\frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(0)}{p}-\frac{1}{N^2}\sum_{i=1}^{N}Y_i^2(0).$$

$$Cov\left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_iY_i}{p},\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}\right]$$

$$=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov\left[\frac{D_iY_i}{p},\frac{(1-D_j)Y_j}{1-p}\right]$$

$$=\frac{1}{N^2}\sum_{i=1}^{N}Cov\left[\frac{D_iY_i}{p},\frac{(1-D_i)Y_i}{1-p}\right]$$

$$=-\frac{1}{N^2}\sum_{i=1}^{N}E\left[\frac{D_iY_i}{p}\right]E\left[\frac{(1-D_i)Y_i}{1-p}\right]=-\frac{1}{N^2}\sum_{i=1}^{N}Y_i(1)Y_i(0).$$

Variance of the Horvitz-Thomspon estimator (*)

- Finally,

$$
\begin{aligned}
Var[\hat{\tau}_{HT}] = {}& Var\left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_i Y_i}{p}\right] + Var\left[\frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}\right] \\
& - 2 * Cov\left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_i Y_i}{p}, \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)Y_i}{1-p}\right] \\
={}& \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(1)}{p} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(0)}{1-p} \\
& - \frac{1}{N^2}\sum_{i=1}^{N}Y_i^2(1) - \frac{1}{N^2}\sum_{i=1}^{N}Y_i^2(0) + \frac{2}{N^2}\sum_{i=1}^{N}Y_i(1)Y_i(0) \\
={}& \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(1)}{p} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(0)}{1-p} - \frac{1}{N^2}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)]^2 \\
\leq{}& \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(1)}{p} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{Y_i^2(0)}{1-p}.
\end{aligned}
$$

Variance estimation of the Horvitz-Thomspon estimator

- When $N \to \infty$, $Var[\hat{\tau}_{HT}] \to 0$.
- The Horvitz-Thomspon estimator is (root-N) consistent.
- We estimate the first two terms of the variance with their sample analogues.
- The last term is essentially the variance of τ_i, which cannot be estimated.
- What we can have is

$$\widehat{Var}[\hat{\tau}_{HT}] = \frac{1}{Np}\frac{\sum_{i=1}^{N} D_i Y_i^2}{Np} + \frac{1}{N(1-p)}\frac{\sum_{i=1}^{N}(1-D_i)Y_i^2}{N(1-p)}$$

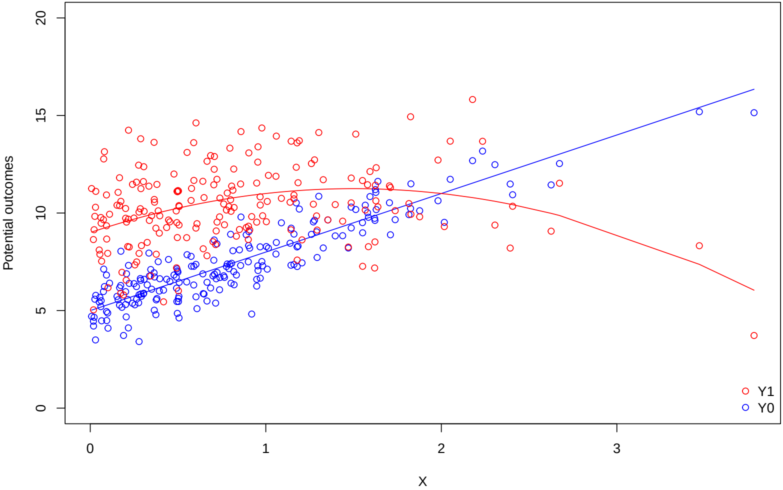
- We can show that $E\left[\widehat{Var}[\hat{\tau}_{HT}]\right] \geq Var[\hat{\tau}_{HT}]$.
- This is known as the Neyman variance estimator.
- The Neyman variance estimator is conservative unless the treatment effects are constant.

Asymptotics of the Horvitz-Thomspon estimator

- Li and Ding (2017) proved that $\sqrt{N}(\hat{\tau}_{HT} - \tau)$ converges to a normal distribution.
- The result is based on a theorem proved by Hoeffding, our next-door neighbor.
- The asymptotic 95% confidence interval for $\hat{\tau}_{HT}$ is as follows:

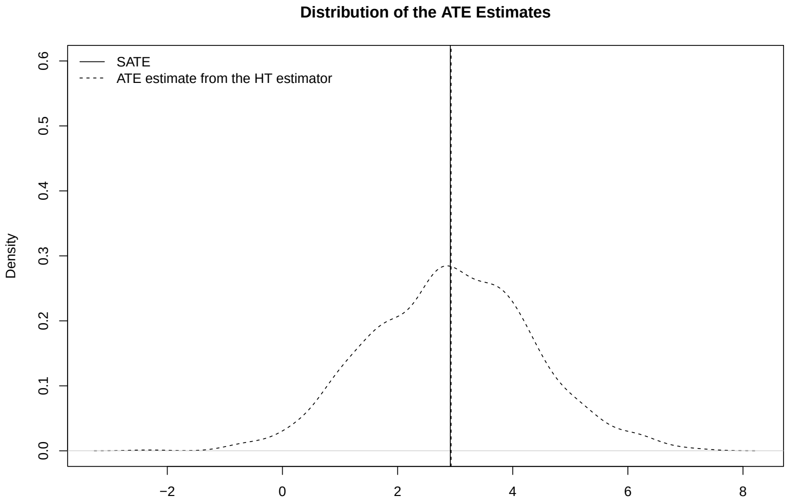
$$\left[\hat{\tau}_{HT} - 1.96 * \sqrt{\widehat{Var}[\hat{\tau}_{HT}]}, \hat{\tau}_{HT} + 1.96 * \sqrt{\widehat{Var}[\hat{\tau}_{HT}]}\right].$$

The Horvitz-Thompson estimator: simulation




```
## The SATE is 2.916262
```

The Horvitz-Thompson estimator: simulation




```
## The average of variance estimates is 1.891
```

The Hajek estimator

- The Hajek estimator for τ is

$$\hat{\tau}_{HA} = \frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - D_i)Y_i.$$

- This estimator is biased: $E\left[\frac{x}{y}\right] \neq \frac{E[x]}{E[y]}$.
- Yet it is root-N consistent and asymptotically normal for τ.
- We can derive its variance and statistical properties using the Delta method (Taylor expansion).
- Let's first consider the term $\frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i = \frac{\sum_{i=1}^{N} D_i Y_i}{\sum_{i=1}^{N} D_i}$.
- This is a ratio estimator.

Variance of the Hajek estimator (*)

- In general, a ratio estimator has the form of $\frac{x}{y}$.
- We can derive the Taylor expansion of this function around the values $(E[x], E[y])$:

$$\frac{x}{y} = \frac{E[x]}{E[y]} + \frac{1}{E[y]}(x - E[x]) - \frac{E[x]}{E^2[y]}(y - E[y]) + R.$$

- This process is known as linearization.
- In the previous example, $x = \frac{1}{Np}\sum_{i=1}^{N} D_i Y_i$, $y = \frac{1}{Np}\sum_{i=1}^{N} D_i$, $E[x] = \frac{1}{N}\sum_{i=1}^{N} Y_i(1) = \bar{Y}(1)$, and $E[y] = 1$.
- Suppose $x \to E[x]$ and $y \to E[y]$ when $N \to \infty$, then we can see that $\frac{x}{y} \to \frac{E[x]}{E[y]}$.
- Hence, $\frac{x}{y}$ is consistent for $\frac{E[x]}{E[y]}$.
- Similarly, $\sqrt{N}\frac{x}{y}$ converges to a normal distribution if $\sqrt{N}x$ and $\sqrt{N}y$ are asymptotically normal.

Variance of the Hajek estimator (*)

- To derive the variance of $\frac{x}{y}$, note that

$$
\begin{aligned}
Var\left[\frac{x}{y}\right] =& E\left[\frac{x}{y} - \frac{E[x]}{E[y]}\right]^2 \\
\rightarrow& \frac{1}{E^2[y]}E[x-E[x]]^2 + \frac{E^2[x]}{E^4[y]}E[y-E[y]]^2 \\
&- \frac{2E[x]}{E^3[y]}E\left[(x-E[x])(y-E[y])\right] \\
=& E\left[\frac{1}{E[y]}\left(x-E[x]\right) - \frac{E[x]}{E^2[y]}\left(y-E[y]\right)\right]^2
\end{aligned}
$$

Variance of the Hajek estimator (*)

- Plugging into the expression above, we have

$$
\begin{aligned}
& Var\left[\frac{\sum_{i=1}^N D_i Y_i}{\sum_{i=1}^N D_i}\right] \\
\to & E\left[\frac{1}{N}\sum_{i=1}^N\left[\frac{D_i Y_i}{p} - Y_i(1)\right] - \bar{Y}(1)\left[\frac{1}{Np}\sum_{i=1}^N D_i - 1\right]\right]^2 \\
= & E\left[\frac{1}{N}\sum_{i=1}^N \frac{D_i}{p}\left[Y_i - \bar{Y}(1)\right]\right]^2 \\
= & \frac{1}{N^2}\frac{\sum_{i=1}^N [Y_i(1) - \bar{Y}(1)]^2}{p} - \frac{1}{N^2}\sum_{i=1}^N [Y_i(1) - \bar{Y}(1)]^2.
\end{aligned}
$$

Variance of the Hajek estimator

- We can do the calculation for all the three terms.
- The identifiable part in the variance equals

$$\frac{1}{N^2}\frac{\sum_{i=1}^{N}[Y_i(1)-\bar{Y}(1)]^2}{p}+\frac{1}{N^2}\frac{\sum_{i=1}^{N}[Y_i(0)-\bar{Y}(0)]^2}{1-p}.$$

- The omitted part equals

$$-\frac{1}{N^2}\sum_{i=1}^{N}[Y_i(1)-Y_i(0)-(\bar{Y}(1)-\bar{Y}(0))]^2.$$

- In complete randomization, the two estimators are equivalent and the variance is the same as the one for the Hajek estimator.

Variance of the Hajek estimator

- We can see that the first term in the variance equals

$$
\begin{aligned}
&\frac{1}{N^2}\frac{\sum_{i=1}^{N}[Y_i(1)-\bar{Y}(1)]^2}{p}\\
=&\frac{1}{N^2}\frac{\sum_{i=1}^{N} Y_i^2(1)}{p}-\frac{1}{N^2}\frac{\sum_{i=1}^{N} 2Y_i(1)\bar{Y}(1)}{p}+\frac{1}{N^2}\frac{\sum_{i=1}^{N}[\bar{Y}(1)]^2}{p}\\
=&\frac{1}{N^2}\frac{\sum_{i=1}^{N} Y_i^2(1)}{p}-\frac{1}{N}\frac{[\bar{Y}(1)]^2}{p}.
\end{aligned}
$$

- The first part is the first term in the variance of the Horvitz-Thompson estimator and the second part is negative.
- The Hajek estimator is always more efficient.
- This is because the Hajek estimator uses "stabilized weights."

Variance estimation of the Hajek estimator

- The Neyman variance estimator in this case is

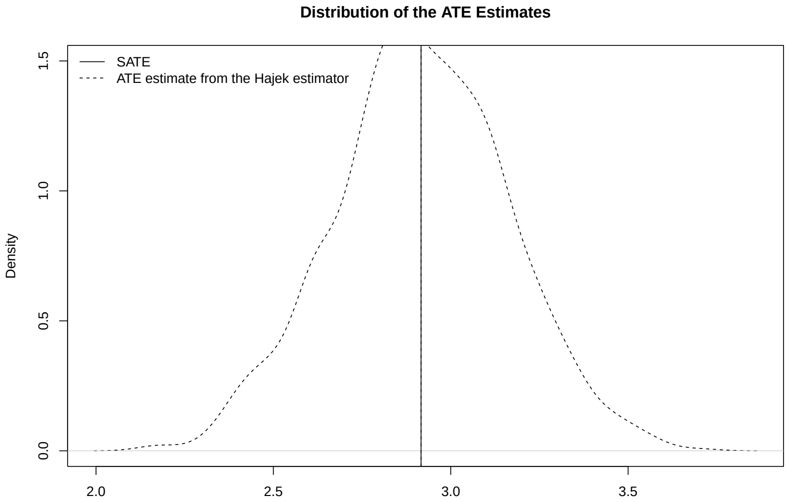
$$\widehat{Var}[\hat{\tau}_{HA}] = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0}, \text{ with}$$

$$S_1^2 = \frac{\sum_{i=1}^{N} D_i (Y_i - \widehat{\bar{Y}}(1))^2}{N_1 - 1} \text{ and}$$

$$S_0^2 = \frac{\sum_{i=1}^{N} (1 - D_i)(Y_i - \widehat{\bar{Y}}(0))^2}{N_0 - 1}.$$

- Here $\widehat{\bar{Y}}(d)$ is an estimate of $\bar{Y}(d)$, like $\widehat{\bar{Y}}(1) = \frac{1}{N_1} \sum_{i=1}^{N} D_i Y_i$.
- S_1^2 and S_0^2 are the sampling variance of Y_i in the treatment group and the control group, respectively.

The Hajek estimator: simulation




```
## The average of variance estimates is 0.09
```

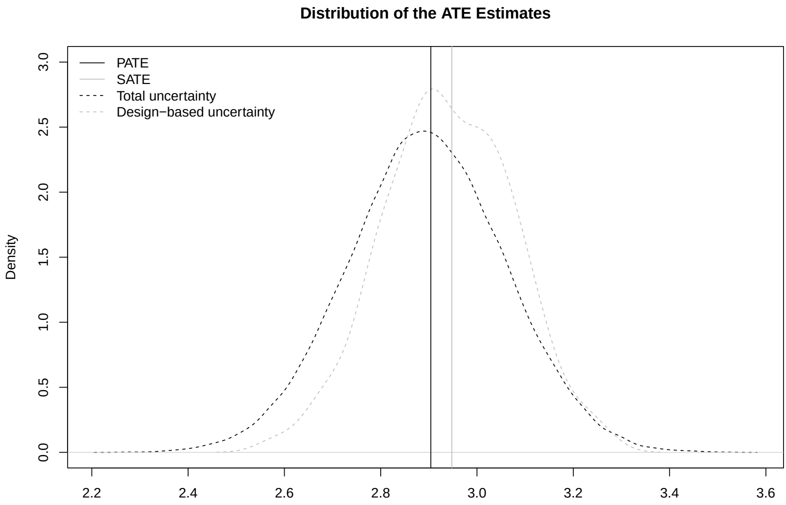
Design-based uncertainty

- What does the variance tell us?
- If we repeat the assignment process and obtain a series of $\hat{\tau}_{HT}$, how large will the variation be?
- The only source of randomness is treatment assignment, or the value of D_i.
- This is known as the design-based uncertainty.
- Conventionally, we believe the variance describes the uncertainty caused by sampling error.
- If so, what does the standard error mean if our analysis is at the population level (e.g., 50 states in the US)?
- We can think the collection of the same population under different treatment assignments as the real population (sometimes known as the **super population**).
- E.g., we draw a sample of 50 states from the super population of 2^{50} possibilities.

Sampling-based uncertainty

- Suppose we are only interested in the average ideal point of American people.
- We randomly draw a sample of 1,000 Americans and calculate the mean of their ideal points.
- The calculated mean will differ from one sample to another.
- This is known as the sampling-based uncertainty.
- There is no design-based uncertainty if we are interested in descriptive quantities.
- Both types of uncertainties may exist in practice.
- If the sample is representative, the unidentifiable part in the variance will be exactly the sampling variance.
- The Neyman variance estimator is then consistent for the combination of both types of uncertainties.

Sampling-based vs. design-based uncertainty



Sampling uncertainty vs. design uncertainty

```
## Total uncertainty = 0.163

## Design-based uncertainty = 0.133

## Sampling-based uncertainty = 0.03
```

Justify the Neyman variance

- In finite sample, the Neyman variance is conservative for the true variance of the SATE (design-based uncertainty).
- The reason is that we cannot estimate the part driven by treatment effect heterogeneity.
- We can construct sharp bounds of it (Aronow et al. 2014; Imbens and Menzel 2018).
- It is consistent when the effect is homogeneous.
- It is also consistent for the true variance of the PATE with representative sampling (design-based uncertainty + sampling-based uncertainty).

References I

Aronow, Peter M, Donald P Green, Donald KK Lee, et al. 2014. "Sharp Bounds on the Variance in Randomized Experiments." *The Annals of Statistics* 42 (3): 850–71.

Imbens, Guido, and Konrad Menzel. 2018. "A Causal Bootstrap." National Bureau of Economic Research.

Li, Xinran, and Peng Ding. 2017. "General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference." *Journal of the American Statistical Association* 112 (520): 1759–69.