A Causal Bootstrap

Guido Imbens and Konrad Menzel (AoS, to appear)

Summer, 2021
Background

- After applying a treatment $W_i \in \{0, 1\}$, the outcome is $Y_i(W_i)$.
- Causal effect: $Y(1) - Y(0)$.

Fundamental problem

We can only observe one realization of $W$ at a time, i.e.,

$$
\text{Causal effect} = \begin{cases} 
Y(1) - ?? & \text{if } W = 1 \\
?? - Y(0) & \text{if } W = 0
\end{cases}.
$$

- Parameter of interest:

$$
\tau_{ATE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0)).
$$
**Assumption**

**Assumption 1.1. (Sampling Experiment)**

The population consists of $N$ units with potential values $(Y_i(0), Y_i(1))_{i=1}^N$ which are i.i.d. draws from the distribution $F_{01}(y_0, y_1)$. The $n$ observed units are sampled at random and without replacement from the population,

$$Y_i(0), Y_i(1) \perp \perp R_i$$

where we denote $q := n/N \in (0, 1]$.

**Assumption 1.2. (Complete Randomization)**

Treatment assignment is completely randomized, that is for each unit with $R_i = 1$ we have

$$(Y_i(0), Y_i(1)) \perp \perp W_i$$

where $W_i = 1$ for $n_1$ units selected at random and without replacement from the $n$ observations with $R_i = 1$, and the propensity score $p := n_1/n$ satisfies $0 < p < 1$. 
Source of Randomness

- **Sampling Uncertainty**: uncertainty arose from $R_1, \ldots, R_N$.
  - captured by conventional standard error.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Actual Sample</th>
<th>Alternative Sample I</th>
<th>Alternative Sample II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y_i$</td>
<td>$W_i$</td>
<td>$R_i$</td>
</tr>
<tr>
<td>1</td>
<td>9.1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>?</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2.3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>-3.6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$N$</td>
<td>?</td>
<td>?</td>
<td>0</td>
</tr>
</tbody>
</table>
Source of Randomness

- Design Uncertainty: uncertainty arised from $W_1, \ldots, W_N$.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Actual Sample</th>
<th>Alternative Sample I</th>
<th>Alternative Sample II</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y_i(1)$</td>
<td>$Y_i(0)$</td>
<td>$W_i$</td>
<td>$Y_i(1)$</td>
</tr>
<tr>
<td>1</td>
<td>3.2</td>
<td>?</td>
<td>1</td>
<td>3.2</td>
</tr>
<tr>
<td>2</td>
<td>-1.6</td>
<td>?</td>
<td>1</td>
<td>-1.6</td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>2.3</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>?</td>
<td>-3.1</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>$N$</td>
<td>-5.7</td>
<td>?</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>
Notation

- Population distribution (with size $N$) of potential outcomes:
  \[ F_{01}^p(y_0, y_1) := \sum_{i=1}^{N} \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\}/N. \]

- Sample distribution of size $n$:
  \[ F_{01}^s(y_0, y_1) := \sum_{i=1}^{N} R_i \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\}/n. \]

- Number of treated units in the sample: $n_1$.

- Number of control units in the sample: $n_0 = n - n_1$.

- Empirical c.d.f. given the randomized treatment:
  \[
  \hat{F}_0(y_0) := \frac{1}{n_0} \sum_{i=1}^{N} R_i (1 - W_i) \mathbb{1}\{Y_i(0) \leq y_0\};
  \]
  \[
  \hat{F}_1(y_1) := \frac{1}{n_1} \sum_{i=1}^{N} R_i W_i \mathbb{1}\{Y_i(1) \leq y_1\}
  \]
The True Variance of the Estimator for the Average Treatment Effect

Denote

\[
S^2_0 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i(0) - \bar{Y}(0))^2
\]

\[
S^2_1 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i(1) - \bar{Y}(1))^2
\]

\[
S^2_{01} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i(1) - Y_i(0) - \tau_{ATE})^2
\]

Then the exact variance of \( \hat{\tau} \) is

\[
\text{Var}(\hat{\tau}) = \frac{S^2_0}{n_0} + \frac{S^2_1}{n_1} - \frac{S^2_{01}}{N}
\]
The True Variance of the Estimator for the Average Treatment Effect

An analytical form of estimator can be

$$\hat{\text{Var}}(\hat{\tau}) = \frac{\hat{S}_0^2}{n_0} + \frac{\hat{S}_1^2}{n_1} - \frac{\hat{S}_{01}^2}{N}$$

where

$$\frac{\hat{S}_j^2}{n_j} = \frac{1}{n_j-1} \sum_{i=1}^{N} R_i \mathbb{1}(W_i = j)(Y_i - \bar{Y}_j)^2$$

and

$$\frac{\hat{S}_{01}^2}{N}$$

is an estimator of the sharp lower bound for $S_{01}^2$. 

The Classical Bootstrap

- Classical Bootstrap approximates the cumulative distribution $F_{YW}$ of $(Y_i, W_i)$ by the empirical distribution

$$
\hat{F}_{YW}(w, y) = \frac{1}{n} \sum_{i=1}^{N} R_i \mathbb{1}(Y_i \leq y, W_i \leq w).
$$

Remarks

In classical bootstrap, there is purely sampling uncertainty. It impute all missing values in the population by replications.
Aim: Bootstrapping in a way that the uncertainty is solely design-based uncertainty.

Idea: Modify the way how we impute the missing values from the observed values. Note that the joint distribution of potential outcomes in population is

\[ F_{01}^{p}(y_0, y_1) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\} = C(F_{0}^{p}(y_0), F_{1}^{p}(y_1)), \]

where \( C : [0, 1]^2 \to [0, 1] \) is a non-decreasing copula function. Hence our target \( \tau \) can be written as a functional of the marginal distributions, which can be estimated from a completely randomized experiment. The resulting task is the choice of coupling \( C \).
Least Favorable Coupling for the Average Treatment Effect

**Assumption 2.1.**
The first four moments of the respective marginal distributions of $Y_i(0)$ and $Y_i(1)$ are bounded.

**Proposition 2.1. (Least Favorable Coupling for the ATE)**
Suppose that Assumption 2.1 holds. Then, given the marginal distributions $F_0, F_1$, the variance bound is uniquely attained at

$$\sigma^2(F_0, F_1) := \lim_{N} n \text{Var}_{F_{01}}(\hat{\tau})$$

where $F_{01}^{iso} := C^{iso}(F_0, F_1)$ is the joint distribution corresponding to the isotone coupling $C^{iso}(u, v) = \min(u, v)$.

**Remarks**
It attains the upper bound for the asymptotic variance.
Generating the Empirical Population

1. Let $Y_j^0, j = 1, \ldots, n_0$ denotes the ordered sample of values with $W_i = 0$, and $Y_k^1, k = 1, \ldots, n_1$ denotes the ordered sample with $W_i = 1$. 

2. Let $N_0 = \lceil n_0 N/n \rceil$ and $N_1 = N - N_0$. Define

$$M^\ell_j := \left\lfloor \frac{j}{n_0} N_\ell \right\rfloor - \left\lfloor \frac{j - 1}{n_0} N_\ell \right\rfloor, \quad \ell = 0, 1.$$ 

3. Generate the empirical population $(\tilde{Y}_i, \tilde{W}_i)_{i=1}^N$ by including $M^0_j$ copies of $Y_j^0$ with $W_j = 0$ and $M^1_j$ copies of $Y_j^1$ with $W_j = 1$. 
Imputing Missing Counterfactuals

Impute the missing counterfactuals according to

\[ \tilde{Y}_i(0) := \begin{cases} 
\tilde{Y}_i & \text{if } \tilde{W}_i = 0 \\ 
\hat{F}_0^{-1}(\hat{F}_1(\tilde{Y}_i)) & \text{otherwise} 
\end{cases} \]

\[ \tilde{Y}_i(1) := \begin{cases} 
\tilde{Y}_i & \text{if } \tilde{W}_i = 1 \\ 
\hat{F}_1^{-1}(\hat{F}_0(\tilde{Y}_i)) & \text{otherwise} 
\end{cases} \]
Resampling Algorithm

1. For $b$th bootstrap replication, draw $n$ units of $(Y_{ib}^*(0), Y_{ib}^*(1))$ from the empirical population at random and without replacement.

2. Generate $W_{1b}^*, \ldots, W_{nb}^*$ by selecting $n_1$ units from the sample without replacement and set $W_{ib}^* = 1$ for the selected units, $W_{ib}^* = 0$ otherwise. Hence we have the bootstrap sample $Y_{ib}^* = Y_{ib}^*(W_{ib}^*)$ for $i = 1, \ldots, n$.

3. Obtain the estimates and the studentized values

$$
\hat{\tau}_{b}^* = \frac{1}{n_1} \sum_{i=1}^{n} W_{ib}^* Y_{ib}^* - \frac{1}{n_0} \sum_{i=1}^{n} (1 - W_{ib}^*) Y_{ib}^*; \\
\hat{\sigma}_{b}^* = \sigma(\hat{F}_{0b}^*, \hat{F}_{1b}^*);
$$

$$
T_{b}^* = \sqrt{n} \frac{\hat{\tau}_{b}^* - \hat{\tau}}{\hat{\sigma}_{b}^*}
$$
Bootstrap Algorithm

1. Create an empirical population \((\tilde{Y}_i, \tilde{W}_i)_{i=1}^N\) by selecting \(M^0_j\) copies of \(Y_j\) with \(W_j = 0\) and \(M^1_j\) copies of \(Y^-1_j\) with \(W_j = 1\).

2. Impute potential values \(\tilde{Y}_i(0), \tilde{Y}_i(1)\) for each \(i = 1, \ldots, N\) where \(\tilde{Y}_i(W_i) = \tilde{Y}_i\) and \(\tilde{Y}_i(1 - W_i)\) is obtained.

3. Simulate the randomized distribution by repeatedly drawing \(n\) units of \(Y^*_i(0)\) and \(Y^*_i(1)\) out of that empirical population without replacement and generating randomization draws \(W^*_1, \ldots, W^*_n\) by setting \(W^*_{ib} = 1\) for \(n_1\) units sampled from \(\{1, \ldots, n\}\) without replacement, and \(W^*_{ib} = 0\) for the remaining \(n - n_1\) units. We then set \(Y^*_{ib} := Y^*_i(W^*_{ib})\).

4. Given \((Y^*_{ib}, W^*_{ib})\), compute bootstrap version of the statistic \(T^*_b\).
The proposed confidence intervals for $\tau$ is

$$\hat{C}_{1-\alpha} := \left[ \hat{\tau} - \hat{\sigma} \hat{c}_{1-\alpha} / \sqrt{n}, \hat{\tau} - \hat{\sigma} \hat{c}_\alpha / \sqrt{n} \right],$$

where $\hat{c}_p$ is the $p$th quantile of bootstrap samples $T_1^*, \ldots, T_B^*$. 
Basic designs

I. No treatment effect, same marginal distribution
   - $Y_i(0) = Y_i(1) \sim N(0, 1)$ and $n_0 = n_1 = 100$.
   - all procedures are expected to do well.

II. Random treatment effect, different marginal distribution
   - $Y_i(0) \sim N(0, 1)$, $Y_i(1) = 0$ and $n_0 = n_1 = 100$.
   - causal standard errors and causal bootstrap should do better.

III. Design II with smaller sample
    - $Y_i(0) \sim N(0, 1)$, $Y_i(1) = 0$ and $n_0 = n_1 = 20$.

IV. Heterogeneous treatment effect, non-Gaussian distribution
    - $Y_i(0) = AZ + (1 - A)4Z$ where $A \sim Bern(0.9)$ and $Z \sim N(0, 1)$.
    - $Y_i(1) = 0$ and $n_0 = n_1 = 20$.
    - this highlights the difference between the bootstrap and Gaussian inference.

Note
The average treatment effects in the simulations are all zero.
### Basic results

**Figure 1:** Adapted from Table 3 in the paper.

<table>
<thead>
<tr>
<th>Variance Estimator</th>
<th>Critical Values</th>
<th>Pivotal Statistic</th>
<th>Design I Cov Rate</th>
<th>Med s.e.</th>
<th>Design II Cov Rate</th>
<th>Med s.e.</th>
<th>Design III Cov Rate</th>
<th>Med s.e.</th>
<th>Design IV Cov Rate</th>
<th>Med s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{\text{Neyman}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9536</td>
<td>0.1412</td>
<td>0.9950</td>
<td>0.0999</td>
<td>0.9870</td>
<td>0.2218</td>
<td>0.9776</td>
<td>0.3330</td>
</tr>
<tr>
<td>$\hat{V}_{\text{AGL}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9528</td>
<td>0.1404</td>
<td>0.9524</td>
<td>0.0706</td>
<td>0.9334</td>
<td>0.1568</td>
<td>0.9116</td>
<td>0.2354</td>
</tr>
<tr>
<td>$\hat{V}_{s\text{-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9518</td>
<td>0.1405</td>
<td>0.9944</td>
<td>0.0994</td>
<td>0.9850</td>
<td>0.2162</td>
<td>0.9744</td>
<td>0.3245</td>
</tr>
<tr>
<td>$\hat{V}_{c\text{-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9512</td>
<td>0.1400</td>
<td>0.9494</td>
<td>0.0704</td>
<td>0.9302</td>
<td>0.1548</td>
<td>0.9084</td>
<td>0.2325</td>
</tr>
<tr>
<td>N/A</td>
<td>Fisher Exact</td>
<td>No</td>
<td>0.9766</td>
<td>0.1411</td>
<td>0.9630</td>
<td>0.0999</td>
<td>0.9626</td>
<td>0.2219</td>
<td>0.9698</td>
<td>0.3332</td>
</tr>
<tr>
<td>$\hat{V}_{\text{Neyman}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9534</td>
<td>0.1421</td>
<td>0.9954</td>
<td>0.1012</td>
<td>0.9900</td>
<td>0.2404</td>
<td>0.9838</td>
<td>0.3865</td>
</tr>
<tr>
<td>$\hat{V}_{\text{AGL}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9528</td>
<td>0.1433</td>
<td>0.9954</td>
<td>0.1012</td>
<td>0.9900</td>
<td>0.2404</td>
<td>0.9838</td>
<td>0.3865</td>
</tr>
<tr>
<td>$\hat{V}_{\text{Neyman}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9526</td>
<td>0.1414</td>
<td>0.9510</td>
<td>0.0715</td>
<td>0.9446</td>
<td>0.1681</td>
<td>0.9434</td>
<td>0.2802</td>
</tr>
<tr>
<td>$\hat{V}_{\text{AGL}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9530</td>
<td>0.1419</td>
<td>0.9510</td>
<td>0.0715</td>
<td>0.9446</td>
<td>0.1681</td>
<td>0.9434</td>
<td>0.2802</td>
</tr>
</tbody>
</table>

| Target | 0.9500 | 0.1414 | 0.9500 | 0.0707 | 0.9500 | 0.1581 | 0.9500 | 0.2500 |

**Note**

The number of replications used in the bootstrap is not stated.
Basic results

Figure 2: Adapted from Table 4 in the paper. $Y_i(1)$’s are shifted by $\tau_n = 2/\sqrt{n}$ to investigate local power. This was NOT included in the arXiv version (I got trapped).

<table>
<thead>
<tr>
<th>Variance Estimator</th>
<th>Critical Values</th>
<th>Pivotal Statistic</th>
<th>Design I n=40</th>
<th>Design I n=100</th>
<th>Design I n=200</th>
<th>Design I n=400</th>
<th>Design II n=40</th>
<th>Design II n=100</th>
<th>Design II n=200</th>
<th>Design II n=400</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\nu}_{\text{Neyman}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.1828</td>
<td>0.1676</td>
<td>0.1748</td>
<td>0.1656</td>
<td>0.2476</td>
<td>0.2382</td>
<td>0.2282</td>
<td>0.2314</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{AGL}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.1934</td>
<td>0.1750</td>
<td>0.1766</td>
<td>0.1676</td>
<td>0.5346</td>
<td>0.5226</td>
<td>0.5192</td>
<td>0.5238</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{s-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.1966</td>
<td>0.1742</td>
<td>0.1770</td>
<td>0.1672</td>
<td>0.2720</td>
<td>0.2464</td>
<td>0.2312</td>
<td>0.2332</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{c-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.1942</td>
<td>0.1790</td>
<td>0.1776</td>
<td>0.1682</td>
<td>0.5442</td>
<td>0.5264</td>
<td>0.5210</td>
<td>0.5224</td>
</tr>
<tr>
<td>N/A</td>
<td>Fisher Exact</td>
<td>No</td>
<td>0.0274</td>
<td>0.0714</td>
<td>0.1018</td>
<td>0.1200</td>
<td>0.1852</td>
<td>0.2230</td>
<td>0.2408</td>
<td>0.2596</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{Neyman}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.1650</td>
<td>0.1630</td>
<td>0.1712</td>
<td>0.1668</td>
<td>0.2030</td>
<td>0.2114</td>
<td>0.2164</td>
<td>0.2190</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{AGL}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.1760</td>
<td>0.1662</td>
<td>0.1720</td>
<td>0.1672</td>
<td>0.4886</td>
<td>0.2642</td>
<td>0.2394</td>
<td>0.2314</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{Neyman}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.1714</td>
<td>0.1644</td>
<td>0.1712</td>
<td>0.1638</td>
<td>0.4800</td>
<td>0.4940</td>
<td>0.4990</td>
<td>0.5146</td>
</tr>
<tr>
<td>$\hat{\nu}_{\text{AGL}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.1652</td>
<td>0.1624</td>
<td>0.1698</td>
<td>0.1628</td>
<td>0.4800</td>
<td>0.4944</td>
<td>0.4992</td>
<td>0.5146</td>
</tr>
</tbody>
</table>
Coupled designs

V. Heterogeneous treatment effect, bivariate Gaussian distribution
   ▸ \( \text{Var}\{Y_i(0)\} = 0.5 \) and \( \text{Var}\{Y_i(1)\} = 2 \).
   ▸ different correlation coefficients \( \varrho_{01} \) and sample sizes \((n_0, n_1)\).
   ▸ expect asymptotically exact coverage under isotonic coupling \( \varrho_{01} = 1 \).
   ▸ expect conservative coverage if \( \varrho_{01} < 1 \).
   ▸ expect Fisher’s exact procedure to underestimate the spread of the randomization distribution.
   ▸ should not expect refinements for the bootstrap relative to Gaussian inference.

VI. Heterogeneous treatment effect, bivariate non-Gaussian distribution
   ▸ \( Y_i(0) = 0 \) (note that the potential outcomes differ from design IV).
   ▸ \( Y_i(1) = AZ + (1 - A)4Z \) where \( A \sim \text{Bern}(0.9) \) and \( Z \sim \text{N}(0, 1) \).
   ▸ expect refinements for the bootstrap relative to Gaussian inference.
**Coupled results**

---

**Coverage of nominal 95% Confidence Intervals, Gaussian Potential Outcomes with Different Couplings**

<table>
<thead>
<tr>
<th>Variance Estimator</th>
<th>Critical Values</th>
<th>Pivotal Statistic</th>
<th>$\varphi_0 = 1$</th>
<th>$(n_0, n_1) = (50, 20)$</th>
<th>$\varphi_0 = 0$</th>
<th>$\varphi_0 = -1$</th>
<th>minimum</th>
<th>$\varphi_0 = 1$</th>
<th>$(n_0, n_1) = (200, 80)$</th>
<th>$\varphi_0 = 0$</th>
<th>$\varphi_0 = -1$</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{V}_{\text{Neyman}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9542</td>
<td>0.9640</td>
<td>0.9830</td>
<td>0.9542</td>
<td></td>
<td>0.9662</td>
<td>0.9810</td>
<td>0.9902</td>
<td>0.9662</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{AGL}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9376</td>
<td>0.9490</td>
<td>0.9740</td>
<td>0.9376</td>
<td></td>
<td>0.9498</td>
<td>0.9700</td>
<td>0.9826</td>
<td>0.9498</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{c-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9486</td>
<td>0.9600</td>
<td>0.9806</td>
<td>0.9486</td>
<td></td>
<td>0.9650</td>
<td>0.9802</td>
<td>0.9900</td>
<td>0.9650</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{c-boot}}$</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9340</td>
<td>0.9450</td>
<td>0.9712</td>
<td>0.9340</td>
<td></td>
<td>0.9482</td>
<td>0.9690</td>
<td>0.9814</td>
<td>0.9482</td>
<td></td>
</tr>
<tr>
<td>$N/A$</td>
<td>Fisher Exact</td>
<td>No</td>
<td>0.9306</td>
<td>0.9058</td>
<td>0.8920</td>
<td>0.8920</td>
<td></td>
<td>0.8742</td>
<td>0.8620</td>
<td>0.8466</td>
<td>0.8466</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{AGL}}$</td>
<td>Fisher Exact</td>
<td>Yes</td>
<td>0.9988</td>
<td>0.9852</td>
<td>0.9542</td>
<td>0.9542</td>
<td></td>
<td>0.9964</td>
<td>0.9776</td>
<td>0.9472</td>
<td>0.9472</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{Neyman}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9646</td>
<td>0.9716</td>
<td>0.9864</td>
<td>0.9646</td>
<td></td>
<td>0.9668</td>
<td>0.9824</td>
<td>0.9906</td>
<td>0.9668</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{AGL}}$</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9632</td>
<td>0.9718</td>
<td>0.9870</td>
<td>0.9632</td>
<td></td>
<td>0.9666</td>
<td>0.9820</td>
<td>0.9910</td>
<td>0.9666</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{Neyman}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9498</td>
<td>0.9584</td>
<td>0.9804</td>
<td>0.9498</td>
<td></td>
<td>0.9520</td>
<td>0.9712</td>
<td>0.9820</td>
<td>0.9520</td>
<td></td>
</tr>
<tr>
<td>$\bar{V}_{\text{AGL}}$</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9488</td>
<td>0.9584</td>
<td>0.9804</td>
<td>0.9488</td>
<td></td>
<td>0.9520</td>
<td>0.9712</td>
<td>0.9816</td>
<td>0.9520</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 3:** Adapted from Table 5 in the paper.
Coupled results

| Variance Estimator | Critical Values | Pivotal Statistic | \((n_0, n_1)\) 
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{V}_{\text{Neyman}})</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9768</td>
</tr>
<tr>
<td>(\hat{V}_{\text{AGL}})</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9186</td>
</tr>
<tr>
<td>(\hat{V}_{s \text{-boot}})</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9752</td>
</tr>
<tr>
<td>(\hat{V}_{c \text{-boot}})</td>
<td>Gaussian</td>
<td>No</td>
<td>0.9144</td>
</tr>
<tr>
<td>N/A</td>
<td>Fisher Exact</td>
<td>No</td>
<td>0.9752</td>
</tr>
<tr>
<td>(\hat{V}_{\text{Neyman}})</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9870</td>
</tr>
<tr>
<td>(\hat{V}_{\text{AGL}})</td>
<td>Std. Bootstrap</td>
<td>Yes</td>
<td>0.9870</td>
</tr>
<tr>
<td>(\hat{V}_{\text{Neyman}})</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9470</td>
</tr>
<tr>
<td>(\hat{V}_{\text{AGL}})</td>
<td>Causal Bootstrap</td>
<td>Yes</td>
<td>0.9470</td>
</tr>
</tbody>
</table>

**Figure 4:** Adapted from Table 6 in the paper.
Two-stage scheme of sampling

The causal bootstrap’s setting can be seen as a two-stage scheme of sampling without replacement from nested finite populations:

1. Draw \( n \) units without replacement from the population of \( N \) units.
2. Draw \( n_1 \) units at random and without replacement to receive the treatment \( W_i = 1 \).
   - the remaining \( n_0 = n - n_1 \) units are assigned \( W_i = 0 \).
   - step 2 is conditionally independent of step 1.

This view allows us to characterize the asymptotic properties of the causal bootstrap.

- An asymptotic Donsker Theorem for empirical processes based on sampling without replacement from a finite population is available from Bickel (1969).
Sampling uncertainty

Define the (joint) distributions of the functional:

\[
F_{01}^p(y_0, y_1) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\},
\]

\[
F_{01}^s(y_0, y_1) := \frac{1}{n} \sum_{i=1}^{n} R_i \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\},
\]

and similarly for the marginals \(F_0^p, F_1^p, F_0^s, F_1^s\).

The sampling uncertainty can be characterized by:

\[
F_{01}^s(y_0, y_1) - F_{01}^p(y_0, y_1) = \frac{1}{n} \sum_{i=1}^{N} (R_i - q) \mathbb{1}\{Y_i(0) \leq y_0, Y_i(1) \leq y_1\}, \tag{1}
\]

where \(q = n/N\).
Design uncertainty

Define

\[
\hat{F}_0(y_0) := \frac{1}{n(1 - p)} \sum_{i=1}^{N} R_i (1 - W_i) \mathbb{1}\{Y_i(0) \leq y_0\},
\]

\[
\hat{F}_1(y_1) := \frac{1}{np} \sum_{i=1}^{N} R_i W_i \mathbb{1}\{Y_i(1) \leq y_1\},
\]

where \( p = n_1/n \); see Section 1.2.

The design uncertainty can be characterized by:

\[
\left( \hat{F}_0(y_0) - F_0^s(y_0) \right) - \left( \hat{F}_1(y_1) - F_1^s(y_1) \right) = \frac{1}{np} \sum_{i=1}^{N} R_i (W_i - p) \left( -p \mathbb{1}\{Y_i(0) \leq y_0\}/(1 - p) \right) \mathbb{1}\{Y_i(1) \leq y_1\}.
\]

(2)

Note

The definitions of \( \hat{F}_0(y_0) \) and \( \hat{F}_1(y_1) \) are different in Sections 1.2 and 5 (probably typos).
Consistency and randomization CLT

**Consistency**

Under regularity conditions, \( \hat{\tau} \) and \( \hat{\sigma} \) are consistent for \( \tau(F^p_0, F^p_1) \) and \( \sigma(F^p_0, F^p_1) \), respectively.

The proof is based on the Glivenko-Cantelli theorem and continuous mapping theorem.

**Randomization CLT**

Under regularity conditions,

\[
\sqrt{n} \frac{\hat{\tau} - \tau}{\hat{\sigma}} \xrightarrow{d} N \left( 0, \frac{\sigma^2(F^p_{01})}{\sigma^2(F^p_0, F^p_1)} \right),
\]

where \( \sigma^2(F_{01}) := \lim_{n \to \infty} n \text{Var}_{F_{01}}(\hat{\tau}) \)

The proof is based on Bickel (1969), the functional Delta method and Slutsky's theorem.
The proof is similar to that of the randomization CLT.

The bootstrap CLT and randomization CLT together

- show that the bootstrap algorithm in Section 3 converges to a “least-favorable” limiting experiment in an appropriate sense because the asymptotic variances are 1 and less than 1 by construction.
- apply to any other functional that satisfy the regularity conditions.
- achieves refinements with the $t$-ratio (self-normalization) under slightly stronger conditions.

The CIs are asymptotically valid by replacing the unidentified randomization variance with an estimate of the bound; see Corollary 5.1.

\[ \sqrt{n} \frac{\hat{T}^* - \hat{T}}{\hat{\sigma}^*} \to N(0, 1). \]
Assumptions

When treatment is not completely randomized

- Observable attributes: \( X_i \).
- Indicator if unit \( i \) is included in the sample sample: \( R_i \).
  - a random sample of size \( n \leq N \) is observed (superpopulation model).

**Unconfoundedness/ignorability**

Treatment assignment is independent across units \( i = 1, \ldots, n \) and strongly ignorable given \( X_i \), i.e., \( \{Y_i(0), Y_i(1)\} \perp \perp W_i \mid \{X_i, R_i\} \).

Under unconfoundedness and independent assignment, the assignment mechanism for a binary treatment is fully described by the propensity score

\[
e(x) := \Pr(W_i = 1 \mid X_i = x).
\]

This paper focuses on the case that \( e(x) \) is known, but it is possible to extend when \( e(x) \) has to be estimated.
Assumptions

Overlap

The propensity score satisfies $0 < \underline{e} \leq e(x) \leq \bar{e} < 1$ for all values of $x$ in the support of $X_i$.

Note

Clearly, overlap assumption is violated if the events are rare.

Given these assumption, a natural estimator for $\tau_{ATE}$ is

$$\hat{\tau}_{ATE} := \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{W_i Y_i}{e(X_i)} - \frac{(1 - W_i)Y_i}{1 - e(X_i)} \right\}.$$
Assumptions

Superpopulation

For each unit \( i = 1, \ldots, N \) in the population, attributes \( X_i \) are i.i.d. draws from the distribution \( F_X(x) \), and potential values of \( Y_i(0), Y_i(1) \) are independent draws from the distribution \( F_{01}(y_0, y_1 \mid x) \). \( F_X \) and \( F_{01} \) have bounded p.d.f.s \( f_X(x) \) and \( f_{01}(y_0, y_1 \mid x) \), respectively, that are twice continuously differentiable in the continuously distributed components of \( x \).

Note

"\( Y_i(0), Y_i(0) \) are independent..." is probably a typo.

This assumption is necessary as

- the quality of asymptotic approximations depends on properties of that underlying meta-population; and
- it permits consistent estimation of conditional distributions.
Balancing Score

Nonparametric estimation may suffer from a curse of dimensionality in the number of attributes. This paper therefore considers

\[
\hat{F}_{0n}(y_0, b) := \frac{1}{n} \sum_{i=1}^{n} \left( \frac{(1 - W_i) \mathbb{1}\{Y_i \leq y_0, b(X_i) \leq b\}}{1 - e(X_i)} \right),
\]

\[
\hat{F}_{1n}(y_1, b) := \frac{1}{n} \sum_{i=1}^{n} \frac{W_i \mathbb{1}\{Y_i \leq y_0, b(X_i) \leq b\}}{e(X_i)},
\]

where \(b(x)\) is a balancing score, i.e., a basis for incorporating the attributes.

Note

Balancing score is a standard tool in propensity score matching; see Wikipedia.
Bootstrap Algorithm

The algorithm when treatment is not completely randomized is

1. **Impute missing counterfactuals.**
   - i. obtain the empirical conditional rank by \( \hat{V}_i := \hat{F}_{W_i}(Y_i \mid b(X_i)) \).
   - ii. impute the values by \( \hat{Y}_{W_i} := Y_i \) and \( \hat{Y}_{(1-W_i)} := \hat{F}_{(1-W_i)}^{-1}(\hat{V}_i \mid b(X_i)) \).
     - note that this coupling preserves the estimated conditional distribution possibly up to a discretization error.

2. **Estimate the randomization distribution of \( \hat{\tau}_{ATE} \).**
   - i. for the \( b \)-th bootstrap sample, draw \( W_{1b}, \ldots, W_{nb} \sim \text{Bern}(e(X_i)) \) independently.
   - ii. Compute the treatment contrast
     \[
     \hat{\tau}_b^* := \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{W_{ib} \hat{Y}_{1i}}{e(X_i)} - \frac{(1 - W_{ib}) \hat{Y}_{0i}}{1 - e(X_i)} \right\}.
     \]
   - iii. for \( B \) independent replications, use the empirical distribution of \( \hat{\tau}_1^*, \ldots, \hat{\tau}_B^* \) as the bootstrap estimator for the randomization distribution of \( \hat{\tau} \).
Conclusion

- Statistical error may come from different sources.
  - causal inference need to deal with sampling and design uncertainty.
- Causal inference should be based on conservative estimation.
  - the joint distribution of potential values is fundamentally underidentified.
- Causal bootstrap
  - base on least favorable randomization distribution.
  - able to handle both sampling and design uncertainty.