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.,

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

- 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 arised from $R_1, \ldots, R_N$.
  - captured by conventional standard error.

TABLE 1
SAMPLING-BASED UNCERTAINTY (? IS MISSING)

Unit	Actual Sample			Alternative Sample I			Alternative Sample II			...
	Y_i	W_i	R_i	Y_i	W_i	R_i	Y_i	W_i	R_i	...
1	9.1	0	1	?	?	0	?	?	0	...
2	?	?	0	?	?	0	−1.6	1	1	...
3	2.3	0	1	1.9	0	1	2.3	0	1	...
4	−3.6	1	1	−3.6	1	1	?	?	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
N	?	?	0	−5.7	1	1	?	?	0	...

Source of Randomness

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

TABLE 2
DESIGN-BASED UNCERTAINTY (? IS MISSING)

Unit	Actual Sample			Alternative Sample I			Alternative Sample II			...
	$Y_i(1)$	$Y_i(0)$	W_i	$Y_i(1)$	$Y_i(0)$	W_i	$Y_i(1)$	$Y_i(0)$	W_i	...
1	3.2	?	1	3.2	?	1	?	9.1	0	...
2	−1.6	?	1	−1.6	?	1	?	7.1	0	...
3	?	2.3	0	0.5	?	1	0.5	?	1	...
4	?	−3.1	0	?	−3.1	0	-3.6	?	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...
N	−5.7	?	1	?	2.8	0	-5.7	?	1	...

Notation

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

$$\widehat{F}_0(y_0) := \frac{1}{n_0} \sum_{i=1}^{N} R_i (1 - W_i) \mathbb{1}\{Y_i(0) \le y_0\};$$

$$\widehat{F}_1(y_1) := \frac{1}{n_1} \sum_{i=1}^{N} R_i W_i \mathbb{1}\{Y_i(1) \le y_1\}$$

The True Variance of the Estimator for the Average Treatment Effect

Denote

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

$$S_1^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(1) - \bar{Y}(1))^2$$

$$S_{01}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i(1) - Y_i(0) - \tau_{ATE})^2$$

Then the exact variance of $\widehat{\tau}$ is

$$\mathrm{Var}(\widehat{\tau}) = \frac{S_0^2}{n_0} + \frac{S_1^2}{n_1} - \frac{S_{01}^2}{N}$$

The True Variance of the Estimator for the Average Treatment Effect

- An analytical form of estimator can be

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

where $\frac{\widehat{S}_j^2}{n_j} = \frac{1}{n_j - 1} \sum_{i=1}^N R_i \mathbb{1}(W_i = j)(Y_i - \bar{Y}_j)^2$ and $\frac{\widehat{S}_{01}^2}{N}$ 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

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

Remarks

In classical bootstrap, there is purely sampling uncertainty. It impute all missing values in the population by replications.

The Causal Bootstrap

- 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

$$F_{01}^{p}(y_0, y_1) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{Y_i(0) \le y_0, Y_i(1) \le y_1\} = C(F_0^p(y_0), F_1^p(y_1)),$$

  where $C : [0,1]^2 \mapsto [0,1]$ is a non-decreasing copula function. Hence our target τ 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

$$\sigma^2(F_0, F_1) := \lim_N n \mathrm{Var}_{F_{01}^{iso}}(\hat{\tau})$$

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

$$M_j^\ell := \left\lceil \frac{j}{n_0} N_\ell \right\rceil - \left\lceil \frac{j-1}{n_0} N_\ell \right\rceil, \qquad \ell = 0, 1.$$

3. Generate the empirical population $(\tilde{Y}_i, \tilde{W}_i)_{i=1}^N$ by including M_j^0 copies of Y_j^0 with $W_j = 0$ and M_j^1 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 bth bootstrap replication, draw n units of $(Y_{ib}^*(0), Y_{ib}^*(1))$ from the empirical population at random and without replacement.
2. Generate $W_{1b}^*, \dots, W_{nb}^*$ 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 $i = 1, \dots, n$.
3. Obtain the estimates and the studentized values

$$\widehat{\tau}_b^* = \frac{1}{n_1}\sum_{i=1}^{n} W_{ib}^* Y_{ib}^* - \frac{1}{n_0}\sum_{i=1}^{n}(1 - W_{ib}^*)Y_{ib}^*;$$

$$\widehat{\sigma}_b^* = \sigma(\hat{F}_{0b^*}, \hat{F}_{1b}^*);$$

$$T_b^* = \sqrt{n}\frac{\hat{\tau}_b^* - \hat{\tau}}{\hat{\sigma}_b^*}$$

Bootstrap Algorithm

1. Create an empirical population $(\widetilde{Y}_i, \widetilde{W}_i)_{i=1}^N$ by selecting M_j^0 copies of Y_j with $W_j = 0$ and M_j^1 copies of Y_j^1 with $W_j = 1$.
2. Impute potential values $\widetilde{Y}_i(0)$, $\widetilde{Y}_i(1)$ for each $i = 1, \dots, N$ where $\widetilde{Y}_i(W_i) = \widetilde{Y}_i$ and $\widetilde{Y}_i(1 - W_i)$ 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 $W_1^*, \dots, W_n^*$ by setting $W_{ib}^* = 1$ for n_1 units sampled from $\{1, \dots, n\}$ 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^*.

Confidence Intervals

Bootstrap Studentized CI

The proposed confidence intervals for τ is

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

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

Basic designs

1. No treatment effect, same marginal distribution
   - $Y_i(0) = Y_i(1) \sim \mathrm{N}(0,1)$ and $n_0 = n_1 = 100$.
   - all procedures are expected to do well.
2. Random treatment effect, different marginal distribution
   - $Y_i(0) \sim \mathrm{N}(0,1)$, $Y_i(1) = 0$ and $n_0 = n_1 = 100$.
   - causal standard errors and causal bootstrap should do better.
3. Design II with smaller sample
   - $Y_i(0) \sim \mathrm{N}(0,1)$, $Y_i(1) = 0$ and $n_0 = n_1 = 20$.
4. Heterogeneous treatment effect, non-Gaussian distribution
   - $Y_i(0) = AZ + (1-A)4Z$ where $A \sim \mathrm{Bern}(0.9)$ and $Z \sim \mathrm{N}(0,1)$.
   - $Y_i(1) = 0$ 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

95% Confidence Intervals And Standard Errors

Variance Estimator	Critical Values	Pivotal Statistic	Design I Cov Rate	Design I Med s.e.	Design II Cov Rate	Design II Med s.e.	Design III Cov Rate	Design III Med s.e.	Design IV Cov Rate	Design IV Med s.e.
$\hat{\mathbb{V}}_{\text{Neyman}}$	Gaussian	No	0.9536	0.1412	0.9950	0.0999	0.9870	0.2218	0.9776	0.3330
$\hat{\mathbb{V}}_{\text{AGL}}$	Gaussian	No	0.9528	0.1404	0.9524	0.0706	0.9334	0.1568	0.9116	0.2354
$\hat{\mathbb{V}}_{\text{s-boot}}$	Gaussian	No	0.9518	0.1405	0.9944	0.0994	0.9850	0.2162	0.9744	0.3245
$\hat{\mathbb{V}}_{\text{c-boot}}$	Gaussian	No	0.9512	0.1400	0.9494	0.0704	0.9302	0.1548	0.9084	0.2325
N/A	Fisher Exact	No	0.9766	0.1411	0.9630	0.0999	0.9626	0.2219	0.9698	0.3332
$\hat{\mathbb{V}}_{\text{Neyman}}$	Std. Bootstrap	Yes	0.9534	0.1421	0.9954	0.1012	0.9900	0.2404	0.9838	0.3865
$\hat{\mathbb{V}}_{\text{AGL}}$	Std. Bootstrap	Yes	0.9528	0.1433	0.9954	0.1012	0.9900	0.2404	0.9838	0.3865
$\hat{\mathbb{V}}_{\text{Neyman}}$	Causal Bootstrap	Yes	0.9526	0.1414	0.9510	0.0715	0.9446	0.1681	0.9434	0.2802
$\hat{\mathbb{V}}_{\text{AGL}}$	Causal Bootstrap	Yes	0.9530	0.1419	0.9510	0.0715	0.9446	0.1681	0.9434	0.2802
	Target		0.9500	0.1414	0.9500	0.0707	0.9500	0.1581	0.9500	0.2500

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

Note

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

Basic results

Power against local Alternatives, $\tau = 2/\sqrt{n}$.

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

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).*

Coupled designs

V. Heterogeneous treatment effect, bivariate Gaussian distribution
   - $\mathrm{Var}\{Y_i(0)\} = 0.5$ and $\mathrm{Var}\{Y_i(1)\} = 2$.
   - different correlation coefficients ϱ_{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 $A \sim \mathrm{Bern}(0.9)$ and $Z \sim \mathrm{N}(0, 1)$.
   - expect refinements for the bootstrap relative to Gaussian inference.

Coupled results

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

Variance Estimator	Critical Values	Pivotal Statistic	$(n_0, n_1) = (50, 20)$				$(n_0, n_1) = (200, 80)$			
			$\varrho_{01} = 1$	$\varrho_{01} = 0$	$\varrho_{01} = -1$	minimum	$\varrho_{01} = 1$	$\varrho_{01} = 0$	$\varrho_{01} = -1$	minimum
$\hat{\mathbb{V}}_{\text{Neyman}}$	Gaussian	No	0.9542	0.9640	0.9830	0.9542	0.9662	0.9810	0.9902	0.9662
$\hat{\mathbb{V}}_{\text{AGL}}$	Gaussian	No	0.9376	0.9490	0.9740	0.9376	0.9498	0.9700	0.9826	0.9498
$\hat{\mathbb{V}}_{\text{s-boot}}$	Gaussian	No	0.9486	0.9600	0.9806	0.9486	0.9650	0.9802	0.9900	0.9650
$\hat{\mathbb{V}}_{\text{c-boot}}$	Gaussian	No	0.9340	0.9450	0.9712	0.9340	0.9482	0.9690	0.9814	0.9482
N/A	Fisher Exact	No	0.9306	0.9058	0.8920	0.8920	0.8742	0.8620	0.8466	0.8466
$\hat{\mathbb{V}}_{\text{AGL}}$	Fisher Exact	Yes	0.9988	0.9852	0.9542	0.9542	0.9964	0.9776	0.9472	0.9472
$\hat{\mathbb{V}}_{\text{Neyman}}$	Std. Bootstrap	Yes	0.9646	0.9716	0.9864	0.9646	0.9668	0.9824	0.9906	0.9668
$\hat{\mathbb{V}}_{\text{AGL}}$	Std. Bootstrap	Yes	0.9632	0.9718	0.9870	0.9632	0.9666	0.9820	0.9910	0.9666
$\hat{\mathbb{V}}_{\text{Neyman}}$	Causal Bootstrap	Yes	0.9498	0.9584	0.9804	0.9498	0.9520	0.9712	0.9820	0.9520
$\hat{\mathbb{V}}_{\text{AGL}}$	Causal Bootstrap	Yes	0.9488	0.9584	0.9804	0.9488	0.9520	0.9712	0.9816	0.9520

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

Coupled results

Coverage of nominal 95% Confidence Intervals, non-Gaussian Potential Values with Isotone Coupling

Variance Estimator	Critical Values	Pivotal Statistic	(n_0, n_1) (20, 20)	(n_0, n_1) (50, 50)	(n_0, n_1) (100, 100)	(n_0, n_1) (200, 200)	(n_0, n_1) (500, 500)
$\hat{\mathbb{V}}_{\text{Neyman}}$	Gaussian	No	0.9768	0.9866	0.9914	0.9932	0.9924
$\hat{\mathbb{V}}_{\text{AGL}}$	Gaussian	No	0.9186	0.9358	0.9396	0.9450	0.9436
$\hat{\mathbb{V}}_{\text{s-boot}}$	Gaussian	No	0.9752	0.9864	0.9912	0.9928	0.9924
$\hat{\mathbb{V}}_{\text{c-boot}}$	Gaussian	No	0.9144	0.9336	0.9378	0.9436	0.9436
N/A	Fisher Exact	No	0.9752	0.9652	0.9672	0.9560	0.9592
$\hat{\mathbb{V}}_{\text{Neyman}}$	Std. Bootstrap	Yes	0.9870	0.9912	0.9940	0.9942	0.9934
$\hat{\mathbb{V}}_{\text{AGL}}$	Std. Bootstrap	Yes	0.9870	0.9912	0.9940	0.9942	0.9934
$\hat{\mathbb{V}}_{\text{Neyman}}$	Causal Bootstrap	Yes	0.9470	0.9532	0.9582	0.9548	0.9482
$\hat{\mathbb{V}}_{\text{AGL}}$	Causal Bootstrap	Yes	0.9470	0.9532	0.9582	0.9548	0.9482

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:

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

$$F_{01}^{s}(y_0, y_1) := \frac{1}{n}\sum_{i=1}^{N} R_i \mathbb{1}\{Y_i(0) \le y_0, Y_i(1) \le y_1\},$$

and similarly for the marginals $F_0^p, F_1^p, F_0^s, F_1^s$.

The sampling uncertainty can be characterized by:

$$F_{01}^{s}(y_0, y_1) - F_{01}^{p}(y_0, y_1) = \frac{1}{n}\sum_{i=1}^{N} (R_i - q) \mathbb{1}\{Y_i(0) \le y_0, Y_i(1) \le y_1\}, \qquad (1)$$

where $q = n/N$.

Design uncertainty

Define

$$\hat{F}_0(y_0) := \frac{1}{n(1-p)} \sum_{i=1}^{N} R_i(1-W_i)\mathbb{1}\{Y_i(0) \le y_0\},$$
$$\hat{F}_1(y_1) := \frac{1}{np} \sum_{i=1}^{N} R_i W_i \mathbb{1}\{Y_i(1) \le y_1\},$$

where $p = n_1/n$; see Section 1.2.

The design uncertainty can be characterized by:

$$\begin{pmatrix} \hat{F}_0(y_0) - F_0^s(y_0) \\ \hat{F}_1(y_1) - F_1^s(y_1) \end{pmatrix} = \frac{1}{np} \sum_{i=1}^{N} R_i(W_i - p) \begin{pmatrix} -p\mathbb{1}\{Y_i(0) \le y_0\}/(1-p) \\ \mathbb{1}\{Y_i(1) \le y_1\} \end{pmatrix}. \quad (2)$$

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 $\tau(F_0^p, F_1^p)$ and $\sigma(F_0^p, F_1^p)$, respectively.

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

Randomization CLT

Under regularity conditions,

$$\sqrt{n}\frac{\hat{\tau} - \tau}{\hat{\sigma}} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma^2(F_{01}^p)}{\sigma^2(F_0^p, F_1^p)}\right),$$

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

The proof is based on Bickel (1969), the functional Delta method and Slutsky's theorem.

Bootstrap CLT

Bootstrap CLT

Under regularity conditions,

$$\sqrt{n}\frac{\hat{\tau}^* - \hat{\tau}}{\hat{\sigma}^*} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

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.

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

$$e(x) := \mathbb{P}(W_i = 1 \mid X_i = x).$$

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 $0 < \underline{e} \leq e(x) \leq \overline{e} < 1$ 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 τ_{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

$$\hat{F}_{0n}(y_0, b) := \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - W_i)\mathbb{1}\{Y_i \le y_0, b(X_i) \le b\}}{1 - e(X_i)},$$

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

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.
   1. obtain the empirical conditional rank by $\hat{V}_i := \hat{F}_{W_i}(Y_i \mid b(X_i))$.
   2. impute the values by $\hat{Y}_{W_i i} := Y_i$ and $\hat{Y}_{(1-W_i)i} := \hat{F}^{-1}_{(1-W_i)}(\hat{V}_i \mid b(X_i))$.
      - ★ note that this coupling preserves the estimated conditional distribution possibly up to a discretization error.
2. Estimate the randomization distribution of $\hat{\tau}_{ATE}$.
   1. for the b-th bootstrap sample, draw $W^*_{1b}, \ldots, W^*_{nb} \sim \mathrm{Bern}(e(X_i))$ independently.
   2. Compute the treatment contrast

$$\hat{\tau}^*_b := \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{W^*_{ib} \hat{Y}_{1i}}{e(X_i)} - \frac{(1 - W^*_{ib}) \hat{Y}_{0i}}{1 - e(X_i)} \right\}.$$

   3. 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.