Randomized Experiments: Neyman Model and Covariate Adjustment

Ed Wu

jameswu@umich.edu

December 6, 2018

The Neyman Model

Randomized Experiments

Consider a randomized experiment with N participants. For now, we'll assume the following:

- There are two treatment groups, treatment and control
- Each participant is **independently** assigned to treatment with probability $p = 1/2$

Potential Outcomes

Associated with each participant are two potential outcomes, t_i and c_i

- Potential outcomes are **fixed** values, not random
- These potential outcomes represent the outcome the i-th participant would experience under treatment and control
- Denote the individual treatment effect as $\tau_i = t_i - c_i$ and the average treatment effect (ATE) as $\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tau_i$

Treatment Assignment

- Each participant is randomly and independently assigned to treatment or control. Let

$$T_i = \begin{cases} 1, & \text{Unit } i \text{ is assigned to treatment} \\ 0, & \text{Unit } i \text{ is assigned to control} \end{cases}$$

- If unit i is assigned to treatment, we observe t_i; otherwise, we observe c_i. In other words, the observed outcome Y_i is:

$$Y_i = T_i t_i + (1 - T_i) c_i.$$

- Let $\mathcal{T}$ and $\mathcal{C}$ denote the indices of the treatment and control units, and let n be the number of elements in $\mathcal{T}$

Did the Treatment Work?

In analyzing the results of our experiment, we want to know if the treatment "helped"

- This might be at the individual level (looking at τ_i's) or the group level (looking at the ATE, $\bar{\tau}$)
- In our case, the primary parameter of interest will be the ATE
- If we could observe t_i and c_i simultaneously, we would know the exact effect of the treatment for observation i

Did the Treatment Work?

In analyzing the results of our experiment, we want to know if the treatment "helped"

- This might be at the individual level (looking at τ_i's) or the group level (looking at the ATE, $\bar{\tau}$)
- In our case, the primary parameter of interest will be the ATE
- If we could observe t_i and c_i simultaneously, we would know the exact effect of the treatment for observation i

Unfortunately, we can only observe either t_i or c_i, but never both

The Virtues of Randomization

- Due to randomization, the only difference (in expectation) between the treatment groups is the treatment itself
- We can obtain an unbiased estimate of the ATE using the simple difference estimator:

$$\hat{\tau}_{sd} = \frac{1}{n} \sum_{i \in \mathcal{T}} Y_i - \frac{1}{N-n} \sum_{i \in \mathcal{C}} Y_i$$

What's Random?

- t_i and c_i are fixed, T_i is random
- The expected value of an estimator is therefore taken with respect to the possible treatment assignments

A Toy Example

Consider an experiment with the following $N = 3$ units:

i	t_i	c_i	τ_i
1)	10	6	4
2)	4	2	2
3)	8	8	0

If $T_1 = 1$, we observe $Y_1 = 10$. If $T_1 = 0$, we observe $Y_1 = 6$. We also see that the average treatment effect is $\frac{4+2+0}{3} = 2$.

A Toy Example

- There are $2^3 = 8$ possible treatment assignment vectors
- For each treatment assignment vector, we would get different estimate of the ATE
- The expected value and standard error of the estimator would be obtained by the average and standard deviation of these estimates
- If the expected value of an estimator is 2, then it is an unbiased estimate of the ATE

A Toy Example

- There are $2^3 = 8$ possible treatment assignment vectors
- For each treatment assignment vector, we would get different estimate of the ATE
- The expected value and standard error of the estimator would be obtained by the average and standard deviation of these estimates
- If the expected value of an estimator is 2, then it is an unbiased estimate of the ATE

Technical note: the simple difference estimator is undefined when the units are all either treatment or control, so we would really only consider the 6 remaining treatment assignment vectors

Hypothesis Testing under the Neyman Model

In reality, we estimate the ATE using the single randomization observed

- We can test whether this estimate is significant under the sharp null (H_0: $t_i = c_i$ for all i)
- The test statistic is the simple difference estimator
- The variance (conditional on n) of the difference in means is estimated as

$$\frac{s_t^2}{n} + \frac{s_c^2}{N-n}$$

  where s_t^2 and s_c^2 are the sample variances of the treatment and control groups
- This is the same procedure as an independent 2-sample t-test

Covariate Adjustment

Covariate Adjustment in Randomized Experiments

- While the treatment groups should be balanced due to randomization, it is possible there may be imbalances between the groups (*e.g.*, the control group might be older on average)
- One to way to adjust for these imbalances is to use variables that have been measured prior to treatment assignment (covariates)
- We can use the LOOP estimator to adjust for covariates
- Let $Z_i \in \mathbb{R}^q$ be a q-dimensional vector of covariates associated with unit i

Covariate Adjustment using OLS

We can make covariate adjustments using OLS

Covariate Adjustment using OLS

We can make covariate adjustments using OLS

- We regress Y onto the treatment assignment T and the covariates Z:

$$Y = \alpha + \tau T + \beta Z + \epsilon$$

- Our estimated coefficient $\hat{\tau}$ would be the OLS estimate for the average treatment effect

Covariate Adjustment using OLS

Is this an unbiased estimate for the ATE?

Covariate Adjustment using OLS

The OLS estimate is biased!

- Under the Neyman model, the standard assumptions of OLS are not satisfied
- For example, OLS assumes that the treatment effect is constant for all units and that the errors are independent of the predictors
- Under OLS, we can model the observed outcome as $Y_i = \alpha + \tau T_i + \beta Z_i + \epsilon_i$
- However, ϵ_i necessarily depends on T_i, which is the only source of randomness in the Neyman model

OLS and the Neyman Model

- The bias of OLS shrinks quickly with N
- OLS can inflate variance relative to the simple difference estimator
- Standard errors can be severely biased
- These last two issues can be ameliorated by including terms between the covariates and the treatment assignment, and using robust standard errors

Design-based Estimators

- Design-based estimators rely only on randomization as the basis for inference
- The simple difference estimator is a design-based estimator, but doesn't allow for covariate adjustments
- We will go over a design-based estimator that allows for covariate adjustment next

LOOP Estimator

Another Unbiased Estimator

- Consider the following estimate of τ_i:

$$\begin{aligned}\hat{\tau}_i &= 2Y_i T_i - 2Y(1 - T_i) \\ &= \begin{cases} 2Y_i, & T_i = 1 \\ -2Y_i, & T_i = 0 \end{cases}\end{aligned}$$

- This is an unbiased estimate of the individual treatment effect

$$\mathrm{E}(\hat{\tau}_i) = 0.5(2t_i) + 0.5(-2c_i) = t_i - c_i = \tau_i$$

and averaging across $\hat{\tau}_i$ will result in an unbiased estimate of the ATE

LOOP Estimator

- Let $m_i = \frac{1}{2}(t_i + c_i)$
- We define an estimate of τ_i as:

$$\begin{aligned}\hat{\tau}_i &= 2(Y_i - \hat{m}_i)T_i - 2(Y_i - \hat{m}_i)(1 - T_i) \\ &= \begin{cases} 2(Y_i - \hat{m}_i), & T_i = 1 \\ -2(Y_i - \hat{m}_i), & T_i = 0 \end{cases}\end{aligned}$$

  where $\hat{m}_i$ is an estimate of m_i
- If $\hat{m}_i = m_i$, then $\hat{\tau}_i$ will have zero variance
- If $\hat{m}_i$ and T_i are independent, then $\hat{\tau}_i$ is unbiased

Estimating $\hat{m}_i$

- If $\hat{m}_i = m_i$, then $\hat{\tau}_i$ will have zero variance
- If $\hat{m}_i$ and T_i are independent, then $\hat{\tau}_i$ is unbiased

Goal: calculate $\hat{m}_i$ as close to m_i as possible, while making sure $\hat{m}_i$ and T_i are independent

Leave-One-Out Estimation of $\hat{m}_i$

- We can ensure the independence of $\hat{m}_i$ and T_i by using a leave-one-out procedure
- We define the LOOP ("Leave-One-Out Potential outcomes") estimator as

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^{N} [2(Y - \hat{m}_i) T_i - 2(Y - \hat{m}_i)(1 - T_i)]$$

in which m_i is estimated using the following procedure

Estimating Potential Outcomes using a Leave-One-Out Procedure

For each i:

1. Leave observation i out
2. Fit a model (*e.g.*, OLS, lasso, random forest, etc.) to the remaining $N-1$ observations
   (For example, regress Y on T and Z using the other $N-1$ observations)
3. Calculate $\hat{t}_i$ by using the fitted model, plugging in Z_i for the covariates and $T_i = 1$ (calculate $\hat{c}_i$ using $T_i = 0$)
4. Set $\hat{m}_i = \frac{1}{2}(\hat{t}_i + \hat{c}_i)$

By leaving out observation i when we fit the model, we ensure that $\hat{m}_i$ and T_i are independent

LOOP with No Covariates

- Suppose we estimate m_i without making use of covariates
- We estimate t_i as the mean of the observed outcomes in the treatment group (excluding observation i):

$$\hat{t}_i = \frac{\sum_{k \in \mathcal{T} \setminus i} Y_k}{n - T_i}$$

- Similarly, to estimate c_i, we have:

$$\hat{c}_i = \frac{\sum_{k \in \mathcal{C} \setminus i} Y_k}{(N - n) - (1 - T_i)}$$

LOOP with No Covariates

- In this case, the LOOP estimator is exactly equal to the simple difference estimator
- We can improve on the simple difference estimator by improving our estimate of m_i: we have some assurance that our adjustment won't hurt performance
- We therefore wish to use a method that will improve over mean imputation

Decision Trees

- Decision trees are a prediction method that can be used either for classification (categorical response) or regression (continuous response)
- Given a response Y and a set of predictors $X_1, \ldots, X_p$, we successively split the data based on the values of the predictors until we reach some stopping criterion
- Each time we split the data, we choose predictor and boundary that gives the "best" split among the candidate predictors. For example, we might pick the split that gives the greatest reduction in variance among the resulting nodes
- The prediction for a new observation is determined by sending the observation through the constructed tree, and taking the average of the Y values for the "leaf" that it ends up at

Bootstrap Aggregation

- Individual decision trees have low bias, but are high variance
- As an extreme example, we could keep splitting our data until each leaf contains exactly one observation

Bootstrap Aggregation

One way to address this issue is by "bagging" (bootstrap aggregating). For a training set of size K:

- Create each of n bootstrap samples by:
  - Sampling $k \leq K$ observations with replacement from the training set
  - Constructing a prediction model (for example, a decision tree) from these k observations
- Given a new observation, average the n predictions obtained from the n models constructed

By bagging, we can average across many trees to reduce the variance of our resulting model

Random Forests

- A random forest is an ensemble method that is essentially bagged decision trees
- When constructing the trees, each split is determined on a random subset of the candidate predictors

Random Forests

Suppose we have a training set of size K with P predictors. We construct n trees (by default $n = 500$). Each tree is built as follows:

1. Sample $k \leq K$ observations (by default $k = K$) with replacement from the training set
2. Construct a decision tree from these k observations
3. At each split, we pick $m \leq P$ variables at random and pick the best split from these m variables

The primary tuning parameters in the random forest are n and m

Out-of-Bag Error

Random forests can take advantage of essentially built-in cross validation in the form of OOB ("out-of-bag") error:

- For any given tree, some number of observations will not be sampled (they will be out-of-bag)
- For each observation i, we can obtain the out-of-bag prediction by using the trees for which i was not sampled

Advantages of Random Forests

Random forests have several other advantages

- Performance is usually very good for both regression and classification
- Model selection is generally not necessary – the algorithm will use the information from the useful variables, while ignoring the irrelevant variables

LOOP with Random Forests

Because of the advantages listed above, we propose the use of random forests for LOOP:

1. Improved precision: random forests have strong performance
2. Automatic variable selection: it is not necessary to select the covariates in advance
3. Computationally efficiency: using out-of-bag predictions means we can fit a single random forest instead of N random forests

Questions?

Results

Simulation 1: Bias of OLS

We generate a single set of covariates and potential outcomes as follows:

- $N = 30$
- A single covariate Z, that takes values 0, 1, or 2
- For each value of Z, there are 10 subjects, generated from a normal distribution with standard deviation 0.1 and the following mean:
  - $Z = 0$: $c = 0, t = 1$
  - $Z = 1$: $c = 1, t = 1$
  - $Z = 2$: $c = 1, t = 2$

Simulation 1: Bias of OLS

Below, we compare the performance of LOOP, Cross Estimation, and OLS. We generate 100,000 treatment assignment vectors to estimate the bias and true standard error of each estimator. We also provide a mean nominal standard error:

Table: Simulation 1 Results: LOOP, Cross Estimation, and OLS

Method	**Bias Estimate**	**Mean Nominal SE**	**Est. of True SE**
LOOP	0.00006	0.0442	0.0384
Cross Estimation	0.00067	0.1060	0.0373
OLS	-0.01415	0.1076	0.0440

Simulation 2: Binary Response

Consider a randomized experiment in which the response is either 0 or 1

- Each of N subjects has one of the following set of potential outcomes (a) $c = 0, t = 0$, (b) $c = 0, t = 1$, or (c) $c = 1, t = 1$
- There is a single informative covariate $Z_1 \sim \mathrm{N}(0, 1)$. Higher values of Z_1 increase the probability that a participant falls in group (b) or (c):
  - Fix the predictive power $c > 0$
  - Let $w_{i1} = 1$, $w_{i2} = \exp(0.5c \times Z_{i1})$, and $w_{i3} = \exp(c \times Z_{i1})$
  - The probability that observation i is assigned to group j is $p_{ij} = w_{ij}/(w_{i1} + w_{i2} + w_{i3})$
- There are k noise covariates

Simulation 2: Binary Response

We perform three different simulations:

1. Fix $N = 200$ and the predictive power $c = 3$; vary the number of noise covariates from $k = 5$ to $k = 100$ in increments of 5
2. Fix $c = 3$ and $k = 50$; vary N from 100 to 1000 in increments of 50
3. Fix $N = 200$ and $k = 50$; vary c from 1 to 5.5 in increments of 0.5

Simulation 2: Binary Response

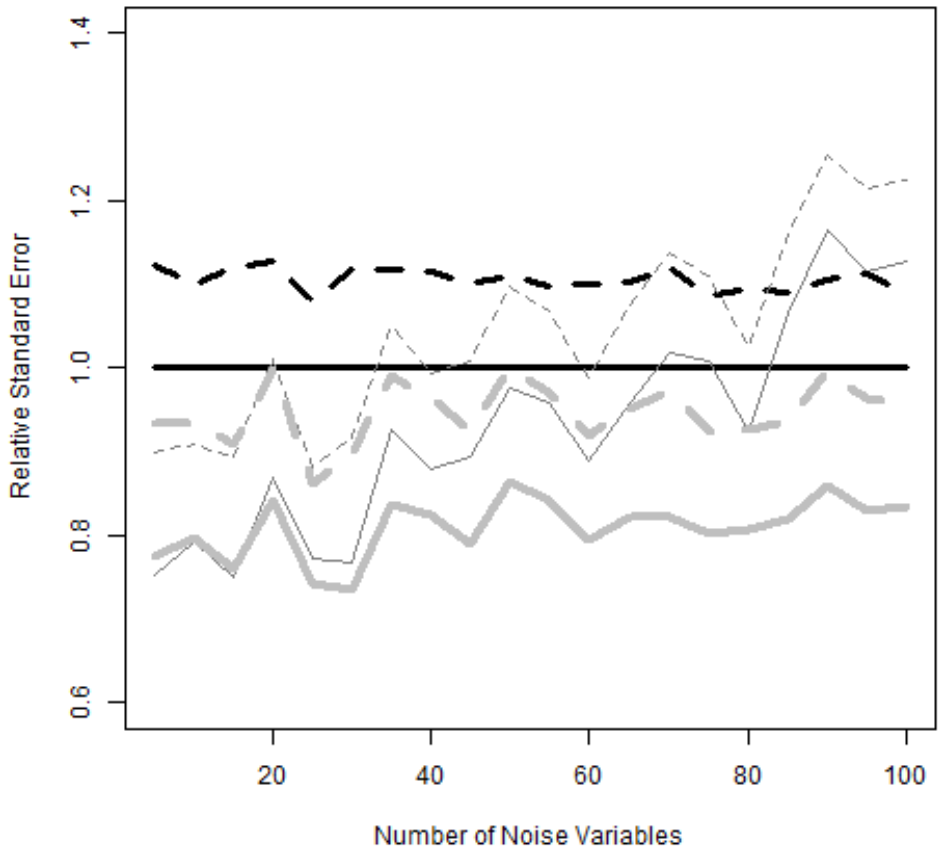

Figure: Solid lines: estimate of the true standard error; dotted lines: nominal standard error. Black: simple difference; thin gray: OLS; bold gray: LOOP

Simulation 2: Binary Response

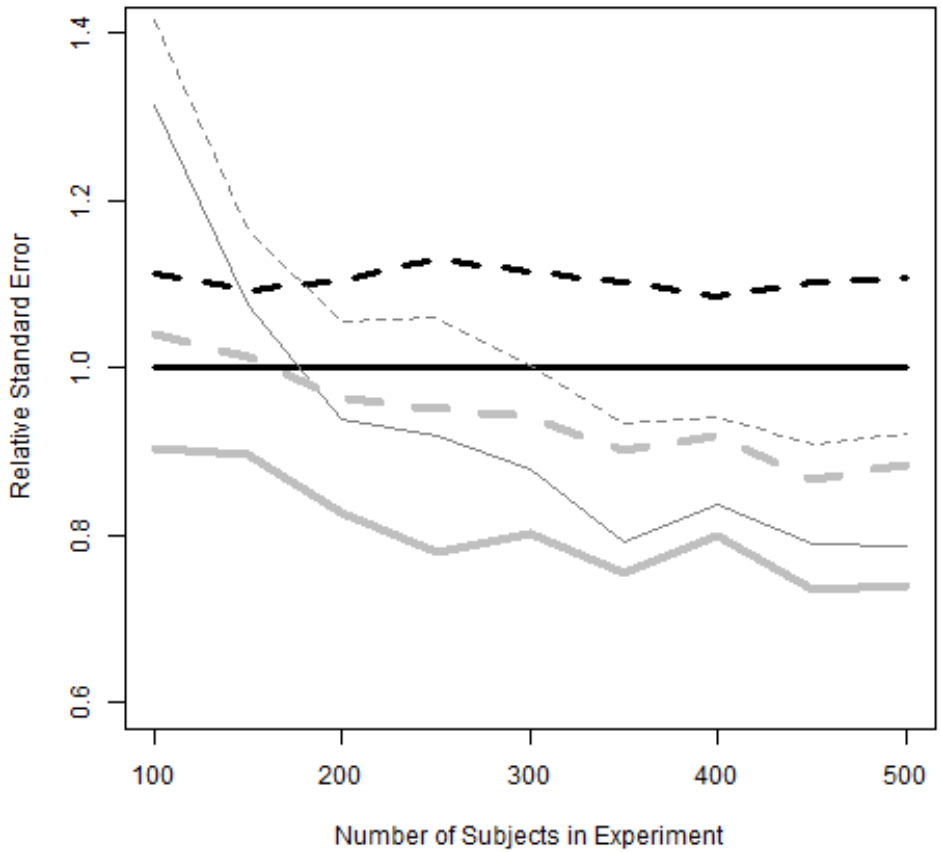

Figure: Solid lines: estimate of the true standard error; dotted lines: nominal standard error. Black: simple difference; thin gray: OLS; bold gray: LOOP

Simulation 2: Binary Response

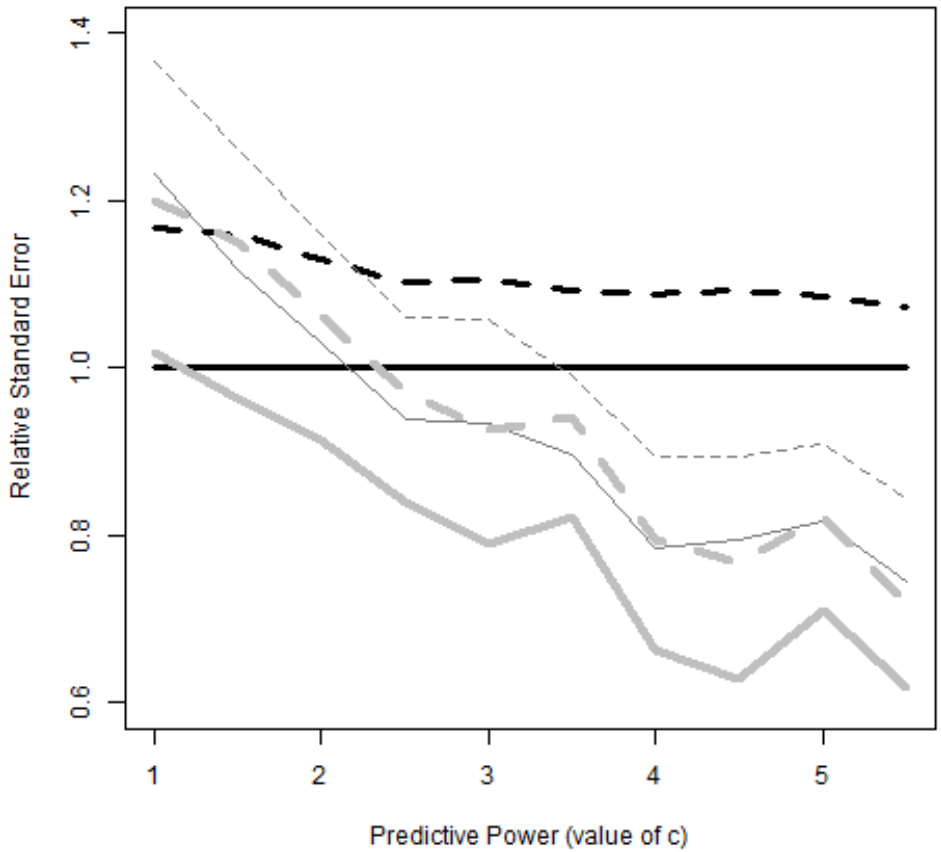

Figure: Solid lines: estimate of the true standard error; dotted lines: nominal standard error. Black: simple difference; thin gray: OLS; bold gray: LOOP

Cash Transfer Programs and Enrollment

Experiment studying the effects of cash transfer programs on educational outcomes in Bogota, Colombia. There were 2 treatments (as well as a control group):

1. Basic: Students receive a bi-monthly payment of roughly 15 USD if they attend school at least 80% of days in each month
2. Savings: Students receive a bi-monthly payment of roughly 10 USD if they attend school at least 80% of days in each month. The remaining third is held in a bank account and paid to the students' families when it is time to re-enroll for the next year

Cash Transfer Programs and Enrollment

- 3,427 students assigned to basic treatment; 3,424 to savings treatment; remaining 4,056 to control
- Re-enrollment status was the outcome variable collected by the researchers (available for roughly 90% of the students)
- We also considered the status of a missing re-enrollment as an outcome

Cash Transfer Programs and Enrollment

Table: Effect of Treatment on Missing Status and Re-enrollment Status

Treatments	Method	Missing Status Est. ($\times 10^{-2}$)	Missing Status Var. ($\times 10^{-5}$)	Re-enrollment Status Est. ($\times 10^{-2}$)	Re-enrollment Status Var. ($\times 10^{-5}$)
Basic vs. Savings	LOOP	-0.11	3.6	-2.56	1.4
	Simple Difference	0.67	5.5	-2.83	1.4
	OLS	0.38	4.0	-2.94	1.3
Basic vs. Control	LOOP	-0.20	3.3	1.56	1.3
	Simple Difference	0.41	5.1	1.71	1.4
	OLS	0.14	3.7	1.58	1.3
Saving vs. Control	LOOP	-0.13	3.3	4.22	1.3
	Simple Difference	-0.25	4.9	4.55	1.3
	OLS	-0.23	3.7	4.63	1.3

Variance of LOOP

Variance Estimation

Having produced an unbiased estimate, we now wish to obtain an estimate for the variance. First, some additional notation:

- We now let the probability of treatment be some $p_i \in (0, 1)$ for each i
- If p_i is the same for all i, we write $p_i = p$
- Define the (signed) inverse probability weights U_i as

$$U_i = \begin{cases} 1/p_i, & T_i = 1 \\ -1/(1-p_i), & T_i = 0 \end{cases}$$

Variance of $\hat{\tau}_i$

Noting that the LOOP estimator is a mean of the individual $\hat{\tau}_i$'s, we need to calculate $\text{Var}(\hat{\tau}_i)$ and $\text{Cov}(\hat{\tau}_i, \hat{\tau}_j)$ $(i \neq j)$

$$\begin{aligned}
\text{Var}(\hat{\tau}_i) &= \text{Var}\left[\mathbb{E}(\hat{\tau}_i|\hat{m}_i)\right] + \mathbb{E}[\text{Var}(\hat{\tau}_i|\hat{m}_i)] \\
&= \text{*****SOME MATH*****} \\
&= \frac{1}{p_i(1-p_i)}\text{MSE}(\hat{m}_i)
\end{aligned}$$

Covariance of $\hat{\tau}_i$ and $\hat{\tau}_j$

Next, denote $\text{Cov}(\hat{\tau}_i, \hat{\tau}_j)$ by γ_{ij}:

$$\begin{aligned}\gamma_{ij} &= \text{*****MORE MATH*****} \\ &= \text{Cov}(\hat{m}_i U_i, \hat{m}_j U_j) \\ &= \rho_{ij} \sqrt{\frac{\text{Var}(\hat{m}_i)\text{Var}(\hat{m}_j)}{p_i p_j (1-p_i)(1-p_j)}}\end{aligned}$$

where

$$\rho_{ij} = \text{Corr}(\hat{m}_i U_i, \hat{m}_j U_j)$$

Variance Expression

Combining the prior results yields:

$$\begin{aligned}
\text{Var}(\hat{\tau}) &= \text{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i\right) \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N}\text{Var}(\hat{\tau}_i) + \sum_{i\neq j}\text{Cov}(\hat{\tau}_i, \hat{\tau}_j)\right] \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N}\frac{1}{p_i(1-p_i)}\text{MSE}(\hat{m}_i) + \sum_{i\neq j}\gamma_{ij}\right]
\end{aligned}$$

Variance Expression

We consider the case where $p_i = p$

$$\text{Var}(\hat{\tau}) = \frac{1}{N}\left[\frac{\overline{\text{MSE}}}{p(1-p)} + (N-1)\bar{\gamma}\right]$$

where

$$\overline{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\text{MSE}(\hat{m}_i)$$

and

$$\bar{\gamma} = \frac{1}{N(N-1)}\sum_{i\neq j}\gamma_{ij}.$$

The Covariance Terms are Negligible

$$\gamma_{ij} = \text{Cov}(\hat{m}_i U_i, \hat{m}_j U_j)$$

We argue that γ_{ij} is often negligible, in the sense that γ_{ij} goes to zero at a sufficiently fast rate

- Recall that treatment assignments are independent and thus so are U_i and U_j
- The only reason for $\hat{m}_i U_i$ and $\hat{m}_j U_j$ to be correlated would be through the dependence of $\hat{m}_i$ on U_j and vice versa
- As N increases, the effect of U_j on $\hat{m}_i$ should shrink, as the number of other U_k's increase

The Covariance Terms are Negligible

$$\gamma_{ij} = \text{Cov}(\hat{m}_i U_i, \hat{m}_j U_j) = \rho_{ij} \sqrt{\frac{\text{Var}(\hat{m}_i)\text{Var}(\hat{m}_j)}{p_i p_j (1-p_i)(1-p_j)}}$$

(Slightly) more formally, we want $\bar{\gamma}$ to go to zero faster than $1/N$

- Suppose $\text{Var}(\hat{m}_i)$ and $\text{Var}(\hat{m}_j)$ go to zero at rate $1/N$. Then if ρ_{ij} goes to zero at any rate, γ_{ij} will go to zero faster than $1/N$
- This is not an unreasonable assumption: intuitively, we expect our estimate of $\hat{m}_i$ to stabilize as N increases

The Covariance Terms are Negligible

- Even more formally, we give conditions under which γ_{ij} goes to zero in the paper
- As a special case, if $\hat{m}_i$ is a polynomial of degree D, then so long as $\text{Var}(\hat{m}_i)$ goes to zero (at any rate) for all i, then $\bar{\gamma}$ will go to zero faster than D/N
- Finally, we provide an unbiased estimator of γ_{ij} in case there is a concern that the covariances are not negligible

The Variance of the LOOP Estimator

Because $\bar{\gamma}$ is negligible, we have:

$$\begin{aligned}\text{Var}(\hat{\tau}) &= \frac{1}{N}\left[\frac{\overline{\text{MSE}}}{p(1-p)} + (N-1)\bar{\gamma}\right] \\ &\approx \frac{\overline{\text{MSE}}}{Np(1-p)}\end{aligned}$$

This last value is what we estimate next

Estimating the Variance using Cross Validation

Having calculated the true variance of LOOP, we now wish to obtain an estimate for the variance. It is straightforward to show:

$$\begin{aligned}\mathrm{Var}(\hat{\tau}) &\approx \frac{\overline{\mathrm{MSE}}}{Np(1-p)} \\ &\leq \frac{1}{N}\left[\frac{1-p}{p}M_t + \frac{p}{1-p}M_c + 2\sqrt{M_t M_c}\right]\end{aligned}$$

where

$$M_t = \frac{1}{N}\sum_{i=1}^{N}\mathrm{MSE}(\hat{t}_i)$$

$$M_c = \frac{1}{N}\sum_{i=1}^{N}\mathrm{MSE}(\hat{c}_i).$$

Estimating the Variance using Cross Validation

We estimate M_t and M_c by leave-one-out cross validation. For example, we can use

$$\hat{M}_t = \frac{1}{n} \sum_{i \in \mathcal{T}} (\hat{t}_i - t_i)^2$$

to estimate

$$M_t = \frac{1}{N} \sum_{i=1}^{N} \text{MSE}(\hat{t}_i)$$

Similarly, we let $\hat{M}_c = \frac{1}{N-n} \sum_{i \in \mathcal{C}} (\hat{c}_i - c_i)^2$

Estimating the Variance using Cross Validation

We plug our estimates into the variance expression to get the final variance estimate:

$$\widehat{\text{Var}}(\hat{\tau}) = \frac{1}{N}\left[\frac{1-p}{p}\hat{M}_t + \frac{p}{1-p}\hat{M}_c + 2\sqrt{\hat{M}_t\hat{M}_c}\right]$$

LOOP vs. Simple Difference Estimator

LOOP with No Covariates

- Suppose we estimate $\hat{m}_i$ without making use of covariates
- To do this, we take the mean of the observed outcomes in the treatment group (excluding observation i) to estimate t_i:

$$\hat{t}_i = \frac{\sum_{k \in \mathcal{T} \setminus i} Y_k}{n - T_i}$$

- Similarly, we take the mean of the observed outcomes in the control group (excluding observation i) to estimate c_i:

$$\hat{c}_i = \frac{\sum_{k \in \mathcal{C} \setminus i} Y_k}{(N - n) - (1 - T_i)}$$

LOOP vs. Simple Difference: ATE Estimate

In this case, the LOOP estimator is exactly equal to the simple difference estimator. The comparison between LOOP and the simple difference estimator has two main implications:

1. We can improve on the simple difference estimator by improving our estimate of m_i: we have some assurance that our adjustment won't hurt performance
2. LOOP can be considered a re-interpretation of the simple difference estimator and the t-test for analyzing randomized experiments

LOOP vs. Simple Difference: Variance Estimate

- In the no covariate case, the variance estimate for the LOOP and simple difference estimators are almost equal
- First, when we estimate the potential outcomes without covariates, we have:

$$\hat{M}_t = \frac{n}{n-1} s_t^2$$

$$\hat{M}_c = \frac{N-n}{N-n-1} s_c^2$$

where s_t^2 and s_c^2 are the sample variances of the treatment and control groups

LOOP vs. Simple Difference: Variance Estimate

- Plugging $\hat{M}_t$ and $\hat{M}_c$ into our variance estimate yields:

$$\widehat{\text{Var}}(\hat{\tau}) \leq \left(\frac{n}{Np}\right)\frac{s_t^2}{n-1} + \left(\frac{N-n}{N(1-p)}\right)\frac{s_c^2}{N-n-1}$$
$$\approx \frac{s_t^2}{n-1} + \frac{s_c^2}{N-n-1}$$

  with equality in the first line when $\hat{M}_t$ and $\hat{M}_c$ are equal
- This last expression is nearly identical to the variance used in the t-test (the denominators differ by 1)

The Simple Difference Estimator is a Special Case of LOOP

- Given that LOOP produces the same ATE estimate and a nearly identical variance estimate, we can view the simple difference estimator as a special case of LOOP
- With LOOP, we also have the option to use covariate information to improve precision

Different Randomization Schemes

Dependent Treatment Assignments

Previously, we assumed that the treatment assignments are independent

- Consider the case where we randomly assign n units to treatment and leave the remaining $N - n$ units to control
- Then the T_i's are no longer independent: if observation 1 is treatment, the rest are more likely to be control

Random Drop Procedure

We can ensure that $\hat{m}_i$ and T_i are still independent using the "Random Drop" procedure. Once again, consider the case where we randomly assign n units to treatment and leave the remaining $N - n$ units to control:

- If observation i is assigned to treatment, we randomly pick one of the control observations and drop that observation (in addition to observation i) when estimating $\hat{m}_i$
- If observation i is assigned to control, we randomly pick one of the treatment observations and drop that observation
- Regardless of whether T_i is 0 or 1, we estimate $\hat{m}_i$ with $N - 2$ of the remaining observations
- $n - 1$ of these will be assigned to treatment, $N - n - 1$ will be assigned to control, and the specific allocation will not depend on T_i

Random Drop Procedure Example

Consider a randomized experiment with $N = 5$ participants, 2 of which will be assigned to treatment and the remaining 3 to control:

Table: Illustration of the Random Drop Procedure

#	Treatment Assignments	Potential Drops when Estimating m_1		
1)	T T C C C	T T \ C C	T T C \ C	T T C C \
2)	T C T C C	T \ T C C	T C T \ C	T C T C \
3)	T C C T C	T \ C T C	T C \ T C	T C C T \
4)	T C C C T	T \ C C T	T C \ C T	T C C \ T
5)	C T T C C	C \ T C C	C T \ C C	
6)	C T C T C	C \ C T C	C T C \ C	
7)	C T C C T	C \ C C T	C T C C \	
8)	C C T T C	C C \ T C	C C T \ C	
9)	C C T C T	C C \ C T	C C T C \	
10)	C C C T T	C C C \ T	C C C T \	

Does LOOP Still Work?

Sort of. The estimate itself remains relatively unchanged:

- Using the random drop procedure ensures we still get an unbiased estimate
- We lose some information by dropping an observation; however, we can perform the random drop procedure many times, getting an unbiased estimate each time
- If we perform enough trials and then average the results, we get an unbiased estimate where we essentially use all information
- In the no covariate case, we can show that the expectation across all random drops is exactly equal to the value had we not performed the random drop in the first case

Does LOOP Still Work?

The variance estimation procedure needs to be modified:

- We make use of the independence of treatment assignments when estimating the variance
- At this point it is unclear how much we need to change the variance estimation under different randomization schemes

Random Drop in Other Randomization Schemes

- The random drop procedure can be used in other randomization schemes
- For example, in a block design, the randomly dropped observation would need to be in the same block as observation i