University of Illinois, Urbana and Champaign

Critique of Misidentified Selection Bias and Redefinement of Sufficient Statistics for Treatment Effects and Causal Inference using Quadruple Randomized Controlled Trials (QRCT)

Eun Chul Choi

Abstract:

In defining selection bias, we have considered only the parallel universe of the treated group or the untreated group rather than including the parallel universe of the untreated group or the treated group. This makes causal inference theories unbalanced because they were developed on one side of treated or untreated group while the other side remained static. In this paper, we redefined selection bias by considering all four existing universes in the real and hypothetical worlds, and we provided sufficient statistics for estimating the effects of treatment with Quadruple Randomized Controlled Trials (QRCT)

Contents

1 Misidentified Selection Bias

1.1 Missing parallel universe for the untreated group in selection bias

Setting up a hypothetical, parallel universe of the treated group, as shown below, has been a fundamental principle for causal inference in statistics and social science.

$$E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)$$

$$= \underbrace{E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1)}_{\text{Causality}} \quad (1)$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\text{Selection Bias in Treated group side}} \quad (2)$$

Explanation of the notations:

1. $Y, X_{individual, hypothetical\ fact(treated, untreated)}$

2. $T : Indicator\ function(T = 1 : treated,\ T = 0 : untreated)$

3. $t : time$

$t = 0 : when\ information\ of\ treatment\ is\ created\ (Random\ Sample\ Selection\ in\ RCT\ setting)$

$t = 1 : when\ treatment\ effects\ are\ completed$

However, even though we defined the Selection Bias, we did not take into account the parallel universe of the untreated group. If we use the definition proposed by Imbens and Rubin, two selection biases exist, as shown below:

$$E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1) \tag{3}$$

$$= \underbrace{E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1)}_{\text{Causality}} \tag{4}$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\text{Selection Bias in Treated group side}} \tag{5}$$

$$E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1) \tag{6}$$

$$= \underbrace{E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,1}|X_{i,1}, T = 0, t = 1)}_{\text{Selection Bias in Untreated group side}} \tag{7}$$

$$+ \underbrace{E(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\text{Causality}} \tag{8}$$

The main point of this paper is these two selection biases do not move symmetrically, so another definition is needed to consider the movements of 4 universes, which can be represented by distributions, such as:

$D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1)$ *Untreated Group in Real World*
$D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1)$ *Treated Group in Real World*
$D_{i,1}(Y_{i,1}|X_{i,1}, T = 0, t = 1)$ *Untreated Group in Hypothetical Parallel Universe*
$D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$ *Treated Group in Hypothetical Parallel Universe*

Then, what will be new selection biases which takes all the universes into account. We suggest the new definition provided as below:

Definition 1.1 (New Selection Bias) *Selection Bias taking into account the four universes that are represented by* $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1), D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1), D_{i,1}(Y_{i,1}|X_{i,1}, T = 0, t = 1), D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$ *such that*

$$
\begin{aligned}
&|\underbrace{E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\text{Difference between Treated group and Untreated group in Real World}} \\
&-(\underbrace{E(Y_{i,0}|X_{i,0}, T = 1, t = 1) - E(Y_{i,1}|X_{i,1}, T = 0, t = 1)}_{\text{Hypothetical Cross-validation of Treatment Effect}})|
\end{aligned}
$$

Before explaining intuition, we would like to clarify notations in this paper. The principle is that, if there is no prime on to distributions and their expectations, then they represent a state in unbiased Randomized Controlled Trials. If there is a prime on to distributions and their expectations, it represents an incomplete state due to the limitations of the real world and the setting of the Randomized Controlled Trials is biased.

According to Imbens and Rubin, if we utilize the definition above, we can take two selection biases into account. It represents the aggregated selection bias of two sides, and we will define Quadruple Randomized Controlled Trials Bias, which stands for skewness of both the Selection Bias and Aggregated Causality (Treatment Effects) in the following chapter.

We describe further below how four universes transform if we implement asymptotically perfect, unbiased Randomized Controlled Trials. First, our intuition is that, over the course of asymptotically perfect RCT, parallel universes merge pair by pair, such as $(D'_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1), D'_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1)), (D'_{i,1}(Y_{i,1}|X_{i,1}, T = 0, t = 1), D'_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1))$. Also, as we increasingly control the bias in the Randomized Controlled Trials, all four universes align symmetrically pair by pair $(D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1), D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1)), (D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1), D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1))$ and in expectation senses as is shown below:

$$\underbrace{E'(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E'(Y_{i,0}|X_{i,0}, T = 1, t = 1)}_{\textit{Causality}\prime \text{ -> } \textit{Causality}} \tag{9}$$

$$= \underbrace{E'(Y_{i,0}|X_{i,0}, T = 1, t = 1) - E'(Y_{i,1}|X_{i,1}, T = 0, t = 1)}_{\text{Hypothetical Cross-validation of Treatment Effect -> } \textit{Causality}} \tag{10}$$

$$= \underbrace{E'(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E'(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\textit{Causality}\prime\prime \text{ -> } \textit{Causality}} \tag{11}$$

1.2 Resummarization of the Effects of Treatment based on the before-treatment condition and the after-treatment condition

If we think about parallel universes for treatment effect and causality, we should think about four types of universes with a model with two time periods and its conditional distributions as described below, where t=0 is when information is created concerning the effect of the treatment, and t=1 is when the effects of the treatment are complete. t=0 in the RCT setting represents when the random selection was implemented):

	Untreated Group	Treated Group
Real World at t=0	$D_{i,0}(Y_{i,0}\|X_{i,0}, T = 0, t = 0)$	$D_{i,1}(Y_{i,1}\|X_{i,1}, T = 1, t = 0)$
Hypothetical World at t=0	$D_{i,1}(Y_{i,1}\|X_{i,1}, T = 0, t = 0)$	$D_{i,0}(Y_{i,0}\|X_{i,0}, T = 1, t = 0)$
Real World at t=1	$D_{i,0}(Y_{i,0}\|X_{i,0}, T = 0, t = 1)$	$D_{i,1}(Y_{i,1}\|X_{i,1}, T = 1, t = 1)$
Hypothetical World at t=1	$D_{i,1}(Y_{i,1}\|X_{i,1}, T = 0, t = 1)$	$D_{i,0}(Y_{i,0}\|X_{i,0}, T = 1, t = 1)$

We can summarize all of the effects, i.e., before treatment effects and after treatment effects as conditional expectations, as shown below:

$$E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 0) \tag{12}$$

$$= \underbrace{E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1)}_{\text{Causality}} \tag{13}$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 1, t = 1) - E(Y_{i,1}|X_{i,1}, T = 0, t = 1)}_{\text{Hypothetical Cross-validation of Treatment Effect}} \tag{14}$$

$$+ \underbrace{E(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1)}_{\text{Causality}} \tag{15}$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 0)}_{\text{Time Effect}} \tag{16}$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 0, t = 0) - E(Y_{i,1}|X_{i,1}, T = 0, t = 0)}_{\text{Pre-reaction to Treatment Effect}} \tag{17}$$

$$+ \underbrace{E(Y_{i,1}|X_{i,1}, T = 0, t = 0) - E(Y_{i,0}|X_{i,0}, T = 1, t = 0)}_{\text{Hypothetical Cross-validation of Pre-reaction to Treatment Effect}} \tag{18}$$

$$+ \underbrace{E(Y_{i,0}|X_{i,0}, T = 1, t = 0) - E(Y_{i,0}|X_{i,0}, T = 0, t = 0)}_{\text{Pre-reaction to Treatment Effect}} \tag{19}$$

In this paper, we assume researchers are able to perfectly control pre-reaction to treatment effects (17), (18), (19), and concentrate on building sufficient statistics for estimation of treatment effects (13),(15) in unbiased RCT setting.

2 Sufficient Statistics for Treatment Effect and Causal Inference

2.1 Sufficiency in the difference of the conditional sample mean between the treated group and the untreated group

Definition 2.1 (Fisher Sufficiency) *Fisher Sufficiency is a statistic $T(\underset{\sim}{x})$ which is a sufficient statistic for θ if the conditional distribution of sample $\underset{\sim}{x}$ given the value of $T(\underset{\sim}{x})$ does not depend on θ.*

In other words, a sufficient statistic for a parameter θ in a statistic $T(\underset{\sim}{x})$ where $\underset{\sim}{x} = (x_1, \cdots, x_n)$ is the statistic which capture all information about θ contained in the sample $\underset{\sim}{x}$.

We have been estimating treatment effects with $\frac{1}{n}\sum_{i\in T} y_i|x_i - \frac{1}{n}\sum_{i\in U} y_i|x_i$ based on a belief that randomized controlled trials would enable us to mimic $E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1)$. That is, we implemented randomized controlled trials in the belief that $\frac{1}{n}\sum_{i\in U} y_i|x_i$ is thought to be generated from $D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$. In other words, randomized controlled trials enable us to assume $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$.

If our assumption is correct, $\frac{1}{n}\sum_{i\in T} y_i|x_i - \frac{1}{n}\sum_{i\in U} y_i|x_i$ is sufficient statistics, as indicated below:

Proof

Claim: If we assume $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$, then $\frac{1}{n}\sum_{i\in T} y_i|x_i - \frac{1}{n}\sum_{i\in U} y_i|x_i$ is a sufficient statistic for parameter $\theta = E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1)$.

If $p(\underset{\sim}{x}|\theta)$ is the joint pdf of $\underset{\sim}{x}$ and $g(T(\underset{\sim}{x})|\theta)$ is the pdf of θ, then θ is a sufficient statistic if $\forall x \in X$ the ratio $\frac{p(\underset{\sim}{x}|\theta)}{g(T(\underset{\sim}{x})|\theta)}$ is constant in θ.

By assumption $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$, parallel universe between $D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1)$ *and* $D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$ are independent of each other, and $(y_{i,1}|x_{i,1}, T = 1, t = 1), (y_{i,0}|x_{i,0}, T = 1, t = 1)$ are *iid*.

$$\frac{f((y_{1,1}|x_{1,1},T=1,t=1),...,(y_{n,1}|x_{n,1},T=1,t=1),(y_{1,0}|x_{1,0},T=1,t=1),...,(y_{n,0}|x_{n,0},T=1,t=1))|\theta)}{g(\frac{1}{n}\sum_{i\in T} y_i|x_i - \frac{1}{n}\sum_{i\in U} y_i|x_i|\theta)}$$

$$= \frac{f((y_{1,1}|x_{1,1},T=1,t=1),...,(y_{n,1}|x_{n,1},T=1,t=1),(y_{1,0}|x_{1,0},T=1,t=1),...,(y_{n,0}|x_{n,0},T=1,t=1)),\theta)}{g(\frac{1}{n}\sum_{i\in T} y_i|x_i - \frac{1}{n}\sum_{i\in U} y_i|x_i,\theta)}$$

$$= \frac{f((y_{1,1}|x_{1,1},T=1,t=1),...,(y_{n,1}|x_{n,1},T=1,t=1),(y_{1,0}|x_{1,0},T=1,t=1),...,(y_{n,0}|x_{n,0},T=1,t=1)),\theta)}{g(\theta)} \qquad \square$$

However, we are utilizing a strong assumption since $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$ implies $D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1) = D_{i,1}(Y_{i,1}|X_{i,0}, T = 0, t = 1)$. Note that there is no prime on the distributions, which means the trials were unbiased Randomized Controlled Trials. In this paper, we attempt to justify that Quadruple Randomized Controlled Trials provide a proper convergence speed over the path toward asymptotically Perfect RCT, and we also build a sufficient statistic for estimating the treatment effects in the unbiased RCT with a linear combination of the treatment effects of the two sides in a biased RCT setting. In the following chapter, we define asymptotically perfect QRCT and QRCT Bias.

3 Redefinement of Sufficient Statistics for Treatment Effects and Causal Inference with Quadruple Randomized Controlled Trials

3.1 Convergence Rate of Causality for the Treated and Untreated Groups

Quadruple Randomized Controlled Trials simply were used to implement two times for the treated group and two times for the untreated group, such as
$\frac{1}{n}\sum_{i,1,t=0\in T_1} y_{i,1}|x_{i,1}, \frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1}, \frac{1}{n}\sum_{i,0,t=0\in U_1} y_{i,0}|x_{i,0}, \frac{1}{n}\sum_{i,0,t=0\in U_2} y_{i,0}|x_{i,0}$.
Two of the samples, e.g., $\frac{1}{n}\sum_{i,1,t=0\in T_1} y_{i,1}|x_{i,1} and, \frac{1}{n}\sum_{i,0,t=0\in U_2} y_{i,0}|x_{i,0}$, stand for samples from the real world, and $\frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1} and, \frac{1}{n}\sum_{i,0,t=0\in U_1} y_{i,0}|x_{i,0}$ represent ancillary samples to trace the movements of samples from hypothetical worlds. It seems that we are implementing exactly the same RCT as before , but it makes a huge difference in the theoretical world by acquring two more pieces of information. Basic intution indicated that we need a sample that can represent four universes, i.e., $D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1), D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1), D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1), and D_{i,1}(Y_{i,1}|X_{i,0}, T = 0, t = 1)$, to trace their non-symmetric movement on the path over the course of the asymptotically perfect RCT. Then, we must ask 'What is the asymptotically perfect RCT?'. First, we describe the asymptotically perfect RCT that we have been using with one side of the treated group.

Definition 3.1 (Asymptotically Perfect RCT with Two samples) *Asymptoically perfect Randomized Controlled Trial is such that:*

$$plim(|\frac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0}| < \epsilon) = 0, \forall \epsilon$$

This definition leads us to analyze the path toward asymptotically perfect RCT. Based on the definition of above, we have presented in Chapter 1 Equations (9),(10), and (11), indicating how four universes converge in a perfectly symmetric world.

Definition 3.2 *Rate of Convergence A sequence converges linearly to L, if there exists a number* $\mu \in (0,1)$ *such that*

$$lim_{k->\infty}\frac{|x_{k+1} - L|}{|x_k - L|} = \mu$$

First, we will utilize $\frac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0} = K- > 0$ as index for convergence following the definition of asymptotically perfect RCT above.

a. The Rate of Convergence for Treated Group

$$|(\tfrac{1}{n}\textstyle\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in T,t=1} y_{i,0}|x_{i,0}) - \underbrace{(E(Y_{i,1}|X_{i,1},T=1,t=1) - E(Y_{i,0}|X_{i,0},T=1,t=1))}_{Causality\ in\ Treated\ Group\ =\ L}|$$

$$= |(\tfrac{1}{n}\textstyle\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in T,t=1} y_{i,0}|x_{i,0} - \tfrac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0})$$

$$+ K - (E(Y_{i,1}|X_{i,1},T=1,t=1) - E(Y_{i,0}|X_{i,0},T=1,t=1))|$$

$$= |\underbrace{(\frac{1}{n}\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ =\ TE_{K,T=1}} + \underbrace{(\frac{1}{n}\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ CT_{K,T=1}}$$

$$- \underbrace{(\frac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ =\ TE_{K,T=0}} - \underbrace{(\frac{1}{n}\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ =\ CT_{K,T=0}}$$

$$+ \underbrace{(\frac{1}{n}\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1})}_{Sample\ Selection\ Bias(Imbens\ and\ Rubin)\ in\ Untreated\ Group\ Side}$$

$$+K - (E(Y_{i,1}|X_{i,1},T=1,t=1) - E(Y_{i,0}|X_{i,0},T=1,t=1))|$$

b. The Rate of Convergence for Untreated Group

$$|(\tfrac{1}{n}\textstyle\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0}) - \underbrace{(E(Y_{i,1}|X_{i,1},T=0,t=1) - E(Y_{i,0}|X_{i,0},T=0,t=1))}_{Causality\ in\ Untreated\ Group\ =\ L}|$$

$$= |(\tfrac{1}{n}\textstyle\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0} - \tfrac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0})$$

$$+K - (E(Y_{i,1}|X_{i,1},T=0,t=1) - E(Y_{i,0}|X_{i,0},T=0,t=1))|$$

$$= |\underbrace{-(\frac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ =\ TE_{K,T=0}} + \underbrace{(\frac{1}{n}\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ =\ CT_{K,T=0}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ =\ TE_{K,T=1}} - \underbrace{(\frac{1}{n}\sum_{i,1\in T,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ CT_{K,T=1}}$$

$$- \underbrace{(\frac{1}{n}\sum_{i,0\in T,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U,t=1} y_{i,0}|x_{i,0})}_{Sample\ Selection Bias(Imbens\ and\ Rubin)\ in\ Treated\ Group\ Side}$$

$$+K - (E(Y_{i,1}|X_{i,1},T=0,t=1) - E(Y_{i,0}|X_{i,0},T=0,t=1))|$$

As you can see, the convergence rate with respect to the asymptotically perfect RCT with two samples is too fast to trace down the speed of causality convergence in both the treated and untreated sides. Now, we will deaccelerate the convergence speed with Quadruple Randomized Controlled Trials and see whether it matches and balances the causality convergence rate in both the treated and untreated groups.

3.2 Quadruple Randomized Controlled Trials (QRCT)

Definition 3.3 (Asymptotically Perfect RCT for Quadruple Randomized Controlled Trials) *The asymptotically perfect Quadruple Randomized Controlled Trial is such that:*

$$plim(|\frac{1}{n}\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0} + \frac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1}| < \epsilon) = 0, \forall \epsilon$$

Now, we will utilize $\frac{1}{n}\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0} + \frac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1}$ =K′→ 0 as the index for the convergence based on the definition of the asymptotically perfect QRCT, and the fixed convergence rate of causality assumed above.

c. The Rate of Convergence for Treated Group

$$|(\tfrac{1}{n}\textstyle\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in T_1,t=1} y_{i,0}|x_{i,0}) - \underbrace{(E(Y_{i,1}|X_{i,1},T=1,t=1) - E(Y_{i,0}|X_{i,0},T=1,t=1))}_{Causality\ in\ Treated\ Group\ =\ L}|$$

$$= |(\tfrac{1}{n}\textstyle\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in T_1,t=1} y_{i,0}|x_{i,0}$$

$$-\tfrac{1}{n}\textstyle\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0} - \tfrac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0})$$

$$+K' - (E(Y_{i,1}|X_{i,1},T=1,t=1) - E(Y_{i,0}|X_{i,0},T=1,t=1))|$$

$$= |\underbrace{(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ =\ TE_{K',T=1}} + \underbrace{(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ CT_{K',T=1}}$$

$$-\underbrace{(\frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ =\ TE_{K',T=0}} - \underbrace{(\frac{1}{n}\sum_{i,1\in U_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ =\ CT_{K',T=0}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in U_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ Side\ =\ CT_{K',T=0}} + \underbrace{(\frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ =\ TE_{K',T=0}}$$

$$-\underbrace{(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T_2,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ TE_{K',T=1}} - \underbrace{(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ =\ TE_{K',T=1}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T_2,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ TE_{K',T=1}}$$

$$+K' - (E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1))|$$

d. The Rate of Convergence for Untreated Group

$$|(\tfrac{1}{n}\textstyle\sum_{i,1\in U_1,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0}) - \underbrace{(E(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1))}_{Causality\ in\ Untreated\ Group\ =\ L}|$$

$$= |(\tfrac{1}{n}\textstyle\sum_{i,1\in U,t=1} y_{i,1}|x_{i,1} - \tfrac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0}$$

$$-\tfrac{1}{n}\textstyle\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1} + \tfrac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0} - \tfrac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0})$$

$$+ K' - (E(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1))|$$

$$= |\underbrace{-(\frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0} - \frac{1}{n}\sum_{i,0\in U_1,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ =\ TE_{K',T=0}} + \underbrace{(\frac{1}{n}\sum_{i,1\in U_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ =\ CT_{K',T=0}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,1\in T_1,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ =\ TE_{K',T=1}} - \underbrace{(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in T_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ =\ CT_{K',T=1}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1}-\frac{1}{n}\sum_{i,0\in T_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Treated\ Group\ Side\ =\ CT_{K',T=1}}-\underbrace{(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1}-\frac{1}{n}\sum_{i,1\in T_2,t=0} y_{i,1}|x_{i,1})}_{Sample\ Time\ Effect\ in\ Treated\ Group\ Side\ =\ TE_{K',T=0}}$$

$$-\underbrace{(\frac{1}{n}\sum_{i,1\in U_2,t=1} y_{i,1}|x_{i,1}-\frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ Side\ =\ CT_{K',T=0}}+\underbrace{(\frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0}-\frac{1}{n}\sum_{i,0\in U_2,t=0} y_{i,0}|x_{i,0})}_{Sample\ Time\ Effect\ in\ Untreated\ Group\ Side\ =\ TE_{K',T=0}}$$

$$+\underbrace{(\frac{1}{n}\sum_{i,1\in U_2,t=1} y_{i,1}|x_{i,1}-\frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0})}_{Sample\ Causality\ in\ Untreated\ Group\ Side\ =\ CT_{K',T=0}}$$

$$+\mathrm{K}'-(E(Y_{i,1}|X_{i,1},T=0,t=1)-E(Y_{i,0}|X_{i,0},T=0,t=1))|$$

It is apparent that if we decelerate the convergence speed with respect to the asymptotically perfect QRCT, then the convergence rates of causality depend only on sample causality with respect to K'.

3.3 Sufficient Statistic in the Quadruple Randomized Controlled Trials

After we roll a regular tetrahedral die and split the groups into four components before treatment kicks in, what should we do with them to estimate Treatment Effects and infer Causality after treatment has been implemented?

Definition 3.4 (Sufficient Statistic in Quadruple Randomized Controlled Trials) *Sufficient Statistics in Quadruple Randomized Controlled Trials:*

$$\alpha * (\frac{1}{n} \sum_{i,1 \in T_1, t=1} y_{i,1}|x_{i,1} - \frac{1}{n} \sum_{i,0 \in U_1, t=1} y_{i,0}|x_{i,0}) + (1-\alpha) * (\frac{1}{n} \sum_{i,1 \in T_2, t=1} y_{i,1}|x_{i,1} - \frac{1}{n} \sum_{i,0 \in U_2, t=1} y_{i,0}|x_{i,0})$$

where α is Quadruple Randomized Controlled Trials Bias(QRCT)

Then what will be Quadruple Randomized Controlled Trials(QRCT) Bias?

Definition 3.5 (Quadruple Randomized Controlled Trials Bias(Still working on)) *Quadruple Randomized Controlled Trials(QRCT) Bias α is defined as:*

$$\begin{aligned} E * (\alpha * (\frac{1}{n} \sum_{i,1 \in T_1, t=1} y_{i,1}|x_{i,1} - \frac{1}{n} \sum_{i,0 \in T_1, t=1} y_{i,0}|x_{i,0}) + (1-\alpha) * (\frac{1}{n} \sum_{i,1 \in U_2, t=1} y_{i,1}|x_{i,1} - \frac{1}{n} \sum_{i,0 \in U_2, t=1} y_{i,0}|x_{i,0} \\ = E(Y_{i,1}|X_{i,1}, T=1, t=1) - E(Y_{i,0}|X_{i,0}, T=1, t=1) \\ = E(Y_{i,1}|X_{i,1}, T=0, t=1) - E(Y_{i,0}|X_{i,0}, T=0, t=1) \\ = \alpha * (E'(Y_{i,1}|X_{i,1}, T=1, t=1) - E'(Y_{i,0}|X_{i,0}, T=1, t=1)) \\ +(1-\alpha) * (E'(Y_{i,1}|X_{i,1}, T=0, t=1) - E'(Y_{i,0}|X_{i,0}, T=0, t=1)) \end{aligned}$$

The QRCT Sufficient Statistic above was not chosen arbitrarily by combining two sample sides of treated group and untreated group causality. Four universes exist in the theoretical world, so causality always exists in both sides, i.e., in the treated and untreated groups. However, due to the restriction of time and space when we implement QRCT, bias will exist in QRCT. The term α represents QRCT biases, and the more we control them, the closer α will be $\frac{1}{2}$. The New Selection Bias defined in Chapter 1 represents the aggregated selection bias of the two sides in the Quadruple Randomized Controlled Trials Bias , and bias stands for the skewness of the Selection Bias and the Aggregated Causality(Treatment Effects)

proof

Claim: Quadruple Randomized Controlled Trials Sufficient Statistic is sufficient in Fisher information sense for Unbiased QRCT causality such as parameter $\theta = E(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E(Y_{i,0}|X_{i,0}, T = 1, t = 1) = E(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E(Y_{i,0}|X_{i,0}, T = 0, t = 1) = \alpha * (E'(Y_{i,1}|X_{i,1}, T = 1, t = 1) - E'(Y_{i,0}|X_{i,0}, T = 1, t = 1)) + (1 - \alpha) * (E'(Y_{i,1}|X_{i,1}, T = 0, t = 1) - E'(Y_{i,0}|X_{i,0}, T = 0, t = 1))$.

By assumption $D'_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D'_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1), D'_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1) = D'_{i,1}(Y_{i,1}|X_{i,0}, T = 0, t = 1)$, parallel universe between $(D'_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1), D'_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1))$, $(D'_{i,1}(Y_{i,1}|X_{i,1}, T = 0, t = 1), D'_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1))$ are independent each others, and $(y_{i,1}|x_{i,1}, T = 1, t = 1), (y_{i,0}|x_{i,0}, T = 1, t = 1), (y_{i,1}|x_{i,1}, T = 0, t = 1), (y_{i,0}|x_{i,0}, T = 0, t = 1)$ are *iid*.

$$\frac{f((y_{1,1}|x_{1,1},1,1),...,(y_{n,1}|x_{n,1},1,1),(y_{1,0}|x_{1,0},1,1),...,(y_{n,0}|x_{n,0},1,1),(y_{1,1}|x_{1,1},0,1),...,(y_{n,1}|x_{n,1},0,1),(y_{1,0}|x_{1,0},0,1),...,(y_{n,0}|x_{n,0},0,1)|\theta)}{g(\alpha*(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0}) + (1-\alpha)*(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0})|\theta)}$$

$$= \frac{f((y_{1,1}|x_{1,1},1,1),...,(y_{n,1}|x_{n,1},1,1),(y_{1,0}|x_{1,0},1,1),...,(y_{n,0}|x_{n,0},1,1),(y_{1,1}|x_{1,1},0,1),...,(y_{n,1}|x_{n,1},0,1),(y_{1,0}|x_{1,0},0,1),...,(y_{n,0}|x_{n,0},0,1),\theta)}{g(\alpha*(\frac{1}{n}\sum_{i,1\in T_1,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_1,t=1} y_{i,0}|x_{i,0}) + (1-\alpha)*(\frac{1}{n}\sum_{i,1\in T_2,t=1} y_{i,1}|x_{i,1} - \frac{1}{n}\sum_{i,0\in U_2,t=1} y_{i,0}|x_{i,0}),\theta)}$$

$$= \frac{f((y_{1,1}|x_{1,1},1,1),...,(y_{n,1}|x_{n,1},1,1),(y_{1,0}|x_{1,0},1,1),...,(y_{n,0}|x_{n,0},1,1),(y_{1,1}|x_{1,1},0,1),...,(y_{n,1}|x_{n,1},0,1),(y_{1,0}|x_{1,0},0,1),...,(y_{n,0}|x_{n,0},0,1),\theta)}{g(\theta)}$$

□

As you can see above, we can release the strong assumption we have utilized

$D_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$

$D_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1) = D_{i,1}(Y_{i,1}|X_{i,0}, T = 0, t = 1)$

to

$D'_{i,0}(Y_{i,0}|X_{i,0}, T = 0, t = 1) = D'_{i,0}(Y_{i,0}|X_{i,0}, T = 1, t = 1)$

$D'_{i,1}(Y_{i,1}|X_{i,1}, T = 1, t = 1) = D'_{i,1}(Y_{i,1}|X_{i,0}, T = 0, t = 1)$.