Adaptive Experimental Design for Efficient Treatment Effect Estimation: Randomized Allocation via Contextual Bandit Algorithm

効率的な処置効果推定のための適応的実験計画：文脈付きバンディットアルゴリズムからのアプローチ

by

Masahiro Kato

加藤真大

A Master Thesis

修士論文

Submitted to

the Graduate School of the University of Tokyo

on January 27, 2020

in Partial Fulfillment of the Requirements

for the Degree of Master of Information Science and Technology

in Computer Science

Thesis Supervisor: Junya Honda　本多淳也

Senior Lecturer of Computer Science

ABSTRACT

The *randomized control trial* (RCT) is a framework of a scientific experiment that aims to estimate the *treatment effect* defined as the difference between the expected outcomes of two or more treatments. In the RCT, the researcher randomly assigns the treatments to research subjects to get an unbiased estimator of the treatment effect. However, the standard RCT has some problems, such as huge costs by using many research subjects. For mitigating these problems, previous research has tried to develop an algorithm that returns an estimator with the smaller asymptotic variance than a standard RCT. The goal of this thesis is to propose a treatment allocation in an RCT for obtaining an estimator with the smaller asymptotic variance.

In this thesis, we consider a situation where a researcher can select a treatment for a research subject based on past observations. This approach is called *adaptive experimental design*. In this approach, it is difficult to apply the standard statistical method to construct an estimator. This difficulty comes from the fact that the observations are not independent and identically distributed if the researcher changes the allocation rule based on past information. In this thesis, to construct an estimator with the smaller asymptotic variance, we overcome this conventional problem by using an algorithm of the *multi-armed bandit problem* and the theory of *martingale*.

In the proposed method, after a researcher observes a feature of a research subject at a period, he/she estimates the conditional mean of the squared outcomes using *k-nearest neighbor regression*. Next, he/she constructs an estimated optimal probability for treatment allocation. Based on this probability, he/she randomly assigns a treatment to the research subject. Finally, for constructing an estimator of the treatment effect, this thesis uses an *inverse probability weighted estimator* to correct the sample selection bias caused by the adjustment of the allocation to the research subject.

In summary, this thesis is devoted to literature of an RCT by proposing a method of an adaptive experimental design for constructing an estimator with the smaller variance. First, we propose a method to construct an estimator of the treatment effect from samples with dependency caused by changing the allocation rule based on past information. Then, we derive a lower bound of the asymptotic variance of the estimator of the treatment effect and propose an algorithm that constructs an estimator with the smaller variance by using the probability of assigning a treatment achieving the lower bound of the asymptotic variance. We also elucidate theoretical properties of the proposed algorithm for both infinite and finite samples. Finally, we experimentally show that the proposed algorithm outperforms the standard RCT in some cases.

論文要旨

ランダム化比較試験（RCT）とは，2 つ以上の処置の効果の差の期待値として定義される平均処置効果を推定することを目的とする研究試験の方法である．RCT では，研究者は被験者に処置をランダムに割り当てることで処置効果の不偏推定量を得ることができる．一方で，RCT には多くの被験者を使用することで莫大な費用が発生するなどの問題がある．このような問題を軽減するために，過去の研究では少ない被験者で処置効果を推定する効率的なアルゴリズムの開発が取り組まれてきた．本論文の目的も，RCT における分散の小さい推定量を得るための処置の割り当ての方法を提案することにある．

本論文では，適応的実験計画と呼ばれる，研究者が過去の観察結果に基づいて被験者への処置の方法を変更できる状況を考える．過去の観察結果に基づいて被験者への処置の方法を選択する場合，観測値が独立同一分布から得られたものでないために，通常の統計的手法を適用して推定量を構築することは困難になる，本論文では，多腕バンディット問題のアルゴリズムとマルチンゲールの理論を用いて観測値の従属性に由来する従来の問題を解決することで，処置効果の分散の小さい推定量を構築する．

提案手法では，研究者はある時刻において被験者の特徴を観察した後に，その被験者の特徴で条件づけた処置効果の二乗の期待値を k-最近傍法による回帰で推定する．次に，その推定量に基づき処置効果の割り当ての最適な確率を推定し，その確率に従って被験者に処置をランダムに割り当てる．最後に，逆確率加重推定量を用いて処置効果の推定量を構築することで，処置を割り振る確率を被験者に応じて変えることで生じる標本選択バイアスを補正する．

以上のように，より漸近分散の小さい推定量を構築するための適応的実験計画の方法を提案することで，この論文は RCT の研究に貢献する．まず，本論文では，過去の情報に応じて被験者への処置の方法を選択することで生じる時間的な依存関係を有する標本から，処置効果の推定量を構築する方法を提案する．次に，その処置効果の推定量の漸近分散の下限を導出するとともに，その下限を達成する処置の割り当ての確率を用いることで，分散の小さい推定量を構築するアルゴリズムを提案する．加えて，無限の標本と有限の標本が与えられた場合それぞれに対して，そのアルゴリズムの理論的性質を解明する．最後に，いくつかの事例において，標準的な RCT よりも提案された手法の方がよりよい性能を示すことを実験的に示す．

Acknowledgements

I would like to express great appreciation to my thesis supervisor, Lecturer Junya Honda, for his helpful instructions, advice, and suggestions. I would also express my deep appreciation to Professor Hidehiko Ichimura, Professor Yusuke Narita, Doctor Takuya Ishihara, and other members of faculty of economics for their supports and discussions. This research started from discussion with Professor Hidehiko Ichimura and Professor Yusuke Narita. My thanks are also extended to Professor Masashi Sugiyama, Lecturer Issei Sato, and all of our lab members, especially Takeshi Teshima, for their kind helps. Finally, I wish to thank my families for their encouragement throughout my study.

Contents

Chapter 1

Introduction

Discovering causality from observations is a fundamental task in information science. In this thesis, we focus on the causality defined by Rubin (Rubin, 1974), which refers the causal effect as a difference between the outcomes of two actions. In general, one of these actions corresponds to the *treatment*, and the other corresponds to the *control* (Imbens and Rubin, 2015). However, we cannot know the difference directly because researchers can only observe the result of one of the two actions. For example, when a researcher gives a medicine to a patient, the researcher can observe an outcome of the patient given the medicine, but the researcher cannot observe an outcome of the patient who was not given the medicine. The seminal work that mathematically formulated this fact as the *potential outcome* is Neyman (1923). In the 1970s, Rubin extended this concept to estimating the causal effect from non-experimental data (Rubin, 1974, 1978). Because we cannot compare outcomes directly, we consider estimating the difference of the mean outcomes called the *average treatment effect*. When we estimate the mean outcomes of treatments from non-experimental data, we often face a problem of the *sample selection bias*, which is an estimation bias caused by the difference between the distributions of the treated and controlled group. To get rid of this bias, researchers use the *randomized control trial* (RCT). In the RCT, we randomly assign one of treatment to a research subject (Kendall, 2003). By using the RCT correctly, we can obtain an unbiased estimator of the average treatment effect (Imbens and Rubin, 2015).

1.1 History of the RCT and Causal Inference

There is a close relationship between the RCT and causal inference. In this section, we summarize the histories of the RCT and causal inference.

Early History of the RCT: The idea of randomized trials involving treatment and control groups is as old as the historical record itself, appearing in the Hebrew Bible and various societies around the world (Bothwell and Podolsky, 2016). The use of the RCT as a tool for the scientific experiment started from the age of Enlightenment (Bothwell and Podolsky, 2016). In 1753, a Scottish surgeon James Lind reported a clinical trial demonstrating that a diet including citrus fruit was effective against scurvy in sailors at sea (Bothwell and Podolsky, 2016). In the 18th and 19th centuries, loosely controlled trials increasingly appeared to test the utility of different remedies ranging from mesmerism to homeopathy. By the late 20th century, the RCT has been accepted as a standard for "rational therapeutics" (Meldrum, 2000). On the other hand, the theoretical justification of the RCT just started in the early 20th century (Imbens and Rubin, 2015).

Mathematical Justification of the RCT: The critical concept of the theory of the RCT is the potential outcome(Imbens and Rubin, 2015). The idea of the potential outcome also has a long history. For example, J.S. Mill, who is one of the greatest philosophers in the 19th century, mentioned the concept in his book (Mill, 1843; Imbens and Rubin, 2015). The seminal mathematical definition of the potential outcome was given in 1923 by Neyman's master thesis (Neyman, 1923; Pearl and Mackenzie, 2018). Combining the formulation of the RCT by Fisher in 1925, Neyman's definition became the standard in an RCT. Neyman's work mainly focused on an RCT with an urn model, which is a method of the RCT, and did not discuss causal inference from observed data. In the 1970s, Rubin extended Neyman's framework to causal inference from observed data with sample selection bias. By Rubin's generalization, we could estimate the causality from both an RCT and observational studies (Rubin, 1978, 1974).

Acceptance of the RCT as an Scientific Tool: Through these historical attempts, researchers have accepted the RCT as a scientific tool with theoretical guarantees. One of the earliest published RCTs in medicine appeared in 1948, which is conducted by the British Medical Research Council (MRC) of streptomycin for the treatment of tuberculosis (Hill, 1990; Yoshioka, 1998). Following medicine, economists have used the RCT at least since the 1960s (Greenberg and Shroder, 2004; Duflo et al., 2008). The RCT was progressively accepted as a tool for policy evaluation in the U.S. from the 1970s to the 1990s. Especially in development economics, the RCT became its integral part after the rapid growth started from the mid-1990s. For example, in 1997, the Mexican government implemented a program called PROGRESA, which combines a traditional cash transfer program with financial incentives for families to invest in the human capital of children (Gertler and Boyce, 2003). As a remarkable event, the Nobel Prize committee awarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019 to Banerjee, Duflo, and Kremer, who promoted the use of the RCT in development economics, "for their experimental approach to alleviating global poverty" (The Royal Swedish Academy of Sciences, 2019). In industry, the RCT also became the standard method to determine whether potential improvements of an alternative method are significant compared to a well-established default (Siroker and Koomen, 2013). In the applied domain, these kinds of RCTs are often referred to as A/B testing (Kohavi and Longbotham, 2017; Yang et al., 2017).

1.2 Adaptive Experimental Design

While the RCT is a reliable framework for scientific experiments, it still has some problems, such as costs and ethics (Nardini, 2014). For example, in clinical trial, the use of placebos concerns the problem of deception. A researcher sometimes prescribes placebos to patients and the patients must be made to believe they are receiving a working treatment, even though they are not, for the placebo effect to play a role at all (Nardini, 2014). Thus, in some cases, clinical trials are not only costly, but also unethical. For mitigating these problems, previous research has tried to develop an algorithm that returns an efficient estimator than a standard RCT. In this thesis, an efficient estimator means that a estimator with the smaller variance. One of the ideas for improving the efficiency of the RCT is *adaptive experimental design*. In adaptive experimental design, we are allowed to use past information of the RCT to improve the next RCT. The concept of the adaptive experimental design can be traced back to the 1970s (Pong and Chow, 2016). Nowadays, researchers recognize the importance of the adaptive experimental design to make an estimator efficient and robust. For example, Food and Drug Administration in the U.S. defined the adaptive experimental design as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial (CDER, 2018). Among some methodologies of the adaptive experimental design, we focus on adaptive randomization (Chow and Chang, 2011). In adaptive randomization, we can change the probability of assigning a treatment using past information. In methodologies of adaptive randomization, researchers try to fix the imbalance of assignments among groups. Hahn et al. (2011) proposed a method of adaptive randomization with situation where a researcher can separate an RCT into two stages. In the first stage, the researcher tries to find a probability of assigning a treatment that maximizes the statistical power. Then, the researcher uses the probability in the RCT of the second stage. This technique can be applied to many scientific experiments that have a pilot phase or some more general multistage or group-sequential structure.

1.3 Adaptive Experimental Design for Efficient Treatment Effect Estimation

Although Hahn et al. (2011) derived a lower bound of the asymptotic variance of the semi-parametric model and proposed an algorithm that achieves the lower bound, their method is hard to realize in practice because it requires infinite samples for estimating the variance of the outcomes before changing the probability. In this thesis, we tackle this problem by using the theory of the *multi-armed bandit problem* and *martingale*. In the multi-armed bandit problem, we model a situation where a player of a slot game tries to select the best slot machine that maximizes the player's profit. The martingale theory enables us to conduct statistical inference from observations with dependency.

In the proposed method, we estimate the probability of assigning a treatment that minimizes the asymptotic variance of an estimator and allocate the treatment following the probability. For constructing an estimator of the treatment effect, this thesis uses an *inverse probability weighted estimator* to correct the sample selection bias caused by the adjustment of the allocation to the research subject. The idea of the proposed method follows Hahn et al. (2011), which first proposed changing the probability of assignment based on covariates to estimate the treatment effect efficiently. Although Hahn et al. (2011) derived a lower bound of the asymptotic variance and proposed an algorithm that achieves the lower bound, their method is hard to realize in practice because it requires infinite samples for estimating the variance of the outcomes before changing the probability. On the other hand, in our proposed method, we integrate the two-step procedure by theoretically solving the problem caused by the dependency. The proposed method does not require infinite samples to construct the probability of assignment and works well in a practical situation.

Multi-armed Bandit Problem: The multi-armed bandit problem is a problem setting of machine learning, which models a situation where a player of a slot game tries to select the best slot machine that maximizes the player's profit. Algorithms of the multi-armed bandit problem have been used as a method of adaptive randomization (Villar et al., 2015). When we use the multi-armed bandit problem for adaptive randomization, we regard treatments as *arms*. In adaptive randomization with algorithms of multi-armed bandit problem, the standard goal of algorithms of multi-armed bandit problem is the maximization of profit obtained from treatments. On the other hand, Yang et al. (2017) proposed a framework to conduct a statistical test to identify arms with better profits recently. In this thesis, we propose an algorithm of the multi-armed bandit problem that estimates the treatment effect efficiently. Our motivation is similar to Yang et al. (2017); Jamieson and Jain (2018). While Yang et al. (2017); Jamieson and Jain (2018) only try to identify arms with better profit, we estimate the treatment effect itself efficiently using an algorithm of the multi-armed bandit problem.

1.4 Our Contributions

Major contributions of this thesis are summarized as follows:

- We establish a framework for statistical causal inference from data gathered from time-dependent algorithm. We can apply the proposed a framework to other problems of causal inference.
- We theoretically derive the lower bound of the asymptotic variance with data with dependency in an adaptive experimental design.
- We propose an algorithm that can achieve the lower bound of the asymptotic variance.
- We elucidate theoretical properties with both infinite and finite samples.
- We experimentally show that the proposed algorithm outperforms the standard RCT in some cases.

This thesis contributes to literature and practitioners of an RCT by proposing an efficient experimental design with theoretical guarantees. Besides, the proposed framework can be applied to other causal inference problems such as an off-policy evaluation.

1.5 Organization of this Thesis

In the following chapters, we introduce the proposed algorithm with its theoretical analysis and experimental results. First, in Chapter 2, we define the problem setting. In Chapter 3, we present a new framework for statistical inference using samples with dependency. With in the framework, we propose an algorithm for constructing an efficient estimator of the treatment effect in Chapter 4 and its theoretical proprieties in Chapter 5. Finally, in Chapter 6, we elucidate empirical performances of the proposed algorithm using synthetic and semi-synthetic datasets.

Chapter 2

Problem Formulation

In this thesis, we consider a setting where a research subject with an m-dimensional feature, $X_t \in \mathcal{X} = [0,1]^m$, visits us at a period, and we can assign a treatment to the research subject based on the feature. After the assignment, we observe the response from the subject immediately. We assume that there are T periods and two treatments $d = 1$ and $d = 0$. This setting is referred as *contextual multi-armed bandit problem*. Let us denote a probability that we select a treatment d as $p(D_t = d|X_t)$. For simplicity, we assumed that $p(D_t = d|X_t)$ is bounded as $p(D_t = d|X_t) \in [\varepsilon, 1-\varepsilon]$, where ε is a positive value less than $\frac{1}{2}$. Let us denote an outcome of treatment at t-th round as the following function:

$$Y_t(\cdot) : \{0,1\} \to \mathbb{R}.$$

We also assume that $0 < \underline{Y} \leq |Y_t| \leq \overline{Y}$, where $\underline{Y}$ and $\overline{Y}$ are positive constant values. Let us denote a set of samples obtained until t-th round as $\{(Y_t, X_t, D_t)\}_{t=1}^{T}$, where $Y_t = I[D_t = 1]Y_t(1) + I[D_t = 0]Y_t(0)$ is an outcome at t-th round, where $I[\cdot]$ is the indicator function.

2.1 Parameter of Interest and IPW Estimator

Our interest is in estimation of the *average treatment effect* (ATE). The ATE $\theta(d)$ between treatment $d = 1$ and $d = 0$ is defined as follows (Imbens and Rubin, 2015):

$$\theta = \mathbb{E}[Y_t(1) - Y_t(0)]. \tag{2.1}$$

Let us assume a traditional assumption called *unconfounded treatment allocation* as $D_t \perp (Y_t(0), Y_t(1))|X_t$. Heckman et al. (1997) pointed out that, for identification of the ATE, this assumption can be weakened to mean independence, i.e., $\mathbb{E}[Y_t(d)|D_t, X_t] = \mathbb{E}[Y_t(d)|X_t]$ for $d = 0, 1$. Under this assumption, the ATE of (2.1) can be rewritten as follows:

$$\theta = \mathbb{E}[Y_t(1) - Y_t(0)] = \mathbb{E}\left[\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)} - \frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right].$$

Following this formulation, if the samples $\{(Y_t(1), Y_t(0), X_t, D_t)\}_{t=1}^{T}$ are i.i.d with allocation probability $p(D = 1|X_t)$, we can construct an estimator of the average treatment effect as follows:

$$\hat{\theta}_T^{\mathrm{IPW}} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)} - \frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right). \tag{2.2}$$

This is called *inverse probability weighting* (IPW) estimator (Imbens and Rubin, 2015). Lemma 2.1.1 below shows the asymptotic distribution of $\hat{\theta}_T^{\mathrm{IPW}}$.

Lemma 2.1.1. The asymptotic distribution of $\hat{\theta}_T^{\text{IPW}}$ is given as follows:

$$\sqrt{T}(\hat{\theta}_T^{\text{IPW}} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(0, \sigma^2)$ denotes the normal distribution with the mean 0 and the variance σ^2, and $\sigma^2 = \mathbb{E}\left[\frac{Y_t^2(1)}{p(D_t=1|X_t)}\right] + \mathbb{E}\left[\frac{Y_t^2(0)}{p(D_t=0|X_t)}\right] - \theta^2$.

Previous research also have discussed this asymptotic variance in informal way such as Hirano et al. (2003). In this thesis, we also show the proof in Section 2.3.1.

Based on Lemma 2.1.1, we can conduct a more efficient experiment by adjusting the probability that makes σ^2 small. We regard σ^2 as a function of $p(D_t = 1|X_t)$ and we minimize σ^2 by selecting appropriate $p(D_t = 1|X_t)$. The optimal $p(D_t = 1|X_t)$, which minimizes the asymptotic variance, is given by the following theorem.

Theorem 2.1.1. Let $\mathcal{P}$ be a function class of $p : \mathcal{X} \to [\varepsilon, 1-\varepsilon]$, and let us define the following function $f : \mathcal{P} \to \mathbb{R}$:

$$f(p) = \mathbb{E}\left[\frac{Y_t^2(1)}{p(X_t)}\right] + \mathbb{E}\left[\frac{Y_t^2(0)}{1-p(X_t)}\right] - \theta^2.$$

Then, if $\varepsilon \leq \frac{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}+\sqrt{\mathbb{E}[Y_t^2(0)|X_t]}} \leq 1-\varepsilon$, the minimizer of the function is given as follows:

$$p^{\text{OPT}}(D_t = 1|X_t) = \frac{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]} + \sqrt{\mathbb{E}[Y_t^2(0)|X_t]}}.$$

The proof is shown in Section 2.3.2. On the other hand, the mean squared error between θ and $\hat{\theta}_T^{\text{IPW}}$ is given as

$$\mathbb{E}\left[\left(\theta - \hat{\theta}_T^{\text{IPW}}\right)^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}\left(\mathbb{E}\left[\frac{Y_t^2(1)}{p(D_t = 1|X_t)}\right] + \mathbb{E}\left[\frac{Y_t^2(0)}{p(D_t = 0|X_t)}\right] - \theta^2\right).$$

Therefore, the minimizer of the asymptotic variance is also the minimizer of the mean squared error. Hence, to minimize the variance of the estimator, we can minimize by setting the probability as $p^{\text{OPT}}(D_t = 1|X_t)$. However, because we do not know the value of $\mathbb{E}\left[Y_t^2(d)|X_t\right]$, we need to estimate the value using the past information.

2.2 Data Generating Process

In this section, we define our data generating process in more detail. Let us assume that outcomes $\{(Y_t(1), Y_t(0))\}_{t=1}^T$ are i.i.d. from a distribution with the mean of outcomes $\mu_d = \mathbb{E}[Y_t(d)]$, the conditional mean of outcomes $\mu_d(X_t) = \mathbb{E}[Y_t(d)|X_t]$, and the conditional mean of squared outcomes $\nu_d(X_t) = \mathbb{E}[Y_t^2(d)|X_t]$ for $d = 0, 1$. Let ξ_t be a random draw from the uniform distribution on $[0, 1]$ at the t-th round. In the proposed algorithm, we change a probability of selecting a treatment using the past information. Let us define a real-valued function $\pi_t(X_t, \mathcal{F}_{t-1}) \in [\varepsilon, 1-\varepsilon]$, where $\mathcal{F}_t$ is the history of t-th round defined as follows:

$$\mathcal{F}_{t-1} = \{Y_{t-1}, X_{t-1}, D_{t-1}, Y_{t-2}, X_{t-2}, D_{t-2}, \dots, Y_1, X_1, D_1\}.$$

At each round t, we select a treatment as

$$D_t = I[\xi_t \leq \pi_t(X_t, \mathcal{F}_{t-1})].$$

Let us define a policy Π as a set of π_t, i.e., $\Pi = \{\pi_1, ..., \pi_T\}$. Throughout this thesis, we use the following two equations:

$$\mathbb{E}_{D_t}[I[D_t = 1]|X_t, \mathcal{F}_{t-1}] = \pi_t(X_t, \mathcal{F}_{t-1})$$

$$\mathbb{E}_{(Y_t(d),D_t)}[I[D_t = d]Y_t(d)|X_t, \mathcal{F}_{t-1}] = \mathbb{E}_{Y_t(d)}[\pi_t(X_t, \mathcal{F}_{t-1})Y_t(d)]$$

where the expectation $\mathbb{E}_{D_t}$ is taken over D_t, $\mathbb{E}_{Y_t(d)}$ is taken over $Y_t(d)$, and $\mathbb{E}_{(Y_t(d),D_t)}$ is taken over $(Y_t(d), D_t)$ for $d = 0, 1$. The second equation follows from the fact that $(Y_t(1), Y_t(0), X_t)$ and D_t is independent given X_t. In the data generating process, while $(Y_t(1), Y_t(0))$ is i.i.d. sampled, D_t is not i.i.d. sampled because the probability depends on the past observations.

Domain of Expectations: From this section, we explicitly write the domain of expectations because the samples has complicated dependencies in our problem setting. $\mathbb{E}_{(Y_t(1),Y_t(0))}$, $\mathbb{E}_{X_t}$, $\mathbb{E}_{D_t}$, and $\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}$ denote the expectation over the distribution of $(Y_t(1), Y_t(0))$, X_t, D_t, and $(Y_t(1), Y_t(0), X_t)$, respectively. $\mathbb{E}_{\mathcal{F}_t}$ denotes the expectation over $\{Y_s(1), Y_s(0), X_s, D_s\}_{s=1}^{t}$. $\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}$ denote the expectation over the joint distribution of X_t and $\{Y_s(1), Y_s(0), X_s, D_s\}_{s=1}^{t-1}$. $\mathbb{E}_{\Pi}$ denotes the expectation over $\{Y_s(1), Y_s(0), X_s, D_s\}_{s=1}^{T}$ generated by a policy Π.

2.3 Proofs

In this section, we prove Lemma 2.1.1 and Theorem 2.1.1.

2.3.1 Proof of Lemma 2.1.1

Proof. Let β_t be $\frac{I[D_t=1]Y_t}{p(D_t=1|X_t)} - \frac{I[D_t=0]Y_t}{p(D_t=0|X_t)}$. Because the mean of β_t is θ and $\hat{\theta}_t^{\mathrm{IPW}} = \frac{1}{t}\sum_{s=1}^{t}\beta_s$, we can calculate the variance of $\hat{\theta}_t^{\mathrm{IPW}}$ by calculating the variance of

β_t. The variance of β_t is calculated as follows:

$$\begin{aligned}
&\mathbb{E}\left[(\beta_t - \mathbb{E}[\beta_t])^2\right] \\
&= \mathbb{E}\left[\left(\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)} - \frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right)^2\right] - \theta^2 \\
&= \mathbb{E}\left[\left(\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)}\right)^2\right] \\
&\quad - 2\mathbb{E}\left[\left(\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)}\right)\left(\frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right)\right] \\
&\quad + \mathbb{E}\left[\left(\frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right)^2\right] - \theta^2 \\
&= \mathbb{E}\left[\left(\frac{I[D_t = 1]Y_t}{p(D_t = 1|X_t)}\right)^2\right] + \mathbb{E}\left[\left(\frac{I[D_t = 0]Y_t}{p(D_t = 0|X_t)}\right)^2\right] - \theta^2 \\
&= \mathbb{E}\left[\left(\frac{I[D_t = 1](I[D_t = 1]Y_t(1) + I[D_t = 0]Y_t(0))}{p(D_t = 1|X_t)}\right)^2\right] \\
&\quad + \mathbb{E}\left[\left(\frac{I[D_t = 0](I[D_t = 1]Y_t(1) + I[D_t = 0]Y_t(0))}{p(D_t = 0|X_t)}\right)^2\right] - \theta^2 \\
&= \mathbb{E}\left[\left(\frac{I[D_t = 1]Y_t(1)}{p(D_t = 1|X_t)}\right)^2\right] + \mathbb{E}\left[\left(\frac{I[D_t = 0]Y_t(0)}{p(D_t = 0|X_t)}\right)^2\right] - \theta^2 \\
&= \mathbb{E}\left[\frac{Y_t^2(1)}{p(D_t = 1|X_t)} + \frac{Y_t^2(0)}{p(D_t = 0|X_t)}\right] - \theta^2.
\end{aligned}$$

Therefore, from the central limit theorem, the asymptotic distribution is given as follows:

$$\sqrt{T}(\hat{\theta}_T^{\mathrm{IPW}} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{Y_t^2(1)}{p(D=1|X_t)} + \frac{Y_t^2(0)}{p(D=0|X_t)} - \theta^2\right]$. □

2.3.2 Proof of Theorem 2.1.1

Proof. Here, we rewrite $f(p)$ as follows:

$$f(p) = \mathbb{E}\left[\mathbb{E}\left[\frac{Y_t^2(1)}{p(X_t)} + \frac{Y_t^2(0)}{1 - p(X_t)}\middle| X_t\right]\right] - \theta^2$$

We consider minimizing $f(p)$ by minimizing $\tilde{f}(q) = \mathbb{E}\left[\frac{Y_t^2(1)}{q} + \frac{Y_t^2(0)}{1-q}\middle| X_t\right]$ for $q \in [\varepsilon, 1 - \varepsilon]$. The first derivative of $\tilde{f}(q)$ with respect to q is given as follows:

$$\tilde{f}'(q) = -\frac{\mathbb{E}\left[Y_t^2(1)|X_t\right]}{q^2} + \frac{\mathbb{E}\left[Y_t^2(0)|X_t\right]}{(1-q)^2}.$$

The second derivative of f is given as follows:

$$\tilde{f}''(q) = 2\frac{\mathbb{E}\left[Y_t^2(1)|X_t\right]}{q^3} + 2\frac{\mathbb{E}\left[Y_t^2(0)|X_t\right]}{(1-q)^3}.$$

For $\varepsilon < q < 1-\varepsilon$, because $\tilde{f}''(q) > 0$, the minimizer q^* of $\tilde{f}$ satisfies the following equation:

$$-\frac{\mathbb{E}\left[Y_t^2(1)|X_t\right]}{(q^*)^2} + \frac{\mathbb{E}\left[Y_t^2(0)|X_t\right]}{(1-q^*)^2} = 0.$$

This equation is equivalent to

$$\begin{aligned}
& -(q^*)^2\mathbb{E}\left[Y_t^2(0)|X_t\right] + (1-q^*)^2\mathbb{E}\left[Y_t^2(1)|X_t\right] = 0 \\
\Leftrightarrow\ & q^*\sqrt{\mathbb{E}\left[Y_t^2(0)|X_t\right]} = (1-q^*)\sqrt{\mathbb{E}\left[Y_t^2(1)\right]} \\
\Leftrightarrow\ & q^* = \frac{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]} + \sqrt{\mathbb{E}[Y_t^2(0)|X_t]}}.
\end{aligned}$$

Therefore,

$$p^{\mathrm{OPT}}(D=1|X_t) = \frac{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]} + \sqrt{\mathbb{E}[Y_t^2(0)|X_t]}}.$$

□

Chapter 3

Statistical Inference from Samples with Dependency

In this chapter, we show how we conduct statistical inference when we admit changing the probability of assignment using past information.

3.1 Asymptotic Distribution of of an Estimator for Samples with Dependency

When observations are not i.i.d. sampled, we cannot apply the standard limit theorems to an estimator constructed with the samples. Therefore, to avoid this problem, we propose the following estimator under a policy Π:

$$\begin{aligned}\hat{\theta}^{\Pi}_T =&\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big/\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\\ &-\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\Big/\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}.\end{aligned} \tag{3.1}$$

For this estimator, we apply limit theorems for martingale. Here, we introduce the following random variables:

$$\tau_t=\begin{pmatrix}\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\\ \\ \frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\end{pmatrix},\ Z_t=t\begin{pmatrix}\mu_1\\ \\ \mu_0\end{pmatrix}-\sum_{s=1}^{t}\tau_s.$$

We show that the sequence $\{Z_s\}_{s=1}^{T}$ is martingale in the following lemma.

Lemma 3.1.1. The sequence $\{Z_s\}_{s=1}^{T}$ is martingale.

The proof is shown in Section 3.4.1. From the martingale sequence $\{Z_s\}_{s=1}^{T}$, we construct the martingale difference sequence as follows:

$$W_{s+1}=Z_{s+1}-Z_s=\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix}-\begin{pmatrix}\frac{I[D_{s+1}=1]Y_{s+1}}{\pi_t(X_{s+1},\mathcal{F}_s)}\\ \\ \frac{I[D_{s+1}=0]Y_{s+1}}{1-\pi_t(X_{s+1},\mathcal{F}_s)}\end{pmatrix}.$$

If we assume that the conditions of Proposition B.2 hold in Appendix, the asymptotic distribution of $\hat{\theta}_T$ is given as the following lemma.

Lemma 3.1.2. Suppose that $\pi_t(X_t, \mathcal{F}_{t-1}) \xrightarrow{a.s.} \alpha(X_t)$, where α is a function $\alpha : \mathbb{R}^d \to [\varepsilon, 1-\varepsilon]$, and $\frac{1}{\pi_t(X_t, \mathcal{F}_{t-1})}$ is uniformly integrable. Then,

$$\sqrt{T}(\hat{\theta}^{\Pi}_T - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$

where $\sigma^2 = \mathbb{E}_{(Y_t(1), X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] + \mathbb{E}_{(Y_t(0), X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] - \theta^2$.

The proof is shown in Section 3.4.2. From Lemma 3.1.2, an estimator defined by (3.1) enables us to derive the asymptotic distribution even in the case where the samples are not i.i.d.

3.2 Optimal Probability for Allocating Treatments

As shown in Section 3, the asymptotic variance is minimized when $\alpha(X_t)$ is given as

$$\alpha(X_t) = \frac{\sqrt{\mathbb{E}_{Y_t(1)}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}_{Y_t(1)}[Y_t^2(1)|X_t]} + \sqrt{\mathbb{E}_{Y_t(0)}[Y_t^2(0)|X_t]}}.$$

The estimator defined by (3.1) is asymptotically the same as the following estimator:

$$\hat{\theta}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})} - \frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t = 0]Y_t}{1 - \pi_t(X_t, \mathcal{F}_{t-1})}. \tag{3.2}$$

However, as it is well known in the research of estimation of the treatment effect, using $\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t, \mathcal{F}_{t-1})}$ and $\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t, \mathcal{F}_{t-1})}$ in denominators of each term improves the performance. Therefore, we use the estimator in (3.1) in practice, not (3.2). However, in the theoretical analysis for finite samples, we use estimator (3.1) for simplicity.

3.3 Doubly Robust Estimator

As a well known fact, when we have an estimator of $\mathbb{E}[Y_t(d)|X_t]$, we can construct a *doubly robust estimator* proposed by Bang and Robins (2005). In this thesis, based on the idea of Bang and Robins (2005), we construct a similar estimator defined as follows:

$$\begin{aligned}\hat{\theta}^{\mathrm{DR},\Pi}_T &= \frac{1}{\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t, \mathcal{F}_{t-1})}}\left(\sum_{t=1}^{T}\left(\frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t = 1]}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)z_1(X_t, \mathcal{F}_{t-1})\right)\right)\\ &- \frac{1}{\sum_{t=1}^{T}\frac{I[D_t=1]}{1-\pi_t(X_t, \mathcal{F}_{t-1})}}\left(\sum_{t=1}^{T}\left(\frac{I[D_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t = 0]}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right)z_0(X_t, \mathcal{F}_{t-1})\right)\right),\end{aligned} \tag{3.3}$$

where $z_d(X_t, \mathcal{F}_{t-1})$ is a function of a treatment $d \in \{0, 1\}$, context X_t, and information until $(t-1)$-th period. Here, we derive the asymptotic distribution of a doubly robust estimator $\hat{\theta}^{\mathrm{DR},\Pi}_T$ in the following lemma.

Lemma 3.3.1. Suppose that $z_d(X_t, \mathcal{F}_{t-1}) \xrightarrow{\text{a.s.}} z_d(X_t)$ and $\pi_t(X_t, \mathcal{F}_{t-1}) \xrightarrow{\text{a.s.}} \alpha(X_t)$ holds, where z is a function $z : \{1, 0\} \times \mathbb{R}^d \to \mathbb{R}$, α is a function $\alpha :$

$\mathbb{R}^d \to [\varepsilon, 1-\varepsilon]$, and $z(d, X_t)$ and $\frac{1}{\pi_t(X_t, \mathcal{F}_{t-1})}$ are uniformly integrable. Then, for an estimator defined in (3.3),

$$\sqrt{T}(\hat{\theta}_T^{\mathrm{DR},\Pi} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\mathrm{DR}}\right),$$

where

$$\begin{aligned}
\sigma^2_{\mathrm{DR}} =& \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] - \mu^2(1) + 2\mu(1)\mu(0) - \mu^2(0) \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1(X_t)z_1(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)z_1^2(X_t)\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0(X_t)z_0(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)z_0^2(X_t)\right] \\
&- 2\mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t)\right] - 2\mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t)\right] + 2\mathbb{E}_{X_t}\left[z_1(X_t)z_0(X_t)\right].
\end{aligned}$$

The proof is shown in Section 3.4.3. The variance σ^2_{DR} can be written as follows:

$$\begin{aligned}
\sigma^2_{\mathrm{DR}} =& \sigma^2 + 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1(X_t)z_1(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)z_1^2(X_t)\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0(X_t)z_0(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)z_0^2(X_t)\right] \\
&- 2\mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t)\right] - 2\mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t)\right] + 2\mathbb{E}_{X_t}\left[z_1(X_t)z_0(X_t)\right] \\
=& \sigma^2 + \tilde{\sigma},
\end{aligned}$$

where σ^2 is the variance of an IPW estimator and

$$\begin{aligned}
\tilde{\sigma} =& 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1(X_t)z_1(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)z_1^2(X_t)\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0(X_t)z_0(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)z_0^2(X_t)\right] \\
&- 2\mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t)\right] - 2\mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t)\right] + 2\mathbb{E}_{X_t}\left[z_1(X_t)z_0(X_t)\right].
\end{aligned}$$

The term $\tilde{\sigma}$ can be both positive and negative. For example, if $z_d(X_t) = 1$, $\mu(d) = 0$, and $\alpha(X_t) = 0.5$ for $d = 0, 1$, then $\tilde{\sigma} = 4 > 0$. However, if $z_d(\cdot) = \mu_1(\cdot)$, then we have

$$\begin{aligned}
\tilde{\sigma} =& \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1^2(X_t)\right] + \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0^2(X_t)\right] \\
&- 2\mathbb{E}_{X_t}\left[\mu_1(X_t)\mu_0(X_t)\right] \\
=& -\mathbb{E}_{X_t}\left[\frac{\mu_1^2(X_t)}{\alpha(X_t)}\right] - \mathbb{E}_{X_t}\left[\frac{\mu_0^2(X_t)}{1-\alpha(X_t)}\right] + \mathbb{E}_{X_t}\left[\left(\mu_1(X_t) - \mu_0(X_t)\right)^2\right]
\end{aligned}$$

and

$$\begin{aligned}
\sigma^2_{\mathrm{DR}} =& \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{\mathrm{Var}(Y_t(1)|X_t)}{\alpha(X_t)}\right] + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{\mathrm{Var}(Y_t(0)|X_t)}{1-\alpha(X_t)}\right] \\
&+ \mathrm{Var}(\mu_1(X_t) - \mu_0(X_t)). \qquad (3.4)
\end{aligned}$$

This asymptotic variance achieves the same value as the semi-parametric efficiency bound of a doubly robust estimator when, given the probability of assigning a treatment $\alpha(X_t)$, samples are i.i.d. as

$$(Y_t, D_t, X_t) \sim p(y|d, x)p(d = 1|x)p(x) = p(y|d, x)\alpha(x)p(x).$$

where $p(y|d,x)$ is the distribution of Y_t conditioned on a treatment d and context x, and $p(x)$ is the distribution of X_t (Robins et al., 1994). Therefore, when we have an estimator $z_d(X_t, \mathcal{F}_{t-1})$ that converges to $\mu_d(X_t)$ almost surely, we should use a doubly robust estimator. In this case, the probability of assigning a treatment that minimizes the asymptotic variance is given as follows:

$$p^{\mathrm{DR}}(D_t = 1|X_t) = \frac{\sqrt{\mathrm{Var}(Y_t(1)|X_t)}}{\sqrt{\mathrm{Var}(Y_t(1)|X_t)} + \sqrt{\mathrm{Var}(Y_t(0)|X_t)}}. \tag{3.5}$$

This means that we can make the asymptotic variance of an efficient estimator smaller by using the probability defined as (3.5). This probability is equal to the probability of assigning a treatment used in Hahn et al. (2011). However, compared with the case where samples are i.i.d. as Hahn et al. (2011), we need a condition such that $z_d(X_t, \mathcal{F}_{t-1}) \xrightarrow{\text{a.s.}} \mu_d(X_t)$. Besides, as explained above, misspecification of $z_d(X_t, \mathcal{F}_{t-1})$ may increase variance. Based on this discussion, we do not use a doubly robust estimator in the proposed algorithm. In practice, we can choose either $\hat{\theta}^{\Pi}_T$ or $\hat{\theta}^{\mathrm{DR},\Pi}_T$ depending on the situation.

Efficient IPW estimator: When samples are i.i.d. and we use an IPW estimator defined as (2.2), we can achieve the same asymptotic variance as (3.4) by replacing inverse probability of the estimator with a non-parametric estimator of the probability of assigning a treatment. Even though we know the true probability of assigning a treatment, we can make the asymptotic variance smaller by using an estimator of the probability of assigning a treatment (Hirano et al., 2003). However, in this thesis, we do not use such an IPW estimator with an estimator of the probability of assigning a treatment because we construct a martingale difference sequence from an IPW estimator with the true value of the probability of assigning a treatment.

3.4 Proofs

In this section, we prove Lemmas 3.1.1 and 3.1.2.

3.4.1 Proof of Lemma 3.1.1

Proof.

$$\begin{aligned}
&\mathbb{E}_{\Pi}\left[Z_{s+1}|\mathcal{F}_s\right]\\
&= \mathbb{E}_{\Pi}\left[(s+1)\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix} - \sum_{t=1}^{s+1}\tau_t \middle| \mathcal{F}_s\right]\\
&= \mathbb{E}_{\Pi}\left[\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix} - \tau_{s+1}\middle|\mathcal{F}_s\right] + s\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix} - \sum_{t=1}^{s}\tau_t\\
&= \mathbb{E}_X\left[\mathbb{E}_{\Pi}\left[\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix} - \tau_{s+1}\middle| X_{s+1}, \mathcal{F}_s\right]\middle|\mathcal{F}_s\right] + s\begin{pmatrix}\mu(1)\\ \\ \mu(0)\end{pmatrix} - \sum_{t=1}^{s}\tau_t\\
&= 0 + Z_s.
\end{aligned}$$

□

3.4.2 Proof of Lemma 3.1.2

In order to prove the lemma, we show the following three lemmas.

Lemma 3.4.1.

$$\mathbb{E}_\Pi\left[\left(\mu(1) - \frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)^2\right] = -\mu^2(1) + \mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t, \mathcal{F}_{t-1})}\right]$$
$$\mathbb{E}_\Pi\left[\left(\mu(0) - \frac{I[D_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right)^2\right] = -\mu^2(0) + \mathbb{E}_\Pi\left[\frac{Y_t^2(0)}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right]$$
$$\mathbb{E}_\Pi\left[\left(\mu(1) - \frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)\left(\mu(0) - \frac{I[d_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right)\right] = -\mu(1)\mu(0),$$

where recall that $\mathbb{E}_\Pi[\cdot]$ denotes the operator of the expectation under a policy Π.

Proof. The first equality is proved as follows:

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\left(\mu(1) - \frac{I[D_t = d]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)^2\right]\\
&= \mu^2(1) - 2\mu(1)\mathbb{E}_\Pi\left[\frac{I[D_t = d]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right] + \mathbb{E}_\Pi\left[\left(\frac{I[D_t = d]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)^2\right]\\
&= \mu^2(1) - 2\mu^2(1) + \mathbb{E}_\Pi\left[\left(\frac{I[D_t = d]Y_t(1)}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)^2\right]\\
&= \mu^2(1) - 2\mu^2(1) + \mathbb{E}_\Pi\left[\frac{I[D_t = d]Y_t^2(1)}{\pi_t^2(X_t, \mathcal{F}_{t-1})}\right]\\
&= \mu^2(1) - 2\mu^2(1) + \mathbb{E}_{(X_t, \mathcal{F}_{t-1})}\left[\frac{\pi_t(X_t, \mathcal{F}_{t-1})\mathbb{E}_\Pi[Y_t^2(1)|X_t, \mathcal{F}_t]}{\pi_t^2(X_t, \mathcal{F}_{t-1})}\right]\\
&= -\mu^2(1) + \mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t, \mathcal{F}_{t-1})}\right].
\end{aligned}$$

Similarly,

$$\mathbb{E}_\Pi\left[\left(\mu(0) - \frac{I[d_t = 0]Y_t}{1-\pi_t(\mathcal{F}_{t-1})}\right)^2\right] = -\mu^2(0) + \mathbb{E}_\Pi\left[\frac{Y_t^2(0)}{1-\pi_t(\mathcal{F}_{t-1})}\right].$$

The third equality is proved as follows:

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\left(\mu(1) - \frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)\left(\mu(0) - \frac{I[D_t = 0]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right)\right]\\
&= \mu(1)\mu(0) - \mu(1)\mathbb{E}_\Pi\left[\frac{I[d_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right]\\
&\quad - \mu(0)\mathbb{E}_\Pi\left[\frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\right] + \mathbb{E}_\Pi\left[\frac{I[D_t = 1]Y_t}{\pi_t(X_t, \mathcal{F}_{t-1})}\frac{I[D_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\right]\\
&= \mu(1)\mu(0) - \mu(1)\mu(0) - \mu(0)\mu(1) + 0\\
&= -\mu(1)\mu(0).
\end{aligned}$$

□

Lemma 3.4.2.

$$\frac{1}{T}\sum_{t=1}^{T}\Omega_t - \frac{1}{T}\sum_{t=1}^{T}W_tW_t^\top \xrightarrow{\text{a.s.}} \mathbf{0},$$

where

$$\Omega_t = \begin{pmatrix} -\mu^2(1) + \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0) + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] \end{pmatrix}.$$

Proof. Let us define $\gamma_t^{(1,1)}$ and $\gamma_t^{(0,0)}$ as follows:

$$\gamma_t^{(1,1)}$$
$$= -\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] - \left(\mu(1) - \frac{I[d_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)^2$$
$$\gamma_t^{(0,0)}$$
$$= -\mu^2(0) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] - \left(\mu(0) - \frac{I[d_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)^2.$$

Here, $\left\{S_t^{(1,1)} = \sum_{s=1}^t \gamma_s^{(1,1)}\right\}_{t=1}^T$ and $\left\{S_t^{(0,0)} = \sum_{s=1}^t \gamma_t^{(0,0)}\right\}_{t=1}^T$ are sequences of martingale. For example, for $S_s^{(1,1)}$ and $S_{s+1}^{(1,1)}$, the following equality holds:

$$\mathbb{E}\left[S_{s+1}^{(1,1)}\Big|\mathcal{F}_s\right]$$
$$= \mathbb{E}\left[\sum_{t=1}^{s+1}\left(-\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] - \left(\mu(1) - \frac{I[d_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)^2\right)\Big|\mathcal{F}_s\right]$$
$$= \sum_{t=1}^{s}\left(-\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right] - \left(\mu(1) - \frac{I[d_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)^2\right) = S_s^{(1,1)}.$$

Then, we apply the strong large numbers for martingale (Proposition B.1 in Appendix) to $\{S_t^{(1,1)}\}_{t=1}^T$. We check the following condition:

$$\sum_{t=1}^{\infty}\mathbb{E}[|\gamma_t^{(1,1)}|^{2a}]/t^{1+a} < \infty \quad \text{for } a \geq 1.$$

From our assumptions, $\gamma_t^{(1,1)}$ is bounded as follows:

$$-\mu^2(1)+\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}|\mathcal{F}_{t-1}\right] - \left(\max\left\{\mu(1), \frac{\overline{Y}(1)}{\varepsilon} - \mu(1)\right\}\right)^2$$
$$\leq \gamma_t^{(1,1)} \leq -\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\Big|\mathcal{F}_{t-1}\right].$$

Therefore, there exists a constant C such that

$$\sum_{t=1}^{\infty}\mathbb{E}\left[\left|\gamma_t^{(1,1)}\right|^{2a}\right]/t^{1+a} \leq \sum_{t=1}^{\infty} C^{2a}/t^{1+a}.$$

Then, we bound $\zeta(a) = \sum_{t=1}^{\infty} 1/t^{1+a}$, which is known as the Riemann zeta function. Given $a > 0$, $\zeta(a)$ is known to be a finite value. Therefore, the condition holds and

$$\sum_{s=1}^{t}\gamma_s^{(1,1)}/t \xrightarrow{\text{a.s.}} 0.$$

Hence,

$$\frac{1}{T}\sum_{t=1}^{T}\left(-\mu^2(1)+\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}|\mathcal{F}_{t-1}\right]-\left(\mu(1)-\frac{I[d_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)^2\right)$$
$$\xrightarrow{\text{a.s.}} 0.$$

Similarly,

$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{(0,0)} \xrightarrow{\text{a.s.}} 0$$
$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t^{(1,0)} \xrightarrow{\text{a.s.}} 0.$$

As a result, because each element of $\frac{1}{T}\sum_{t=1}^{T}\Omega_t - \frac{1}{T}\sum_{t=1}^{T}W_tW_t^{\top}$ converges almost surely to the matrix with each element being 0, the lemma holds. □

Lemma 3.4.3.

$$\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})} \xrightarrow{p} 1$$
$$\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})} \xrightarrow{p} 1.$$

Proof. As we showed in the proof of Lemma 3.4.2, because a sequence of random variables

$$\tau_t(1) = 1 - \frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}$$

satisfies the definition of martingale and $\tau_t(1)$ is bounded, we can apply Proposition B.1 in Appendix to $\tau_t(1)$. Therefore, $\frac{1}{T}\sum_{t=1}^{T}\tau_t(1) \xrightarrow{a.s.} 0$. Similarly, we define $\tau_t(0)$ as

$$\tau_t(1) = 1 - \frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}.$$

Then, $\frac{1}{T}\sum_{t=1}^{T}\tau_t(0) \xrightarrow{a.s.} 0$. □

Using Lemmas 3.4.1–3.4.3, and Proposition B.2 in Appendix, we prove Lemma 3.1.2.

Proof of Lemma 3.1.2. From Lemma 3.4.1,

$$\mathbb{E}_{\Pi}\left[W_t^{\top}W_t\right] = \begin{pmatrix} -\mu^2(1)+\mathbb{E}_{\Pi}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0)+\mathbb{E}_{\Pi}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right] \end{pmatrix}.$$

Let us denote

$$\widetilde{\Omega}_t = \begin{pmatrix} -\mu^2(1)+\mathbb{E}_{\Pi}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0)+\mathbb{E}_{\Pi}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right] \end{pmatrix}.$$

Let us define

$$\Omega = \begin{pmatrix} -\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] \end{pmatrix}.$$

First, we show that the condition (a) of Proposition B.2 in Appendix, i.e., as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[W_t^\top W_t\right] = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\Omega}_t \to \Omega.$$

Because of the assumption $\pi_t(\boldsymbol{x}, \mathcal{F}_{t-1}) \xrightarrow{a.s.} \alpha(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$, we can derive $\frac{1}{\pi_t(\boldsymbol{x}, \mathcal{F}_{t-1})} \xrightarrow{a.s.} \frac{1}{\alpha(\boldsymbol{x})}$ from the continuous mapping theorem (Proposition A.1 in Appendix). From the assumption that $\frac{1}{\pi_t(X_t, \mathcal{F}_{t-1})}$ is uniformly integrable and Proposition A.4 in Appendix, $\mathbb{E}_{\Pi}\left[\frac{1}{\pi_t(\boldsymbol{x}, \mathcal{F}_{t-1})}\right] \to \mathbb{E}_{\Pi}\left[\frac{1}{\alpha(\boldsymbol{x})}\right]$ as $t \to \infty$. Therefore, as $T \to \infty$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Omega}_t \\
&= \begin{pmatrix} -\mu^2(1) + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0) + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right] \end{pmatrix} \\
&\to \begin{pmatrix} -\mu^2(1) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0) + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] \end{pmatrix}.
\end{aligned}$$

Therefore, $\frac{1}{T}\sum_{t=1}^{T}\widetilde{\Omega}_t \to \Omega$.

Then, for $\{W_t\}_{t=1}^{T}$, the assumption (a) of Proposition B.2 in Appendix holds from Lemma 3.4.1 and the above result, the assumption (b) holds because the element of W_t is bounded, and the assumption (c) holds from Lemma 3.4.2 and the above result. Therefore,

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}W_t = \sqrt{T}\left(\begin{pmatrix}\mu(1) \\ \mu(0)\end{pmatrix} - \begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} \\ \frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Omega).$$

Next, we use the following result shown in Lemma 3.4.3:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})} &\xrightarrow{p} 1 \\
\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})} &\xrightarrow{p} 1.
\end{aligned}$$

From Proposition A.2 in Appendix,

$$\begin{aligned}
&\begin{pmatrix}\frac{1}{\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}} & -\frac{1}{\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}}\end{pmatrix}\sqrt{T}\left(\begin{pmatrix}\mu(1) \\ \mu(0)\end{pmatrix} - \begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} \\ \frac{1}{T}\sum_{t=1}^{T}\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\end{pmatrix}\right) \\
&\xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),
\end{aligned}$$

where the asymptotic variance is calculated as follows:

$$\sigma^2 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} -\mu^2(1) + \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] & -\mu(1)\mu(0) \\ -\mu(1)\mu(0) & -\mu^2(0) + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$
$$= \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] - \mu^2(1) + 2\mu(1)\mu(0) - \mu^2(0).$$

□

3.4.3 Proof of Lemma 3.3.1

We can use the same procedure of the proof of Lemma 3.1.2. Let a random variable τ_t^{DR} be

$$\tau_t^{\mathrm{DR}} = \begin{pmatrix} \frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_1(X_t,\mathcal{F}_{t-1}) \\ \frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_0(X_t,\mathcal{F}_{t-1}) \end{pmatrix}, \quad Z_t = t\begin{pmatrix} \mu_1 \\ \mu_0 \end{pmatrix} - \sum_{s=1}^{t} \tau_s^{\mathrm{DR}}$$

The sequence $\{Z_s^{\mathrm{DR}}\}_{s=1}^T$ is martingale. From the martingale sequence $\{Z_s^{\mathrm{DR}}\}_{s=1}^T$, we construct the martingale difference sequence as follows:

$$W_t^{\mathrm{DR}} = Z_t - Z_{t-1} = \begin{pmatrix} \mu(1) \\ \mu(0) \end{pmatrix} - \begin{pmatrix} \frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_1(X_t,\mathcal{F}_{t-1}) \\ \frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})} + \left(1 - \frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_0(X_t,\mathcal{F}_{t-1}) \end{pmatrix}$$

Using these sequences and assumptions shown in Lemma 3.3.1, we prove Lemma 3.3.1. In order to prove the lemma, we use the following two lemmas.

Lemma 3.4.4.

$$\mathbb{E}_\Pi\left[\left(\mu(1) - \frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} - \left(1 - \frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_1(X_t,\mathcal{F}_{t-1})\right)^2\right]$$
$$= -\mu^2(1) + \mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]$$
$$+ 2\mathbb{E}_\Pi\left[\left(1 - \frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_1(X_t)z_1(X_t,\mathcal{F}_{t-1})\right] - \mathbb{E}_\Pi\left[\left(1 - \frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1^2(X_t,\mathcal{F}_{t-1})\right]$$

$$\mathbb{E}_\Pi\left[\left(\mu(0) - \frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})} - \left(1 - \frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_0(X_t,\mathcal{F}_{t-1})\right)^2\right]$$
$$= -\mu^2(0) + \mathbb{E}_\Pi\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right]$$
$$+ 2\mathbb{E}_\Pi\left[\left(1 - \frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_0(X_t)z_0(X_t,\mathcal{F}_{t-1})\right]$$
$$- \mathbb{E}_\Pi\left[\left(1 - \frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0^2(X_t,\mathcal{F}_{t-1})\right]$$

$$\mathbb{E}_\Pi\Bigg[\left(\mu(1) - \frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})} - \left(1 - \frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_1(X_t,\mathcal{F}_{t-1})\right)$$
$$\left(\mu(0) - \frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})} - \left(1 - \frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right) z_0(X_t,\mathcal{F}_{t-1})\right)\Bigg]$$
$$= -\mu(1)\mu(0) + \mathbb{E}_\Pi\left[\mu_1(X_t)z_0(X_t,\mathcal{F}_{t-1})\right] + \mathbb{E}_\Pi\left[\mu_0(X_t)z_1(X_t,\mathcal{F}_{t-1})\right] - \mathbb{E}\left[z_1(X_t,\mathcal{F}_{t-1})z_0(X_t,\mathcal{F}_{t-1})\right],$$

where recall that $\mathbb{E}_\Pi[\cdot]$ denotes the operator of the expectation under a policy Π.

Proof. In this proof, we use the following properties:

$$\mathbb{E}_\Pi\left[\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]=0,$$
$$\mathbb{E}_\Pi\left[\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right]=0.$$

Then,

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\left(\mu(1)-\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}-\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right)^2\right]\\
&=\mu^2(1)+\mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]+\mathbb{E}_\Pi\left[\left(\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right)^2\right]\\
&\quad-2\mu(1)\mathbb{E}_\Pi\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]+2\mathbb{E}_\Pi\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]\\
&\quad-2\mathbb{E}_\Pi\left[\mu(1)\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]\\
&=\mu^2(1)+\mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]+\mathbb{E}_\Pi\left[\left(\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right)^2\right]\\
&\quad-2\mu^2(1)+2\mathbb{E}_\Pi\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]-0\\
&=\mu^2(1)+\mathbb{E}_\Pi\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&\quad+2\mathbb{E}_\Pi\left[\left(1-\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_1(X_t)z_1(X_t,\mathcal{F}_{t-1})\right]-\mathbb{E}_\Pi\left[\left(1-\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1^2(X_t,\mathcal{F}_{t-1})\right].
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\left(\mu(0)-\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}-\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right)^2\right]\\
&=-\mu^2(0)+\mathbb{E}_\Pi\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&\quad+2\mathbb{E}_\Pi\left[\left(1-\frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_0(X_t)z_0(X_t,\mathcal{F}_{t-1})\right]\\
&\quad-\mathbb{E}_\Pi\left[\left(1-\frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0^2(X_t,\mathcal{F}_{t-1})\right].
\end{aligned}$$

Finally,

$$
\begin{aligned}
&\mathbb{E}_{\Pi}\Bigg[\left(\mu(1)-\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}-\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right)\\
&\qquad\left(\mu(0)-\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}-\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right)\Bigg]\\
&=\mu(1)\mu(0)-\mathbb{E}_{\Pi}\left[\mu(1)\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right]-\mathbb{E}_{\Pi}\left[\mu(1)\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right]\\
&\quad-\mathbb{E}_{\Pi}\left[\mu(0)\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\right]+\mathbb{E}_{\Pi}\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\frac{I[D_t=0]Y_t}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&\quad+\mathbb{E}_{\Pi}\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right]\\
&\quad-\mathbb{E}_{\Pi}\left[\mu(0))\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]\\
&\quad+\mathbb{E}_{\Pi}\left[\frac{I[D_t=1]Y_t}{\pi_t(X_t,\mathcal{F}_{t-1})}\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\right]\\
&\quad+\mathbb{E}_{\Pi}\left[\left(1-\frac{I[D_t=1]}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1(X_t,\mathcal{F}_{t-1})\left(1-\frac{I[D_t=0]}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0(X_t,\mathcal{F}_{t-1})\right]\\
&=\mu(1)\mu(0)-\mu(1)\mu(0)-0-\mu(1)\mu(0)+0\\
&\quad+\mathbb{E}_{\Pi}\left[\mu_1(X_t)z_0(X_t,\mathcal{F}_{t-1})\right]-0+\mathbb{E}_{\Pi}\left[\mu_0(X_t)z_1(X_t,\mathcal{F}_{t-1})\right]+\mathbb{E}_{\Pi}\left[-z_1(X_t,\mathcal{F}_{t-1})z_0(X_t,\mathcal{F}_{t-1})\right]\\
&=-\mu(1)\mu(0)+\mathbb{E}_{\Pi}\left[\mu_1(X_t)z_0(X_t,\mathcal{F}_{t-1})\right]+\mathbb{E}_{\Pi}\left[\mu_0(X_t)z_1(X_t,\mathcal{F}_{t-1})\right]-\mathbb{E}\left[z_1(X_t,\mathcal{F}_{t-1})z_0(X_t,\mathcal{F}_{t-1})\right],
\end{aligned}
$$

□

Lemma 3.4.5.

$$
\frac{1}{T}\sum_{t=1}^{T}\Omega_t^{\mathrm{DR}}-\frac{1}{T}\sum_{t=1}^{T}W_t^{\mathrm{DR}}W_t^{\mathrm{DR}\top}\xrightarrow{\text{a.s.}}\mathbf{0},
$$

where

$$
\begin{aligned}
\Omega_t^{\mathrm{DR}}&=\begin{pmatrix}E & F\\ F & G\end{pmatrix},\\
E&=-\mu^2(1)+\mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}\middle|\mathcal{F}_{t-1}\right]\\
&\quad+2\mathbb{E}_{X_t}\left[\left(1-\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_1(X_t)z_1(X_t,\mathcal{F}_{t-1})\middle|\mathcal{F}_{t-1}\right]\\
&\quad-\mathbb{E}_{X_t}\left[\left(1-\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_1^2(X_t,\mathcal{F}_{t-1})\middle|\mathcal{F}_{t-1}\right]\\
F&=-\mu^2(0)+\mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\middle|\mathcal{F}_{t-1}\right]\\
&\quad+2\mathbb{E}_{X_t}\left[\left(1-\frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)\mu_0(X_t)z_0(X_t,\mathcal{F}_{t-1})\middle|\mathcal{F}_{t-1}\right]\\
&\quad-\mathbb{E}_{X_t}\left[\left(1-\frac{1}{1-\pi_t(X_t,\mathcal{F}_{t-1})}\right)z_0^2(X_t,\mathcal{F}_{t-1})\middle|\mathcal{F}_{t-1}\right]\\
G&=-\mu(1)\mu(0)+\mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t,\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\right]\\
&\quad+\mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t,\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\right]-\mathbb{E}_{X_t}\left[z_1(X_t,\mathcal{F}_{t-1})z_0(X_t,\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\right],
\end{aligned}
$$

Proof. Here, in addition to the assumptions of Lemma 3.1.2, we use the assumption $z_d(X_t, \mathcal{F}_{t-1}) \xrightarrow{\text{a.s}} z(d, X_t)$ and $z(d, X_t)$ is uniformly integrable. Then, we can prove the lemma using the same procedure as the proof of Lemma 3.4.2. □

Then, we prove Lemma 3.3.1 using Lemmas 3.4.3, 3.4.4, and 3.4.5.

Proof of Lemma 3.3.1. Let $\Omega^{\mathrm{DR}} = \begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix}$, where

$$\begin{aligned}
\tilde{E} =& -\mu^2(1) + \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1(X_t)z_1(X_t)\right] \\
&- \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)z_1^2(X_t)\right] \\
\tilde{F} =& -\mu^2(0) + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0(X_t)z_0(X_t)\right] \\
&- \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)z_0^2(X_t)\middle|\mathcal{F}_{t-1}\right] \\
\tilde{G} =& -\mu(1)\mu(0) + \mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t)\right] \\
&+ \mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t)\right] - \mathbb{E}_{X_t}\left[z_1(X_t)z_0(X_t)\right],
\end{aligned}$$

Under the assumptions and Lemmas 3.4.3, 3.4.4, and 3.4.5, we can use the same procedure as the proof of Lemma 3.1.2. We have

$$\sqrt{T}(\hat{\theta}_T^{\Pi,\mathrm{DR}} - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2_{\mathrm{DR}}),$$

where

$$\begin{aligned}
\sigma^2_{\mathrm{DR}} =& \begin{pmatrix} 1 & -1 \end{pmatrix}\Omega^{\mathrm{DR}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
=& \mathbb{E}_{(Y_t(1),X_t)}\left[\frac{Y_t^2(1)}{\alpha(X_t)}\right] + \mathbb{E}_{(Y_t(0),X_t)}\left[\frac{Y_t^2(0)}{1-\alpha(X_t)}\right] - \mu_1^2(X_t) + 2\mu_1(X_t)\mu_0(X_t) - \mu_0^2(X_t) \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)\mu_1(X_t)z_1(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{\alpha(X_t)}\right)z_1^2(X_t)\right] \\
&+ 2\mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)\mu_0(X_t)z_0(X_t)\right] - \mathbb{E}_{X_t}\left[\left(1 - \frac{1}{1-\alpha(X_t)}\right)z_0^2(X_t)\right] \\
&- 2\mathbb{E}_{X_t}\left[\mu_1(X_t)z_0(X_t)\right] - 2\mathbb{E}_{X_t}\left[\mu_0(X_t)z_1(X_t)\right] + 2\mathbb{E}_{X_t}\left[z_1(X_t)z_0(X_t)\right].
\end{aligned}$$

□

Chapter 4

Algorithm for Efficient Adaptive Randomization

Following the discussion of the previous chapters, we propose an algorithm to achieve an estimation with efficiency in this chapter. As we showed above, even when we use samples with dependency, we can minimize the asymptotic variance by setting the probability of selecting treatment 1 as $\frac{\sqrt{\mathbb{E}_{Y_t(1)}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}_{Y_t(1)}[Y_t^2(1)|X_t]}+\sqrt{\mathbb{E}[Y_t^2(0)|X_t]}}$. However, because we do not know the conditional means of squared outcomes

$$\mathbb{E}_{Y_t(1)}[Y_t^2(1)|X_t],\ \text{and}\ \mathbb{E}_{Y_t(0)}[Y_t^2(0)|X_t],$$

we consider estimating them using sequentially arriving samples. In the following parts, we show methods for estimating the conditional variance and the main algorithm.

4.1 Estimation of the Conditional Mean of the Squared Outcomes

For constructing the optimal choice probability, we consider estimating $\nu_d(X_t) = \mathbb{E}[Y_t^2(d)|X_t]$ non-parametrically. Let $\hat{\nu}_{t,d}(X_t)$ be an estimator of $\nu_d(X_t)$ at t-th period. In this thesis, we apply k_n-NN regression to construct the estimator $\hat{\nu}_{t.d}(X_t)$ with n samples by the following way proposed in Yang and Zhu (2002).

First, we fix $\boldsymbol{x} \in \mathbb{R}^d$. Let $k_n > 0$ be a value depending on the sample size n. Let $N_{t,d}$ be $\sum_{s=1}^{t} \mathbb{1}[D_s = d]$. At t-th round, we construct an estimator using the $k_{N_{t,d}}$-NN regression and samples $\{(X_{t'}, Y_{t'})\}_{t'=1}^{N_{t,d}}$ gathered when $D_{t'} = d$ as follows:

$$\hat{\nu}_{t,d}(\boldsymbol{x}) = \frac{1}{k_{N_{t,d}}} \sum_{i=1}^{k_n} Y'^2_{\pi(\boldsymbol{x},i)}, \tag{4.1}$$

where π is the permutation of $\{1, 2, ..., N_{t,d}\}$ such that

$$\|X'_{\pi(\boldsymbol{x},1)} - \boldsymbol{x}\| \leq \|X'_{\pi(\boldsymbol{x},2)} - \boldsymbol{x}\| \leq \cdots \leq \|X'_{\pi(\boldsymbol{x},N_{t,d})} - \boldsymbol{x}\|.$$

4.2 Algorithm for Adaptive Experimental Design via Bandit Feedback

When $\nu_d(X_t)$ takes values near 0 or 1, the proposed estimator takes extremely large values. Therefore, to stabilize the algorithm, we restrict the range of $\pi_t(X_t, \mathcal{F}_{t-1})$ as follows. We assumed that $Y_t(d)$ is bounded by $\underline{Y}$ and $\overline{Y}$. Under this assumption, the range of $\nu_d(X_t)$ becomes $[\underline{Y}^2, \overline{Y}^2]$. As we discussed above,

Algorithm 1

Parameter: $\rho \geq 0$, which is the number of samples that we assign treatments with equal probability.
Initialization:
At $t = 1, 2$, select $D_t = t$. Set $\pi_t(X_t, \mathcal{F}_{t-1}) = 1/2$.
for $t = 3$ to T **do**
  if $t < \rho$ **then**
    Set $\pi_t(X_t, \mathcal{F}_{t-1}) = 0.5$.
  else
    Construct estimators $\hat{\nu}_{t,1}(X_t)$ and $\hat{\nu}_{t,0}(X_t)$ with $k_{N_{t-1,d}}$-NN regression with $k_t = \sqrt{t}$.
    Let us modify $\hat{\nu}_{t,1}(X_t)$ as $\tilde{\nu}_{t.d}(\boldsymbol{x})$ in (4.2).
    Using $\tilde{\nu}_{t.d}(\boldsymbol{x})$, construct an estimator $\hat{p}^{\mathrm{OPT}}_t(D_t = 1|X_t)$ with (4.3).
    Set $\pi_t(X_t, \mathcal{F}_{t-1}) = \hat{p}^{\mathrm{OPT}}_t(D_t = 1|X_t)$.
  end if
  Draw ξ_t from the uniform distribution on $[0, 1]$.
  $D_t = \mathbb{1}[\xi_t \leq \pi_t(X_t, \mathcal{F}_{t-1})]$.
end for

the ideal value of $\pi_t(X_t, \mathcal{F}_{t-1})$ is $p^{\mathrm{OPT}}(D_t = 1|X_t) = \frac{\sqrt{\nu_1(X_t)}}{\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)}} \in [\varepsilon, 1-\varepsilon]$, where $\varepsilon = \frac{\overline{Y}}{\underline{Y}+\overline{Y}}$.

Using the above assumption, we introduce an algorithm for efficient estimation of the treatment effect. At t-th round, we estimate $\mathbb{E}[Y^2_t(d)|\boldsymbol{x}] = \nu_d(X_t)$ with $k_{N_{t,d}}$ nearest neighbors defined in (4.1), where $k_{N_{t,d}} = \sqrt{N_{t,d}}$. Because the estimators $\hat{\nu}_{t.d}(X_t)$ for $d = 1, 2$ can take extremely large or small values, we modify them using the prior knowledge about the upper and lower bound. Let us define modified estimators as

$$\tilde{\nu}_{t.d}(X_t) = \begin{cases} \overline{Y}^2 & \hat{\nu}_{t.d}(X_t) \geq \overline{Y}^2 \\ \hat{\nu}_{t.d}(X_t) & \underline{Y}^2 \leq \hat{\nu}_{t.d}(X_t) < \overline{Y}^2 \\ \underline{Y}^2 & \hat{\nu}_{t.d}(X_t) < \underline{Y}^2. \end{cases} \tag{4.2}$$

Then, using the modified estimators $\tilde{\nu}_{t.d}(X_t)$, we estimate the optimal probability as follows:

$$\hat{p}^{\mathrm{OPT}}_t(D_t = 1|X_t) = \frac{\sqrt{\tilde{\nu}_{t.1}(X_t)}}{\sqrt{\tilde{\nu}_{t.1}(X_t)} + \sqrt{\tilde{\nu}_{t.0}(X_t)}}. \tag{4.3}$$

Following this probability, we select a treatment of the t-th round.

Besides, we can consider assigning treatments to the first some samples with equal probability before using the proposed algorithm. This heuristics may stabilize our algorithm because it enables us to avoid critical estimation error of the mean of the squared outcome.

With this heuristics, we show the detail of the proposed algorithm in Algorithm 1.

Chapter 5

Theoretical Analysis

In this thesis, we give the theoretical analysis for the algorithm shown in the previous chapter from two viewpoints: asymptotic theory and regret analysis. In the following part, Theorem 5.2.1 gives us the asymptotic distribution of the estimator, which enables us to conduct a statistical test. The regret analysis in Theorem 5.3.1 guarantees the behavior when the the number of trials is finite.

5.1 Convergence of an Estimator of the Conditional Variance

In addition to the data generating process we defined above, we put the following three assumptions:

Assumption 5.1.1 (Yang and Zhu (2002), Eq. (5))**.** The function $\nu_d(\boldsymbol{x}_t)$ be continuous in X_t.

Let $\psi(z;\nu_d)$ be a modulus of continuity defined by

$$\psi(z;\nu_d)=\sup\left\{\left|\nu_d(x')-\nu_d(x'')\right|:|x'-x''|_\infty\leq z\right\}.$$

The term ψ represents the smoothness of the function ν_d.

Assumption 5.1.2 (Yang and Zhu (2002), Assumption 2)**.** The probability $p(x)$ is uniformly bounded above and away from 0 on $[0,1]^d$, i.e., $\underline{c}\leq p(x)\leq\overline{c}$.

Let us assume $(Y_t(d))^2=\nu_d(X_t)+\epsilon_{d,t}$, where $\epsilon_{d,t}$ is a random variable with mean 0 and a finite variable.

Assumption 5.1.3 (Yang and Zhu (2002), Assumption 3)**.** The error term $\epsilon_{d,t}$ also satisfies the moment condition such that there exist positive constants v and w satisfying, for all $m\geq 2$,

$$\mathbb{E}_{(Y_t(d),X_t)}[|\epsilon_{d,t}|^m]\leq\frac{m!}{2}v^2w^{m-2}.$$

Under these assumptions, we can show the following lemma from the result of Yang and Zhu (2002).

Lemma 5.1.1 (Yang and Zhu (2002), Eq. (4))**.** For $\kappa>0$, let $\eta_\kappa=\sup\{z:\psi(z;v_d)\leq\kappa\}$. There exists a constant $M>0$ such that, for $\kappa>0$, $h<\eta_{\kappa/4}$, and $k_t\leq\underline{c}th^k/2$,

$$\begin{aligned}&\mathbb{P}\left(|\hat{\nu}_{t,d}(\boldsymbol{x})-\nu_d(\boldsymbol{x})|\geq\kappa\right)\\&\leq M\exp\left(-\frac{3k_t}{14}\right)+\left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right)+\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right).\end{aligned}$$

According to Yang and Zhu (2002), for k_t such that $k_t\varepsilon^2/\log t \to \infty$ and $k_t = o(t)$, we can choose $h \to 0$ satisfying $h \geq (2k_t/(\underline{c}t))^{1/d}$. From the this discussion and the Borel-Cantelli lemma, we can show the following corollary (Yang and Zhu, 2002).

Corollary 5.1.1 (Yang and Zhu (2002))**.** For k_t such that $k_t\varepsilon^2/\log t \to \infty$ and $k_t = o(t)$, with probability 1,

$$|\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})| \to 0.$$

Besides, because our algorithm used $k_t = \sqrt{t}$ in our algorithm, which satisfies $k_t\varepsilon^2/\log t \to \infty$ and $k_t = o(t)$, the following corollary holds.

Corollary 5.1.2. For $k_t = \sqrt{t}$, there exists a constant $M > 0$ such that, for $t > \left(\frac{2}{\underline{c}\eta^k_{\kappa/4}}\right)^2$,

$$\begin{aligned}&\mathbb{P}\left(|\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})| \geq \kappa\right)\\ &\leq M\exp\left(-\frac{3k_t}{14}\right) + \left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right) + \exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right).\end{aligned}$$

Using there results, we can bound $\mathbb{E}_{\mathcal{F}_t}\left[\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})\right]$ by the following lemma.

Lemma 5.1.2. For $\kappa > 0$, $\eta_\kappa = \sup\{z : \psi(z; v_d) \leq \kappa\}$, $k_t = \sqrt{t}$, and $t > \left(\frac{2}{\underline{c}\eta^k_{\kappa/4}}\right)^2$, there exists a constant $M > 0$ such that

$$\begin{aligned}&\left|\mathbb{E}_\Pi\left[\sqrt{\hat{\nu}_{t,d}(\boldsymbol{x})} - \sqrt{\nu_d(\boldsymbol{x})}\right]\right|\\ &\leq \kappa + \frac{\overline{Y}^2}{2\underline{Y}}\Bigg(M\exp\left(-\frac{3k_t}{14}\right)\\ &\quad + \left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right) + \exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\Bigg).\end{aligned}$$

The proof of Lemma 5.1.2 is shown in Section 5.4.1.

5.2 Analysis for Infinite Samples

In order to derive the asymptotic distribution, we need to check whether $\pi_t(X_t, \mathcal{F}_{t-1})$ converges to $p^{\mathrm{OPT}}(D = 1|X_t)$ almost surely and $\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}$ is uniformly integrable.

We showed $\hat{\nu}_{t,d}(\boldsymbol{x}) \xrightarrow{a.s.} \nu_d(\boldsymbol{x})$ in Corollary 5.1.1. Therefore, we can derive the following corollary.

Corollary 5.2.1.

$$\pi_t(X_t, \mathcal{F}_{t-1}) \xrightarrow{\text{a.s.}} p^{\mathrm{OPT}}(D_t = 1|X_t).$$

Then, we show that $\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}$ is uniformly integrable.

Lemma 5.2.1. $\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}$ is uniformly integrable.

Proof. Because $\hat{\nu}_{t,1}(X_t)$ and $\hat{\nu}_{t,0}(X_t)$ are bounded,

$$\frac{1}{\pi_t(X_t, \mathcal{F}_{t-1})} = \frac{\sqrt{\hat{\nu}_{t-1,1}(X_t)} + \sqrt{\hat{\nu}_{t-1,0}(X_t)}}{\sqrt{\hat{\nu}_{t-1,1}(X_t)}}$$

is also bounded. From Proposition A.3 in Appendix, $\frac{1}{\pi_t(X_t,\mathcal{F}_{t-1})}$ is uniformly integrable. $\square$

From Corollary 5.2.1 and Lemma 5.2.1, we have the following theorem.

Theorem 5.2.1. The asymptotic distribution of $\hat{\theta}_T$ is given as follows:

$$\sqrt{T}(\hat{\theta}^{\Pi}_T - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$

where $\sigma^2 = \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y^2_t(1)}{p^{\mathrm{OPT}}(D_t=1|X_t)}\right] + \mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y^2_t(0)}{p^{\mathrm{OPT}}(D_t=0|X_t)}\right] - \theta^2$.

Proof. From Lemma 5.2.1, the condition of Lemma 3.1.2 hold. Therefore, we can directly use the asymptotic distribution that we derived in Lemma 3.1.2 directly. $\square$

5.3 Analysis for Finite Samples

In this section, we analyze the behavior of the estimator with finite samples. In this scenario, we cannot apply the asymptotic theory. Therefore, we introduce the framework of *regret analysis*, which is used in the research of on-line algorithm such as multi-armed bandit problem. Throughout this section, for the convenience of theoretical analysis, we simplify the estimator as

$$\hat{\theta}^{\Pi}_T = \sum_{t=1}^{T}\frac{I[D_t = 1]Y_t}{\pi_t(X_t\mathcal{F}_{t-1})}\Big/T - \sum_{t=1}^{T}\frac{I[D_t = 0]Y_t}{1-\pi_t(X_t, \mathcal{F}_{t-1})}\Big/T.$$

This omit the denominators, which converges 1 in probability, from the estimator (3.1).

Regret Analysis: As well as the standard analysis of multi-armed bandit problem, we define the regret between algorithms. In this thesis, we define the regret based on the mean squared error (MSE). Let us define the MSE of an algorithm Π as $\mathbb{E}_{\Pi}\left[\left(\theta - \theta^{\Pi}\right)^2\right]$, where $\mathbb{E}_{\Pi}[\cdot]$ denotes the operator of the expectation under a policy Π. Then, we can define the regret between the MSEs of the optimal policy Π^{OPT} and a policy Π. Let us denote an estimator given by the optimal policy as $\theta^{\Pi^{\mathrm{OPT}}}_T$. Then, the MSE of $\theta^{\Pi^{\mathrm{OPT}}}_T$ is calculated as follows:

$$\begin{aligned}&\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta - \hat{\theta}^{\Pi^{\mathrm{OPT}}}_T\right)^2\right]\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\frac{Y^2_t(1)}{p^{\mathrm{OPT}}(D_t = 1|X_t)} + \frac{Y^2_t(0)}{1-p^{\mathrm{OPT}}(D_t = 1|X_t)} - \theta^2\right].\end{aligned}$$

The regret of a policy Π is defined as follows:

$$\texttt{regret} = \mathbb{E}_{\Pi}\left[\left(\theta - \hat{\theta}^{\Pi}_T\right)^2\right] - \mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta - \hat{\theta}^{\Pi^{\mathrm{OPT}}}_T\right)^2\right].$$

Under our algorithm, the regret is bounded as the following theorem.

Theorem 5.3.1. Assume that there exist positive constants v and w satisfying, for all $m \geq 2$,

$$\mathbb{E}_{(Y_t(d),X_t)}[|\epsilon_{d,t}|^m] \leq \frac{m!}{2}v^2 w^{m-2}.$$

Then, for a constant $\kappa > 0$ and $\eta_\kappa = \sup\{z : \psi(z; v_d) \leq \kappa\}$, there exists M such that

$$\begin{aligned}
&\texttt{regret} = \mathbb{E}_\Pi\left[\left(\theta - \hat{\theta}_T^{\Pi}\right)^2\right] - \mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta - \hat{\theta}_T^{\Pi^{\mathrm{OPT}}}\right)^2\right]\\
&\leq \frac{2\overline{Y}^2\kappa'}{4T\underline{Y}^4} + \frac{2\overline{Y}^4}{4T^2\underline{Y}^4}\Bigg(\frac{392M}{9} + 2^{4d+13}\left(\tfrac{7}{3}\right)^{2(d+3)}\varepsilon^{-2(d+3)}\Gamma(2d+6)\\
&\quad + \frac{1568}{9\varepsilon^2} + 2^{8d+25}3^{-2(d+3)}\Gamma(2d+6)\left(\tfrac{\varepsilon^2\kappa^2}{v^2+w\varepsilon\kappa/4}\right)^{-2(d+3)} + \frac{512(v^2+w\varepsilon\kappa/4)^2}{9\varepsilon^4\kappa^4}\Bigg),
\end{aligned}$$

where $\kappa' = \frac{1}{T}\left[\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right] + \left(1 - \frac{1}{T}\left[\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right]\right)\kappa$

The proof is shown in Section 5.4.2.

Remark 1. This result tells us that the regret is bounded by $\mathrm{o}(1/T)$ by taking $\kappa = T^{-1/4}$ under some regularity conditions of the smoothness of ν_d. In contrast, if we use a constant value for the probability of assigning a treatment, the regret becomes $\mathrm{O}(1/T)$.

5.4 Proofs

In this section, we prove Lemma 5.1.2 and Theorem 5.3.1.

5.4.1 Proof of Lemma 5.1.2

Proof. For $\kappa > 0$, $\eta_\kappa = \sup\{z : \psi(z; v_d) \leq \kappa\}$, and $t > \left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2$,

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\sqrt{\hat{\nu}_{t,d}(\boldsymbol{x})} - \sqrt{\nu_d(\boldsymbol{x})}\right]\\
&= \mathbb{E}_\Pi\left[\frac{1}{\sqrt{\hat{\nu}_{t,d}(\boldsymbol{x})} + \sqrt{\nu_d(\boldsymbol{x})}}\left(\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})\right)\right]\\
&\leq \mathbb{E}_\Pi\left[\frac{1}{2\underline{Y}}\left|\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})\right|\right]\\
&\leq \kappa + \frac{\overline{Y}^2}{2\underline{Y}}\mathbb{P}\left(\left|\hat{\nu}_{t,d}(\boldsymbol{x}) - \nu_d(\boldsymbol{x})\right| \geq \kappa\right)\\
&\leq \kappa + \frac{\overline{Y}^2}{2\underline{Y}}\left(M\exp\left(-\frac{3k_T}{14}\right) + \left(T^{d+2}+1\right)\left(\exp\left(-\frac{3k_T\varepsilon}{28}\right) + \exp\left(-\frac{k_T\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}_\Pi\left[\sqrt{\nu_d(\boldsymbol{x})} - \sqrt{\hat{\nu}_{t,d}(\boldsymbol{x})}\right]\\
&\leq \kappa + \frac{\overline{Y}^2}{2\underline{Y}}\Bigg(M\exp\left(-\frac{3k_T}{14}\right)\\
&\quad + \left(T^{d+2}+1\right)\left(\exp\left(-\frac{3k_T\varepsilon}{28}\right) + \exp\left(-\frac{k_T\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\Bigg).
\end{aligned}$$

$\square$

5.4.2 Proof of Theorem 5.3.1

Proof.

$$(\theta - \hat{\theta}_T^{\Pi})^2 = \left(\frac{1}{T}\theta - \frac{1}{T}\tau_1 + \cdots + \frac{1}{T}\theta - \frac{1}{T}\tau_T\right)^2 = \frac{1}{T^2}\left(\theta - \tau_1 + \cdots + \theta - \tau_T\right)^2$$

Let δ_t be $\theta - \tau_t$. Then,

$$\begin{aligned}
&\mathbb{E}\left[(\theta - \hat{\theta}_T)^2\right] \\
&= \frac{1}{T^2}\mathbb{E}_{\Pi}\left[\left(\sum_{t=1}^{T}\delta_t\right)^2\right] \\
&= \frac{1}{T^2}\mathbb{E}_{\Pi}\left[\sum_{t=1}^{T}\delta_t^2 + 2\sum_{t=1}^{T}\sum_{s=1}^{t-1}\delta_t\tau_s\right] \\
&= \frac{1}{T^2}\mathbb{E}_{\Pi}\left[\sum_{t=1}^{T}\delta_t^2\right] + 2\mathbb{E}_{\Pi}\left[\sum_{t=1}^{T}\sum_{s=1}^{t-1}\delta_t\delta_s\right].
\end{aligned}$$

We use the following result:

$$\begin{aligned}
&\mathbb{E}_{\Pi}\left[\sum_{t=1}^{T}\sum_{s=1}^{t-1}\delta_t\delta_s\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{t-1}\mathbb{E}_{\mathcal{F}_{t-1}}\left[\mathbb{E}_{\Pi|\mathcal{F}_{t-1}}\left[\delta_t\delta_s|\mathcal{F}_{t-1}\right]\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{t-1}\mathbb{E}_{\mathcal{F}_{t-1}}\left[\mathbb{E}_{\Pi|\mathcal{F}_{t-1}}\left[\delta_t|\mathcal{F}_{t-1}\right]\delta_s\right] \\
&= \sum_{t=1}^{T}\sum_{s=1}^{t-1}\mathbb{E}_{\mathcal{F}_{t-1}}\left[0\delta_s\right] = 0.
\end{aligned}$$

Therefore,

$$\mathbb{E}_{\Pi}\left[(\theta - \hat{\theta}_T^{\Pi})^2\right] = \frac{1}{T^2}\mathbb{E}_{\Pi}\left[\sum_{t=1}^{T}\delta_t^2\right] = \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[\delta_t^2\right].$$

Then,

$$\begin{aligned}
&\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta - \hat{\theta}_T^{\mathrm{OPT}}\right)^2\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t = 1|X_t)} + \frac{Y_t^2(0)}{1 - p^{\mathrm{OPT}}(D_t = 1|X_t)} - \theta^2\right].
\end{aligned}$$

Because the randomness of $\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t=1|X_t)} + \frac{Y_t^2(0)}{1-p^{\mathrm{OPT}}(D_t=1|X_t)} - \theta^2$ only depends on Y_t and X_t,

$$\begin{aligned}
&\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta - \hat{\theta}_T^{\mathrm{OPT}}\right)^2\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t = 1|X_t)} + \frac{Y_t^2(0)}{1 - p^{\mathrm{OPT}}(D_t = 1|X_t)} - \theta^2\right].
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}_{\Pi}\left[\left(\theta-\hat{\theta}_T\right)^2\right]-\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta-\hat{\theta}_T^{\mathrm{OPT}}\right)^2\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[\delta_t^2\right]\\
&\quad-\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t=1|X_t)}+\frac{Y_t^2(0)}{1-p^{\mathrm{OPT}}(D_t=1|X_t)}-\theta^2\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\Pi}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}+\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}-\theta^2\right]\\
&\quad-\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(Y_t(1),Y_t(0),X_t)}\left[\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t=1|X_t)}+\frac{Y_t^2(0)}{1-p^{\mathrm{OPT}}(D_t=1|X_t)}-\theta^2\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\mathbb{E}_{(Y_t(1),Y_t(0))}\left[\frac{Y_t^2(1)}{\pi_t(X_t,\mathcal{F}_{t-1})}+\frac{Y_t^2(0)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}-\theta^2\Big|X_t,\mathcal{F}_{t-1}\right]\right]\\
&\quad-\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\mathbb{E}_{(Y_t(1),Y_t(0))}\left[\frac{Y_t^2(1)}{p^{\mathrm{OPT}}(D_t=1|X_t)}+\frac{Y_t^2(0)}{1-p^{\mathrm{OPT}}(D_t=1|X_t)}-\theta^2\Big|X_t,\mathcal{F}_{t-1}\right]\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\nu_1(X_t)}{\pi_t(X_t,\mathcal{F}_{t-1})}+\frac{\nu_0(X_t)}{1-\pi_t(X_t,\mathcal{F}_{t-1})}-\theta^2\right]\\
&\quad-\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)}+\frac{\nu_0(X_t)}{1-p^{\mathrm{OPT}}(D_t=1|X_t)}-\theta^2\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&\quad+\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\left((1-\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=0|X_t)\right)\nu_0(X_t)}{p^{\mathrm{OPT}}(D_t=0|X_t)(1-\pi_t(X_t,\mathcal{F}_{t-1}))}\right].
\end{aligned}
$$

Now, we focus on bounding $\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right]$. The term $\left(\frac{(\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right)$ is bounded as

$$
\begin{aligned}
&\frac{(\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(\mathcal{F}_{t-1})}\\
&=\frac{\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)}\left(\frac{\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t)}{\pi_t(\mathcal{F}_{t-1})}\right)\\
&\le\frac{\nu_1(X_t)}{\varepsilon p^{\mathrm{OPT}}(D_t=1|X_t)}\left(\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t)\right).
\end{aligned}
$$

Then, we bound $\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right]$ as follows:

$$
\begin{aligned}
&\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&\le\frac{\nu_1(X_t)}{\varepsilon p^{\mathrm{OPT}}(D_t=1|X_t)}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t)\right].
\end{aligned}
$$

Here, $\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\pi_t(X_t,\mathcal{F}_{t-1}) - p^{\mathrm{OPT}}(D_t=1|X_t)\right]$ is bounded as follows:

$$
\begin{aligned}
&\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\pi_t(X_t,\mathcal{F}_{t-1}) - p^{\mathrm{OPT}}(D_t=1|X_t)\right]\\
&=\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})\sqrt{\tilde{\nu}_{t,1}(X_t)} - (\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})\sqrt{\nu_1(X_t)}}{(\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})}\right]\\
&=\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\sqrt{\nu_0(X_t)}\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)}\sqrt{\nu_1(X_t)}}{(\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})}\right]\\
&=\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\sqrt{\nu_0(X_t)}\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\nu_0(X_t)}\sqrt{\nu_1(X_t)} + \sqrt{\nu_0(X_t)}\sqrt{\nu_1(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)}\sqrt{\nu_1(X_t)}}{(\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})}\right]\\
&=\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\sqrt{\nu_0(X_t)}(\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\nu_1(X_t)}) + \sqrt{\nu_1(X_t)}(\sqrt{\nu_0(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)})}{(\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})}\right]
\end{aligned}
$$

Because, from the assumption, $\sqrt{\nu_d(X_t)} = \sqrt{\mathbb{E}[Y_t^2(d)|X_t]} \in [\underline{Y}, \overline{Y}]$, where $\underline{Y} > 0$,

$$
\begin{aligned}
&\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\sqrt{\nu_0(X_t)}(\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\nu_1(X_t)}) + \sqrt{\nu_1(X_t)}(\sqrt{\nu_0(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)})}{(\sqrt{\tilde{\nu}_{t,1}(X_t)}+\sqrt{\tilde{\nu}_{t,0}(X_t)})(\sqrt{\nu_1(X_t)}+\sqrt{\nu_0(X_t)})}\right]\\
&\leq \mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\overline{Y}((\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\nu_1(X_t)}) + (\sqrt{\nu_0(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)}))}{4\underline{Y}^2}\right].
\end{aligned}
$$

From the convergence rate of k_n-NN estimator shown in Lemma 5.1.1, for a constant $\kappa > 0$,

$$
\begin{aligned}
&\leq \mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\overline{Y}((\sqrt{\tilde{\nu}_{t,1}(X_t)} - \sqrt{\nu_1(X_t)}) + (\sqrt{\nu_0(X_t)} - \sqrt{\tilde{\nu}_{t,0}(X_t)}))}{4\underline{Y}^2}\right]\\
&\leq \frac{\overline{Y}}{4\underline{Y}^2}\Bigg(\kappa + \frac{\overline{Y}^2}{\underline{Y}}\Bigg(M\exp\left(-\frac{3k_T}{14}\right)\\
&\quad + \left(T^{d+2}+1\right)\left(\exp\left(-\frac{3k_T\varepsilon}{28}\right) + \exp\left(-\frac{k_T\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\Bigg)\Bigg).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\left((1-\pi_t(X_t,\mathcal{F}_{t-1}) - p^{\mathrm{OPT}}(D_t=0|X_t)\right)\nu_0(X_t)}{p^{\mathrm{OPT}}(D_t=0|X_t)(1-\pi_t(X_t,\mathcal{F}_{t-1}))}\right]\\
&\leq \frac{\overline{Y}}{4\underline{Y}^2}\Bigg(\kappa + \frac{\overline{Y}^2}{\underline{Y}}\Bigg(M\exp\left(-\frac{3k_t}{14}\right)\\
&\quad + \left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right) + \exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\Bigg)\Bigg).
\end{aligned}
$$

Therefore,

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{(\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1|X_t))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1|X_t)\pi_t(X_t,\mathcal{F}_{t-1})}\right]\\
&+\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\frac{\left((1-\pi_t(X_t,\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=0|X_t)\right)\nu_0(X_t)}{p^{\mathrm{OPT}}(D_t=0|X_t)(1-\pi_t(X_t,\mathcal{F}_{t-1}))}\right]\\
\leq&\frac{2\overline{Y}}{4T^2\underline{Y}^2}\sum_{t=1}^{\left\lceil\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right\rceil}1+\frac{2\overline{Y}}{4T\underline{Y}^2}\sum_{t=\left\lceil\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right\rceil}^{T}\left(\kappa+\frac{\overline{Y}^2}{2\underline{Y}}\left(M\exp\left(-\frac{3k_t}{14}\right)\right.\right.\\
&\left.\left.+\left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right)+\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\right)\right)\\
\leq&\frac{2\overline{Y}}{4T\underline{Y}^2}\kappa'+\frac{2\overline{Y}}{4T^2\underline{Y}^2}\sum_{t=1}^{T}\left(\frac{\overline{Y}^2}{2\underline{Y}}\left(M\exp\left(-\frac{3k_t}{14}\right)\right.\right.\\
&\left.\left.+\left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right)+\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\right)\right)\\
\leq&\frac{2\overline{Y}}{4T\underline{Y}^2}\kappa'+\frac{2\overline{Y}}{4T^2\underline{Y}^2}\sum_{t=1}^{\infty}\left(\frac{\overline{Y}^2}{2\underline{Y}}\left(M\exp\left(-\frac{3k_t}{14}\right)\right.\right.\\
&\left.\left.+\left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right)+\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\right)\right),
\end{aligned}$$

where $\kappa'=\frac{1}{T}\left\lceil\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right\rceil+\left(1-\frac{1}{T}\left\lceil\left(\frac{2}{\underline{C}\eta^k_{\kappa/4}}\right)^2\right\rceil\right)\kappa$.

Here, because $k_t=\sqrt{t}$,

$$\sum_{t=1}^{\infty}\exp\left(-\frac{3\sqrt{t}}{14}\right)\leq\int_0^\infty\exp\left(-\frac{3\sqrt{t}}{14}\right)dt=\frac{392}{9}$$

$$\begin{aligned}
\sum_{t=1}^{\infty}t^{d+2}\exp\left(-\frac{3\sqrt{t}\varepsilon}{28}\right)&\leq\int_0^\infty t^{d+2}\exp\left(-\frac{3\sqrt{t}\varepsilon}{28}\right)dt\\
&=2^{4d+13}\left(\tfrac{7}{3}\right)^{2(d+3)}\varepsilon^{-2(d+3)}\Gamma(2d+6),
\end{aligned}$$

$$\sum_{t=1}^{\infty}\exp\left(-\frac{3\sqrt{t}\varepsilon}{28}\right)\leq\int_0^\infty\exp\left(-\frac{3\sqrt{t}\varepsilon}{28}\right)dt=\frac{1568}{9\varepsilon^2},$$

$$\begin{aligned}
\sum_{t=1}^{\infty}t^{d+2}\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)&\leq\int_0^\infty t^{d+2}\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)dt\\
&=2^{8d+25}3^{-2(d+3)}\Gamma(2d+6)\left(\tfrac{\varepsilon^2\kappa^2}{v^2+w\varepsilon\kappa/4}\right)^{-2(d+3)},
\end{aligned}$$

$$\begin{aligned}
\sum_{t=1}^{\infty}\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)&\leq\int_0^\infty\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)dt\\
&=\frac{512(v^2+w\varepsilon\kappa/4)^2}{9\varepsilon^4\kappa^4}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_{\Pi}\left[\left(\theta-\hat{\theta}_T^{\Pi}\right)^2\right]-\mathbb{E}_{\Pi^{\mathrm{OPT}}}\left[\left(\theta-\hat{\theta}_T^{\mathrm{OPT}}\right)^2\right]\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\left(\frac{(\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=1))\nu_1(X_t)}{p^{\mathrm{OPT}}(D_t=1)\pi_t(\mathcal{F}_{t-1})}\right)\right]\\
&\quad+\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{(X_t,\mathcal{F}_{t-1})}\left[\left(\frac{\left((1-\pi_t(\mathcal{F}_{t-1})-p^{\mathrm{OPT}}(D_t=0)\right)\nu_0(X_t)}{p^{\mathrm{OPT}}(D_t=0)(1-\pi_t(\mathcal{F}_{t-1}))}\right)\right]\\
&\le\frac{2\overline{Y}}{4T\underline{Y}^2}\kappa'+\frac{2\overline{Y}^3}{4T^2\underline{Y}^3}\left(M\frac{392}{9}+\left(t^{d+2}+1\right)\left(\exp\left(-\frac{3k_t\varepsilon}{28}\right)+\exp\left(-\frac{k_t\varepsilon^2\kappa^2}{16(v^2+w\varepsilon\kappa/4)}\right)\right)\right)\\
&\le\frac{2\overline{Y}\kappa'}{4T\underline{Y}^2}\\
&\quad+\frac{2\overline{Y}^3}{4T^2\underline{Y}^3}\left(\frac{392M}{9}+2^{4d+13}\left(\tfrac{7}{3}\right)^{2(d+3)}\varepsilon^{-2(d+3)}\Gamma(2d+6)\right.\\
&\quad\left.+\frac{1568}{9\varepsilon^2}+2^{8d+25}3^{-2(d+3)}\Gamma(2d+6)\left(\tfrac{\varepsilon^2\kappa^2}{v^2+w\varepsilon\kappa/4}\right)^{-2(d+3)}+\frac{512(v^2+w\varepsilon\kappa/4)^2}{9\varepsilon^4\kappa^4}\right).
\end{aligned}$$

□

Chapter 6

Experiments

In this chapter, we show the performance of the proposed algorithm empirically. In the experiments, we show how the proposed algorithm works compared with an RCT with $p(D_t = 1|X_t) = 0.5$ and two ideal algorithms with information of the population. We use two types of dataset: synthetic datasets and semi-synthetic datasets.

Experiments with Synthetic Dataset: First, we used the following synthetic datasets. We generated a covariate $\boldsymbol{X}_t \in \mathbb{R}^5$ at each round as follows:

$$\boldsymbol{X}_t = (X_{t1}, X_{t2}, X_{t3}, X_{t4}, X_{t5})^\top,$$
$$X_{tk} \sim \mathcal{N}(0, 10) \quad \text{for } k = 1, 2, 3, 4, 5,$$

where $\mathcal{N}(0, \sigma^2)$ denotes the normal distribution with the mean 0 and the variance σ^2. In this experiment, we considered the following model of a potential outcome:

$$Y_t(d) = \mu_d + \sum_{k=1}^{5} X_{tk} + e_{td},$$

where μ_d is a constant and e_{td} is the error term. The error term e_{td} follows the normal distribution and we denote the standard deviation as std_d. We used six datasets with different μ_d and std_d, Datasets 1–6, with 200 rounds (samples). For Datasets 1–3, we set $\mu_0 = \mu_1 = 0$. For Dataset 1, we set $\mathrm{std}_0 = 60$ and $\mathrm{std}_1 = 80$. For Dataset 2, we set $\mathrm{std}_0 = 40$ and $\mathrm{std}_1 = 80$. For Dataset 3, we set $\mathrm{std}_0 = 20$ and $\mathrm{std}_1 = 80$. For Datasets 4–5, we set $\mathrm{std}_0 = 60$ and $\mathrm{std}_1 = 80$. For Dataset 4, we set $\mu_0 = 100$ and $\mu_1 = 100$. For Dataset 5, we set $\mu_0 = 80$ and $\mu_1 = 100$. For Dataset 6, we set $\mu_0 = 60$ and $\mu_1 = 100$. From the samples generated from the above settings, we estimated the treatment effect using the following five policies: an RCT with $p(D_t = 1|X_t) = 0.5$ for all rounds, the proposed algorithm with $\rho = 10$, the proposed algorithm with $\rho = 50$, an RCT with $p(D_t = 1|X_t) = \frac{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}}{\sqrt{\mathbb{E}[Y_t^2(1)|X_t]}+\sqrt{\mathbb{E}[Y_t^2(0)|X_t]}}$ for all rounds, and an RCT with $p(D_t = 1|X_t) = \frac{\sqrt{\mathrm{Var}(Y_t(1)|X_t)}}{\sqrt{\mathrm{Var}(Y_t(1)|X_t)}+\sqrt{\mathrm{Var}(Y_t(0)|X_t)}}$ for all rounds. We refer the first naive randomized trail as RCT, the second adaptive randomization as ADPT1, the third adaptive randomization as ADPT2, the fifth randomized trial as OPT1, and sixth randomized trial as OPT2 in tables and figures of the result. We ran 1000 independent trials for each setting. The results of experiment for Datasets 1–3 are shown in Table 6.1 and Figure 6.1. The results for Datasets 4–6 are shown in Table 6.2 and Figure 6.2. In Tables 6.1 and 6.2, we show the mean of the error $\theta - \hat{\theta}$, and the standard deviation of $\hat{\theta}$ at the 100th (mid) round and the 200th (final) round. In Figure 6.1 and Figure 6.2, we plotted the mean squared error of $\hat{\theta}$ and the 5% and 95% quantiles of the squared error.

Table 6.1: Experimental results using Datasets 1–3. The mean of the error $\theta - \hat{\theta}$, and the standard deviation of $\hat{\theta}$ are shown at the 100-th (mid) round and the 200-th (final) round. The lowest standard deviations among RCT, ADPT1, and ADPT2 are bold.

		Dataset 1				Dataset 2				Dataset 3			
	ρ	$T = 100$		$T = 200$		$T = 100$		$T = 200$		$T = 100$		$T = 200$	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
RCT	-	-0.277	12.156	-0.155	8.504	-0.490	12.607	0.129	8.993	0.331	13.790	0.110	**9.839**
ADPT1	10	0.189	**10.505**	0.223	7.608	0.477	12.663	0.549	**8.664**	0.254	13.998	0.204	9.975
ADPT2	50	0.136	10.577	-0.095	**7.572**	-0.156	**12.277**	-0.095	8.685	-0.230	**13.671**	-0.341	10.047
OPT1	-	-0.166	10.106	-0.090	7.146	-0.404	11.845	-0.112	8.502	0.097	13.016	-0.104	9.346
OPT2	-	-0.149	9.828	-0.269	6.843	-0.138	11.575	-0.142	7.787	0.286	13.324	-0.059	9.492

Table 6.2: Experimental results using Datasets 4–6. The mean of the error $\theta - \hat{\theta}$, and the standard deviation of $\hat{\theta}$ are shown at the 100-th (mid) round and the 200-th (final) round. The lowest standard deviations among RCT, ADPT1, and ADPT2 are bold-face.

		Dataset 3				Dataset 4				Dataset 5			
	ρ	$T = 100$		$T = 200$		$T = 100$		$T = 200$		$T = 100$		$T = 200$	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
RCT	-	0.356	12.872	-0.163	8.865	-0.440	12.747	-0.126	9.143	-0.327	12.795	-0.099	**9.168**
ADPT1	10	0.617	**12.343**	0.292	**8.603**	-0.224	**12.727**	0.099	9.249	-0.382	12.993	-0.126	9.317
ADPT2	50	0.036	12.363	0.186	8.825	-0.049	12.883	0.036	**8.913**	-0.343	**12.784**	-0.377	9.236
OPT1	-	0.582	11.988	0.224	8.408	-0.603	12.384	-0.413	8.780	-0.246	12.459	-0.199	9.106
OPT2	-	0.882	11.503	0.333	8.182	-0.759	11.535	-0.194	7.931	-1.063	11.641	-0.492	8.179

Experiments with Semi-Synthetic Datasets: In evaluation of algorithms for estimating the treatment effect, it is difficult to find 'real-world' data that can be used for the evaluation. Previous research made semi-synthetic datasets with artificially simulated outcomes and covariate data from a real study. These datasets allow us to know the true treatment effect. We use data from the Infant Health and Development Program (IHDP). We follow a setting of simulation proposed by Hill (2011). In the setting of Hill (2011), she used 747 samples with 6 continuous covariates and 19 binary covariates. As a function of these covariates, Hill (2011) generated the outcomes as a function of these covariates. Hill (2011) considered two scenario: response surface A and response surface B. In response surface A, Hill (2011) generated $Y_t(1)$ and $Y_t(0)$ as follows:

$$Y_t(0) \sim \mathcal{N}(\boldsymbol{X}_t\boldsymbol{\beta}_A, \sigma_A),\ Y_t(1) \sim \mathcal{N}(\boldsymbol{X}_t\boldsymbol{\beta}_A + 4, 1),$$

where $\sigma^2_A = 1$ and elements of $\boldsymbol{\beta}_A \in \mathbb{R}^{25}$ are randomly sampled values from $(0, 1, 2, 3, 4)$ with probabilities $(0.5, 0.2, 0.15, 0.1, 0.05)$. In our experiment, in addition to the dataset of Hill (2011), we generated two more different data by changing σ_A. We call a dataset proposed by Hill (2011) Dataset A1, a dataset proposed with $\sigma^2_A = 2$ Dataset A2, and a dataset proposed with $\sigma^2_A = 4$ Dataset A3. In response surface B, Hill (2011) generated $Y_t(1)$ and $Y_t(0)$ as follows:

$$Y_t(0) \sim \mathcal{N}(\exp\left(\boldsymbol{X}_t\boldsymbol{\beta}_B + \boldsymbol{W}\right), \sigma^2_B),\ Y_t(1) \sim \mathcal{N}(\boldsymbol{X}_t\boldsymbol{\beta}_B - q, 1)$$

where $\sigma^2_B = 1$, $\boldsymbol{W}$ is an offset matrix of the same dimension as $\boldsymbol{X}_t$ with every value equal to 0.5, s was a constant to normalize the average treatment effect conditional on $d = 1$ to be 4, and elements of $\boldsymbol{\beta}_B \in \mathbb{R}^{25}$ were randomly sampled values $(0, 0.1, 0.2, 0.3, 0.4)$ with probabilities $(0.6, 0.1, 0.1, 0.1, 0.1)$. As well as the experiment with data of the surface A, in addition to the dataset of Hill (2011), we generated two more different dataset by changing σ_B. We call a dataset proposed by Hill (2011) Dataset B1, a dataset proposed with $\sigma_A = 5$ Dataset B2,

Table 6.3: Experimental results using Datasets A1–A3. The mean of the error $\theta - \hat{\theta}$, and the standard deviation of $\hat{\theta}$ are shown at the 100th (mid) round and the 200th (final) round. The lowest standard deviations are bold-face.

		Dataset 1				Dataset 2				Dataset 3			
	ρ	$T = 100$		$T = 200$		$T = 100$		$T = 200$		$T = 100$		$T = 200$	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
RCT	-	1.668	6.718	0.485	2.617	2.038	8.923	0.590	2.495	2.850	10.465	0.686	**1.815**
ADPT1	10	1.988	8.841	0.531	3.054	1.881	6.805	0.460	**1.411**	2.595	9.239	0.670	2.096
ADPT2	100	1.484	**4.762**	0.430	**1.726**	1.788	**6.523**	0.421	1.479	2.505	**8.627**	0.753	4.617

Table 6.4: Results of experiments using Datasets B1–B3. The mean of the error $\theta - \hat{\theta}$, and the standard deviation of $\hat{\theta}$ are shown at the 100th (mid) round and the 200th (final) round. The lowest standard deviations are bold.

		Dataset 1				Dataset 2				Dataset 3			
	ρ	$T = 100$		$T = 200$		$T = 100$		$T = 200$		$T = 100$		$T = 200$	
	ρ	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
RCT	-	1.725	7.712	0.436	1.617	1.425	**4.794**	0.418	**1.269**	2.443	9.812	0.652	1.390
ADPT1	10	1.759	8.324	0.410	1.622	1.569	5.714	0.483	2.513	2.854	17.438	0.675	2.729
ADPT2	100	1.678	**7.176**	0.408	**1.389**	1.649	7.196	0.400	1.998	2.168	**7.119**	0.547	**1.352**

and a dataset proposed with $\sigma_B = 10$ Dataset B3. From these samples, we estimated the treatment effect using the following three policies: an RCT with $p(D_t = 1|X_t) = 0.5$ for all rounds, the proposed algorithm with $\rho = 10$, and the proposed algorithm with $\rho = 100$. We refer these three policies as RCT, ADPT1, ADPT2. We simulated 1000 times. The results of experiment for Datasets A1–A3 are shown in Table 6.3. The results for Datasets B1–B3 are shown in Table 6.4. In Table 6.3 and Table 6.4, we showed the mean of the error $\theta - \hat{\theta}$, and the estimated standard deviation of $\hat{\theta}$ at the 373th (mid) round and the 747th (final) round.

Interpretations: First, we can confirm that the reported mean errors are near 0. These mean errors should become 0 with the sufficient number of trial because the estimators are unbiased, and the results show that the estimators satisfy this property. In many datasets, the proposed algorithm achieves the lower standard deviations than an RCT. When the difference of the variances of the potential outcomes is large, the proposed algorithm tends to outperform an RCT. This is because the the proposed algorithm gained the efficiency from the difference between $\mathbb{E}\left[Y_t^2(1)|X_t\right]$ and $\mathbb{E}\left[Y_t^2(0)|X_t\right]$. On the other hand, when the difference of variances of the potential outcomes is not large, an RCT tend to show better performance than the proposed algorithm. We consider that this is because of the existence of the inverse of $\mathbb{E}\left[Y_t^2(d)|X_t\right]$ in an IPW estimator makes an estimator of the proposed algorithm unstable.

When the samples are i.i.d., Hahn et al. (2011) showed that, if the probability of assigning a treatment is (3.5), an IPW estimator using an estimator of the probability of assigning a treatment for the inverse probability in the IPW estimator achieves the lowest asymptotic variance. However, as shown in experiments using synthetic datasets, we can find some cases where an estimator with the probability of assigning a treatment defined as (3.5) do not show better performance than the other algorithms in the case of small sample size. We consider that this is because, to achieve the lowest asymptotic variance, we need to replace the inverse probability with its estimator. However, when the sample size is small, we cannot obtain a non-parametric estimator of the probability of assigning a treatment that approximates the true probability of assigning a treatment well, and the estimation error of the probability of assigning a treatment makes the estimator of the treatment effect worse by replacing the inverse of the probability of assigning with its estimator.

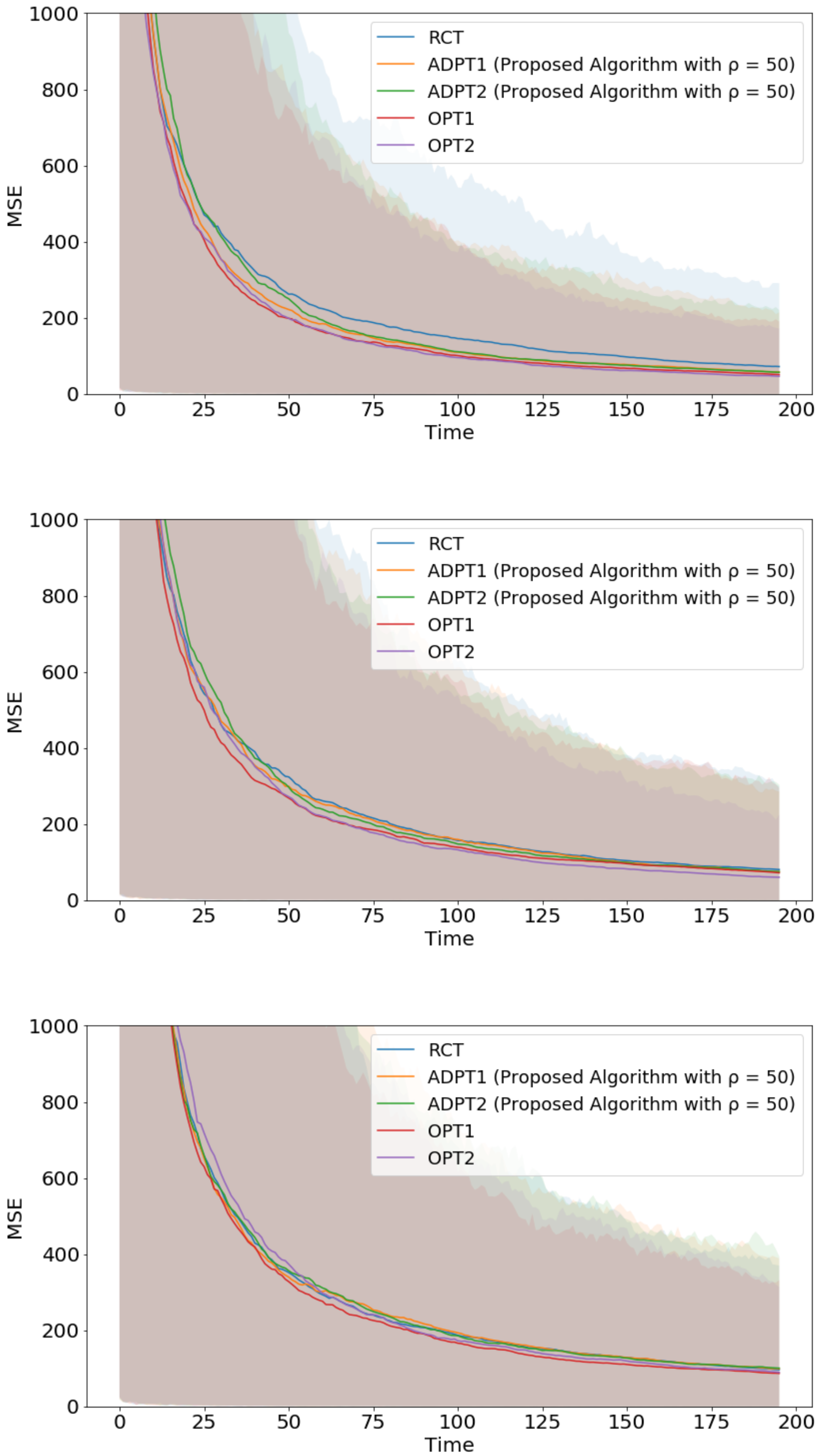

Figure 6.1: Results of experiments using Datasets 1–3. The mean squared error and the 5% and 95% quantiles of squared error.

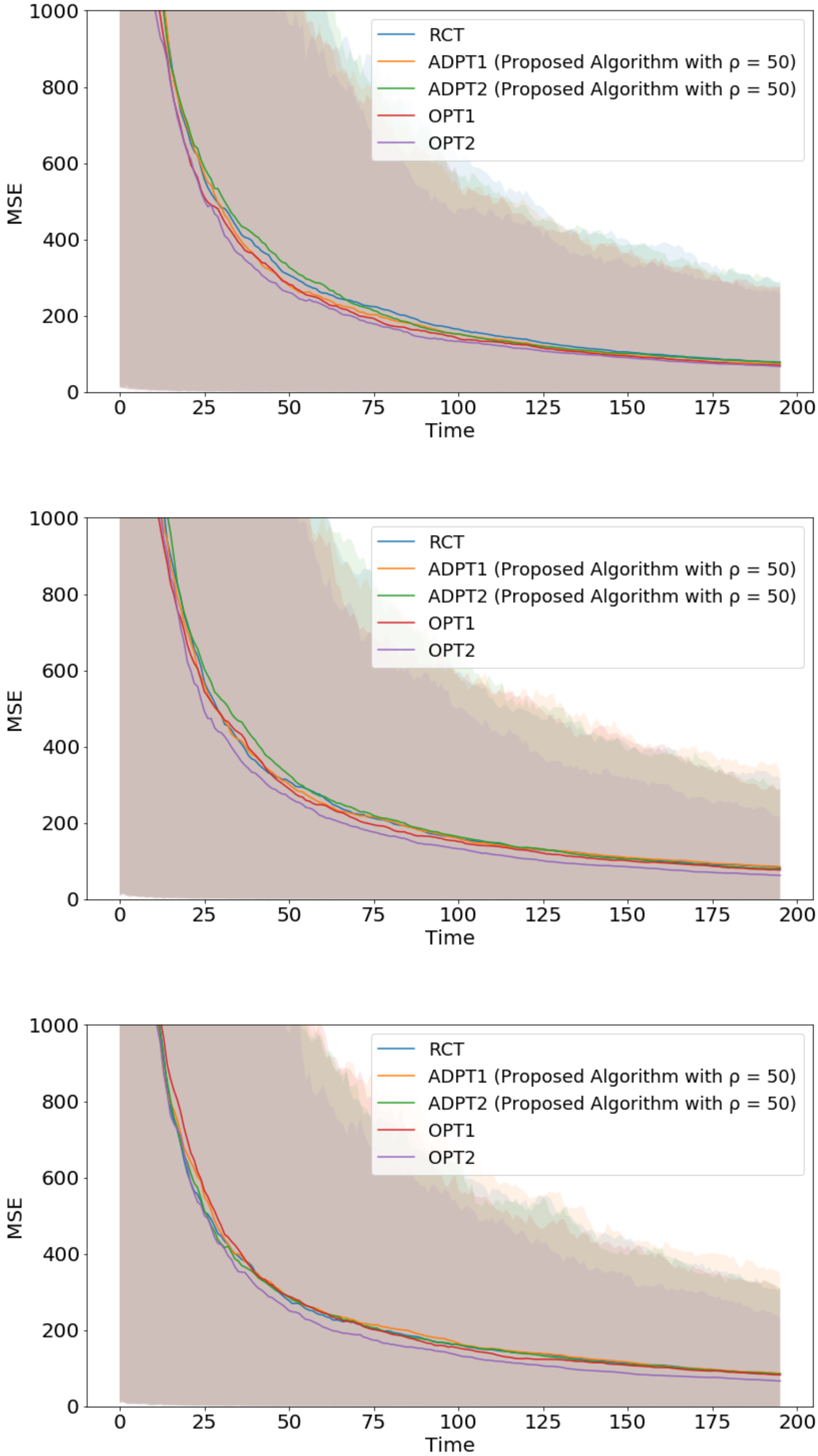

Figure 6.2: Results of experiments using Datasets 4–6. The mean squared error and the 5% and 95% quantiles of squared error.

Chapter 7

Conclusion

In this thesis, we presented a novel algorithm for estimating the treatment effect efficiently in the RCT under the setting of the adaptive experimental design. In the adaptive experimental design, we are allowed to change the probability of assigning a treatment using past information. In this setting, Hahn et al. (2011) proposed an algorithm that separates an experiment into two stages. In the first stage, they estimate the optimal probability of assigning a treatment that achieves the theoretically lowest asymptotic variance of an estimator of the treatment effect constructed using the probability of assigning a treatment. In the second stage, they conduct an RCT following the optimal probability of assigning a treatment. The drawback of this algorithm is to assume that we have infinite samples in a pilot phase, which is not practical in the real world. We extended a method of Hahn et al. (2011) by assuming a situation such that a research subject visits a researcher, and the researcher can change the probability of assigning a treatment for each subject. This problem setting is the same as the setting of the multi-armed bandit problem. We proposed an algorithm of the multi-armed bandit problem that gives us an efficient estimator of the treatment effect and showed the empirical performance of the proposed algorithm. To derive the theoretical properties of the proposed algorithm, we applied the martingale theory to construct an estimator of the proposed algorithm. By using the martingale theory, we analyzed the theoretical properties of the case with infinite and finite samples. From the experimental result, we confirmed that the proposed algorithm works well in practice.

Appendix

In this appendix, we show mathematical tools and theoretical results related with martingales.

A Mathematical Tools

Proposition A.1. [Continuous Mapping Theorem, Greene (2003), Theorem D. 12, p. 1113] For a continuous function of $g(a_n)$ that is not a function of n, if $a_n \xrightarrow{p} a$, then

$$g(a_n) \xrightarrow{\text{p}} g(a).$$

If $a_n \xrightarrow{a.s.} a$, then

$$g(a_n) \xrightarrow{\text{a.s.}} g(a).$$

Proposition A.2. [Slutsky Theorem, Greene (2003), Theorem D. 16 1, p. 1117] If $a_n \xrightarrow{d} a$ and $b_n \xrightarrow{p} b$, then

$$a_n b_n \xrightarrow{d} ba.$$

Definition A.1. [Uniformly Integrable, Hamilton (1994), p. 191] A sequence $\{a_t\}$ is said to be uniformly integrable if for every $\epsilon > 0$ there exists a number $c > 0$ such that

$$\mathbb{E}[|a_t| \cdot I[|a_t \geq c|]] < \epsilon$$

for all t.

Proposition A.3. [Sufficient Conditions for Uniformly Integrable, Hamilton (1994), Proposition 7.7, p. 191] (a) Suppose there exist $r > 1$ and $M < \infty$ such that $\mathbb{E}[|a_t|^r] < M$ for all t. Then $\{a_t\}$ is uniformly integrable. (b) Suppose there exist $r > 1$ and $M < \infty$ such that $\mathbb{E}[|b_t|^r] < M$ for all t. If $a_t = \sum_{j=-\infty}^{\infty} h_j b_{t-j}$ with $\sum_{j=-\infty}^{\infty} |h_j| < \infty$, then $\{a_t\}$ is uniformly integrable.

Proposition A.4. [Uniform Integrability and Convergence in Probability, Folland (2013)] If $a_n \xrightarrow{p} a$, (i)–(iii) in the following are equivalent.

(i) $\{a_n | n \in \mathbb{N}\}$ is uniformly integrable.

(ii) $a_n \to a$ in L^1.

(iii) $\mathbb{E}[|a_n|] \to \mathbb{E}[|a|] < \infty$.

B Limit Theorems and Concentration Inequality for Dependent Variables

Proposition B.1. [Strong Law of Large Numbers for Martingale, Chow (1967)] If $\{S_t = \sum_{s=1}^{t} A_s\}_{t=1}^{T}$ is a martingale such that

$$\sum_{t=1}^{\infty} \mathbb{E}[|A_t|^{2a}]/t^{1+a} < \infty \quad \text{for } a \geq 1,$$

then $S_t/t \xrightarrow{a.s.} 0$.

Proposition B.2. [Central Limit Theorem for a Martingale Difference Sequence, Hamilton (1994), Proposition 7.9, p. 194] Let $\{\boldsymbol{B}_t\}_{t=1}^{\infty}$ be an n-dimensional vector martingale difference sequence with $\overline{\boldsymbol{B}}_T = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{B}_t$. Suppose that (a) $\mathbb{E}[\boldsymbol{B}_t \boldsymbol{B}_t^\top] = \Omega_t$, a positive definite matrix with $(1/T)\sum_{t=1}^{T} \Omega_t \to \Omega$, a positive definite matrix; (b) $\mathbb{E}[B_{it}B_{jt}B_{it}B_{mt}] < \infty$ for all t and all i, j, l, and m (including $i = j = l = m$), where B_{it} is the i-th element of vector $\boldsymbol{B}_t$; and (c) $(1/T)\sum_{t=1}^{T} \boldsymbol{B}_t \boldsymbol{B}_t^\top \xrightarrow{p} \Omega$. Then $\sqrt{T}\overline{\boldsymbol{B}}_T \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Omega)$.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973.

Bothwell, L. E. and Podolsky, S. H. (2016). The emergence of the randomized, controlled trial. *New England Journal of Medicine*, 375(6):501–504. PMID: 27509097.

CDER, C. (2018). Adaptive designs for clinical trials of drugs and biologics guidance for industry draft guidance.

Chow, S.-C. and Chang, M. (2011). *Adaptive Design Methods in Clinical Trials*. Chapman and Hall/CRC, 2 edition.

Chow, Y. S. (1967). On a strong law of large numbers for martingales. *Ann. Math. Statist.*, 38(2):610.

Duflo, E., Glennerster, R., and Kremer, M. (2008). *Using Randomization in Development Economics Research: A Toolkit*, volume 4. North Holland, Amsterdam and New York.

Folland, G. (2013). *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley.

Gertler, P. J. and Boyce, S. (2003). An Experiment in Incentive-Based Welfare: The Impact of PROGRESA on Health in Mexico. Royal Economic Society Annual Conference 2003 85, Royal Economic Society.

Greenberg, D. and Shroder, M. (2004). *The Digest of Social Experiments, 3rd edition.*

Greene, W. (2003). *Econometric Analysis*. Pearson Education.

Hahn, J., Hirano, K., and Karlan, D. (2011). Adaptive experimental design using the propensity score. *Journal of Business and Economic Statistics*, 29(1):96–108.

Hamilton, J. (1994). *Time series analysis*. Princeton Univ. Press, Princeton, NJ.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Hill, S. A. B. (1990). Memories of the british streptomycin trial in tuberculosis: The first randomized clinical trial. *Controlled Clinical Trials*, 11(2):77 – 79.

Hirano, K., Imbens, G., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71:1161–1189.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jamieson, K. G. and Jain, L. (2018). A bandit approach to sequential experimental design with false discovery control. In *Advances in Neural Information Processing Systems 31*, pages 3664–3674. Curran Associates, Inc.

Kendall, J. M. (2003). Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, 20(2):164–168.

Kohavi, R. and Longbotham, R. (2017). *Online Controlled Experiments and A/B Testing*, pages 922–929.

Meldrum, M. L. (2000). A brief history of the randomized controlled trial: From oranges and lemons to the gold standard. *Hematology/Oncology Clinics of North America*, 14(4):745 – 760.

Mill, J. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. John W. Parker.

Nardini, C. (2014). The ethics of clinical trials. *Ecancermedicalscience*, 8:387.

Neyman, J. (1923). Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. *Statistical Science*, 5:463–472.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc., New York, NY, USA, 1st edition.

Pong, A. and Chow, S.-C. (2016). Handbook of adaptive designs in pharmaceutical and clinical development. pages 1–475.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. 89:846–866.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58.

Siroker, D. and Koomen, P. (2013). A/b testing: The most powerful way to turn clicks into customers.

The Royal Swedish Academy of Sciences (2019). The prize in economic sciences 2019.

Villar, S., Bowden, J., and Wason, J. (2015). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*, 30:199–215.

Yang, F., Ramdas, A., Jamieson, K. G., and Wainwright, M. J. (2017). A framework for multi-a(rmed)/b(andit) testing with online fdr control. In *Advances in Neural Information Processing Systems 30*, pages 5957–5966. Curran Associates, Inc.

Yang, Y. and Zhu, D. (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *Ann. Statist.*, 30(1):100–121.

Yoshioka, A. (1998). Use of randomisation in the medical research council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ*, 317(7167):1220–1223.