Active Adaptive Experimental Design for Treatment Effect Estimation with Covariate Choice

Masahiro Kato [1]   **Akihiro Oga** [1]   **Wataru Komatsubara** [1]   **Ryo Inokuchi** [1]

Abstract

This study designs an adaptive experiment for efficiently estimating *average treatment effects* (ATEs). In each round of our adaptive experiment, an experimenter sequentially samples an experimental unit, assigns a treatment, and observes the corresponding outcome immediately. At the end of the experiment, the experimenter estimates an ATE using the gathered samples. The objective is to estimate the ATE with a smaller asymptotic variance. Existing studies have designed experiments that adaptively optimize the propensity score (treatment-assignment probability). As a generalization of such an approach, we propose optimizing the covariate density as well as the propensity score. First, we derive the efficient covariate density and propensity score that minimize the semiparametric efficiency bound and find that optimizing both covariate density and propensity score minimizes the semiparametric efficiency bound more effectively than optimizing only the propensity score. Next, we design an adaptive experiment using the efficient covariate density and propensity score sequentially estimated during the experiment. Lastly, we propose an ATE estimator whose asymptotic variance aligns with the minimized semiparametric efficiency bound.

1. Introduction

Experimental approaches play a pivotal role in uncovering causality in various fields of science and industrial applications, such as epidemiology and economics (Shein-Chung Chow, 2005; Hahn et al., 2011; Chow & Chang, 2011; FDA, 2019). This study focuses on cases with *binary treatments*. Typically, one of these treatments corresponds

---
[1]Mizuho-DL Financial Technology Co., Ltd., Tokyo, Japan. Correspondence to: Masahiro Kato <masahiro-kato@fintec.co.jp>.

to the *treatment*, and the other to the *control* (Imbens & Rubin, 2015). Among the quantities representing causal effects, this study focuses on an *average treatment effect* (ATE), which defines a causal effect as the difference between the expected outcomes of binary treatments.

Treatments are an abstraction of an experimenter's choices, including drugs, online advertisements, and economic policies, which are also referred to differently, such as actions and arms. A difference in outcomes of binary treatments (treatment effect) is often considered a causal effect (Imbens & Rubin, 2015). However, due to the counterfactual property, *individual treatment effects* cannot be observed. Therefore, we use ATEs to capture causality.

A fundamental experimental approach for estimating ATEs is a *randomized control trial* (RCT). In an RCT, one of the binary treatments is randomly assigned to each experimental unit to obtain an unbiased estimator of ATEs. While RCTs are the gold standard and reliable (Hariton & Locascio, 2018), they often require large sample sizes, which can be costly. To reduce the sample size as much as possible, or to reduce an estimation error given a sample size, *adaptive experimental designs* have gained increasing attention across various fields (van der Laan, 2008; Hahn et al., 2011; Chow & Chang, 2011; Villar et al., 2015; FDA, 2019).

For estimating ATEs efficiently (with smaller asymptotic variances), van der Laan (2008) and Hahn et al. (2011) design adaptive experiments that optimize the propensity score (treatment-assignment probability). They show that by optimizing the propensity score, we can minimize the asymptotic variance of ATE estimators. Note that an asymptotic variance can be interpreted as an asymptotic mean squared error (MSE) of ATE estimation. Furthermore, in hypothesis testing, as an asymptotic variance decreases, the sample size required for hypothesis testing also decreases. Tabord-Meehan (2022), Kato et al. (2020), and Cook et al. (2023) design refined experiments based on this approach.

As a generalization of this approach, we consider a novel setting where an experimenter can sample experimental units based on their covariates and covariate density set by the experimenter. We show that by optimizing the covariate density as well as the propensity score, we can further minimize asymptotic variances compared to adaptive experiments that

only optimize the propensity score. Then, we design an adaptive experiment that optimizes efficient covariate density and propensity score by using past observations.

Firstly, we derive the semiparametric efficiency bound, a lower bound of the asymptotic variance, which is a functional of a covariate density and a propensity score. Based on this efficiency bound, we compute the efficient covariate density and propensity score that minimize the lower bound. Then, we design an adaptive experiment that sequentially estimates the efficient covariate density and propensity score. In the proposed experiment, an experimental unit is sampled following an adaptively estimated efficient covariate density and randomly assigned treatments following an estimated efficient propensity score. At the experiment's end, the experimenter obtains an estimator whose asymptotic variance aligns with the semiparametric efficiency bound. We employ the *Augmented Inverse Probability Weighting* (AIPW) estimator with covariate shift adaptation using Importance Weighting (AIPWIW). We refer to the setting of an adaptive experiment as *active adaptive experimental design with covariate choice* and our designed experiment as the *Active-Adaptive-Sampling (AAS)-AIPWIW experiment*.

Our method is inspired by three lines of work: adaptive experimental design for efficient ATE estimation, active learning using a covariate shift, and off-policy evaluation under a covariate shift. van der Laan (2008), Hahn et al. (2011), Tabord-Meehan (2022), and Kato et al. (2020) design adaptive experiments for estimating ATEs with smaller asymptotic variances by optimizing propensity scores. Our result generalizes their work by incorporating covariate density optimization. In active learning of regression models, Sugiyama (2006) demonstrates that the asymptotic variance of regression coefficient estimators can be reduced by appropriately shifting the covariate density. That study deals with a non-adaptive experiment and regression with importance weighting (Shimodaira, 2000). In contrast, our target is ATE estimation, and we adaptively optimize the covariate density to estimate ATEs. Note that this approach for active learning is an application of importance sampling (Owen, 2013). In causal inference under a covariate shift, Uehara et al. (2020) presents an efficient estimator of policy values, a generalization of ATEs, using double machine learning (DML, Chernozhukov et al., 2018).

Intuitively, we can reduce the asymptotic variance of the ATE estimator by using more units for a treatment with a larger variance. In existing studies, an experimenter assigns a treatment to an experimental unit with a higher propensity score if it has a larger variance. In this study, we also optimize the choice of experimental units based on their covariates in addition to the propensity score. Under this approach, we use more experimental units if their conditional variances of outcomes are large, conditioned on their covari-

ates. See Figure 2 in Section 4. This covariate choice allows us to reduce the asymptotic variance of an ATE estimator.

Our framework introduces a novel aspect of optimizing covariate density, which has not been fully explored in the literature on adaptive experiments for ATE estimation. While active learning focuses on optimizing the covariate density, adaptive experiments typically focus on optimizing the propensity score. We refer to our setting as an *active adaptive experiment* for ATE estimation.

Contributions   Compared to existing work, our contributions are summarized as follows:

- We propose a novel framework for an adaptive experiment by optimizing the covariate density, as well as the propensity score (Section 2);

- We develop the semiparametric efficiency bound for the ATE in our context (Section 3);

- Based on this lower bound, we derive efficient covariate density and propensity score (Section 4);

- Using these efficient probabilities, we design an adaptive experiment for efficient ATE estimation (Section 5).

2. Problem Setting

This section provides our problem formulation. We define potential outcomes and observations separately by following the Neyman–Rubin causal model (Neyman, 1923; Rubin, 1974). Then, we discuss estimating the ATE from the observations gathered from an adaptive experiment.

2.1. Potential Outcomes

There is a binary treatment $a \in \{1, 0\}$. Let us define the corresponding potential outcome by $Y(a)$. Let $X \in \mathcal{X} \subset \mathbb{R}^d$ be a d-dimensional covariate, where $\mathcal{X}$ is the space.

For any $x \in \mathcal{X}$, let $P_0(x)$ be the true conditional distribution of (Y_1, Y_0) given $X = x$ whose density is given as $r^1(y(1), y(0) \mid x)$. Let $\mathbb{E}[\cdot \mid X = x]$ and $\mathrm{Var}(\cdot \mid X = x)$ be expectation and variance operators over $P_0(x)$. Let us denote the true mean and variance of $Y(a)$ conditioned by $X = x \in \mathcal{X}$ by $\mu_0(a)(x) = \mathbb{E}[Y(a) \mid X = x]$ and $\sigma_0^2(a)(x) = \mathrm{Var}(Y(a) \mid X = x)$, respectively.

We cannot observe both $Y(1)$ and $Y(0)$ simultaneously. We only observe $Y(A)$ for a treatment-assignment indicator $A \in \{1, 0\}$. We define how we observe data in Section 2.3.

We make the following regularity assumption.

Assumption 2.1. There exist known universal constants $\underline{C}$ and $\overline{C}$ such that $0 < \underline{C} < \overline{C} < \infty$ and for any $x \in \mathcal{X}$, under $P_0(x)$, $\left|\mu_0(a)(x)\right| < \overline{C}$ and $\underline{C} < \sigma_0^2(a)(x) < \overline{C}$ hold.

2.2. Average Treatment Effect

As a causal effect, we are interested in the difference $Y(1) - Y(0)$, referred to as an ITE in some cases. However, because one of $Y(1)$ and $Y(0)$ is unobservable owing to the counterfactual property, we cannot observe it directly. Therefore, we focus on the expected value of $Y(1) - Y(0)$.

Let $q(x)$ be the density of an evaluation covariate distribution. Our interest is in the ATE over $q(x)$, defined as

$$\theta_0 := \mathbb{E}_{X \sim q(x), Y(a) \sim r^a(y|X)}[Y(1) - Y(0)],$$

where $\mathbb{E}_{X \sim q(x), Y(a) \sim r^a(y|X)}$ is expectation over X generated from $q(x)$ and $Y(a)$ given X generated from $r^a(y \mid X)$. Note that $\theta_0 = \mathbb{E}_{X \sim q(x)}[\mathbb{E}[Y(1) - Y(0) \mid X]] = \mathbb{E}_{X \sim q(x)}[\mu_0(1)(X) - \mu_0(0)(X)]$ holds, where $\mathbb{E}_{X \sim q(x)}$ denotes expectation over X generated from $q(x)$.

Our interest is to estimate the ATE with a smaller asymptotic variance via the following adaptive experiment.

For simplicity, $q(x)$ is assumed to be known to the experimenter. This assumption can be mitigated by considering different settings. In Appendix E.1, we introduce another setting where we can observe covariates X_t generated from $q(x)$ and decide whether we use it by the rejection sampling.

Assumption 2.2. The covariate density $q(x)$ is known.

2.3. Active Adaptive Experiments

Consider an adaptive experiment with T rounds, $[T] := \{1, 2, \ldots, T\}$. In the experiment, in each $t \in [T]$, an experimenter decides a covariate density $p_t(x)$ and observes an experimental unit with covariates $X_t \sim p_t(x)$. Given the observed covariates X_t, the experimenter decides propensity score $w_t(a \mid X_t)$ and assigns treatment $A_t \in \{1, 0\}$, where $A_t \in \{1, 0\}$ is a *treatment indicator*. Then, the experimenter observes an outcome $Y_t = \mathbb{1}[A_t = 1]Y_t(1) + \mathbb{1}[A_t = 0]Y_t(0)$, where $(Y_t(1), Y_t(0))$ is an i.i.d. copy of $(Y(1), Y(0))$. After round T, the experimenter estimates θ_0 by using the observations $\{(Y_t(1), Y_t(0))\}_{t=1}^T$.

In our adaptive experiment, an experimenter can optimize the probabilities of $p_t(x)$ and $w_t(a \mid X_t)$ to obtain an estimator of θ_0 with a smaller asymptotic variance.

In summary, the experimenter takes the following steps:

Step 1: based on past observations, an experimenter decides a covariate density p_t and a propensity score w_t;

Step 2: the experimenter observes covariate $X_t \sim p_t(x)$;

Step 3: based on past information, the experimenter decides a propensity score $w_t(a \mid X_t)$;

Step 4: the experimenter assigns treatment $A_t = a$ with probability $w_t(a \mid X_t)$;

Step 5: the experimenter observes outcome $Y_t = \mathbb{1}[A_t = 1]Y_t(1) + \mathbb{1}[A_t = 0]Y_t(0)$, where $Y_t(a) \sim r^a(y \mid X_t)$;
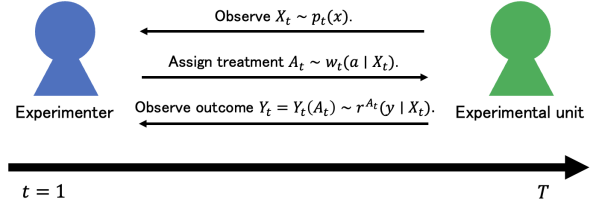

Figure 1. Active adaptive experiment.

Step 6: after iterating Steps 1–5 for $t = 1, 2, \ldots, T$, in round T, we estimate the ATE.

We assume that $r^a(y \mid x)$ is invariant across rounds, and $Y_t(a)$ are drawn from the density independently of the other round's random variables; that is, $\{(Y_t(1), Y_t(0))\}_{t=1}^T$ is i.i.d., but $p_t(a)$ and $w_t(a \mid x)$ can take different values across rounds based on past observations.

Thus, the adaptive experiments yields observations $\{(Y_t, A_t, X_t)\}_{t=1}^T$, where (Y_t, A_t, X_t) and (Y_s, A_s, X_s) are correlated over time for $t \neq s$, that is, the samples are not i.i.d. Let $\mathcal{F}_{t-1} = \{X_{t-1}, A_{t-1}, Y_{t-1}, \ldots, X_1, A_1, Y_1\}$ be the history. The probability $w_t(a \mid x)$ is a function of a covariate X_t, an action A_t, and history $\mathcal{F}_{t-1}$.

We refer to an experiment with this setting as an *active adaptive experiment with covariate choice*.

Remark 2.3 (Stable unit treatment value assumption (SUTVA)). The DGP implies SUTVA; that is, $r^a(y \mid x)$ is invariant against A_t across $t \in [T]$ (Angrist et al., 1996).

3. Semiparametric Efficiency Bound

This section provides a lower bound for the asymptotic variance of ATE estimators. We focus on semiparametric efficiency bound for a class of regular estimators. The semiparametric efficiency bound generalize the Cramer-Rao lower bound, where the latter considers parametric models, while the former considers non- and semiparametric models.

The semiparametric efficiency bound is a lower bound for a class of regular estimators, which are $\sqrt{T}$-consistent estimators for all DGPs and remain so under perturbations of size $1/\sqrt{T}$ to the DGP (see van der Vaart (1998)).

For simplicity, suppose that (Y_t, A_t, X_t) is an i.i.d. copy of (Y, A, X). Let $r^a(y \mid x)$ be a density of $Y(a)$ conditioned on $X = x$, $w(d \mid x)$ be a probability of A conditioned on $X = x$, and $p(x)$ be a density of X. Suppose that (Y, A, X) is generated as $(Y, A, X) \sim f(y, a, x)$, where

$$f(y, a, x) = \prod_{d \in \{1,0\}} \left(r^d(y(d) \mid x) w(d \mid x) \right)^{a=d} p(x),$$

where $r^d(y(d) \mid x)$ is the marginal distribution of $Y(d)$. Note that in our setting, we can manipulate $w(a \mid x)$ and $p(x)$ and update them during an experiment.

3

Furthermore, we *hypothetically* assume access to a set of covariates sampled from $q(x)$. Then, to derive the semiparametric efficiency bound, we consider a stratified sampling scheme, where two independent datasets $\mathcal{D}_T := \{(Y_t, A_t, X_t)\}_{t=1}^T$ and $\widetilde{\mathcal{D}}_S := \{\widetilde{X}_s\}_{s=1}^S$ are available. Note that $(Y_t, A_t, X_t) \overset{\text{i.i.d}}{\sim} \prod_{d \in \{1,0\}} \left\{ r^d(y \mid x)w(d \mid x) \right\}^{a=d} p(x)$ and $\widetilde{X}_s \overset{\text{i.i.d}}{\sim} q(x)$. First, we derive the semiparametric efficiency bound under a situation where $\mathcal{D}_T$ and $\widetilde{\mathcal{D}}_S$ are available. Then, since we know $q(x)$ and can generate infinitely many samples from $q(x)$, we focus on the semiparametric efficiency bound at a limit of $S \to \infty$.

Under this DGP and $q(x)$, we obtain the following theorem. The proof is shown in Appendix A.

Theorem 3.1. *Under Assumption 2.2, given $\mathcal{D}_T$ and $\widetilde{\mathcal{D}}_S$, as $S \to \infty$, the semiparametric efficiency bound for θ_0 with fully nonparametric models is*

$$\tau(w,p) = \mathbb{E}_{X \sim q(x)} \left[\left(\frac{\sigma_0^2(1)(X)}{w(1 \mid X)} + \frac{\sigma_0^2(0)(X)}{w(0 \mid X)} \right) \frac{q(X)}{p(X)} \right].$$

Theorem 3.1 is a lower bound; that is, a theoretically best asymptotic variance. It means that even though we minimize the lower bound, it is still unclear whether an estimator whose asymptotic variance aligns with the lower bound exists. In Section 5, we actually design an adaptive experiment under which an ATE estimator's asymptotic variance aligns with the lower bound minimized over w and p.

This semiparametric efficiency bound can be regarded as a functional of p and w. During our adaptive experiments, we can minimize the bound regarding p and w.

4. Efficient Probabilities

This section investigates w and p that minimizes $\tau(w,p)$ by considering the following functional optimization problem:

$$(w^*, p^*) := \underset{(w,p) \in \mathcal{W} \times \mathcal{P}}{\arg \min} \tau(w,p), \qquad (1)$$

where $\mathcal{W}$ is a set of all measurable functions $w : \mathcal{X} \to \Delta := \left\{ u = (u_1 \ u_2 \ \dots \ u_K)^\top \in [0,1]^K \mid \sum_{k=1}^K u_k = 1 \right\}$ with its a-th element $w(a \mid x)$, and $\mathcal{P}$ is a set of all measurable functions $p : \mathcal{X} \to \mathbb{R}^+$ such that $\int_{x \in \mathcal{X}} p(x)\mathrm{d}x = 1$. We refer to the minimizers (stationary point) p^* and w^* as *efficient* covariate density and propensity score since the semiparametric efficiency bound is minimized under them.

4.1. Efficient Propensity Score

First, we find an efficient propensity score that minimizes the semiparametric efficiency bound given some $p(x)$. Given p, a stationary point regarding w is given as the point-wise optimizer for each $x \in \mathcal{X}$. That is, given

x, for $w(x) := \{w(a \mid x)\}_{a \in [K]} \in (0,1)^K$, we have $w^*(x) = \arg \max_{w \in \Delta} \left\{ \frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)} \right\} \frac{q^2(x)}{p(x)}$. Then, for each $x \in \mathcal{X}$, the efficient propensity score is

$$w^*(a \mid x) = \frac{\sigma_0(a)(x)}{\sigma_0(1)(x) + \sigma_0(0)(x)} \qquad \forall a \in \{1,0\}.$$

This efficient propensity score is referred to as the *Neyman allocation*, which has been traditionally employed in designing experiments with binary treatments without covariate choice (Neyman, 1934; van der Laan, 2008; Hahn et al., 2011; Kato et al., 2020; Adusumilli, 2022; Dai et al., 2023).

4.2. Efficient Covariate Density

Next, we find an efficient covariate density that minimizes the semiparametric efficiency bound for fixed w.

Theorem 4.1. *For each $x \in \mathcal{X}$, the minimizer p^* of the semiparametric efficiency bound $\tau(w,p)$ with a fixed w is*

$$p^*(x) = \frac{\sqrt{\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)}}}{\mathbb{E}_{X \sim q(x)} \left[\sqrt{\frac{\sigma_0^2(1)(X)}{w(1|X)} + \frac{\sigma_0^2(0)(X)}{w(0|X)}} \right]} q(x).$$

By substituting w^* obtained in Section 4.1, we have

$$p^*(x) = \frac{\sigma_0(1)(x) + \sigma_0(0)(x)}{\mathbb{E}_{X \sim q(x)} [\sigma_0(1)(X) + \sigma_0(0)(X)]} q(x).$$

This result implies that we can reduce the efficiency bound by drawing samples whose sums of conditional standard deviations of potential outcomes, $\sigma_0(1)(x) + \sigma_0(0)(x)$, are large. That is, we assign treatments to experimental units with high-variance outcomes more than those with less-variance outcomes to accurately estimate the ATE.

For example, in clinical trials, to reduce the ATE, we gather elderly with high blood pressure more than healthy youth because the former may have more volatile outcomes. We illustrate this intuition in Figure 2.

4.3. Minimized Semiparametric Efficiency Bound

The minimizer of $\tau(w,p)$ is given as w^* and p^*. By substituting them, we can obtain the minimized semiparametric efficiency bound $\tau^* := \tau(w^*, p^*)$ as follows:

$$\tau^* = \left(\mathbb{E}_{X \sim q(x)} \left[\sqrt{\frac{\sigma_0^2(1)(X)}{w^*(1 \mid X)} + \frac{\sigma_0^2(0)(X)}{w^*(0 \mid X)}} \right] \right)^2$$
$$= \left(\mathbb{E}_{X \sim q(x)} [\sigma_0(1)(X) + \sigma_0(0)(X)] \right)^2.$$

4.4. Efficiency Gain

This section investigates how the efficient probability assignment probability and covariate density improve the semiparametric efficiency bound (efficiency gain).
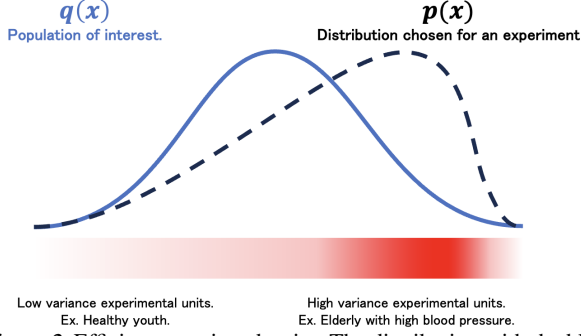
4

Figure 2. Efficient covariate density. The distribution with the blue line represents the covariate density of interest, while that with the dotted black line represents the covariate density in an experiment.

Efficient propensity score   As a result of the optimization, for each $x \in \mathcal{X}$, for any $w \in \mathcal{W}$, we have

$$\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)}$$
$$\geq \frac{\sigma_0^2(1)(x)}{w^*(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w^*(0 \mid x)} = \left(\sigma_0(1)(x) + \sigma_0(0)(x)\right)^2.$$

We confirm how the lower bound is minimized by considering the uniform sampling $w^{\mathrm{Unif}}(a \mid x) = 1/2$ for each $a \in \{1, 0\}$ and $x \in \mathcal{X}$. We have

$$\frac{\sigma_0^2(1)(x)}{w^{\mathrm{Unif}}(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w^{\mathrm{Unif}}(0 \mid x)} = 2\left(\sigma_0^2(1)(x) + \sigma_0^2(0)(x)\right)$$
$$\geq \left(\sigma_0(1)(x) + \sigma_0(0)(x)\right)^2 = \frac{\sigma_0^2(1)(x)}{w^*(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w^*(0 \mid x)}.$$

The difference between them is $\left(\sigma_0(1)(x) - \sigma_0(0)(x)\right)^2$. Thus, as the difference between $\sigma_0(1)(x)$ and $\sigma_0(0)(x)$ becomes large, the efficiency gain also becomes large.

Efficient covariate density   Let us fix the propensity score as $w^*(a \mid x) = \frac{\sigma_0(1)(x)}{\sigma_0(1)(x) + \sigma_0(0)(x)}$ and consider a case where the covariate density is fixed at $q(x)$. When (Y, A, X) is generated from a density $\prod_{d \in [K]} \left\{ r^a(y \mid x) w^*(d \mid x) \right\}^{d=a} q(x)$, the semiparametric efficiency bound becomes

$$\widetilde{\tau} := \mathbb{E}_{X \sim q(x)} \left[\frac{\sigma_0^2(1)(X)}{w^*(1 \mid X)} + \frac{\sigma_0^2(0)(X)}{w^*(0 \mid X)} \right]$$
$$= \mathbb{E}_{X \sim q(x)} \left[\left(\sigma_0(1)(X) + \sigma_0(0)(X)\right)^2 \right].$$

That is, any regular estimators cannot obtain a smaller asymptotic variance than the lower bound when samples are generated from $\prod_{d \in [K]} \left\{ r^a(y \mid x) w^*(d \mid x) \right\}^{d=a} q(x)$.

Next, we compare the efficiency bound τ^* when samples are generated from $\prod_{d \in [K]} \left\{ r^a(y \mid x) w^*(d \mid x) \right\}^{d=a} p^*(x)$.

From the Jensen inequality, it follows that

$$\tau^* = \left(\mathbb{E}_{X \sim q(x)} \left[\sigma_0(1)(X) + \sigma_0(0)(X) \right] \right)^2$$
$$\leq \mathbb{E}_{X \sim q(x)} \left[\left(\sigma_0(1)(X) + \sigma_0(0)(X)\right)^2 \right] = \widetilde{\tau}.$$

The equality holds when $\left(\sigma_0(1)(x) + \sigma_0(0)(x)\right)^2$ is the same across $x \in \mathcal{X}$. Therefore, when $\left(\sigma_0(1)(x) + \sigma_0(0)(x)\right)^2$ varies across x, we can reduce the semiparametric efficiency bound by optimizing $p(x)$; that is $\tau^* < \widetilde{\tau}$.

5. Active Adaptive Experimental Design for Treatment Effect Estimation

We design an adaptive experiment with covariate choice, referred to as the *Active-Adaptive-Sampling-and-Augmented-Inverse-Probability-Weighting* (AAS-AIPWIS) experiment because our approach resembles active learning, which adaptively optimizes covariate density and the propensity score, and then estimate the ATE by the AIPWIW estimator. We first define our experiment in Section 5.1 and then show the asymptotic property in Section 5.2.

5.1. AAS-AIPWIS Experiment

This section provides the definition of our experiment, which consists of the *sampling phase* and *estimation phase*. In round $t \in [T]$, an experimenter engages in the sampling phase, where the experimenter optimizes the covariate density and the propensity score using past observations up to round t; then, the experimenter observes covariates and assigns a treatment following the efficient probabilities. After round T, the estimation phase starts, where we estimate the ATE using $\{(Y_t, A_t, X_t)\}_{t=1}^T$. We describe the details below, and its pseudo-code is provided in Algorithm 1.

Sampling phase   Define the following estimators:

- **Conditional mean estimator.** Let $\widehat{\mu}_t(a)(x)$ be an $\mathcal{F}_{t-1}$-measurable estimator of $\mu_0(a)(x)$ for each $a \in \{1, 0\}$ and $x \in \mathcal{X}$ such that $|\widehat{\mu}_t(a)(x)| < \overline{C}$ holds, and it converges to $\mu_0(a)(x)$ almost surely.

- **Conditional second-moment estimator.** Let $\widehat{\nu}_t(a)(x)$ be an $\mathcal{F}_{t-1}$-measurable estimator of $\nu_0(a)(x) := \mathbb{E}[Y_t^2(a) \mid x]$ for each $a \in \{1, 0\}$ and $x \in \mathcal{X}$.

- **Conditional variance estimator.** Let $\widehat{\sigma}_t^2(a)(x)$ be an $\mathcal{F}_{t-1}$-measurable bounded estimator of $\sigma_0^2(a)(x)$ for each $a \in \{1, 0\}$ and $x \in \mathcal{X}$, defined as

$$\widehat{\sigma}_t^2(a)(x) = \mathrm{thre}\left(\widehat{\nu}_t(a)(x) - \widehat{\mu}_t^2(a)(x); \underline{C}, \overline{C}\right),$$

where $\mathrm{thre}(A; a, b) = A$ if $a \leq A \leq b$, $\mathrm{thre}(A; a, b) = a$ if $A < a$, and $\mathrm{thre}(A; a, b) = b$ if $A > b$.

Specifically, we construct $\widehat{\mu}_t(a)(x)$ and $\widehat{\nu}_t(a)(x)$ only by using past observations $\mathcal{F}_{t-1} = \{(Y_s, A_s, X_s)\}_{s=1}^{t-1}$.

For $\widehat{\mu}_t(a)(x)$ and $\widehat{\nu}_t(a)(x)$, we can use nonparametric estimators, such as the nearest neighbor regression estimator and kernel regression estimator, which have been proven to converge to the true function almost surely under a bounded sampling probability $\widehat{w}_t$ by Yang & Zhu (2002) and Qian & Yang (2016). In our simulation studies, we employ the Nadaraya-Watson kernel regression (Nadaraya, 1964; Watson, 1964), whose consistency has been proved by Qian & Yang (2016) in an adaptive experiment when the propensity score is lower bounded by a universal constant.

Then, we estimate the efficient propensity score as

$$\widehat{p}_t(x) := \frac{\widehat{\sigma}_t(1)(x) + \widehat{\sigma}_t(0)(x)}{\mathbb{E}_{X \sim q(x)}\left[\widehat{\sigma}_t(1)(x) + \widehat{\sigma}_t(0)(x)\right]} q(x)$$

$$\widehat{w}_t(a \mid x) := \frac{\widehat{\sigma}_t(a)(x)}{\widehat{\sigma}_t(1)(x) + \widehat{\sigma}_t(0)(x)}, \qquad (2)$$

where $\widehat{\theta}_t(x) := \widehat{\mu}_t(1)(x) - \widehat{\mu}_t(0)(x)$.

We assumed that we can directly compute $\mathbb{E}_{X \sim q(x)}\left[\widehat{\sigma}_t(1)(X) + \widehat{\sigma}_t(0)(X)\right]$ by using the known $q(x)$. For example, we generate enough samples from $q(x)$ and compute the expectation.

Instead of the true expectation, by using past observations, we can approximate it as $\mathbb{E}_{q(x)}\left[\widehat{\sigma}_t(1)(X) + \widehat{\sigma}_t(0)(X)\right] \approx$

$$\frac{1}{t-1}\sum_{s=1}^{t-1}\left(\widehat{\sigma}_t(1)(X_s) + \widehat{\sigma}_t(0)(X_s)\right)\frac{q(X_s)}{\widehat{p}_s(X_s)}.$$

Estimation phase In the estimation phase, we estimate the ATE θ_0 by using the AIPWIW estimator, defined as

$$\widehat{\theta}_T := \frac{1}{T}\sum_{t=1}^{T}\Psi_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t), \qquad (3)$$

where $\Psi_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t) :=$

$$\left(\frac{\mathbb{1}[A_t = 1]\left(Y_t(1) - \widehat{\mu}_t(1)(X_t)\right)}{\widehat{w}_t(1 \mid X_t)}\right.$$

$$\left. - \frac{\mathbb{1}[A_t = 0]\left(Y_t(0) - \widehat{\mu}_t(0)(X_t)\right)}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q(X_t)}{\widehat{p}_t(X_t)}$$

$$+ \mathbb{E}_{\widetilde{X} \sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right],$$

where $\mathbb{E}_{\widetilde{X} \sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right]$ is computed using the known $q(x)$. This estimator is a variant of the *Adaptive AIPW (A2IPW) estimator* proposed in Kato et al. (2020) with importance weighting (Uehara et al., 2020).

This estimator removes a sample selection bias caused by the propensity score, which assigns treatments based on covariates. Furthermore, by employing conditional mean estimators $\widehat{\mu}_t$ constructed from past observations, we can employ $\mathbb{E}\left[\Psi_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t) \mid \mathcal{F}t - 1\right] = \theta_0$. This method is a form of sample-splitting technique, as discussed in (van der Laan, 2008), frequently utilized in semiparametric analysis and recently refined as DML (Chernozhukov et al., 2018).

5.2. Asymptotic Distribution and Efficiency

This section provides an asymptotic distribution of $\widehat{\theta}_T$. First, we make the following assumptions.

Assumption 5.1. Random variables $Y_t(a) - \mu_a(a)(x)$ and $X_t - \mathbb{E}[X_t]$ are mean-zero sub-Gaussian.

Assumption 5.2. For all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$,

$$\widehat{\mu}_t(a)(x) \xrightarrow{\text{a.s.}} \mu_0(a)(x), \quad \widehat{\sigma}_t^2(a)(x) \xrightarrow{\text{a.s.}} \sigma_0^2(a)(x)$$

hold as $t \to \infty$.

This assumption also implies $\widehat{p}_t(x) \xrightarrow{\text{a.s.}} p^*(x)$.

Theorem 5.3. *Consider the AAS-AIPWIS experiment. Suppose that Assumptions 2.1–2.2 and 5.1–5.2 hold. Then,*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{\text{d}} \mathcal{N}(0, \tau^*)$$

holds as $T \to \infty$.

Thus, the asymptotic variance of $\widehat{\theta}_T$ aligns with the semiparametric efficiency bound minimized concerning the covariate density $p(x)$ and the propensity score $w(a \mid x)$. This result implies that our proposed experiment successfully reduces the asymptotic variance, and we cannot achieve a smaller variance within a class of regular estimators.

In the proof, we use the sample-splitting technique developed by Klaassen (1987), van der Laan (2008), Luedtke & van der Laan (2016), Chernozhukov et al. (2018), and Kato et al. (2021). With this technique, we can derive the asymptotic distribution only with the convergence rate condition in Assumption 5.2, without assuming the Donsker condition.

We also note that $\widehat{\theta}_T$ is consistent for θ_0, even without assuming Assumption 5.1.

Corollary 5.4. *Suppose that Assumptions 2.1–2.2 and 5.1 hold. Then, $\widehat{\theta}_T - \theta_0$ holds as $T \to \infty$.*

This property is based on the double robustness of the AIPW estimator $\widehat{\theta}_T$, which guarantees the consistency of $\widehat{\theta}_T$ if either $\mu_0(a)(x)$ or $w^*(a \mid x)$ is consistently estimated. In our case, since we know the true value of $w^*(a \mid x)$, the consistency is guaranteed by the double robustness. For this property, even if we cannot estimate $\mu_0(a)(x)$ consistently, we can still consistently estimate θ_0.

Algorithm 1 AAS-AIPWIS experiment

Parameter: Positive constants C_μ and C_{σ^2}.

Initialization:

for $t \in \{1, 2\}$ **do**

    Assign $A_t = 1$ at $t = 1$ and $A_t = 0$ at $t = 2$. For each $a \in \{1, 0\}$, set $\widehat{w}_t(a \mid x) = 1/2$.

end for

for $t = K + 1$ to T **do**

    Construct $\widehat{p}_t(x)$ following (2).

    Draw X_t from $\widehat{p}_t(x)$.

    Construct $\widehat{w}_t(1 \mid X_t)$ by using the variance estimators.

    Draw ξ_t from the uniform distribution on $[0, 1]$.

    $A_t = 1$ if $\xi_t \leq \widehat{w}_t(1 \mid X_t)$ and $A_t = 0$ otherwise.

end for

Estimate θ_0 by using the AIPWIW estimator $\widehat{\theta}_T$ in (3).
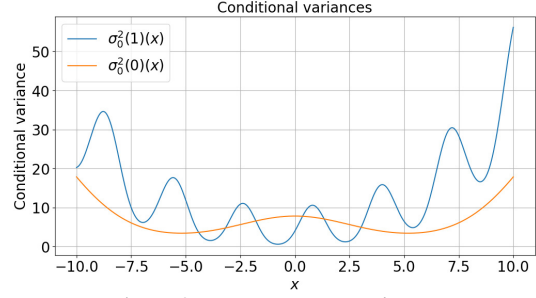

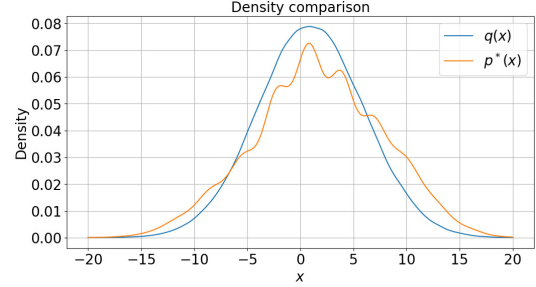
Figure 3. Heterogeneous variances.

Figure 4. Comparison between $q(x)$ and $p^*(x)$ in the experiment with the covariates following the Gaussian distribution.
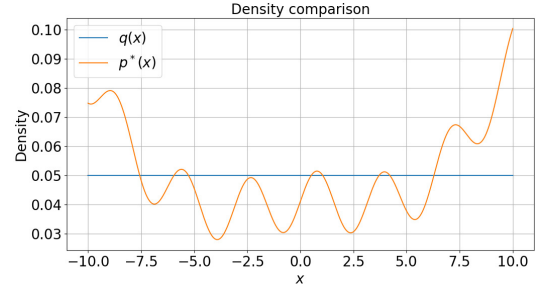

Figure 5. Comparison between $q(x)$ and $p^*(x)$ in the experiment with the covariates following the Uniform distribution.

Note that the asymptotic variances in the lower bound and our estimator in Theorem 5.3 do not depend on the dimension d. The dimensionality may affect whether the nuisance estimators $\widehat{\mu}_t$ and $\widehat{\sigma}_t^2$ satisfy Assumption 5.2. For example, if d is significantly large, we might use arguments about sparsity or benign overfitting to guarantee that Assumption 5.2 holds. Once Assumption 5.2 holds, the asymptotic variance of $\widehat{\theta}_T$ does not depend on d.

6. Simulation Studies

This section presents simulation studies. There are two objectives for the simulation. First, through simulation studies, we clarify the problem setting. Second, for the problem, we confirm the soundness of our designed experiment.

Basic setup Let $T = 2000$. We specify the means as $\mathbb{E}_{X \sim q(x)}[Y(1)] = 10$ and $\mathbb{E}_{X \sim q(x)}[Y(0)] = 7$, under which the ATE is $\theta_0 = 3$. Let covariates x be one-dimensional. For any x, the conditional means are

$$\mu_0(1)(x) := C_1\left(-x + 3x^2 - 1\right),$$
$$\mu_0(0)(x) := C_0\left(0.1x + 0.2\right),$$

where C_1 and C_0 are parameters so that $\mathbb{E}_{X \sim q(x)}[\mu_0(1)(X)] = \mathbb{E}_{X \sim q(x)}[Y(1)] = 10$ and $\mathbb{E}_{X \sim q(x)}[\mu_0(0)(X)] = \mathbb{E}_{X \sim q(x)}[Y(0)] = 7$.

Heterogeneous variances We define $\sigma_0^2(a)(x)$ as follows:

$$\sigma_0^2(1)(x) := 2 + 1.2\sin(2x) + (x + x^2)/25$$
$$\sigma_0^2(0)(x) := 2 + 0.8\cos(x/2) + x^2/50.$$

For $x \in [-10, 10]$, we plot $\sigma_0^2(a)(x)$ in Figure 3.

Alternative experiments We compare our designed experiment with the deterministic RCT (DRCT), the RCT, and the Adaptive-Sampling-and-AIPW (AS-AIPW) experiment. In the deterministic RCT, we deterministically assign treatment 1 if t is odd and assign treatment 0 if t is even. In the RCT, we randomly assign treatments with probability $w(1 \mid x) = w(0 \mid x) = 0.5$ for any x. We estimate the ATE by using the sample average in the DRCT and the RCT. In the AS-AIPW experiment, we optimize the propensity score but do not optimize the covariate density (experimental units are just generated from $q(x)$); then, we estimate the ATE with the AIPW estimator without importance weighting. The AS-AIPW estimator is basically the same method proposed in van der Laan (2008), Hahn et al. (2011), and Kato et al. (2020), though we simplified them for simplicity.

Additionally, we conduct the AS-AIPW and AAS-AIPWIW experiments under the true w^* and p^*; that is, they are known at the beginning of an experiment. We refer to the experiments as the AS-AIPW (Oracle) and the AAS-AIPWIW (Oracle). Although we use the true values for w^* and p^* in the sampling phase, we estimate $\mu(a)(x)$ in the estimation phase to construct the AIPW estimator.
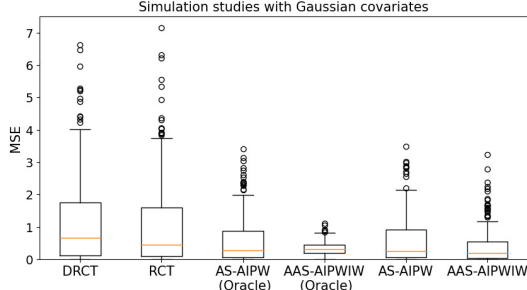
Figure 6. Results of simulation studies with covariates following with a Gaussian distribution $\mathcal{N}(1, 25)$.
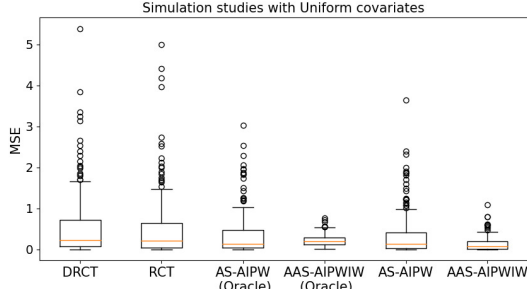

Figure 7. Results of simulation studies with covariates following with a Uniform distribution $\mathcal{U}(-10, 10)$.

Performance measure. In each of our simulations, we conduct 200 experiments and compute the performance measures. In this section, in the y-axis of Figures 6–7, we report the empirical MSEs for θ_0; that is, $\frac{1}{200} \sum_{i=1}^{200} (\theta_0 - \hat{\theta}_T^{(i)})^2$, where $\hat{\theta}_T^{(i)}$ denotes the estimator of ATE in the i-th trial. Note that the asymptotic MSEs correspond to the asymptotic variances.

Covariates following with a Gaussian distribution. Let $q(x)$ be the density of the Gaussian distribution $\mathcal{N}(1, 25)$ with mean 1 and variance 25. For $q(x)$, we plot $q(x)$ and $p^*(x)$ in Figure 4. We show the results in Figure 6.

Covariates following with a Uniform distribution Let $q(x)$ be a density of the Uniform distribution with support $[-10, 10]$. For $q(x)$, we plot $q(x)$ and $p^*(x)$ in Figure 5. We show the results of simulation studies in Figure 7.

Results From Figures 6 and 7, we find that our experiments successfully reduce the MSE compared to the others. Surprisingly, the AS-AIPW and the AAS-AIPW show almost identical or better performances than the oracle ones. Such a paradox often occurs in semiparametric analysis, as pointed out by (Henmi & Eguchi, 2004; Hitomi et al., 2008).

7. Related Work and Discussion

In this section, we discuss related topics.

7.1. Related Work

We provide a literature review of related work in the field of adaptive experimental design for efficiently estimating ATEs.

Adaptive sampling has long been investigated with various purposes (Wald, 1945; Robbins, 1952; Solomon & Zacks, 1970; Rosenberger et al., 2001; Bai et al., 2024). Pioneering contributions for ATE estimation by van der Laan (2008) and Hahn et al. (2011) lay the foundation for this field.

Building upon van der Laan (2008) and Hahn et al. (2011), Kato et al. (2020) and Tabord-Meehan (2022) propose refined methods by extending the existing methodologies. These studies derive the semiparametric efficiency bound for ATE estimators and specify the efficient propensity score that minimizes this bound.

This approach is further refined by Cook et al. (2023), who proves the asymptotic normality of the A2IPW estimator under weaker assumptions than those of Kato et al. (2020). They also propose an adaptive experiment that ensures finite-sample stability by truncating the propensity score (Waudby-Smith et al., 2024). Furthermore, they derive tighter concentration inequalities using results from Howard et al. (2021), Waudby-Smith & Ramdas (2023), and Waudby-Smith et al. (2023). Gupta et al. (2021) designs an adaptive experiment involving instrumental variables. There are other studies discussing related topics (Blackwell et al., 2022; Offer-Westort et al., 2021).

Recent work by Dai et al. (2023) introduces an approach based on online stochastic projected gradient descent for experimental design when the optimal treatment allocation rule depends on unknown parameters. Their study rigorously addresses the issue of variance estimation, previously considered negligible at the order of $O(1/\sqrt{T})$ under asymptotic approximation. They propose a method to effectively estimate variance and conduct experimental design while accounting for the impact of variance estimation errors. Zhao (2023) also investigates the variance estimation problem in Neyman allocation.

The AIPW estimator is popular in adaptive experiments. Utilizing its variance reduction and martingale properties, the A2IPW estimator is introduced in Kato et al. (2020) based on the arguments in van der Laan (2008) and Chernozhukov et al. (2018). Kato et al. (2021) and Li & Owen (2023) propose DML for adaptive experiments. Hadad et al. (2019) and Zhan et al. (2021) also focus on this topic, presenting different estimators under a different setting. There are also several approaches for statistical inference in adaptive experiments (Zhang et al., 2020; Bibaut et al., 2021).

In the context of active learning, Sugiyama (2006) demonstrates that an appropriate covariate density choice can re-

duce the asymptotic variance of regression coefficients. This setting is extended to the pool-based setting by Sugiyama & Nakajima (2009), which is relevant to our context. Wu (2019) further extends these approaches to adaptive experiments, contrasting with earlier non-adaptive experiments where observations are gathered without parameter estimation during the experiment. Independently, Zrnic & Candès (2024) explores active learning for causal inference.

Uehara et al. (2020) considers off-policy evaluation, a generalization of ATE estimation, under a covariate shift. They show that ATE estimation with importance weighting, as proposed by Shimodaira (2000), with nonparametrically estimated density-ratio, may not yield an estimator whose asymptotic variance aligns with the semiparametric efficiency bound.

Best-Arm Identification (BAI), another related topic, is a type of stochastic multi-armed bandit (MAB) problem (Thompson, 1933; Lai & Robbins, 1985), aimed at identifying the best treatment (arm) (Audibert et al., 2010; Kaufmann et al., 2016; Russo, 2020). BAI has two settings: fixed-confidence (Garivier & Kaufmann, 2016) and fixed-budget. The fixed-confidence setting is related to sequential testing (Bechhofer et al., 1968; Garivier & Kaufmann, 2016), while the fixed-budget setting is BAI with a fixed sample size (Bubeck et al., 2009; 2011; Degenne, 2023). Kato (2024b;c;a) investigate and generalize the Neyman allocation in fixed-budget BAI, proving its asymptotic optimality.

In the MAB problem, covariates are referred to as contexts. While various algorithms with contexts have been investigated in the MAB problem, it is still under exploration how to utilize contexts in BAI. In the fixed-confidence setting, Russac et al. (2021), Kato & Ariu (2021), and Simchi-Levi et al. (2024) investigate the BAI with covariates, which is referred to as contextual information. In the fixed-budget setting, Kato et al. (2024) designs an adaptive experiment for policy learning.

7.2. Sample Size Calculation

In statistical hypothesis testing, sample size calculation is important. Here, we discuss how to decide T in our method. Let the null and alternative hypotheses be $H_0 : \theta_0 = \mu$ and $H_1(\Delta) : \theta_1 = \mu + \Delta$, where $\mu \in \mathbb{R}$ and $\Delta > 0$ are some constants. Let $\alpha \in (0, 1)$ be a significance level such that $\mathbb{P}(\text{reject } H_0) = \alpha$, and $\beta \in (0, 1)$ be the Type II error level. Let R_T^α be a rejection region defined based on α. Then, we are interested in deciding the minimum sample size for hypothesis testing, which is defined as $T_{\alpha,\beta}^*(\Delta) :=$ $\min \{T : \mathbb{P}_{H_1(\Delta)}(\widehat{\theta}_T \in R_T^\alpha) \geq 1 - \beta\}$ for some estimator $\widehat{\theta}_T$ of θ_0. Let us assume that $\sqrt{T}(\widehat{\theta}_T - \mu) \xrightarrow{\text{d}} \mathcal{N}(0, V)$ under H_0 and $\sqrt{T}(\widehat{\theta}_T - \mu - \Delta) \xrightarrow{\text{d}} \mathcal{N}(0, V)$ under H_1.

Then, when using t-statistics for hypothesis testing, we have $T_{\alpha,\beta}^*(\Delta) = \frac{V}{\Delta^2}(z_{1-\alpha/2} - z_\beta)$, where z_c is the c-quantile of the standard normal distribution (Balsubramani & Ramdas, 2016; Kato et al., 2020). For example, in the AAS-AIPWIW, we substitute τ^* into V. This result implies that we can reduce the sample size T for hypothesis testing by reducing the asymptotic variance. For the details of sample size calculation, see Appendix G in Kato et al. (2020).

7.3. Extensions

We can consider several extensions of our designed experiment. In Appendix E, we propose three extensions: (i) rejection sampling, (ii) batch design, and (iii) experiments with multiple treatments. In the rejection sampling setting, we do not assume that $q(x)$ is known, but we sequentially observe an experimental unit generated from the unknown $q(x)$. By employing rejection sampling, we can decide whether to accept the observed experimental unit and if we accept it, we assign a treatment to it. This setting is feasible without knowing $q(x)$. In (ii), we can update $\widehat{p}_t$ and $\widehat{w}_t$ at certain rounds. For example, we can update them at $t = 100, 200, 300, \dots$ and need to keep using the same $\widehat{p}_t$ and $\widehat{w}_t$ between the two update timings. In (iii), we consider there are treatments more than or equal to three.

In this study, we assumed that $q(x)$ is known. For example, if the population we are interested in consists of the people of a certain country, we can obtain information about their distribution through government statistics. Then, in an experiment, we consider gathering people with different distributions using the method we propose. In applications, we can also consider estimating $q(x)$ using observations from a pilot experiment. If there are M samples from $q(x)$, it is possible to estimate $q(x)$ using those samples. In this case, if M diverges independently of T, the current results can be directly applicable (asymptotics of $M \to \infty$ and $T \to \infty$). However, if the divergence rate of M depends on T, we need to slightly modify our results and methods for each setting of applications.

8. Conclusion

In this study, we designed an adaptive experiment to estimate ATEs efficiently. In our adaptive experiment, we can optimize covariate densities and propensity scores in each round based on past observations obtained until that round. To derive the efficient covariate density and propensity score, we focused on the semiparametric efficiency bound for the ATE. Based on the derived efficient probabilities, we designed an adaptive experiment using the AIPW estimator with importance weighting. Under the designed experiment, we demonstrated that the asymptotic variance of the AIPW estimator aligns with the efficiency bound, minimized with respect to the covariate density and propensity score.

Impact Statement

This paper presents an innovative approach to machine learning, specifically focusing on the efficient estimation of ATEs through adaptive experimental design. Our work has the potential to significantly impact various sectors, including healthcare, economics, and social sciences. However, if not applied carefully, the optimization techniques could lead to biased sampling or unethical manipulation of experimental conditions, potentially harming subjects or leading to incorrect conclusions. Furthermore, the adaptive nature of the experiments, relying on sequential sampling and treatment assignment based on past observations, may raise data privacy and consent issues, especially in sensitive fields like healthcare. In conclusion, while our work aims to advance the field of machine learning by providing a novel framework for efficient ATE estimation through adaptive experiments, it is crucial to consider the ethical implications and societal impacts of its application. Researchers and practitioners should be mindful of the potential risks and strive to apply these methods in a responsible and equitable manner.

References

Adusumilli, K. Neyman allocation is minimax optimal for best arm identification with two arms, 2022. arXiv:2204.05527.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434): 444–455, 1996.

Audibert, J.-Y., Bubeck, S., and Munos, R. Best arm identification in multi-armed bandits. In *Conference on Learning Theory (COLT)*, pp. 41–53, 2010.

Bai, Y., Shaikh, A. M., and Tabord-Meehan, M. A primer on the analysis of randomized experiments and a survey of some recent advances, 2024. arXiv:2405.03910.

Balsubramani, A. and Ramdas, A. Sequential nonparametric testing with the law of the iterated logarithm. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 42–51, 2016.

Bechhofer, R., Kiefer, J., and Sobel, M. *Sequential Identification and Ranking Procedures: With Special Reference to Koopman-Darmois Populations*. University of Chicago Press, 1968.

Bibaut, A., Dimakopoulou, M., Kallus, N., Chambaz, A., and van der Laan, M. Post-contextual-bandit inference. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Blackwell, M., Pashley, N. E., and Valentino, D. Batch adaptive designs to improve efficiency in social science experiments, 2022. URL https://mattblackwell.org/research/batch_adaptive/. Working paper.

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pp. 23–37. Springer Berlin Heidelberg, 2009.

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 2011.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21:C1–C68, 2018.

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

Cook, T., Mishler, A., and Ramdas, A. Semiparametric efficient inference in adaptive experiments. In *NeurIPS 2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World*, 2023. URL https://openreview.net/forum?id=xfj5jjpOaL. a]rXiv:2311.18274.

Dai, J., Gradu, P., and Harshaw, C. CLIP-OGD: An experimental design for adaptive neyman allocation in sequential experiments. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

Degenne, R. On the existence of a complexity in fixed budget bandit identification. In *Conference on Learning Theory*, volume 195, pp. 1131–1154. PMLR, 2023.

FDA. Adaptive designs for clinical trials of drugs and biologics: Guidance for industry. Technical report, U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), 2019.

Garivier, A. and Kaufmann, E. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory (COLT)*, 2016.

Gelfand, I. M., Silverman, R. A., et al. *Calculus of variations*. Courier Corporation, 2000.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2004.

Gupta, S., Lipton, Z. C., and Childers, D. Efficient online estimation of causal effects by deciding what to observe. In *Advances in Neural Information Processing Systems*, 2021.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. Confidence intervals for policy evaluation in adaptive experiments, 2019.

Hahn, J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331, 1998.

Hahn, J., Hirano, K., and Karlan, D. Adaptive experimental design using the propensity score. *Journal of Business and Economic Statistics*, 2011.

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

Hariton, E. and Locascio, J. J. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *International Journal of Obstetrics & Gynaecology*, 125(13), 2018.

Henmi, M. and Eguchi, S. A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, 2004.

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

Hirano, K., Imbens, G., and Ridder, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 2003.

Hitomi, K., Nishiyama, Y., and Okui, R. A puzzling phenomenon in semiparametric estimation problems with infinite-dimensional nuisance parameters. *Econometric Theory*, 24(6):1717–1728, 2008.

Howard, S. R., Ramdas, A., McAuliffe, J. D., and Sekhon, J. S. Time-uniform, nonparametric, nonasymptotic confidence sequences. *Annals of Statistics*, 2021.

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

Kato, M. Adaptive generalized neyman allocation: Local asymptotic minimax optimal best arm identification, 2024a. arXiv:2405.19317.

Kato, M. Locally optimal fixed-budget best arm identification in two-armed gaussian bandits with unknown variances, 2024b. arXiv:2312.12741.

Kato, M. Generalized neyman allocation: Minimax optimal fixed-budget gaussian best arm identification, 2024c. arXiv:2310.19788.

Kato, M. and Ariu, K. The role of contextual information in best arm identification, 2021. arXiv:2106.14077v1.

Kato, M., Ishihara, T., Honda, J., and Narita, Y. Adaptive experimental design for efficient treatment effect estimation: Randomized allocation via contextual bandit algorithm. *arXiv:2002.05308*, 2020.

Kato, M., McAlinn, K., and Yasui, S. The adaptive doubly robust estimator and a paradox concerning logging policy. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Kato, M., Okumura, K., Ishihara, T., and Kitagawa, T. Adaptive experimental design for policy learning, 2024. arXiv:2401.03756.

Kaufmann, E., Cappé, O., and Garivier, A. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

Klaassen, C. A. J. Consistent estimation of the influence function of locally asymptotically linear estimators. *Annal of Statistics*, 1987.

Lai, T. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 1985.

Li, H. H. and Owen, A. B. Double machine learning and design in batch adaptive experiments, 2023. arXiv:2309.15297.

Loeve, M. *Probability Theory*. Graduate Texts in Mathematics. Springer, 1977.

Luedtke, A. R. and van der Laan, M. J. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713–742, Apr 2016.

Nadaraya, E. A. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

Narita, Y., Yasui, S., and Yata, K. Efficient counterfactual learning from bandit feedback. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019.

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

Neyman, J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:123–150, 1934.

Offer-Westort, M., Coppock, A., and Green, D. P. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, 65(4):826–844, 2021.

Owen, A. B. *Monte Carlo theory, methods and examples*. https://artowen.su.domains/mc/, 2013.

Qian, W. and Yang, Y. Kernel estimation and model combination in a bandit problem with covariates. *Journal of Machine Learning Research*, 2016.

Robbins, H. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527 – 535, 1952.

Rockafellar, R. T. *Convex analysis*. Princeton university press, 2015.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N., and Ricks, M. L. Optimal adaptive designs for binary response trials. *Biometrics*, 57(3):909–913, 2001.

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

Russac, Y., Katsimerou, C., Bohle, D., Cappé, O., Garivier, A., and Koolen, W. M. A/b/n testing with control in the presence of subpopulations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Russo, D. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6):1625–1647, 2020.

Shein-Chung Chow, Mark Chang, A. P. Statistical consideration of adaptive methods in clinical development. *Journal of Biopharmaceutical Statistics*, 2005.

Shimodaira, H. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.

Simchi-Levi, D., Wang, C., and Xu, J. On experimentation with heterogeneous subgroups: An asymptotic optimal δ-weighted-pac design, 2024. SSRN:4721755.

Solomon, H. and Zacks, S. Optimal design of sampling from finite populations: A critical review and indication of new research areas. *Journal of the American Statistical Association*, 65(330):653–677, 1970.

Sugiyama, M. Active learning in approximately linear regression based on conditional expectation of generalization error. *Journal of Machine Learning Research*, 7(6):141–166, 2006.

Sugiyama, M. and Nakajima, S. Pool-based active learning in approximate linear regression. *Machine Learning*, 75(3):249–274, Jun 2009.

Tabord-Meehan, M. Stratification Trees for Adaptive Randomisation in Randomised Controlled Trials. *The Review of Economic Studies*, 90(5):2646–2673, 12 2022.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 1933.

Uehara, M., Kato, M., and Yasui, S. Off-policy evaluation and learning for external validity under a covariate shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 49–61, 2020.

van der Laan, M. J. The construction and analysis of adaptive group sequential designs, 2008. URL https://biostats.bepress.com/ucbbiostat/paper232. U.C. Berkeley Division of Biostatistics Working Paper Series.

van der Vaart, A. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

Vershynin, R. *High-dimensional probability*, volume 7. Cambridge University Press, Cambridge, 2018.

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

Wald, A. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

Watson, G. S. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372, 1964.

Waudby-Smith, I. and Ramdas, A. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2023.

Waudby-Smith, I., Wu, S., and Ramdas, A. Nonparametric extensions of randomized response for private confidence sets. In *International Conference on Machine Learning (ICML)*, volume 202, pp. 36748–36789. PMLR, 2023.

Waudby-Smith, I., Wu, L., Ramdas, A., Karampatziakis, N., and Mineiro, P. Anytime-valid off-policy inference for contextual bandits. *ACM / IMS J. Data Sci.*, 2024.

White, H. Chapter v - central limit theory. In *Asymptotic Theory for Econometricians*, Economic Theory, Econometrics, and Mathematical Economics, pp. 107–131. 1984.

Wu, D. Pool-based sequential active learning for regression. *IEEE Transactions on Neural Networks and Learning Systems*, 30(5):1348–1359, 2019.

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

Zhan, R., Hadad, V., Hirshberg, D. A., and Athey, S. Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Conference on Knowledge Discovery & Data Mining (KDD)*, 2021.

Zhang, K., Janson, L., and Murphy, S. Inference for batched bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Zhao, J. Adaptive neyman allocation, 2023. arXiv:2309.08808.

Zrnic, T. and Candès, E. J. Active statistical inference, 2024. arXiv:2403.03208.

A. Proof of Theorem 3.1

In this section, we prove Theorem 3.1, the semiparametric efficiency bound for regular estimators of θ_0. Our derivation of the semiparametric efficiency bound follows the approaches of Hahn (1998), Hirano et al. (2003), Narita et al. (2019), and Uehara et al. (2020).

Their proof considers a nonparametric model for the distribution of potential outcomes and defines regular subparametric models. Then, (i) we characterize the tangent set for all regular parametric submodels, (ii) verify that the parameter of interest is pathwise differentiable, (iii) verify that a guessed semiparametric efficient influence function lies in the tangent set, and (iv) calculate the expected square of the influence function.

A.1. Stratified Sampling

As discussed in Section 3, we consider a stratified sampling scheme to derive the semiparametric efficiency bound. This sampling scheme assumes the availability of the following two datasets:

$$\mathcal{D}_T \coloneqq \{(Y_t, A_t, X_t)\}_{t=1}^T,$$
$$\widetilde{\mathcal{D}}_S \coloneqq \{\widetilde{X}_s\}_{s=1}^S,$$

where $(Y_t, A_t, X_t) \overset{\text{i.i.d}}{\sim} \prod_{d\in\{1,0\}} \left\{ r^d(y \mid x)w(d \mid x) \right\}^{d=a} p(x)$ and $\widetilde{X}_s \overset{\text{i.i.d}}{\sim} q(x)$. We refer to $\mathcal{D}_T$ as the observation data and $\widetilde{\mathcal{D}}_S$ as the evaluation data. Because we know $q(x)$, we can obtain infinite samples from $q(x)$. Therefore, we first derive the semiparametric efficiency bound under the sampling scheme and then consider $S \to \infty$ and then consider the semiparametric efficiency bound regarding the asymptotics of T.

Our derivation of the semiparametric efficiency bound is inspired by the technique used in Narita et al. (2019) and Uehara et al. (2020), but we modify it since we need to consider the asymptotics of T and S separately.

A.2. Parametric Submodels

We consider regular parametric submodels with a parameter $\beta \in \mathbb{R}$ to discuss the semiparametric efficiency bound under the stratified sampling scheme.

We denote the probability densities under the parametric submodels by

$$f(y, a, x; \beta) = \prod_{d\in\{1,0\}} \left(r^d(y \mid x; \beta)w(d \mid x; \beta) \right)^{a=d} p(x; \beta) \tag{4}$$

and $q(\widetilde{x}; \beta)$, where $r^d(y \mid x; \beta)$ is the conditional density of $Y(d)$ given $X = x$, $w(d \mid x; \beta) = \Pr(A_t = d \mid X = x; \beta)$, $p(x; \beta)$ is the density of X in observation data, and $q(\widetilde{x}; \beta)$ is the density of $\widetilde{X}$ in evaluation data.

We also denote $\prod_{d\in\{1,0\}} \left\{ r^d(y \mid x; \beta)w(d \mid x; \beta) \right\}^{a=d}$ by $f(y, a \mid x; \beta)$.

Then, the log densities under the parametric submodels are given as

$$\log \left(f(y, a, x; \beta) \right)$$
$$= \mathbb{1}[a = 1]\Big(\log \left(r^1(y \mid x) \right) + \log \left(w(1 \mid x; \beta) \right) \Big) + \mathbb{1}[a = 0]\Big(\log \left(r^0(y \mid x) \right) + \log \left(w(0 \mid x; \beta) \right) \Big) + \log \left(p(x; \beta) \right),$$

and $\log \left(q(\widetilde{x}; \beta) \right)$.

We consider the joint log-likelihood of $\mathcal{D}_T$ and $\mathcal{D}_S$, which is defined as

$$\sum_{t=1}^T \log \left(f(Y_t, A_t, X_t; \beta) \right) + \sum_{s=1}^S \log \left(q(\widetilde{X}_s; \beta) \right).$$

By taking the derivative of $\sum_{t=1}^T \log \left(f(Y_t, A_t, X_t; \beta) \right) + \sum_{s=1}^S \log \left(q(\widetilde{X}_s; \beta) \right)$ with respect to β, we can obtain the corresponding score as

$$S(W; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} \left(\sum_{t=1}^T \log \left(f(Y_t, A_t, X_t; \beta) \right) + \sum_{s=1}^S \log \left(q(\widetilde{X}_s; \beta) \right) \right)$$

$$= \sum_{t=1}^{T} \left(\mathbb{1}[A_t = 1] \left\{ s^1(Y_t \mid X_t; \beta) + \frac{\dot{w}(1 \mid X_t; \beta)}{w(1 \mid X_t; \beta)} \right\} + \mathbb{1}[A_t = 0] \left(s^0(Y_t \mid X_t; \beta) + \frac{\dot{w}(0 \mid X_t; \beta)}{w(0 \mid X_t; \beta)} \right) + \zeta(X_t; \beta) \right)$$
$$+ \sum_{s=1}^{S} \widetilde{\zeta}(\widetilde{X}_s; \beta),$$

where s^d, $\dot{\omega}$, ζ, and $\widetilde{\zeta}$ are defined as

$$s^d(y \mid x; \beta) := \frac{\mathrm{d}}{\mathrm{d}\beta} \log r^d(y \mid x; \beta),$$

$$\dot{\omega}(d \mid x; \beta) := \frac{\mathrm{d}}{\mathrm{d}\beta} w(d \mid x; \beta),$$

$$\zeta(x; \beta) := \frac{\mathrm{d}}{\mathrm{d}\beta} \log p(x; \beta),$$

$$\widetilde{\zeta}(\widetilde{x}; \beta) := \frac{\mathrm{d}}{\mathrm{d}\beta} \log q(\widetilde{x}; \beta).$$

A.3. ATE Function

Next, we consider the ATE under the parametric submodel. Under our defined parametric submodel, let $\theta(\beta)$ be our parameter of interest as a function of β, which is given as

$$\theta(\beta) := \mathbb{E}_{X \sim q(x;\beta),(Y,A) \sim f(y,a|X;\beta)} \left[Y(1) - Y(0) \right].$$

We have the following lemma for the derivative of $\theta(\beta)$.

Lemma A.1. *Under the parametric submodel defined in (4), we have*

$$\frac{\partial \theta(\beta)}{\partial \beta} = \mathbb{E}_W \left[Y(1)s^1(Y(1) \mid X; \beta) - Y(0)s^0(Y(0) \mid X; \beta) + \theta(\widetilde{X}; \beta)\widetilde{\zeta}(\widetilde{X}; \beta) \right],$$

where

$$\theta(X; \beta) := \mathbb{E}_{Y(1) \sim r^1(y(1)|X;\beta)}[Y(1) \mid X] - \mathbb{E}_{Y(0) \sim r^0(y(0)|X;\beta)}[Y(0) \mid X].$$

The statement follows from Appendix B in Narita et al. (2019).

A.4. Cramér-Rao Lower Bound

To obtain an intuition about the semiparametric efficiency bound, we discuss the Cramér-Rao lower bound under a specification of the parametric submodels, which simplifies an argument about the efficiency bound.

To consider the Cramér-Rao lower bound, we specify the score functions as follows:

$$s^1(y \mid x; \beta) = \frac{1}{T} \frac{\left(y - \mathbb{E}_{Y(1) \sim r^d(y(1)|x;\beta)}[Y(1) \mid X = x] \right)}{w(1 \mid x)} \tag{5}$$

$$s^0(y \mid x; \beta) = -\frac{1}{T} \frac{\left(y - \mathbb{E}_{Y(0) \sim r^d(y(0)|x;\beta)}[Y(0) \mid X = x] \right)}{w(0 \mid x)}$$

$$\dot{w}(d \mid X; \beta_0) = 0,$$

$$\zeta(x; \beta) = 0,$$

$$\widetilde{\zeta}(\widetilde{x}; \beta) = \frac{1}{S} \left\{ \theta(\widetilde{x}; \beta) - \theta(\beta) \right\}.$$

We consider the Cramér-Rao lower bound when these score functions are true ones in the DGP.

We fix T and S and consider a case where m i.i.d. copes of

$$W := \left(X_1, \cdots, X_T, A_1, \cdots, A_T, Y_1, \cdots, Y_T, \widetilde{X}_1, \cdots, \widetilde{X}_S \right)$$

are available, where W is generated from $\prod_{t=1}^T f(Y_t, A_t, X_t; \beta) \prod_{s=1}^S q(\widetilde{x}; \beta)$. Then, we consider the Cramér-Rao lower bound for estimating θ_0.

Let $\mathbb{E}_W$ is an expectation operator over (Y, A, X) and $\widetilde{X}$.

Given a distribution whose density is given as $\prod_{t=1}^T f(Y_t, A_t, X_t; \beta) \prod_{s=1}^S q(\widetilde{X}_s; \beta)$, the Fisher information is given as

$$\mathbb{E}_W\left[\left(\nabla_\beta\left(\sum_{t=1}^T \log\left(f(Y_t, A_t, X_t; \beta)\right) + \sum_{s=1}^S \log\left(q\left(\widetilde{X}_s; \beta\right)\right)\right)\right)^2\right]^{-1}.$$

Then, the Cramér-Rao lower bound for unbiased estimators of the function $\theta(\beta)$ under the parametric submodel is

$$\mathrm{CR}(\theta) := \nabla_\beta\theta(\beta)\mathbb{E}_W\left[\left(\nabla_\beta\left(\sum_{t=1}^T \log\left(f(Y_t, A_t, X_t; \beta)\right) + \sum_{s=1}^S \log\left(q\left(\widetilde{X}_s; \beta\right)\right)\right)\right)^2\right]^{-1}\nabla_\beta\theta(\beta).$$

Theorem A.2. *Under the specification of (5), the Cramér-Rao lower bound $T\mathrm{CR}(\theta)$ is $\tau^*(\omega, p)$ as $S \to \infty$.*

Proof of Theorem A.2. We defined the ATE as

$$\theta(\beta) := \int \left(y(1) - y(0)\right)r^1\left(y(1) \mid \widetilde{x}; \beta\right)r^0\left(y(0) \mid \widetilde{x}; \beta\right)q(\widetilde{x}; \beta)\mathrm{d}\mu\left(y(1), y(0), \widetilde{x}\right).$$

where μ is a baseline measure such as Lebesgue or a counting measure. The Cramér-Rao lower bound for parametric models defined in (5) is given as follows:

$$T\mathrm{CR}(\theta) = TCD^{-1}C^\top,$$

where

$$C := \mathbb{E}_W\left[Y(1)s^1(Y(1) \mid X; \beta_0) - Y(0)s^0(Y(0) \mid X; \beta_0) + \theta(\widetilde{X}; \beta_0)\widetilde{\zeta}(\widetilde{X}; \beta_0)\right],$$

$$D := \mathbb{E}_W\left[\left(\nabla_\beta\left(\sum_{t=1}^T \log\left(f(Y_t, A_t, X_t; \beta_0)\right) + \sum_{s=1}^S \log\left(q(\widetilde{X}_s; \beta_0)\right)\right)\right)^2\right].$$

Here, we used Lemma A.1.

Under the specification in (5), we have

$$T\mathrm{CR}(\theta)$$
$$= T\mathbb{E}_W\left[\frac{1}{T}\left(Y(1)\frac{(Y(1) - \mathbb{E}[Y(1) \mid X])}{w(1 \mid X)} + Y(0)\frac{(Y(0) - \mathbb{E}[Y(0) \mid X])}{w(0 \mid X)}\right) + \frac{1}{S}\theta(\widetilde{X}; \beta_0)\left(\theta(\widetilde{X}; \beta_0) - \theta(\beta_0)\right)\right]$$
$$\times \mathbb{E}_W\left[T\left(\nabla_\beta \log\left(f(Y, A, X; \beta_0)\right)\right)^2 + S\left(\nabla_\beta \log\left(q(\widetilde{X}; \beta_0)\right)^2\right]^{-1}$$
$$= \mathbb{E}_W\left[\left(\frac{(Y(1) - \mathbb{E}[Y(1) \mid X])^2}{w(1 \mid X)} + \frac{(Y(0) - \mathbb{E}[Y(0) \mid X])^2}{w(0 \mid X)}\right) + \frac{T}{S}\left(\theta^2(\widetilde{X}; \beta_0)\right)\right].$$

Letting $S \to \infty$, we have

$$T\mathrm{CR}(\theta) = \mathbb{E}_W\left[\left(\frac{(Y(1) - \mathbb{E}[Y(1) \mid X])^2}{w(1 \mid X)} + \frac{(Y(0) - \mathbb{E}[Y(0) \mid X])^2}{w(0 \mid X)}\right) + \frac{T}{S}\left(\theta^2(\widetilde{X}; \beta_0)\right)\right] = \tau^*(\omega, p).$$

$\square$

A.5. Derivation of the Semiparametric Efficiency Bound

We extend the Cramér-Rao lower for a case with semiparametric models. As we did in Appendix A.4, we regard the whole $T + S$ data at hand as one observation and consider the case where we observe m observations. Then, as m goes to infinity, the total sample size $T' := Tm$ goes to infinity. Since we consider a case where $T + S$ samples are repeatedly observed from the same distribution independently, we can employ the conventional approach for deriving the semiparametric efficiency bound. Therefore, to employ this strategy, only in this section, we regard T' as the total "hypothetical" sample size in the asymptotic analysis regarding $m \to \infty$ while fixing T.

We explain the definition of the semiparametric efficient influence function. This is a function for one observation

$$O = \left(X_1, \cdots, X_T, A_1, \cdots, A_T, Y_1, \cdots, Y_T, \widetilde{X}_1, \cdots, \widetilde{X}_S \right).$$

The tangent set for this parametric submodel at $\beta = \beta_0$ is given as

$$\mathcal{T} := \left\{ \sum_{t=1}^{T} \left(\sum_{d \in \{1,0\}} \mathbb{1}[A_t = d] \left(s^d(Y_t \mid X_t; \beta_0) + \frac{\dot{w}(d \mid X_t; \beta_0)}{w(d \mid X_t; \beta_0)} \right) + \zeta(X_t; \beta_0) \right) + \sum_{s=1}^{S} \widetilde{\zeta}\left(\widetilde{X}_s; \beta_0 \right) \in L_2(W) \right\},$$

where we have

$$\int s^d(y \mid x; \beta_0) r^d(y \mid x; \beta_0) \mathrm{d}y = 0 \qquad \forall x \in \mathcal{X},$$

$$\dot{w}(1 \mid x; \beta_0) + \dot{w}(0 \mid x; \beta_0) = 0 \qquad \forall x \in \mathcal{X},$$

$$\int \zeta(x; \beta_0) p(x; \beta_0) \mathrm{d}x = 0,$$

$$\int \widetilde{\zeta}(\widetilde{x}; \beta_0) p(\widetilde{x}; \beta_0) \mathrm{d}\widetilde{x} = 0.$$

Here, $\theta(\beta)$ is pathwise differentiable at β_0 if there exists a function $\psi(W) \in \mathcal{T}$ such that $\mathbb{E}[\psi^\top(W)\psi(W)] < \infty$ and for all regular parametric submodels

$$\frac{\partial \theta(\beta_0)}{\partial \beta} := \mathbb{E}_W \left[\psi(W) S(W; \beta_0) \right]. \tag{6}$$

We refer to $\psi(W)$ as the semiparametric efficient influence function. As the following proposition, the semiparametric efficient influence function gives a lower bound for the asymptotic variance of regular estimators.

Proposition A.3 (From Theorem 25.20 in van der Vaart (1998).). *The semiparametric efficient influence function $\psi(W)$ is the gradient of θ w.r.t the model $\mathcal{M}$, which has the smallest L_2-norm. It satisfies that for any regular estimator $\widehat{\theta}$ of θ_0 regarding a given parametric submodel, $\mathrm{AMSE}(\widehat{\theta}) \geq \mathrm{Var}(\psi(W))$, where $\mathrm{AMSE}(\widehat{\theta})$ is the second moment of the limiting distribution of $\sqrt{m}\left(\widehat{\theta} - \theta_0 \right)$.*

This states that $\mathrm{Var}(\psi(W))$ is a lower bound in estimating θ_0. We call $T\mathrm{Var}(\psi(W))$ the semiparametric efficiency bound because what we want to consider is a lower bound of $\sqrt{T}\left(\widehat{\theta} - \theta_0 \right)$.

Lastly, we derive the semiparametric efficient influence function in the following lemma.

Lemma A.4. *Under the parametric submodel defined in (4), we have*

$$\psi(W) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\mathbb{1}[A_t = 1]\left(Y_t - \mathbb{E}[Y_t(1) \mid X_t]\right)}{w(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\left(Y_t - \mathbb{E}[Y_t(0) \mid X]\right)}{w(0 \mid X_t)} \right) \frac{q(X_t)}{p(X_t)} + \frac{1}{S} \sum_{s=1}^{S} \left(\theta_0(\widetilde{X}_s) - \theta_0 \right).$$

The (scaled) efficiency bound $T\mathrm{Var}(\psi(W))$ is

$$\mathbb{E}_{X \sim p(x)} \left[\left(\frac{\sigma_0^2(1)(X)}{w(1 \mid X)} + \frac{\sigma_0^2(0)(X)}{w(0 \mid X)} \right) \frac{q^2(X)}{p^2(X)} \right] + \frac{T}{S} \mathrm{Var}_{\widetilde{X} \sim q(x)} \left(\theta(\widetilde{X}) \right).$$

From this lemma, as $S \to \infty$, we obtain τ^* in Theorem 3.1.

A.6. Proof of Lemma A.4

Following existing studies, such as Hahn (1998), Narita et al. (2019), and Uehara et al. (2020), we consider a nonparametric model for the distribution of potential outcomes and define regular subparametric models. Then, we calculate a gradient (a candidate of the semiparametric efficient influence function) of the target function θ_0, characterize the tangent set for all regular parametric submodels, verify that the parameter of interest is pathwise differentiable, verify that the efficient influence function lies in the tangent space and calculates the expected square of the influence function.

As discussed in Appendices A.2, we define parametric submodels as

$$p(W; \beta) = \prod_{t=1}^{T} \prod_{d \in \{1,0\}} \left(r^d(Y_t \mid X_t; \beta) w(d \mid X_t; \beta) \right)^{A_t = d} p(X_t; \beta) \prod_{s=1}^{S} q(\widetilde{X}_s; \beta),$$

$$W = \left\{ \{Y_t, A_t, X_t\}_{t=1}^{T}, \{\widetilde{X}_s\}_{s=1}^{S} \right\}.$$

Then, recall that in Appendix A.2, we define the score function as

$$S(W; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} \left(\sum_{t=1}^{T} \log \left(f(Y_t, A_t, X_t; \beta) \right) + \sum_{s=1}^{S} \log \left(q(\widetilde{X}_s; \beta) \right) \right)$$

$$= \sum_{t=1}^{T} \left(\mathbb{1}[A_t = 1] \left(s^1(Y_t \mid X_t; \beta) + \frac{\dot{w}(1 \mid X_t; \beta)}{w(1 \mid X_t; \beta)} \right) + \mathbb{1}[A_t = 0] \left(s^0(Y_t \mid X_t; \beta) + \frac{\dot{w}(0 \mid X_t; \beta)}{w(0 \mid X_t; \beta)} \right) + \zeta(X_t; \beta) \right)$$

$$+ \sum_{s=1}^{S} \widetilde{\zeta}(\widetilde{X}_s; \beta),$$

where recall that $s^d, \dot{w}, \zeta,$ and $\widetilde{\zeta}$ are defined as

$$s^d(y \mid x; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} \log r^d(y \mid x; \beta),$$

$$\dot{w}(d \mid x; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} w(d \mid x; \beta),$$

$$\zeta(x; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} \log p(x; \beta),$$

$$\widetilde{\zeta}(\widetilde{x}; \beta) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\beta} \log q(\widetilde{x}; \beta).$$

As shown in Lemma A.1, we have

$$\frac{\partial \theta(\beta)}{\partial \beta} = \mathbb{E}_W \left[Y(1) s^1(Y(1) \mid X; \beta) - Y(0) s^0(Y(0) \mid X; \beta) + \theta(\widetilde{X}; \beta) \widetilde{\zeta}(\widetilde{X}; \beta) \right],$$

where recall that

$$\theta(X; \beta) \coloneqq \mathbb{E}_{(Y,A) \sim f(y,a \mid X; \beta)}[Y(1) \mid X] - \mathbb{E}_{(Y,A) \sim f(y,a \mid X; \beta)}[Y(0) \mid X].$$

Then, based on the above preparation, we prove Lemma A.4 by guess and verify.

Proof of Lemma A.4. Let us denote $\mathbb{E}_{\widetilde{X} \sim q(x), (Y,A) \sim f(y,a \mid X)}$ by $\mathbb{E}_W$. We first verify that $\mathbb{E}_W[\psi^2(W)] < \infty$. Here, we have

$$\mathbb{E}_W \left[\left(\frac{\mathbb{1}[A = 1] \left(Y - \mathbb{E}[Y(1) \mid \widetilde{X}] \right)}{w(1 \mid \widetilde{X})} - \frac{\mathbb{1}[A = 0] \left(Y - \mathbb{E}[Y(0) \mid \widetilde{X}] \right)}{w(0 \mid \widetilde{X})} \right) \frac{q(\widetilde{X})}{p(\widetilde{X})} \mid \widetilde{X} \right] = 0.$$

Therefore, we have

$$\mathbb{E}_W\left[\left(\left(\frac{\mathbb{1}[A=1]\left(Y-\mathbb{E}[Y(1)\mid\widetilde{X}]\right)}{w(1\mid\widetilde{X})}-\frac{\mathbb{1}[A=0]\left(Y-\mathbb{E}[Y(0)\mid\widetilde{X}]\right)}{w(0\mid\widetilde{X})}\right)\frac{q(\widetilde{X})}{p(\widetilde{X})}\right)\left(\theta_0(\widetilde{X})-\theta_0\right)\right]$$

$$=\mathbb{E}_W\left[\left(\theta_0(\widetilde{X})-\theta_0\right)\mathbb{E}_W\left[\left(\frac{\mathbb{1}[A=1]\left(Y-\mathbb{E}[Y(1)\mid\widetilde{X}]\right)}{w(1\mid\widetilde{X})}-\frac{\mathbb{1}[A=0]\left(Y-\mathbb{E}[Y(0)\mid\widetilde{X}]\right)}{w(0\mid X)}\right)\frac{q(\widetilde{X})}{p(\widetilde{X})}\mid\widetilde{X}\right]\right]$$

$$=0.$$

Hence, we have

$$\mathbb{E}_W\left[\psi^2(W)\right]$$

$$=\mathbb{E}_W\left[\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbb{1}[A_t=1]\left(Y_t-\mathbb{E}[Y_t(1)\mid\widetilde{X}_t]\right)}{w(1\mid\widetilde{X}_t)}-\frac{\mathbb{1}[A_t=0]\left(Y_t-\mathbb{E}[Y_t(0)\mid\widetilde{X}_t]\right)}{w(0\mid\widetilde{X}_t)}\right)\frac{q(\widetilde{X}_t)}{p(\widetilde{X}_t)}\right.\right.$$

$$\left.\left.+\frac{1}{S}\sum_{s=1}^{S}\left(\theta_0(\widetilde{X}_s)-\theta_0\right)\right)^2\right]$$

$$=\mathbb{E}_W\left[\frac{1}{T}\left(\frac{\sigma_0^2(1)(\widetilde{X}_t)}{w(1\mid\widetilde{X}_t)}+\frac{\sigma_0^2(0)\widetilde{X}_t)}{w(0\mid\widetilde{X}_t)}\right)\frac{q^2(X_t)}{p^2(\widetilde{X}_t)}+\frac{1}{S}\left\{\theta_0(\widetilde{X}_s)-\theta_0\right\}^2\right]<\infty.$$

We next verify that (6) holds ($\frac{\partial\theta(\beta_0)}{\partial\beta}:=\mathbb{E}_W\left[\psi(W)S(W;\beta_0)\right]$) by showing

$$\mathbb{E}_W\left[\psi(W)S(W;\beta_0)\right]=\mathbb{E}_W\left[Y(1)s^1(Y(1)\mid X;\beta_0)-Y(0)s^0(Y(0)\mid X;\beta_0)+\theta(\widetilde{X};\beta_0)\widetilde{\zeta}(\widetilde{X};\beta_0)\right],\qquad(7)$$

from Lemma A.1.

Under our choice of $\psi(W)$, the LHS of (7) becomes

$$\mathbb{E}_W\left[\psi(W)S(W;\beta)\right]$$

$$=\mathbb{E}_W\left[\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\mathbb{1}[A_t=1]\left(Y_t-\mathbb{E}[Y_t(1)\mid X]\right)}{w(1\mid X_t)}-\frac{\mathbb{1}[A_t=0]\left(Y_t-\mathbb{E}[Y_t(0)\mid X_t]\right)}{w(0\mid X_t)}\right)\frac{q(X_t)}{p(X_t)}\right.\right.$$

$$\left.+\frac{1}{S}\sum_{s=1}^{S}\left\{\theta_0(\widetilde{X}_s)-\theta_0\right\}\right)$$

$$\times\left(\sum_{t=1}^{T}\left(\mathbb{1}[A_t=1]\left(s^1(Y_t(1)\mid X_t;\beta)+\frac{\dot{w}(1\mid X;\beta)}{w(1\mid X_t;\beta)}\right)+\mathbb{1}[A_t=0]\left(s^0(Y_t(0)\mid X_t;\beta)+\frac{\dot{w}(0\mid X_t;\beta)}{w(0\mid X_t;\beta)}\right)\right)\right.$$

$$\left.\left.+\sum_{s=1}^{S}\widetilde{\zeta}(\widetilde{X}_s;\beta)\right)\right]$$

$$=\mathbb{E}_W\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\mathbb{1}[A_t=1]\left(Y_t-\mathbb{E}[Y_t(1)\mid X_t]\right)}{w(1\mid X_t)}\frac{q(X_t)}{p(X_t)}\left(s^1(Y_t(1)\mid X_t;\beta)+\frac{\dot{w}(1\mid X_t;\beta)}{w(1\mid X_t;\beta)}\right)\right.$$

$$\left.+\frac{1}{S}\sum_{s=1}^{S}\theta_0(\widetilde{X}_s)\sum_{t=1}^{T}\mathbb{1}[A_t=1]\left(s^1(Y_t(1)\mid X_t;\beta)+\frac{\dot{w}(1\mid X_t;\beta)}{w(1\mid X_t;\beta)}\right)\right]$$

$$-\mathbb{E}_W\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\mathbb{1}[A_t=0]\left(Y_t-\mathbb{E}[Y_t(0)\mid X]\right)}{w(0\mid X_t)}\frac{q(X)}{p(X_t)}\left(s^0(Y_t(1)\mid X_t;\beta)+\frac{\dot{w}(0\mid X_t;\beta)}{w(0\mid X_t;\beta)}\right)\right.$$

$$\left.+\frac{1}{S}\sum_{s=1}^{S}\theta_0(\widetilde{X}_s)\sum_{t=1}^{T}\mathbb{1}[A_t=0]\left(s^0(Y_t(0)\mid X_t;\beta)+\frac{\dot{w}(0\mid X_t;\beta)}{w(0\mid X_t;\beta)}\right)\right]$$

$$- \theta_0 \mathbb{E}_W \left[\frac{1}{T} \sum_{t=1}^{T} \left(\mathbb{1}[A_t = 1] \left(s^1(Y_t(1) \mid X_t; \beta) + \frac{\dot{w}(1 \mid X_t; \beta)}{w(1 \mid X_t; \beta)} \right) + \mathbb{1}[A_t = 0] \left(s^0(Y_t(0) \mid X_t; \beta) + \frac{\dot{w}(0 \mid X_t; \beta)}{w(0 \mid X_t; \beta)} \right) \right) \right.$$

$$\left. + \sum_{s=1}^{S} \widetilde{\zeta}(\widetilde{X}_s; \beta) \right]$$

$$+ \mathbb{E}_W \left[\left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\mathbb{1}[A_t = 1] (Y_t - \mathbb{E}[Y_t(1) \mid X_t])}{w(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0] (Y_t - \mathbb{E}[Y_t(0) \mid X_t])}{w(0 \mid X_t)} \right) \frac{q(X_t)}{p(X_t)} + \frac{1}{S} \sum_{s=1}^{S} \theta_0(\widetilde{X}_s) \right) \right.$$

$$\left. \times \sum_{s=1}^{S} \widetilde{\zeta}(\widetilde{X}_s; \beta) \right],$$

where we used $\mathbb{1}[A_t = 1]\mathbb{1}[A_t = 0] = 0$.

Here, it holds that

$$\mathbb{E}_W \left[\mathbb{1}[A = d] \left(s^d(Y(d) \mid X; \beta) + \frac{\dot{w}(d \mid X; \beta)}{w(d \mid X; \beta)} \right) \mid X \right]$$

$$= \mathbb{E}_W \left[\mathbb{1}[A = d] \mid X \right] \left(s^d(Y(d) \mid X; \beta) + \frac{\dot{w}(d \mid X; \beta)}{w(d \mid X; \beta)} \right)$$

$$= w(d \mid X; \beta) s^d(Y(d) \mid X; \beta) + \dot{w}(d \mid X; \beta) = 0,$$

and

$$\mathbb{E}_W \left[\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\mathbb{1}[A_t = 1] (Y_t - \mathbb{E}[Y_t(1) \mid X_t])}{w(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0] (Y_t - \mathbb{E}[Y_t(0) \mid X_t])}{w(0 \mid X_t)} \right) \frac{q(X_t)}{p(X_t)} + \frac{1}{S} \sum_{s=1}^{S} \left(\theta_0(\widetilde{X}_s) - \theta_0 \right) \right]$$

$$= 0.$$

Therefore, we have

$$\mathbb{E}_W \left[\psi(W) S(W; \beta_0) \right] = \mathbb{E}_W \left[Y(1) \frac{q(X)}{p(X)} s^1(Y(1) \mid X; \beta) - Y(0) \frac{q(X)}{p(X)} s^0(Y(1) \mid X; \beta) + \theta_0(\widetilde{X}) \zeta(\widetilde{X}; \beta) \right].$$

Here, from Lemma A.1, we have

$$\mathbb{E}_W \left[Y(1) \frac{q(X)}{p(X)} s^1(Y(1) \mid X; \beta) - Y(0) \frac{q(X)}{p(X)} s^0(Y(1) \mid X; \beta) + \theta_0(\widetilde{X}) \zeta(\widetilde{X}; \beta) \right]$$

$$= \mathbb{E}_W \left[Y(1) s^1(Y(1) \mid X; \beta) - Y(0) s^0(Y(0) \mid X; \beta) + \theta(\widetilde{X}; \beta) \widetilde{\zeta}(\widetilde{X}; \beta) \right].$$

Lastly, we verify that $\psi \in \mathcal{T}$. We define the functions as

$$s^d(y \mid x; \beta_0) = \frac{1}{T} \frac{(y - \mathbb{E}[Y(d) \mid X = x])}{w(d \mid x)},$$

$$\dot{w}(d \mid X; \beta_0) = 0,$$

$$\zeta(x; \beta_0) = 0,$$

$$\widetilde{\zeta}(\widetilde{x}; \beta_0) = \frac{1}{S} \left\{ \theta_0(\widetilde{x}) - \theta_0 \right\}.$$

Then, $\psi \in \mathcal{T}$. The proof is complete. $\qquad\square$

B. Proof of Theorem 4.1

Proof. We solve the following constraint optimization problem:

$$\min_{z \in \mathcal{P}} \int \left(\left(\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)} \right) \frac{q^2(x)}{p(x)} \right) \mathrm{d}x$$

$$\text{s.t.} \int p(x)dx = 1.$$

Then, by using the Lagrangian variables $\alpha \in (-\infty, +\infty)$, we consider its Lagrange functional

$$\mathcal{L}(z; \alpha) = \int \left(\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)} \right) \frac{q^2(x)}{p(x)} \mathrm{d}x + \alpha \left(\int p(x)\mathrm{d}x - 1 \right).$$

Then we consider the infimum of $\mathcal{L}$ with respect to z and maximization of $\mathcal{L}$ with respect to α, that is, $\inf_{p \in \mathcal{P}} \sup_{\alpha \in (-\infty, +\infty)} \mathcal{L}(p; \alpha)$. Next, we consider the dual problem defined as $\sup_{\alpha \in (-\infty, +\infty)} \inf_{z \in \mathcal{P}} \mathcal{L}(z; \alpha)$. Owing to the convexity of $\int \left(\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)} \right) \frac{q^2(x)}{p(x)} dx$ for $p(\cdot) \in \mathbb{R}^+$, the following equality is obtained (Rockafellar, 2015):

$$\inf_{p \in \mathcal{P}} \sup_{\alpha \in (0, +\infty)} \mathcal{L}(p; \alpha) = \sup_{\alpha \in (0, +\infty)} \inf_{p \in \mathcal{P}} \mathcal{L}(p; \alpha).$$

Hence, we discuss the solution to the dual problem. Given α, we apply the Euler-Lagrange equation for calculating $\inf_p \mathcal{L}(p; \alpha)$ (Gelfand et al., 2000). The minimizer p^* satisfies the following equation:

$$-\left(\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)} \right) \frac{q^2(x)}{p^{*2}(x)} + \alpha = 0.$$

Therefore, we obtain the following KKT condition:

$$-\left(\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)} \right) \frac{q^2(x)}{p^{*2}(x)} + \alpha^* = 0,$$

$$0 < p^*(x),$$

$$\alpha^* \left(\int p(x)dx - 1 \right) = 0.$$

Then, we obtain the optimal solutions as follows:

- if $\alpha^* = 0$, we have $-\left(\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)} \right) \frac{q^2(x)}{p^{*2}(x)} = 0$. In this case, for $0 < p^*(x)$, there is no feasible solution.

- if $\alpha^* \neq 0$, we have $-\left(\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)} + \theta_0(x) \right) \frac{q^2(x)}{p^{*2}(x)} + \theta_0 + \alpha^* = 0$ and $\alpha^* \left(\int p(x)dx - 1 \right) = 0$. From these results, we have

$$p^*(x) = \sqrt{\frac{\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)} + \theta_0(x)}{\alpha^* - \theta_0}} q(x),$$

$$\alpha^* - \theta_0 = \left(\mathbb{E}_q \left[\sqrt{\frac{\sigma_0^2(1)(x)}{w(1 \mid x)} + \frac{\sigma_0^2(0)(x)}{w(0 \mid x)} + \theta_0(x)} \right] \right)^2.$$

Therefore, we have

$$p^*(x) = \frac{\sqrt{\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)}}}{\mathbb{E}_q \left[\sqrt{\frac{\sigma_0^2(1)(x)}{w(1|x)} + \frac{\sigma_0^2(0)(x)}{w(0|x)}} \right]} q(x).$$

$\square$

C. Proof of Theorem 5.3

Let $\mathbb{E}$ be an expectation operator over (Y_t, A_t, X_t) and $\widetilde{X}_s$; that is,

$$\mathbb{E}_{P_t} = \mathbb{E}_{\widetilde{X}_s \sim q(x), X_t \sim \widehat{p}_t(x), A_t \sim \widehat{w}_t(a|X_t), Y_t(A_t) \sim r^{A_t}(y|X_t)}.$$

Let us define

$$
\begin{aligned}
\xi_t &= \left(\psi_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t) - \theta_0 \right) / \sqrt{\tau^*} \\
&= \left(\left(\frac{\mathbb{1}[A_t = 1] \left(Y_t - \widehat{\mu}_t(1)(X_t) \right)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0] \left(Y_t - \widehat{\mu}_t(0)(X_t) \right)}{\widehat{w}_t(0 \mid X_t)} \right) \frac{q(X_t)}{\widehat{p}_t(X_t)} + \mathbb{E}_{\widetilde{X} \sim q(x)} \left[\widehat{\theta}_t(\widetilde{X}) \right] - \theta_0 \right) / \sqrt{\tau^*},
\end{aligned}
$$

where

$$\tau^* = \mathbb{E}_{X \sim p^*(x)} \left[\left(\frac{\sigma_0^2(1)(X)}{w^*(1 \mid X)} + \frac{\sigma_0^2(0)(X)}{w^*(0 \mid X)} \right) \frac{q^2(X)}{p^{*2}(X)} \right].$$

First, the sequence $\{\xi_t\}_{t=1}^T$ is a martingale difference sequence (MDS).

Lemma C.1. *Under the AAS-AIPWIS experiment, $\mathbb{E}_{Y_t(a) \sim r^a(y|x), A_t \sim \widehat{w}_t(a|x), X_t \sim \widehat{p}_t(x)}[\xi_t | \mathcal{F}_{t-1}] = 0$ holds.*

Proof of Lemma C.1. For each $a \in \{1, 0\}$, we have

$$
\begin{aligned}
&\mathbb{E}_{P_t} \left[\frac{\mathbb{1}[A_t = a] \left(Y_t - \widehat{\mu}_t(a)(X_t) \right)}{\widehat{w}_t(a \mid X_t)} \mid X_t, \mathcal{F}_{t-1} \right] \\
&= \frac{\mathbb{E}_{P_t} \left[\mathbb{1}[A_t = a] \left(Y_t(a) - \widehat{\mu}_t(a)(X_t) \right) \mid X_t, \mathcal{F}_{t-1} \right]}{\widehat{w}_t(a \mid X_t)} \\
&= \frac{\widehat{w}_t(a|X_t) \left(\mu(a)(X_t) - \widehat{\mu}_t(a)(X_t) \right)}{\widehat{w}_t(a|X_t)} \\
&= \mu(a)(X_t) - \widehat{\mu}_t(a)(X_t).
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\mathbb{E}_{P_t} \left[\xi_t \mid X_t, \mathcal{F}_{t-1} \right] \\
&= \mathbb{E}_{P_t} \left[\left(\psi_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t) - \theta_0 \right) / \sqrt{\tau^*} \mid X_t, \mathcal{F}_{t-1} \right] \\
&= \left(\left(\mu(1)(X_t) - \mu(0)(X_t) - \widehat{\mu}_t(1)(X_t) + \widehat{\mu}_t(0)(X_t) \right) \frac{q(X_t)}{\widehat{p}_t(X_t)} + \mathbb{E}_{\widetilde{X} \sim q(x)} \left[\widehat{\theta}_t(\widetilde{X}) \right] - \theta_0 \right) / \sqrt{\tau^*}.
\end{aligned}
$$

Note that $\mathbb{E}_{P_t} \left[\widehat{\mu}_t(1)(X_t) - \widehat{\mu}_t(0) \mid \mathcal{F}_{t-1} \right] - \mathbb{E}_{P_t} \left[\widehat{\theta}_t(\widetilde{X}) \right] = 0$. Therefore,

$$\mathbb{E}_{P_t} \left[\xi_t \mid \mathcal{F}_{t-1} \right] = 0.$$

$\square$

Then, we show Theorem 5.3 by using the martingale central limit theorem (CLT), which is given as follows.

Proposition C.2 (Central Limit Theorem for an MDS; from Proposition 7.9, p. 194, Hamilton (1994); also see White (1984))**.** *Let $\{W_t\}_{t=1}^\infty$ be a scalar MDS with some random variable W_t. Let $\overline{W} = \frac{1}{T} \sum_{t=1}^T W_t$. Suppose that*

(a) $\mathbb{E}[W_t^2] = \sigma_t^2$, *a positive value with $\frac{1}{T} \sum_{t=1}^T \sigma_t^2 \to \sigma^2$, a positive value;*

(b) $\mathbb{E}[|W_t|^r] < \infty$ *for some $r > 2$ and all $t \in \mathbb{N}$;*

(c) $\frac{1}{T}\sum_{t=1}^{T} W_t^2 \xrightarrow{p} \sigma^2$.

Then $\sqrt{T}\overline{W} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2)$.

Thus, to apply the martingale CLT, it is enough to verify

(a) $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{Y_t(a)\sim r^a(y|x), A_t\sim \widehat{w}_t(a|x), X_t \sim \widehat{p}_t(x)}[\xi_t^2] \to 1$;

(b) $\mathbb{E}_{Y_t(a)\sim r^a(y|x), A_t\sim \widehat{w}_t(a|x), X_t \sim \widehat{p}_t(x)}[|\xi_t|^r] < \infty$ for some $r > 2$ and all $t \in \mathbb{N}$;

(c) $\frac{1}{T}\sum_{t=1}^{T} \xi_t^2 \xrightarrow{P} 1$.

Under this proof strategy, we show Theorem 5.3 below. Our proof is inspired by Kato et al. (2020). We denote $\mathbb{E}_{Y_t(a)\sim r^a(y|x), A_t\sim \widehat{w}_t(a|x), X_t \sim \widehat{p}_t(x)}$ by $\mathbb{E}_{P_t}$. In the proof, we use the following results.

Definition C.3. [Uniform integrability, Hamilton (1994), p. 191] Let $W_t \in \mathbb{R}$ be a random variable with a probability measure P. A sequence $\{W_t\}$ is said to be uniformly integrable if for every $\epsilon > 0$ there exists a number $c > 0$ such that

$$\mathbb{E}[|W_t| \cdot \mathbb{1}[|W_t| \geq c]] < \epsilon$$

for all t.

Proposition C.4 (Sufficient conditions for uniform integrability; Proposition 7.7, p. 191. Hamilton (1994)). *Let $W_t, Z_t \in \mathbb{R}$ be random variables. Let P be a probability measure of Z_t. (a) Suppose there exist $r > 1$ and $M < \infty$ such that $\mathbb{E}_P[|W_t|^r] < M$ for all t. Then $\{W_t\}$ is uniformly integrable. (b) Suppose there exist $r > 1$ and $M < \infty$ such that $\mathbb{E}_P[|Z_t|^r] < M$ for all t. If $W_t = \sum_{j=-\infty}^{\infty} h_j Z_{t-j}$ with $\sum_{j=-\infty}^{\infty} |h_j| < \infty$, then $\{W_t\}$ is uniformly integrable.*

Proposition C.5 (L^r convergence theorem, p 165, Loeve (1977)). *Let W_t be a random variable with probability measure P and w be a constant. Let $0 < r < \infty$, suppose that $\mathbb{E}[|W_t|^r] < \infty$ for all t and that $W_t \xrightarrow{P} z$ as $n \to \infty$. The following are equivalent:*

(i) $W_t \to w$ in L^r as $t \to \infty$;

(ii) $\mathbb{E}[|W_t|^r] \to \mathbb{E}_P[|w|^r] < \infty$ as $t \to \infty$;

(iii) $\{|W_t|^r, t \geq 1\}$ is uniformly integrable.

Then, we prove Theorem 5.3 as follows.

Proof of Theorem 5.3. We check conditions (a)–(b) in the following Steps 1–3.

Step 1: check of condition (a) Note that $\sqrt{\tau^*}$ is non-random variable. Hence, we focus on the expectation of

$$\left(\frac{\mathbb{1}[A_t = 1]\big(Y_t - \widehat{\mu}_t(1)(X_t)\big)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\big(Y_t - \widehat{\mu}_t(0)(X_t)\big)}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q(X_t)}{\widehat{p}_t(X_t)} + \mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0$$

conditioned on $\mathcal{F}_{t-1}$.

Instead of $\sum_{t=1}^{T} \mathbb{E}_P[\xi 2_t(P)]$, we first consider the convergence of $\sum_{t=1}^{T} \mathbb{E}_P[(xi_t^2(P)|\mathcal{F}_{t-1}] = \Omega_t(P)$; that is, we show $\Omega_t(P) - 1 \xrightarrow{P} 0$. Then, by using the L^r-convergence theorem (Proposition C.5), we show $\sum_{t=1}^{T} \mathbb{E}_P[\xi_t^2(P)] - 1 \to 0$.

The conditional expectation is computed as follows:

$$\mathbb{E}_{P_t}\left[\left(\left(\frac{\mathbb{1}[A_t = 1]\big(Y_t - \widehat{\mu}_t(1)(X_t)\big)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\big(Y_t - \widehat{\mu}_t(0)(X_t)\big)}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q(X_t)}{\widehat{p}_t(X_t)} + \mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2 \Big| \mathcal{F}_{t-1}\right]$$

$$= \mathbb{E}_{P_t}\left[\left(\frac{\mathbb{1}[A_t = 1]\big(Y_t - \widehat{\mu}_t(1)(X_t)\big)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\big(Y_t - \widehat{\mu}_t(0)(X_t)\big)}{\widehat{w}_t(0 \mid X_t)}\right)^2 \frac{q^2(X_t)}{\widehat{p}_t^2(X_t)} \Big| \mathcal{F}_{t-1}\right]$$

23

$$+ 2\mathbb{E}_{P_t}\left[\left(\frac{\mathbb{1}[A_t = 1]\left(Y_t - \widehat{\mu}_t(1)(X_t)\right)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\left(Y_t - \widehat{\mu}_t(0)(X_t)\right)}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q(X_t)}{\widehat{p}_t(X_t)}\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right) \mid \mathcal{F}_{t-1}\right]$$

$$+ \mathbb{E}_{P_t}\left[\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2 \mid \mathcal{F}_{t-1}\right]$$

$$= \mathbb{E}_{P_t}\left[\left(\frac{\mathbb{1}[A_t = 1]\left(Y_t - \widehat{\mu}_t(1)(X_t)\right)^2}{\widehat{w}_t^2(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\left(Y_t - \widehat{\mu}_t(0)(X_t)\right)^2}{\widehat{w}_t^2(0 \mid X_t)}\right)\frac{q^2(X_t)}{\widehat{p}_t^2(X_t)} \mid \mathcal{F}_{t-1}\right]$$

$$+ 2\mathbb{E}_{P_t}\left[\left(\frac{\mathbb{1}[A_t = 1]\left(Y_t - \widehat{\mu}_t(1)(X_t)\right)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\left(Y_t - \widehat{\mu}_t(0)(X_t)\right)}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q(X_t)}{\widehat{p}_t(X_t)}\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right) \mid \mathcal{F}_{t-1}\right]$$

$$+ \mathbb{E}_{P_t}\left[\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2 \mid \mathcal{F}_{t-1}\right]$$

$$= \mathbb{E}_{P_t}\left[\left(\frac{\left(Y_t(1) - \widehat{\mu}_t(1)(X_t)\right)^2}{\widehat{w}_t(1 \mid X_t)} - \frac{\left(Y_t(0) - \widehat{\mu}_t(0)(X_t)\right)^2}{\widehat{w}_t(0 \mid X_t)}\right)\frac{q^2(X_t)}{\widehat{p}_t^2(X_t)} \mid \mathcal{F}_{t-1}\right]$$

$$+ 2\mathbb{E}_{P_t}\left[\left(\theta_0(X_t) - \widehat{\theta}_t(X_t)\right)\frac{q(X_t)}{\widehat{p}_t(X_t)}\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right) \mid \mathcal{F}_{t-1}\right]$$

$$+ \mathbb{E}_{P_t}\left[\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2 \mid \mathcal{F}_{t-1}\right].$$

From $\widehat{\mu}_t(a)(x) \xrightarrow{\text{a.s.}} \mu^a(P)(x)$ and $\widehat{w}_t(a|x) \xrightarrow{\text{a.s.}} w^*(a|x)$, for all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$, we have

$$\left(\frac{\left(Y_t(1) - \widehat{\mu}_t(1)(x)\right)^2}{\widehat{w}_t(1 \mid x)} - \frac{\left(Y_t(0) - \widehat{\mu}_t(0)(x)\right)^2}{\widehat{w}_t(0 \mid x)}\right)\frac{q^2(x)}{\widehat{p}_t^2(x)}$$

$$+ 2\left(\theta_0(x) - \widehat{\theta}_t(x)\right)\frac{q(x)}{\widehat{p}_t(x)}\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right) + \left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2$$

$$\xrightarrow{\text{a.s}} \left(\frac{\left(Y_t(1) - \mu(1)(x)\right)^2}{w^*(1 \mid x)} - \frac{\left(Y_t(0) - \mu(0)(x)\right)^2}{w^*(0 \mid x)}\right)\frac{q^2(x)}{p^{*2}(x)}.$$

Let us define

$$W_t(x) := \left(\frac{\left(Y_t(1) - \widehat{\mu}_t(1)(x)\right)^2}{\widehat{w}_t(1 \mid x)} - \frac{\left(Y_t(0) - \widehat{\mu}_t(0)(x)\right)^2}{\widehat{w}_t(0 \mid x)}\right)\frac{q^2(x)}{\widehat{p}_t^2(x)}$$

$$+ 2\left(\theta_0(x) - \widehat{\theta}_t(x)\right)\frac{q(x)}{\widehat{p}_t(x)}\left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right) + \left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2$$

$$- \left(\frac{\left(Y_t(1) - \mu(1)(x)\right)^2}{w^*(1 \mid x)} - \frac{\left(Y_t(0) - \mu(0)(x)\right)^2}{w^*(0 \mid x)}\right)\frac{q^2(x)}{p^{*2}(x)} - \left(\mathbb{E}_{\widetilde{X}\sim q(x)}\left[\widehat{\theta}_t(\widetilde{X})\right] - \theta_0\right)^2.$$

Therefore, we obtain

$$\mathbb{E}_{P_t}[W_t(X_t) \mid X_t, \mathcal{F}_{t-1}] \xrightarrow{\text{a.s.}} 0.$$

Next, we show $\mathbb{E}_{P_t}[W_t(X_t)] \to 0$ by using the L^r-convergence theorem (Proposition C.5). Note that

$$\mathbb{E}_{P_t}\left[\left(\frac{\left(Y_t(1) - \mu(1)(X_t)\right)^2}{w^*(1 \mid X_t)} - \frac{\left(Y_t(0) - \mu(0)(X_t)\right)^2}{w^*(0 \mid X_t)}\right)\frac{q^2(X_t)}{p^{*2}(X_t)}\right] = \tau^*.$$

To apply the L^r-convergence theorem, we check that $W_t(X_t)$ is uniformly integrable. Here, ξ_t is conditionally sub-Gaussian. From Lemma 2.7.6 of Vershynin (2018), the squared value ξ_t^2 is conditionally sub-exponential. Therefore, $W_t(X_t)$ is a sum of the sub-exponential random variable. This implies that $W_t(X_t)$ is uniformly integrable. As a result, from the L^r-convergence theorem, $\mathbb{E}_{P_t}[W_t(X_t)] \to 0$ holds.

Finally, $\mathbb{E}_{P_t}[W_t(X_t)]/\tau^* \to 0$ implies

$$\mathbb{E}_{P_t}[\xi_t^2] - 1 \to 0.$$

Step 2: check of condition (b) From Assumption 5.1, $\xi_t^{a,b}$ is sub-Gaussian. When ξ_t is sub-Gaussian, the condition holds from Proposition 2.5.2 (ii) of Vershynin (2018).

Step 3: check of condition (c) Let us define u_t as

$$
\begin{aligned}
u_t &= \xi_t^2 - \mathbb{E}\big[\xi_t^2 \mid \mathcal{F}_{t-1}\big] \\
&= \Bigg(\bigg(\frac{\big(Y_t(1) - \widehat{\mu}_t(1)(x)\big)^2}{\widehat{w}_t(1 \mid x)} - \frac{\big(Y_t(0) - \widehat{\mu}_t(0)(x)\big)^2}{\widehat{w}_t(0 \mid x)} \bigg) \frac{q^2(x)}{\widehat{p}_t^2(x)} \\
&\quad + 2\big(\theta_0(x) - \widehat{\theta}_t(x)\big) \frac{q(x)}{\widehat{p}_t(x)} \Big(\mathbb{E}_{\widetilde{X} \sim q(x)}\big[\widehat{\theta}_t(\widetilde{X})\big] - \theta_0 \Big) + \Big(\mathbb{E}_{\widetilde{X} \sim q(x)}\big[\widehat{\theta}_t(\widetilde{X})\big] - \theta_0 \Big)^2 \Bigg)/\tau^* \\
&\quad - \mathbb{E}_{P_t}\Bigg[\bigg(\bigg(\frac{\big(Y_t(1) - \widehat{\mu}_t(1)(x)\big)^2}{\widehat{w}_t(1 \mid x)} - \frac{\big(Y_t(0) - \widehat{\mu}_t(0)(x)\big)^2}{\widehat{w}_t(0 \mid x)} \bigg) \frac{q^2(x)}{\widehat{p}_t^2(x)} \\
&\quad + 2\big(\theta_0(x) - \widehat{\theta}_t(x)\big) \frac{q(x)}{\widehat{p}_t(x)} \Big(\mathbb{E}_{\widetilde{X} \sim q(x)}\big[\widehat{\theta}_t(\widetilde{X})\big] - \theta_0 \Big) + \Big(\mathbb{E}_{\widetilde{X} \sim q(x)}\big[\widehat{\theta}_t(\widetilde{X})\big] - \theta_0 \Big)^2 \bigg)/\tau^* \; \Big| \; \mathcal{F}_{t-1} \Bigg].
\end{aligned}
$$

Note that u_t is an MDS.

From the boundedness of each variable in u_t, we can apply the weak law of large numbers for an MDS and obtain

$$\frac{1}{T} \sum_{t=1}^{T} u_t = \frac{1}{T} \sum_{t=1}^{T} \Big(\xi_t^2 - \mathbb{E}\big[\xi_t^2 \mid \mathcal{F}_{t-1}\big] \Big) \xrightarrow{\mathrm{P}} 0.$$

From Assumption 5.1 and the result of Step 1, as well as the proof in Lemma 10 in Hadad et al. (2019), we obtain

$$\frac{1}{T} \sum_{t=1}^{T} \xi_t^2 - 1 \xrightarrow{\mathrm{P}} 0.$$

Then, from the L^r-convergence theorem, we have

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big[\xi_t^2 \mid \mathcal{F}_{t-1}\big] - 1 \xrightarrow{\mathrm{P}} 0.$$

As a conclusion, we obtain

$$\frac{1}{T} \sum_{t=1}^{T} \xi_t^2 - 1 = \frac{1}{T} \sum_{t=1}^{T} \Big(\xi_t^2 - \mathbb{E}\big[\xi_t^2 \mid \mathcal{F}_{t-1}\big] + \mathbb{E}\big[\xi_t^2 \mid \mathcal{F}_{t-1}\big] - 1 \Big) \xrightarrow{\mathrm{P}} 0.$$

Conclusion In Steps 1–3. we checked the conditions required for applying the martingale CLT hold. Therefore, from the martingale CLT, we obtain the statement. $\qquad \square$

D. Details of the Simulation Studies

This section provides details of simulation studies. The simulation studies were conducted by using the MacBook Pro and a Linux ubuntu18 server with 64 cores and 251 GB of RAM.

D.1. Details of Alternative Experiments

We explain the details of alternative experiments:

DRCT. We sample experimental units from $q(x)$. In each odd round $t = 1, 3, 5, \ldots$, we assign treatment $A_t = 1$. In each even round $t = 2, 4, 6, \ldots$, we assign treatment $A_t = 0$. Because there is no selection bias, we estimate the ATE by using the sample average; that is,

$$\widehat{\theta}_T := \frac{1}{\sum_{t=1}^{T} \mathbb{1}[A_t = 1]} \sum_{t=1}^{T} \mathbb{1}[A_t = 1] Y_t - \frac{1}{\sum_{t=1}^{T} \mathbb{1}[A_t = 0]} \sum_{t=1}^{T} \mathbb{1}[A_t = 0] Y_t.$$

RCT. We sample experimental units from $q(x)$. In each round, we assign treatment $A_t = a$ with probability $w(a \mid x)$ for all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$. Because there is no selection bias, we estimate the ATE by using the sample average; that is,

$$\widehat{\theta}_T := \frac{1}{\sum_{t=1}^{T} \mathbb{1}[A_t = 1]} \sum_{t=1}^{T} \mathbb{1}[A_t = 1] Y_t - \frac{1}{\sum_{t=1}^{T} \mathbb{1}[A_t = 0]} \sum_{t=1}^{T} \mathbb{1}[A_t = 0] Y_t.$$

AS-AIPW (Oracle). We sample experimental units from $q(x)$. In each round, we assign treatment $A_t = a$ with probability $w^*(a \mid x)$ for all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$. Because there is selection bias, we estimate the ATE by using the AIPW estimator; that is,

$$\widehat{\theta}_T := \frac{1}{T} \sum_{t=1}^{T} \frac{\mathbb{1}[A_t = 1] (Y_t - \widehat{\mu}_t(1)(X_t))}{w^*(1 \mid X_t)} - \frac{1}{T} \sum_{t=1}^{T} \frac{\mathbb{1}[A_t = 0] (Y_t - \widehat{\mu}_t(0)(X_t))}{w^*(0 \mid X_t)}$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{X \sim q(x)} \left[\widehat{\mu}_t(1)(X) - \widehat{\mu}_t(0)(X) \right],$$

where $\widehat{\mu}_t(a)(x)$ and $\mathbb{E}_{X \sim q(x)} \left[\widehat{\mu}_t(1)(X) - \widehat{\mu}_t(0)(X) \right]$ are constructed as well as the AAS-AIPWIW experiment.

AAS-AIPWIW (Oracle). We sample experimental units from the efficient covariate density $p^*(x)$. In each round, we assign treatment $A_t = a$ with probability $w^*(a \mid x)$ for all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$. Then, we estimate the ATE by using the AIPWIW estimator; that is,

$$\widehat{\theta}_T := \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\mathbb{1}[A_t = 1] (Y_t - \widehat{\mu}_t(1)(X_t))}{w^*(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0] (Y_t - \widehat{\mu}_t(0)(X_t))}{w^*(0 \mid X_t)} \right) \frac{q(X_t)}{p^*(X_t)}$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{X \sim q(x)} \left[\widehat{\mu}_t(1)(X) - \widehat{\mu}_t(0)(X) \right].$$

AS-AIPW. We sample experimental units from $q(x)$. In each round, we assign treatment $A_t = a$ with probability $w_t(a \mid x)$ for all $a \in \{1, 0\}$ and any $x \in \mathcal{X}$, which is constructed by using past observations, as well as the AA-AIPWIW experiment. Then, we estimate the ATE by using the AIPW estimator; that is,

$$\widehat{\theta}_T := \frac{1}{T} \sum_{t=1}^{T} \frac{\mathbb{1}[A_t = 1] (Y_t - \widehat{\mu}_t(1)(X_t))}{w_t(1 \mid X_t)} - \frac{1}{T} \sum_{t=1}^{T} \frac{\mathbb{1}[A_t = 0] (Y_t - \widehat{\mu}_t(0)(X_t))}{w_t(0 \mid X_t)}$$
$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{X \sim q(x)} \left[\widehat{\mu}_t(1)(X) - \widehat{\mu}_t(0)(X) \right].$$

D.2. MSE with Various T

In Figures 8–9, we plot the MSEs at $T = 500, 750, 1,000, 1,250, 1,500, 1,750, 2,000, 2,250, 2,500, 2,750, 3,000$. The AAS-AIPWIW experiment successfully reduces the MSEs. In some cases, the AAS-AIPWIW outperforms the AAS-AIPWIW (Oracle), which seems paradoxical since the former utilizes the true p^*. Such a phenomenon is often reported in the literature of semiparametric analysis.
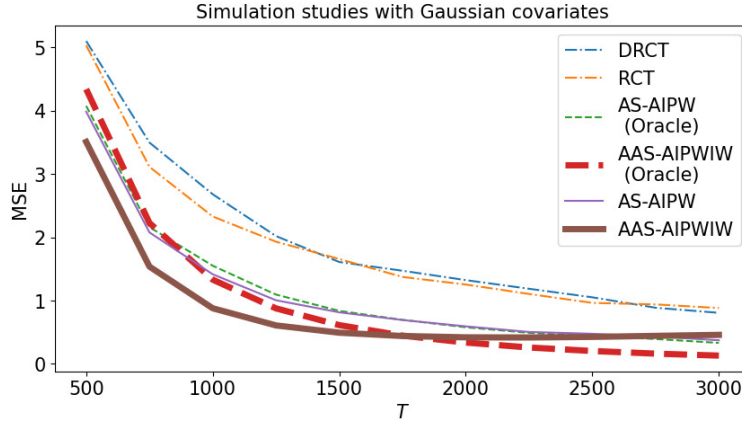
Figure 8. MSEs at $T = 500, 750, 1,000, 1,250, 1,500, 1,750, 2,000, 2,250, 2,500, 2,750, 3,000$ in simulation studies with Gaussian covariates.
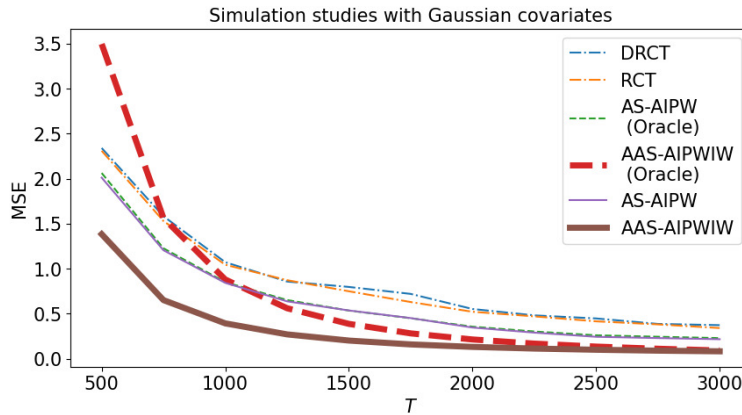

Figure 9. MSEs at $T = 500, 750, 1,000, 1,250, 1,500, 1,750, 2,000, 2,250, 2,500, 2,750, 3,000$ in simulation studies with Uniform covariates.

	RCT	AS-AIPW	AAS-AIPWIW
Gaussian covariates	95.5%	38.0%	26.5%
Uniform covariates	95.5%	48.5%	44.0%

Table 1. Coverage ratios in simulation studies.

D.3. Asymptotic Normality

In Figures 10–11, we plot empirical distributions of $\widehat{\theta}_T$ obtained via the RCT, and the AS-AIPW, and the AAS-AIPWIW experiments. While the RCT keeps a form of a Gaussian distribution, the shapes of empirical distributions of the AS-AIPW and the AAS-AIPWIW experiments seem to have some bias for the asymptotic distribution. From the theoretical viewpoint, such a bias will vanish as the sample size T increases.

D.4. Coverage Ratio

Based on the asymptotic normality, we construct confidence intervals of $\widehat{\theta}_T$ in the RCT, the AS-AIPW, and the AAS-AIPWIW experiments. Then, we confirm how many times 95% confidence intervals include the true θ_0. The ratio that the true θ_0 drops into a confidence interval is referred to as a coverage ratio.

As a result of simulation studies with 200 iterations, we obtain coverage ratios for the RCT, AS-AIPW, and AAS-AIPWIW experiments with Gaussian and Uniform covariates. The results are shown in Table 1.

Although the RCT returns accurate confidence intervals, it seems that the confidence intervals of the AS-AIPW and the
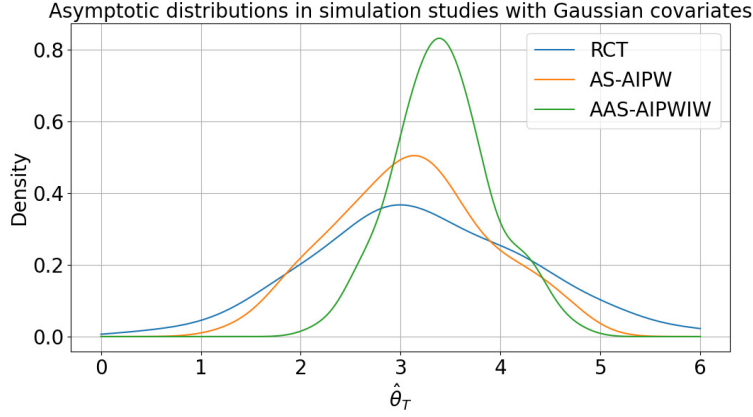
Figure 10. Empirical distributions of ATE estimators in simulation studies with covariates following with a Gaussian distribution $\mathcal{N}(1, 25)$.
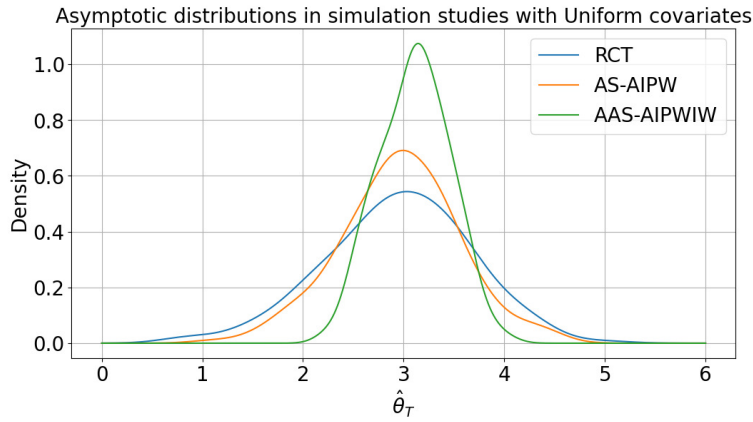

Figure 11. Empirical distributions of ATE estimators in simulation studies with covariates following with a Uniform distribution $\mathcal{U}(-10, 10)$.

AAS-AIPWIW experiments have some bias.

D.5. Simulation Studies with Different Settings

Lastly, we perform simulation studies with different settings.

Simulation studies with homogeneous variances   We investigate performances of the experiments in a setting where $\sigma_0(1)(x) = \sigma_0(0(x)$ for any $x \in \mathcal{X}$. We set

$$\sigma_0^2(1)(x) = 2 + 1.2\sin(2x) + (x + x^2)/25,$$
$$\sigma_0^2(0)(x) = 2 + 1.2\sin(2x) + (x + x^2)/25.$$

We show the results in Figures 14–15.

Although our designed experiment still reduces the MSEs, the degree is less significant than in a case where variances are heterogeneous.

Simulation studies with homogeneous means   Next, we investiate a setting where $\mu_0(1)(x) = \mu_0(0)(x)$ for any $x \in \mathcal{X}$. We define

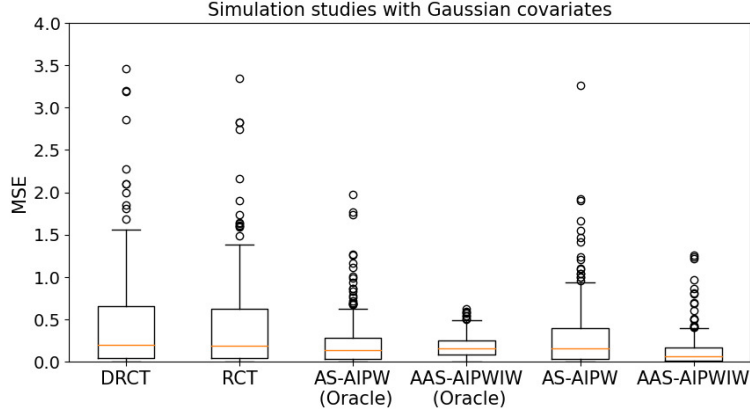$$\mu_0(1)(x) = C_1 x,$$
$$\mu_0(0)(x) = C_0 x,$$

28

Figure 12. Results of simulation studies with covariates following with a Gaussian distribution $\mathcal{N}(1, 25)$ and the homogeneous variances.
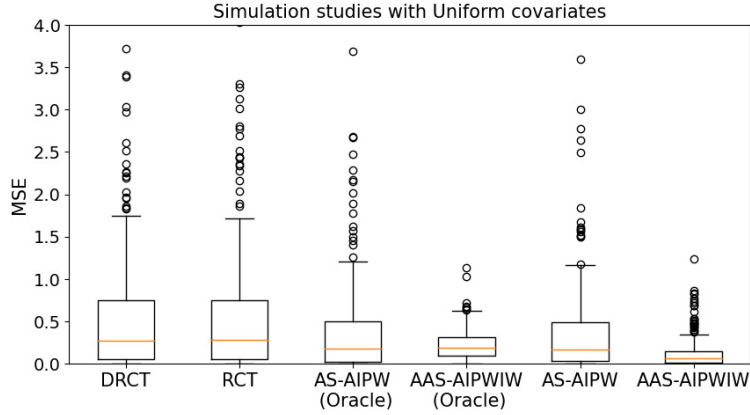

Figure 13. Results of simulation studies with covariates following with a Uniform distribution $\mathcal{U}(-10, 10)$ and the homogeneous variances.

where C_1 and C_0 are parameters so that $\mathbb{E}_{X \sim q(x)}[\mu_0(1)(X)] = \mathbb{E}_{X \sim q(x)}[Y(1)] = 10$ and $\mathbb{E}_{X \sim q(x)}[\mu_0(0)(X)] = \mathbb{E}_{X \sim q(x)}[Y(0)] = 7$. In Figures 12–13, we show the results. Our designed experiment still improves the performances compared to the RCTs, but the MSEs are almost the same as those in the AS-AIPW experiment. We consider that this is because $\mu_0(a)(x)$ has a simple form, and the conditional mean estimator in the AIPW estimator estimates it well, which makes the performances of the experiments insignificant.

Simulation studies with higher dimensional covariates Lastly, we investigate a case with higher dimensional covariates. In each round t, we generate 3 and 20 dimensional covariates. Let $\mathcal{N}(\mathbf{1}, \nu)$ be a multivariate normal distribution with the d-dimensional vector $\mathbf{1} = (1, 1, \ldots, 1)$ and a $(d \times d)$-dimensional diagonal matrix ν whose (i, i)-element is generated from a uniform distribution with support $(1, 5)$ in advance of the experiment. Let $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})$ be the covariates in round t.

As in the experiment in the main section, we specify the means as $\mathbb{E}_{X \sim q(x)}[Y(1)] = 10$ and $\mathbb{E}_{X \sim q(x)}[Y(0)] = 7$, under which the ATE is $\theta_0 = 3$. For any X_t, the conditional means are

$$\mu_0(1)(X_t) := C_1 \left(-X_{1,t} + 3X_{1,t}^2 - 1\right),$$
$$\mu_0(0)(X_t) := C_0 \left(0.1X_{1,t} + 0.2\right),$$

where C_1 and C_0 are parameters so that $\mathbb{E}_{X \sim q(x)}[\mu_0(1)(X)] = \mathbb{E}_{X \sim q(x)}[Y(1)] = 10$ and $\mathbb{E}_{X \sim q(x)}[\mu_0(0)(X)] = \mathbb{E}_{X \sim q(x)}[Y(0)] = 7$.
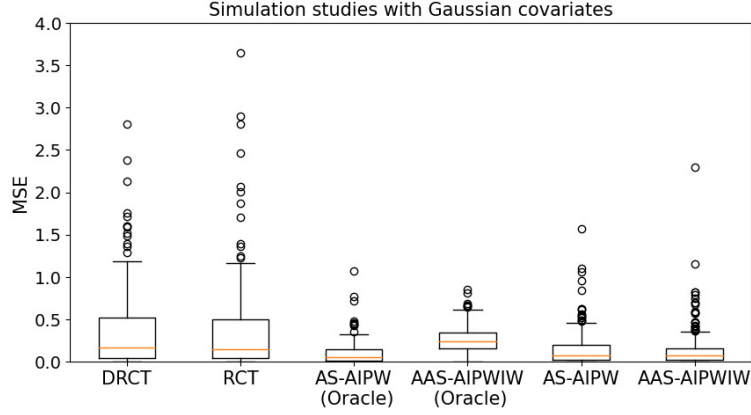
Figure 14. Results of simulation studies with covariates following with a Gaussian distribution $\mathcal{N}(1, 25)$ and the homogeneous means.
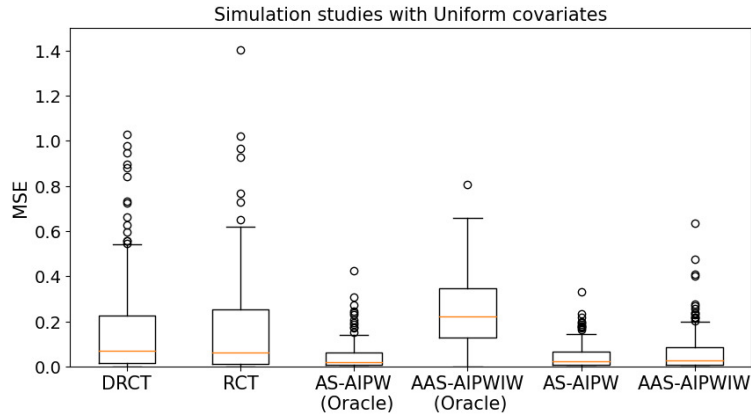

Figure 15. Results of simulation studies with covariates following with a Uniform distribution $\mathcal{U}(-10, 10)$ and the homogeneous means.

We define $\sigma_0^2(a)(X_t)$ as follows:

$$\sigma_0^2(1)(X_t) := 2 + 1.2\sin(2X_{1,t}) + \frac{X_{1,t} + 2X_{1,t}^2}{25},$$

$$\sigma_0^2(0)(X_t) := 2 + 0.8\cos\left(\frac{X_{1,t}}{2}\right) + \frac{X_{1,t}^2}{50}.$$

Then, as in Figures 6–7, we report the empirical MSEs for settings with each dimension $d = 3, 10, 20$. The results are shown in Figure 16.

D.6. Simulation Studies with the Semi-Synthetic Dataset

This section provides simulation studies using the semi-synthetic dataset called the Infant Health and Development Program (IHDP).

The IHDP dataset has simulated outcomes and real-world covariates (Hill, 2011). The total sample size is 747, and the covariates consist of 6 continuous variables and 19 binary variables.

The outcomes are generated artificially, while the covariates are real-world data. In (Hill, 2011), there are two scenarios called response surface A and response surface B. In response surface A, the potential outcomes $Y_t(1)$ and $Y_t(0)$ are generated as

$$Y_t(0) \sim \mathcal{N}(X_t^\top \boldsymbol{\gamma}_A, 1),$$
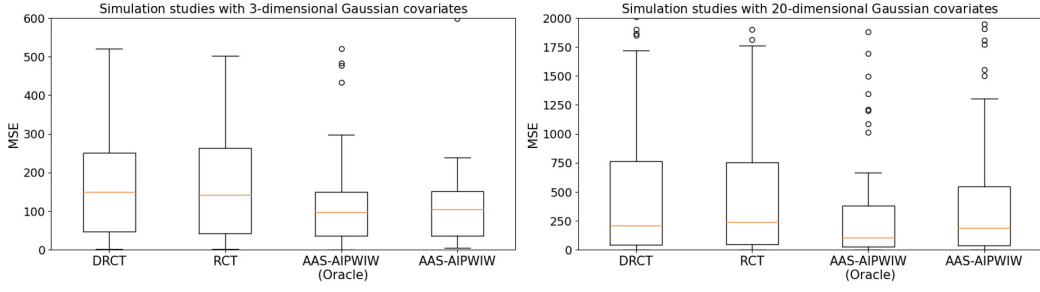
30

Figure 16. Results of simulation studies with covariates following 3-dimensional and 20-dimensional Gaussian distributions.
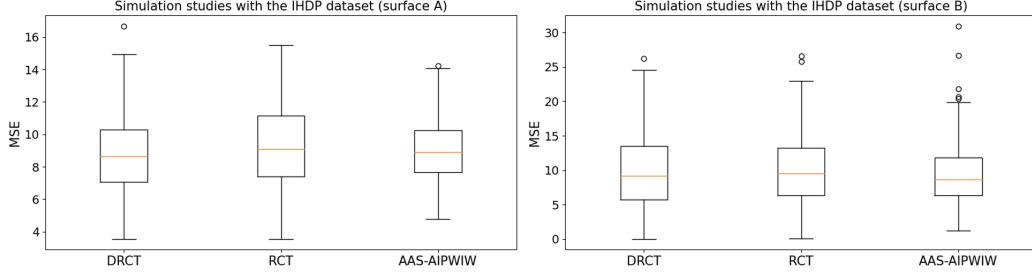

Figure 17. Results of simulation studies with the IHDP dataset (response surface A and B).

$$Y_t(1) \sim \mathcal{N}(X_t^\top \boldsymbol{\gamma}_A + 4, 1),$$

where each element of $\boldsymbol{\gamma}_A \in \mathbb{R}^{25}$ is randomly generated from $\{0, 1, 2, 3, 4\}$ with probabilities $(0.5, 0.2, 0.15, 0.1, 0.05)$.

In response surface B, the potential outcomes $Y_t(1)$ and $Y_t(0)$ are generated as

$$Y_t(0) \sim \mathcal{N}\left(\exp\left((X_t + W)^\top \boldsymbol{\gamma}_B\right), 1\right),$$
$$Y_t(1) \sim \mathcal{N}(X_t^\top \boldsymbol{\gamma}_B - q, 1),$$

where W is an offset matrix of the same dimension as X_t with every value equal to 0.5, q is a constant to normalize the average treatment effect conditional on $d = 1$ to be 4, and each element of $\boldsymbol{\gamma}_B \in \mathbb{R}^{25}$ is randomly generated from $\{0, 0.1, 0.2, 0.3, 0.4\}$ with probabilities $(0.6, 0.1, 0.1, 0.1, 0.1)$.

Our experiment fixes $T = 2000$. In each round t, we sample covariate X_t with replacement from the 747 samples. Then, following each setting of the response surfaces, we generate the corresponding outcomes.

As in Figures 6–7, we report the empirical MSEs for settings with each dimension $d = 3, 10, 20$. The results are shown in Figure 17.

E. Extensions

We introduce several extensions of our proposed framework.

E.1. Rejection Sampling from $q(x)$

This section introduces a setting differing from that defined in Section 2.

Problem setting In this section, we assume that covariate $\widetilde{X}_t$ is generated from $q(x)$ in each round t, and we observe it. Then, we decide whether to assign a treatment to the sample. We employ the rejection sampling (Gelman et al., 2004) for this choice. If we accept $\widetilde{X}_t$, we assign treatment $a \in \{1, 0\}$ and observe a corresponding outcome. If we do not accept $\widetilde{X}_t$, we do not assign any treatments and do not observe any outcomes. In this setting, we can remove Assumption 2.2, which assumes that $q(x)$ is known. We illustrate the procedure in Figure 18.
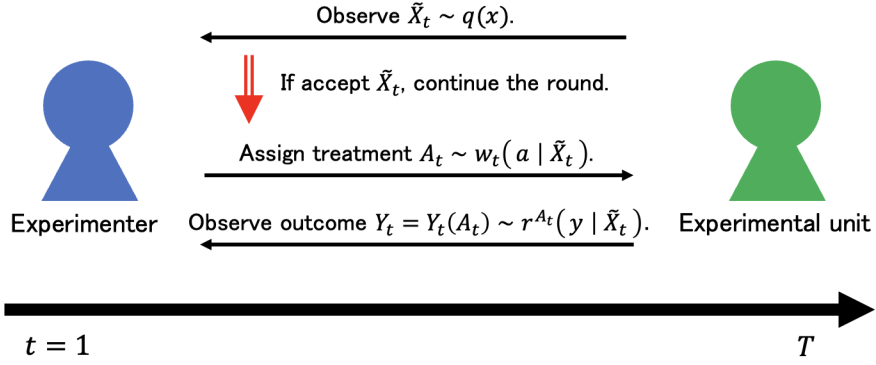
31

Figure 18. Rejection sampling.

Rejection sampling We first explain the rejection sampling. From assumptions, there exists an universal constant $M > 0$ such that $e_t(x) := \widehat{p}_t(x)/q(x) < M$ for all $x \in \mathcal{X}$, where

$$e_t(x) = \frac{\sqrt{\left(\widehat{\sigma}_t(1)(x) + \widehat{\sigma}_t(0)(x)\right)^2}}{\frac{1}{t-1}\sum_{s=1}^{t-1}\sqrt{\left(\widehat{\sigma}_t(1)(X_s) + \widehat{\sigma}_t(0)(X_s)\right)^2}}.$$

holds. We accept $\widehat{X}_t$ if $U < e_t(\widehat{X}_t)/M$, and we assign a treatment to the sample. In contrast, we reject $\widehat{X}_t$ if $U \geq e_t(\widehat{X}_t)/M$, and we do not assign any treatments and cannot observe any outcomes. Then, the following lemma holds.

Lemma E.1. *Let $F_t(x)$ be a cumulative density function of $\widehat{p}_t(x)$. Let U be a random variable generated from the uniform distribution with support $[0, 1]$. Then, it holds that*

$$\mathbb{P}\left(\widehat{X}_t \leq x \mid U < e_t(\widehat{X}_t)/M\right) = F_t(x);$$

that is, accepted $\widehat{X}_t$ follows a density $\widehat{p}_t(x)$.

We omit the proof. See Gelfand et al. (2000).

Procedure By using this lemma, we consider the following procedure.

Step 1: based on past information, an experimenter decides $p_t(x)$ and $w_t(a \mid x)$ for $a \in \{1, 0\}$ and $x \in \mathcal{X}$;

Step 2: an experimenter observes covariate $\widetilde{X}_t \sim q(x)$;

Step 3: by using the rejection sampling, we use $X_t = \widetilde{X}_t$ if $U_t \leq e_t(X_t)/M$, where U_t is a uniform distribution with support $[0, 1]$;

Step 3: based on past information, we choose $A_t = a$ with probability $w_t(a \mid x)$;

Step 4: an experimenter observes outcome $Y_t = \mathbb{1}[A_t = 1]Y_t(1) + \mathbb{1}[A_t = 0]Y_t(0)$, where $Y_t(a) \sim r^a(y \mid x)$,

Step 5: after iterating Step 1–4 for $t = 1, 2, \ldots, T$, in round T, we estimate the ATE using the estimator defined in (8).

ATE estimator The estimator is defined as

$$\widehat{\theta}_T := \frac{1}{T}\sum_{t=1}^{T}\widetilde{\Psi}_t(Y_t, A_t, X_t; w_t, p_t), \tag{8}$$

where

$$\widetilde{\Psi}_t(Y_t, A_t, X_t; \widehat{w}_t, \widehat{p}_t) := \left(\frac{\mathbb{1}[A_t = 1]\left(Y_t(1) - \widehat{\mu}_t(1)(X_t)\right)}{\widehat{w}_t(1 \mid X_t)} - \frac{\mathbb{1}[A_t = 0]\left(Y_t(0) - \widehat{\mu}_t(0)(X_t)\right)}{\widehat{w}_t(0 \mid X_t)} + \widehat{\theta}_t(X_t)\right)\frac{1}{e_t(X_t)}.$$

We can show that the estimator has the same asymptotic variance of $\widehat{\theta}_T$ in Theorem 5.3. The result is summarized in the following proposition. Note that we do not use Assumption 2.2. We omit the proof since it is almost same as that for Theorem 5.3.

Proposition E.2. *Consider the AAS-AIPWIS experiment. Suppose that Assumptions 2.1 and 5.1–5.2 hold. Then,*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tau^*)$$

holds as $T \to \infty$, where recall that τ^* is the semiparametric efficiency bound.

E.2. Batch Design

We introduce a setting where each experimental unit visits sequentially. We can directly extend our experiment to a setting where sets of experimental units visit.

E.3. Multiple Treatments

We can consider a setting where there multiple treatments more $[K] := \{1, 2, \ldots, K\}$ for $K \geq 2$, instead of the binary treatments $\{1, 0\}$. In this case, it is unclear how we define the treatment effect. One way is to define the policy value by using a deterministic evaluation policy $\pi : [K] \times \mathcal{X} \to \mathbb{R}$. In many cases, it is assumed that $\sum_{a \in [K]} \pi(a \mid x)$ for any $x \in \mathcal{X}$, but we do not impose the restriction to derive a more general result, including a case with the restriction as a specific case.

Under the evaluation policy, we define the policy value as $\theta_{0,K} := \mathbb{E}\left[\sum_{a \in [K]} \pi(a \mid X) Y(a)\right]$.

For $\theta_{0,K}$, the semiparametric efficiency bound is given as $\tau_K(w, p) = \mathbb{E}_{X \sim p(x)}\left[\sum_{a \in [K]} \frac{\sigma_0^2(a)(X)}{w(a|X)} \frac{q^2(X)}{p^2(X)}\right] = \mathbb{E}_{X \sim q(x)}\left[\sum_{a \in [K]} \frac{\sigma_0^2(a)(X)}{w(a|X)} \frac{q(X)}{p(X)}\right]$. Because the proof is almost the same as that for Theorem 3.1, we omit the proof.

Then, the efficient probabilities are obtained as

$$w^*(a \mid x) = \frac{\sigma_0(a)(x)}{\sum_{a \in [K]} \sigma_0(a)(x)}$$

$$p^*(x) = \frac{\sqrt{\sum_{a \in [K]} \frac{\sigma_0^2(a)(x)}{w^*(a|x)}}}{\mathbb{E}_{X \sim q(x)}\left[\sqrt{\sum_{a \in [K]} \frac{\sigma_0^2(1)(X)}{w^*(1|X)}}\right]} q(x).$$

We then just directly apply the AAS-AIPWIS experiment.