

# ABC3: Active Bayesian Causal Inference with Cohn Criteria in Randomized Experiments

Taehun Cha and Donghun Lee\*

Korea University  
{cth127, holy}@korea.ac.kr

## Abstract

In causal inference, randomized experiment is a de facto method to overcome various theoretical issues in observational study. However, the experimental design requires expensive costs, so an efficient experimental design is necessary. We propose ABC3, a Bayesian active learning policy for causal inference. We show a policy minimizing an estimation error on conditional average treatment effect is equivalent to minimizing an integrated posterior variance, similar to Cohn criteria (Cohn, Ghahramani, and Jordan 1994). We theoretically prove ABC3 also minimizes an imbalance between the treatment and control groups and the type 1 error probability. Imbalance-minimizing characteristic is especially notable as several works have emphasized the importance of achieving balance. Through extensive experiments on real-world data sets, ABC3 achieves the highest efficiency, while empirically showing the theoretical results hold.<sup>1</sup>

## 1 Introduction

The major goal of causal inference is to estimate the treatment effect which is a relative effect on the treatment group compared to the control group. In randomized experiments, practitioners allocate treatment to subjects to estimate the treatment effect. Randomized experiments free practitioners from various theoretical problems prevailing in an observational study, e.g. unmeasured confounders. However, a randomized experiment is usually more expensive than an observational study, as a result, an efficient experiment design is desirable.

To achieve efficiency in randomized experiments, Efron (1971) first introduced a biased-coin design, and several works tried to minimize the estimation bias by achieving a balance between treatment and control groups (Atkinson 2014). Antognini and Zagoraiou (2011) extended this to covariate-adaptive design to achieve a balance, not only between treatment-control groups but also within sampling strata. Several recent works target the same goal using an adaptive Neyman allocation (Dai, Gradu, and Harshaw

2023) or Pigeonhole design (Zhao and Zhou 2024). However, these lines of work assume the experimental subjects are *given*, not actively *choosable*.

Active learning is a framework where a practitioner can choose unlabeled data points and ask an oracle to label them (Settles 2009). We can further rationalize the experimental design by adopting an active learning framework. For example, internet companies can choose a member whom they implement the A/B test, by utilizing the personal information they have gathered. Pharmaceuticals can choose a subject based on their personal information after the applicants are gathered to save a budget.

To develop a sound active learning method for randomized experiments, (1) it should not violate the standard assumptions in causal inference. Also, (2) the method should achieve a balance between observation and control groups to make a sound conclusion. We introduce ABC3, a novel active learning policy for randomized experiments to remedy these issues. Using the Gaussian process, our policy targets minimizing the error of individual treatment effect estimation. Our contributions are three folds:

- We theoretically show minimizing the estimation error on individual treatment effect is equivalent to minimizing the integrated predictive variance in a Bayesian sense.
- ABC3, the policy minimizing the variance, theoretically minimizes the imbalance between treatment and control groups and the type 1 error rate.
- With extensive experiments, we empirically verify ABC3 outperforms other methods while showing theoretical properties hold.

After examining related works in Section 2, we formalize our problem in Section 3. Then we introduce ABC3 and its theoretical properties in Section 4. We empirically verify the performance and properties of ABC3 in Section 5. Then we conclude our paper with several discussions and limitations in Section 6 and 7.

## 2 Related Works

There are several works exploring an active learning policy for observational data (Sundin et al. (2019), Jesson et al. (2021), and Toth et al. (2022)). Especially, Sundin et al. (2019) proposed an active learning policy for decision making, when treatment-control groups are imbalanced. They

\* corresponding author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>Source codes are available on <https://github.com/AIML-K/ActiveBayesianCausal/>. Technical Appendix is on arXiv version.

theoretically showed the imbalance can result in a type S error, where a practitioner estimates the treatment effect with a different sign. Likewise, Shalit, Johansson, and Sontag (2017) showed that the generalization error is bounded by the imbalance when estimating the treatment effect. However, their work focused on obtaining a balanced representation from observational data, not a randomized experiment setting.

For randomized experiments, Deng, Pineau, and Murphy (2011) suggests an active learning policy sampling a point with the highest predictive variance which is similar to Mackay’s criteria (MacKay 1992). Zhu et al. (2024) also proposed Mackay’s criteria-like method under network interference structure. Song, Mak, and Wu (2023) suggested ACE which targets maximizing the covariance between observed and test data sets. We will compare the effectiveness of our policy with these policies.

Sample-constrained causal inference setting shares a similar goal with active learning: achieving lower estimation error with fewer samples. Addanki et al. (2022) and Ghadiri et al. (2023) proposed an efficient sampling and estimation method in a randomized experiment setting. Harshaw et al. (2023) suggested a sampling method balancing covariates. However, their work assumes a linear relationship between covariates and potential outcomes, which is vulnerable in real-world scenarios. We will also compare the effectiveness of these policies.

### 3 Problem Formulation

Let  $X \in \mathcal{X}$ ,  $Y \in \mathcal{Y}$  and  $A \in \{0, 1\}$  be random variables.  $X$  is a covariate representing each subject,  $Y$  is an outcome, and  $A$  represents a binary treatment. Following the Neyman-Rubin causal model (Rubin 1974), additionally define  $Y^0$  and  $Y^1$ , potential outcomes for either control or treatment. Unlike the usual supervised learning settings, a practitioner can observe only one of  $Y^0$  and  $Y^1$  in a causal inference setting.  $x, y, a, y^0$  and  $y^1$  denotes the realizations of each random variable.

Let  $D_\Omega = \{(x_i, y_i^0, y_i^1)\}_{i=1}^N$  be a subject pool with covariate information  $x_i$  and potential outcomes  $y_i^0, y_i^1 \in \mathbb{R}$ . Let  $D_t^0 = \{(x_i, y_i^0)\}_{i \in I_t^0}$  be an observed control group data set at time  $t$  with an index set  $I_t^0$ . Likewise, define  $D_t^1 = \{(x_i, y_i^1)\}_{i \in I_t^1}$  for a treatment group. Let  $X_\Omega, X_t^0$  and  $X_t^1$  be sets of  $x$ s in each data set,  $D_\Omega, D_t^0$  and  $D_t^1$ .

Our quantity of interest is the *conditional average treatment effect (CATE)*,  $CATE(x) = \mathbb{E}[Y^1 - Y^0 | X = x]$  for each subject  $x$ . Then we can train an estimator  $\hat{CATE}_t(x) = \hat{y}_t^1(x) - \hat{y}_t^0(x)$ , where  $\hat{y}_t^a$ s are regressors trained on each data set  $D_t^a, a \in \{0, 1\}$ .

To evaluate the trained estimator, we use a standard metric, *expected precision in estimation of heterogeneous effect (PEHE)*, Hill (2011)),

$$\epsilon_{PEHE}(C\hat{A}TE_t) = \int_{\mathcal{X}} (C\hat{A}TE_t(x) - CATE(x))^2 d\mathbb{P}(x),$$

where  $\mathbb{P}$  is a probability distribution over whole covariates. In the usual case, we can treat  $\mathbb{P}$  as a discrete distribution

corresponding to the covariates in  $D_\Omega$ . Then we can write  $\mathbb{P} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ , where  $\delta$  is Dirac-delta function.

Meanwhile, in Bayesian statistics, we do not hypothesize the existence of population parameters, e.g.  $CATE(x)$ . Instead, we update our belief with observed data points. To derive a Bayesian policy, we need to define the optimization target in the Bayesian sense.

**Definition 3.1.**

$$\epsilon_{PEHE}^\Omega(C\hat{A}TE_t) = \int_{\mathcal{X}} (C\hat{A}TE_t(x) - C\hat{A}TE_\Omega(x))^2 d\mathbb{P}(x)$$

where  $C\hat{A}TE_\Omega(x) = \hat{y}_\Omega^1(x) - \hat{y}_\Omega^0(x)$ ,  $\hat{y}_\Omega^a$ s are regressors trained on whole data set  $D_\Omega$ .

Intuitively,  $C\hat{A}TE_\Omega$  represents the best oracle estimation observing all factual and counterfactual data points. At any  $t$ , our active learning problem is to choose an optimal (yet unobserved) subject  $x^*$  and its treatment  $a^*$  which can achieve the lowest expected  $\epsilon_{PEHE}^\Omega(C\hat{A}TE_{t+1})$ , i.e.

$$x^*, a^* = \arg \min_{x_{t+1} \in X_\Omega \setminus (X_t^1 \cup X_t^0), a_{t+1} \in \{0, 1\}} \mathbb{E}_{t+1} \left[ \epsilon_{PEHE}^\Omega(C\hat{A}TE_{t+1}) \right],$$

where  $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t]$ ,  $\mathcal{F}_t$  is a filtration at  $t$ . If  $x^* = x_j$  and  $a^* = a$ , we observe  $y_j^a$  and append it to the observed data set  $D_{t+1}^a = D_t^a \cup \{(x_j, y_j^a)\}$  to be used at  $t + 1$ .

While optimizing the expected error, an active learning policy should not violate the standard assumptions for causal inference. A standard randomized experiment in causal inference requires several assumptions to identify the causal effects (Rubin 1974).

**Assumption 3.2.** (Consistency)  $Y = AY^1 + (1 - A)Y^0$

**Assumption 3.3.** (Positivity)  $\mathbb{P}(A = a | X = x) > 0, \forall x$

**Assumption 3.4.** (Randomization)  $(Y^0, Y^1) \perp\!\!\!\perp A | X$

Assumption 3.2. assumes that the observed outcome  $Y$  under treatment  $A$  is equivalent to its potential outcome  $Y^a$ . Assumption 3.3. is required to well-define the conditional expectation. Assumption 3.4. implies that **the treatment  $A$  should be assigned independently from the potential outcome values**. The assumptions guarantee an unbiased estimation of CATE, i.e.  $CATE(x) = \mathbb{E}[Y^1 - Y^0 | X = x] = \mathbb{E}[\mathbb{E}[Y | A = 1, X = x] - \mathbb{E}[Y | A = 0, X = x]]$ .

Assumption 3.4. is crucial in an active learning setting since several active learning algorithms consider  $y$  values when querying  $x$  and  $a$ . For instance, Song, Mak, and Wu (2023) introduced ACE-UCB, a bandit-like policy that targets a subject with the highest individual treatment effect. However, ACE-UCB exploits previously observed outcomes when choosing a subject and treatment. As a result, ACE-UCB can generate confounding between treatment and potential outcomes and may violate Assumption 3.4. So querying  $x$  and  $a$  without considering  $ys$  is crucial when adopting an active learning framework for a causal inference.

## 4 Proposed Method

### ABC3: Active Bayesian Causal Inference with Cohn Criteria

Gaussian process ( $\mathcal{GP}$ , Rasmussen and Williams (2006)) is a non-parametric machine learning model based on Bayesian statistics. It is a powerful tool as it allows flexible function estimation depending on a pre-defined kernel function. Set priors on  $f \sim \mathcal{GP}(m(x), k(x, x'))$ , where  $m(x)$  is a mean prior and  $k$  is a kernel function. Assume a data set  $D = \{(x_i, y_i)\}_{i=1}^N$  with noisy observation  $y_i = Y(x_i) + \epsilon_i$ ,  $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ . We can obtain a posterior distribution of  $f(x^*)$  given data set  $D$  by computing  $f(x^*)|D$  as

$$\begin{aligned} f(x^*)|D &\sim \mathcal{N}(m'(x^*), \sigma^2(x^*)) \text{ where} \\ m'(x^*) &= m(x^*) + \mathbf{k}_*(x^*)^T [\mathbb{K} + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{y}, \\ \text{cov}(x, x^*) &= k(x, x^*) - \mathbf{k}_*(x)^T [\mathbb{K} + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_*(x^*), \\ \sigma^2(x^*) &= \text{cov}(x^*, x^*), \\ \mathbf{k}_*(x^*) &= [k(x_i, x^*)]_{i=1}^N, \mathbb{K} = [k(x_i, x_j)]_{i,j=1}^N, \\ \mathbf{y} &= [y_i]_{i=1}^N. \end{aligned}$$

We can observe that the posterior variance  $\sigma^2(x^*)$  does not depend on  $\mathbf{y}$ . Adopting the notations from above, we define  $\mathbf{k}_{t,*}^a, \mathbb{K}_t^a, \mathbf{y}_t^a$  and  $\text{cov}_t^a$  with observed data set  $D_t^a$  at time  $t$ . We also assume zero-mean prior, i.e.  $m(x) = 0, \forall x$ .

The following theorem identifies our active learning policy with only posterior variance terms.

**Theorem 4.1.** *Assume  $|k(x, x')| < \infty$  and  $|y_i^a| < \infty$  for all  $x, x' \in \mathcal{X}$  and  $a, i$ , as a result,  $\epsilon_{PEHE}^\Omega(\widehat{CATE}_t) < \infty, \forall t$ . Let our estimator  $\widehat{CATE}_t(x) = \hat{y}_t^1(x) - \hat{y}_t^0(x)$ , where  $\hat{y}_t^a(x) = \mathbb{E}_t[Y^a(x)]$  is a mean posterior distribution of gaussian process  $Y^a$  trained on data set  $D_t^a$ . Assume two Gaussian processes  $Y^1$  and  $Y^0$  are independent. Then*

$$\begin{aligned} \arg \min_{x_{t+1}, a_{t+1}} \mathbb{E}_{t+1} \left[ \epsilon_{PEHE}^\Omega(\widehat{CATE}_{t+1}) \right] = \\ \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) \end{aligned}$$

Proof of Theorem 4.1. is in Appendix A.1. Theorem 4.1. states that minimizing the error on CATE estimation is equivalent to minimizing the integrated posterior variance of the estimator. In active learning literature, the active learning policy minimizing the integrated predictive variance is called *Active Learning Cohn* (Cohn, Ghahramani, and Jordan (1994) and Gramacy (2020)). Seo et al. (2000) proposed Active Learning Cohn policy utilizing the Gaussian process in a usual supervised learning setting. We extend this line of work to causal inference and name our policy after his name, **ABC3: Active Bayesian Causal Inference with Cohn Criteria**.

Computing the inverse of all hypothetical covariance matrix  $\mathbb{K}_{t+1}^a$  for all  $x$  is computationally infeasible. However, we can efficiently find the minimizer as  $\mathbb{K}_t^a$  is a principal submatrix of  $\mathbb{K}_{t+1}^a$ .

### Proposition 4.2.

$$\begin{aligned} \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) \\ = \arg \max_{x_{t+1}, a} \\ \frac{\int_{\mathcal{X}} \left[ (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x) \right]^2 d\mathbb{P}(x)}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{k}}_{t+1}^a} \end{aligned} \quad (1)$$

where  $\tilde{\mathbf{k}}_{t+1}^a = [k(x_i, x_{t+1})]_{i \in I_t^a}$

Proof of Proposition 4.2. is in Appendix A.2. From the proposition, we successfully eliminate the dependency on  $\mathbb{K}_{t+1}^a$ , and there is no need to compute an inverse matrix for every  $x$ . We summarize our policy in Algorithm 1.

---

#### Algorithm 1: ABC3

---

**Input:** Current time step  $t$ , whole covariate set  $X_\Omega$ , covariates distribution  $\mathbb{P}$ , previous observations  $X_t^1$  and  $X_t^0$ , kernel  $k$ , noise parameter  $\sigma_\epsilon$

**Output:**  $x_{t+1}, a_{t+1}$

- 1:  $\mathcal{V}_0, \mathcal{V}_1 = \phi, \phi$
  - 2: **for**  $x \in X_\Omega \setminus (X_t^0 \cup X_t^1)$  **do**
  - 3:   Compute  $\tilde{\mathbf{k}}_{t+1}^1$  and  $\tilde{\mathbf{k}}_{t+1}^0$  assuming  $x_{t+1} = x$
  - 4:    $v^0, v^1 =$  Equation (1) for each  $a \in \{0, 1\}$
  - 5:    $\mathcal{V}_0 = \mathcal{V}_0 \cup \{v^0\}, \mathcal{V}_1 = \mathcal{V}_1 \cup \{v^1\}$
  - 6: **end for**
  - 7:  $i, a_{t+1} = \arg \max(\mathcal{V}_0 || \mathcal{V}_1)$
  - 8:  $x_{t+1} = X_\Omega \setminus (X_t^0 \cup X_t^1) [i]$
  - 9: **return**  $x_{t+1}, a_{t+1}$
- 

## Theoretical Analysis

**Balancing Treatment-Control Groups** Unlike usual supervised learning, causal inference requires precise estimation of both functions for treatment and control groups. As a result, the balance between the two groups is crucial to obtain a sound estimation. For example, consider a study on the causal effect of online lectures. Assume our treatment group is concentrated on undergraduate students while our control group is concentrated on graduate students. Then it would be difficult to make a sound conclusion with statistical tools, as the two groups are highly imbalanced.

Several researchers theoretically analyzed the effect of the imbalance on causal inference. Shalit, Johansson, and Sonntag (2017) showed that the generalization error on CATE is upper bounded by the imbalance. Sundin et al. (2019) defined Type S Error, assigning a different sign (+ or -) to CATE, and showed the probability of Type S Error is bounded by the imbalance. Both works utilized *Maximum Mean Discrepancy* (MMD, Gretton et al. (2012)) to quantify and measure the imbalance.

### Definition 4.3.

$$\begin{aligned} \text{MMD}(P, Q, \mathcal{F}) = \\ \sup_{f \in \mathcal{F}} \mathbb{E}_{x \sim P(x)} [f(x)] - \mathbb{E}_{y \sim Q(y)} [f(y)] \end{aligned}$$

where  $\mathcal{F}$  is a unit ball of Reproducing Kernel Hilbert Space (RKHS) induced by a kernel  $k(\cdot, \cdot)$ .

Gretton et al. (2012) showed that the MMD is equivalent to the distance between mean embeddings in RKHS,  $\mu_P$  and  $\mu_Q$ , and can be computed with the kernel function  $k$ .

*Remark 4.4.*

$$\begin{aligned} \text{MMD}(P, Q, \mathcal{F})^2 &= \|\mu_P - \mu_Q\|^2 \\ &= \mathbb{E}_{x, x' \sim P(x)} [k(x, x')] + \mathbb{E}_{y, y' \sim Q(y)} [k(y, y')] \\ &\quad - 2\mathbb{E}_{x \sim P(x), y \sim Q(y)} [k(x, y)] \end{aligned}$$

We show our active learning policy approximately minimizes the upper bound of MMD.

Assume the probability over whole covariates is a discrete distribution corresponding to the covariates in  $D_\Omega$ , i.e.  $\mathbb{P} = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ . For notational convenience, we assume the noise-less observation case  $\sigma_\epsilon^2 = 0$ , but we can easily extend the result to the noisy case. Let  $\mathbb{P}_t^a = \frac{1}{|I_t^a|} \sum_{i \in I_t^a} \delta_{x_i}$ ,  $a \in \{0, 1\}$  be an empirical distribution for treatment-control group up to time  $t$ . Then we obtain the following theorem.

**Theorem 4.5.** *Assume  $|k(x, x')| < \infty$  and  $|y_i^a| < \infty$  for all  $x, x' \in \mathcal{X}$  and  $a, i$ . Let  $\lambda^*$  be a maximum eigenvalue of  $\mathbb{K}_\Omega$ , i.e. covariance matrix of whole covariates. Let  $M = \frac{1}{N^2} \sum_{i, j=1}^N k(x_i, x_j)$ . Define functions  $\delta^*(I_n) = \frac{1}{n} \sum_{i \in I_n} \int_{\mathcal{X}} k(x_i, x) d\mathbb{P}(x)$ , and  $\epsilon^*(I_n) = M - \frac{1}{n^2} \sum_{i, j \in I_n} k(x_i, x_j)$ , where  $I_n \subset \{1, \dots, N\}$  is an  $n$ -element index set. Assume  $\epsilon^*(I_n) \leq 2\delta^*(I_n), \forall I_n$ . Then*

$$\begin{aligned} \text{MMD}(\mathbb{P}_t^1, \mathbb{P}_t^0, \mathcal{F})^2 &\leq 4 \frac{\lambda^*}{|I_t^1|} + 4 \frac{\lambda^*}{|I_t^0|} \\ &\quad + 2 \int_{\mathcal{X}} \mathbb{V}_t [Y^1(x)] + \mathbb{V}_t [Y^0(x)] d\mathbb{P}(x) \end{aligned}$$

Proof of Theorem 4.5. is in Appendix A.3. We can observe that the first two terms decrease as we observe more subjects. The third term is our exact optimization target as introduced in Theorem 4.1. Empirical achievability and intuitive meaning of the assumption are covered in Section 5.6 and Appendix B.

By combining this result with the previous theoretical analysis (Shalit, Johansson, and Sontag (2017) and Sundin et al. (2019)), our active learning policy minimizes the upper bounds of both generalization error and type S error.

**Type 1 Error Minimization** The precise estimation of causal effect is an important object of causal inference. However, **testing the existence of causal effect** is another key component. R. A. Fisher first advocated the randomized experiment to test the existence of causal effects (Fisher 1970). He proposed *Fisher's sharp null hypothesis*, where  $y_i^1 = y_i^0, \forall i = 1, \dots, N$ . A randomized experiment method should not reject the null hypothesis if it finds no causal effect. Type 1 error occurs when we reject a null hypothesis while it is true.

As active learning is a sequential procedure, a practitioner may want to test the statistical significance of the treatment effect sequentially. However, Ham et al. (2023) stated that applying statistical tests sequentially can result in a high type

1 error probability. An active learning policy should avoid the type 1 error to make a sound conclusion.

Here we present our theoretical result that our policy minimizes the upper bound of the type 1 error probability. First, we define the type 1 error.

**Definition 4.6.** (Type 1 Error at  $t$ ) Under Fisher's Sharp null hypothesis, as it implies  $\text{CATE}(x) = 0, \forall x$ ,

$$\mathbb{P}_t [\text{Type 1 Error}(x)] = \mathbb{P}_t [ |Y^1(x) - Y^0(x)| > \alpha ],$$

where  $\alpha$  is a decision threshold.

Then we can show that our policy minimizes the upper bound of the type 1 error rate over the whole  $x$ . Proof of Theorem 4.7. is in Appendix A.4.

**Theorem 4.7.** *Under Fisher's Sharp null hypothesis,  $\arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x)$  also minimizes the upper bound of the  $\int_{\mathcal{X}} \mathbb{P}_{t+1} [\text{Type 1 Error}(x)] d\mathbb{P}(x)$*

## 5 Experiments

In this section, we empirically analyze the theoretical results introduced in Section 4. For the comparison, we utilize **IHDP** (Brooks-Gunn, Liaw, and Klebanov (1992) and Hill (2011)), **Boston** (Harrison and Rubinfeld 1978), **ACIC** (Gruber et al. 2019), and **Lalonde** (LaLonde 1986) data sets. IHDP and ACIC data sets are semi-real data sets where counterfactual outcomes are simulated. Boston and Lalonde data sets do not contain counterfactual outcomes. Following Addanki et al. (2022), we set  $Y^0 = Y^1$  to simulate the null hypothesis circumstance (i.e.  $\text{CATE}(x) = 0, \forall x$ ). All data sets have continuous outcome values as we assume the potential outcome follows a Gaussian process. More detailed explanations of data sets are in Appendix C.

We randomly divide each data set in half for every trial to construct train and test data sets. Each active learning policy selects what to observe from the train data set and regressors are trained on the observed train data set. Policies are evaluated on the test data set for each pre-defined time step. We utilize the following baseline policies.

- **Naive:** This policy randomly selects  $x_{t+1}$  from the subject pool and then decides treatment  $a_{t+1}$  with Bernoulli distribution with probability 0.5.
- **Mackay** (MacKay 1992): This policy selects  $x_{t+1}, a_{t+1} = \arg \max_{x, a} \mathbb{V}_t [Y^a(x)]$ , i.e. chooses the most uncertain point at  $t$ .
- **Leverage** (Addanki et al. 2022): Under linearity assumption between covariates and outcomes, this policy exploits leverage score and shows theoretical guarantees on nearly optimal RMSE on CATE estimation. Unlike other policies, this policy is not sequential, as a result, the observed points for  $t$  and  $t + 1$  can be different.
- **ACE** (Song, Mak, and Wu 2023): Unlike other policies, this policy assumes access to the test data set and select  $x_{t+1}, a_{t+1} = \arg \max_{x, a} \left[ \frac{1}{|n_{test}|} \sum_{i=1}^{|n_{test}|} \text{cov}_t^a [x_i^{test}, x] \right]^2 / \mathbb{V}_t [Y^a(x)]$ . It maximizes the covariance between observed and test data sets while minimizing the predictive variance.

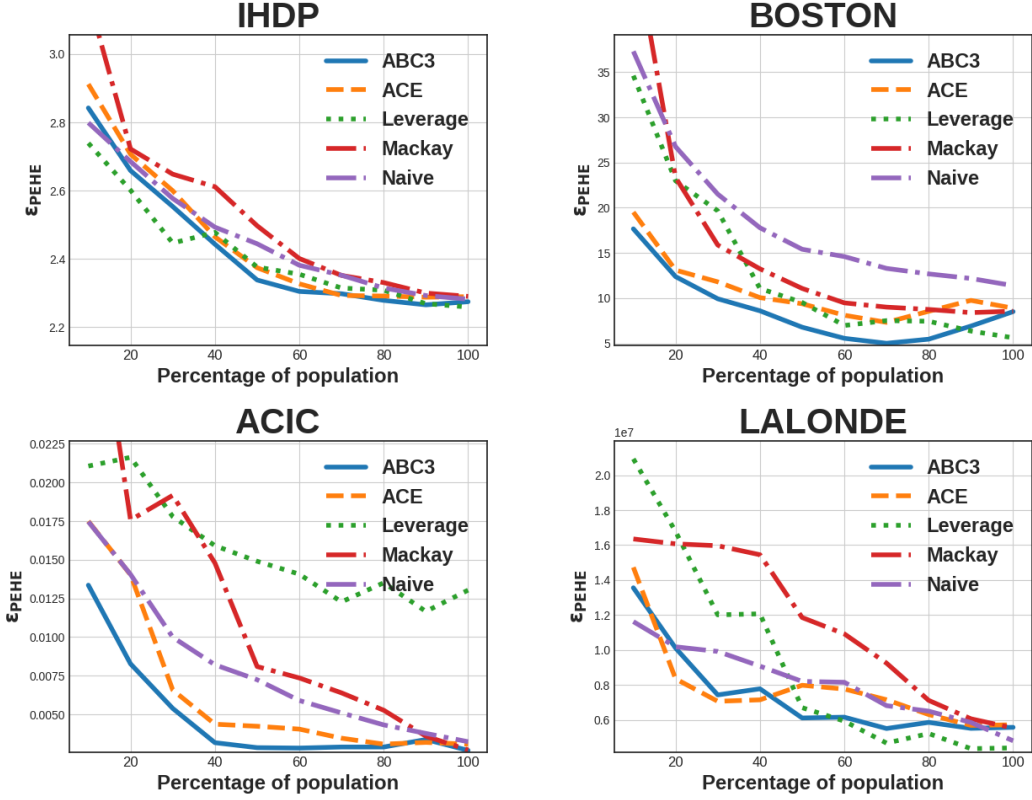


Figure 1: Mean of  $\epsilon_{PEHE}$ .  $x$ -axis represents the observed percentage of the population. We measure  $\epsilon_{PEHE}$  for every 10% observation.

After selecting  $x_{t+1}$  and  $a_{t+1}$ , all policies fit the Gaussian process for regression. We apply feature-wise normalization and  $y$ -standardization for all regressors. (except Leverage, which requires item-wise normalization) We fit two Gaussian process models with Constant Kernel \* Radial Basis Function (RBF) Kernel + White Kernel. We optimize the kernel hyperparameters using *scikit-learn* package.

All the uncertainty-aware policies (ABC3, Mackay and ACE) use the Gaussian process to quantify the uncertainty. For the uncertainty-quantifying kernels, we use RBF kernel with length scale 1.0 with  $\sigma_\epsilon^2 = 1$ . (We check the hyperparameter sensitivity in Section 5.4.)

### Minimizing Error

We measure the  $\epsilon_{PEHE}$  (not  $\epsilon_{PEHE}^\Omega$ ) when observing every 10% of population. We run 100 experiments for every data set. We mark the mean of measured  $\epsilon_{PEHE}$ . The results are in Figure 1. Appendix D presents the standard deviation of measured  $\epsilon_{PEHE}$ .

ABC3 shows the best performance, i.e. the lowest  $\epsilon_{PEHE}$  for most time steps. We can verify ABC3 successfully minimizes the population  $\epsilon_{PEHE}$ , though optimization target of ABC3 is  $\epsilon_{PEHE}^\Omega$ . In most cases, when ABC3 observes only half of the population, it achieves  $\epsilon_{PEHE}$  level which is achieved with full observation by other policies. Especially for Boston, after 20%, ABC3 achieves  $\epsilon_{PEHE}$  which Naive

policy cannot achieve even with full observation.

ACE policy shows comparable results to ABC3. ACE slightly outperforms ABC3 for a 20-40% interval of Lalonde data set. However, ABC3 outperforms ACE in most cases, though ABC3 has no access to the test data set, unlike ACE.

Leverage temporarily outperforms ABC3 for the beginning part of IHDP and the last part of Lalonde. However, for ACIC, Leverage significantly underperforms other policies. The result may imply the vulnerability of linearity assumption in real-world data sets.

Mackay underperforms even Naive policy most times. It is interesting as Mackay utilizes the same uncertainty information as ABC3. The result may imply the importance of *proper* utilization of the same information.

In summary, ABC3 is a promising Bayesian active learning policy, which efficiently and robustly achieves the best performance among the others.

### Balancing Treatment-Control Groups

Theorem 4.5 states that ABC3 theoretically minimizes the maximum mean discrepancy between treatment and control groups. A policy can benefit by minimizing MMD from several theoretical aspects, as introduced in Section 4.2. We empirically verify the property. We measure MMD between  $\mathbb{P}_t^1$  and  $\mathbb{P}_t^0$  for every 10% observation and applied the same setting as the previous section. We report the mean and standard

deviation in Figure 2.

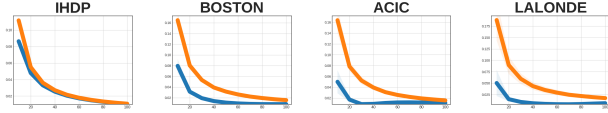


Figure 2: Mean and standard deviation of MMD between observed treatment and control groups,  $\mathbb{P}_t^1$  and  $\mathbb{P}_t^0$ . The blue line is for ABC3, and the orange line is for Naive policy.  $x$ -axis is for the sampled ratio and  $y$ -axis is MMD.

ABC3 achieves substantially lower MMD compared to Naive policy on all data sets. As noted in Theorem 4.5, the upper bound of MMD is minimized as the number of observations increases. As a result, Naive policy also shows a decrease in MMD as time proceeds. However, the MMD gap between ABC3 and Naive is significant at the beginning stage of experiments. The gap is especially large for ACIC and Lalonde data sets. The result empirically supports Theorem 4.5. holds, and ABC3 achieves a balance between treatment-control groups.

### Minimizing Type 1 Error

Theorem 4.7. states ABC3 minimizes the upper bound of the integrated type 1 error probability,  $\int_{\mathcal{X}} \mathbb{P}_{t+1}[\text{Type 1 Error}(x)] d\mathbb{P}(x)$ . Here we verify the property empirically.

We use Boston and Lalonde data sets which assume no treatment effect, i.e.  $Y^0 = Y^1$ . To measure the type 1 error rate, we compute a mean and standard deviation of  $\hat{CATE}(x)$  for every  $x$  in the test data set. Then we implement the Z test with a significance level of 5%. (i.e.  $\alpha = 1.96$ ). Type 1 error occurs when the absolute value of the Z-statistic is bigger than  $\alpha$ . We compute the percentage of  $x$  where the type 1 error occurs. The result is in Figure 3.

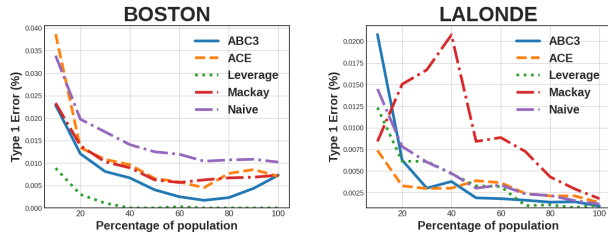


Figure 3: Mean of the type 1 error rate.

As time proceeds, ABC3 achieves a lower type 1 error rate, as expected in Theorem 4.7. ABC3 consistently shows a lower Type 1 error rate than Naive. However, Mackay shows the highest error rate on Lalonde. The result may again imply the importance of *proper* utilization of the uncertainty information.

Meanwhile, Leverage shows a significantly lower error rate for Boston. This aspect was not presented in the original paper (Addanki et al. 2022). The result may imply the strong linearity in Boston data set and the power of the method when the linearity assumption holds.

### Hyperparameter Sensitivity

To implement ABC3, we need to determine uncertainty-quantifying kernel, kernel parameters, and observation error  $\sigma_\epsilon^a$  as hyperparameters. As usual machine learning models utilizing the kernel method, selecting an appropriate kernel and parameters is crucial for obtaining precise estimation. We present hyperparameter sensitivity analysis for ABC3.

For the kernel, we test **RBF** kernel (utilized throughout this paper), **Matern** kernel, and Exp-Sine-Squared kernel (**Sine**). RBF kernel has one parameter, lengthscales, which determines how ‘local’ the output function would be. The smaller the lengthscales, the more local and wiggler the resulting function is. Matern kernel is a generalization of RBF kernel and has two parameters, lengthscales, and smoothness. Sine kernel assumes that our data shows a periodic pattern. it has two parameters, lengthscales, and periodicity. Here we test only lengthscales, with setting periodicity as 1. We present  $\epsilon_{PEHE}$  for each setting in Figure 4.

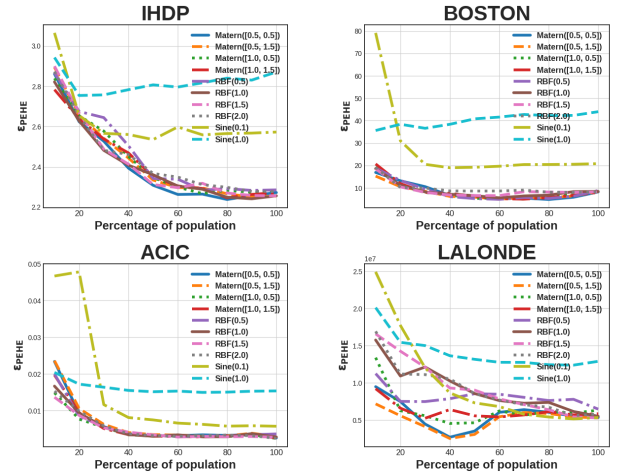


Figure 4:  $\epsilon_{PEHE}$  for every kernel and kernel parameter. The numbers in parenthesis are kernel parameters.

For IHDP, ACIC, and Boston, Matern and RBF kernels show robust performance along different kernel parameters. However, Sine kernel significantly deteriorates in the three data sets. The result implies no (or weak) periodicity in the three data sets. Meanwhile, Sine kernel with a length scale of 0.1 shows the best performance on Lalonde data set. It may imply a periodicity in the Lalonde data set. The result gives a similar lesson: selecting an appropriate kernel and parameters is crucial for obtaining precise estimation. Overall, RBF and Matern kernels with ‘reasonable’ parameters are safe options for general data sets.

We also test the hyperparameter sensitivity to  $\sigma_\epsilon^a$ . The result is in Figure 5. Except the extreme case ( $\sigma_\epsilon^0 : \sigma_\epsilon^1 = 1.0 : 0.1$ ), ABC3 shows robust performance for different  $\sigma_\epsilon^a$ s. Interestingly, 1.0 : 0.1 shows the best performance for IHDP data set, which may imply the noisiness in the control group. However, 1.0 : 0.1 significantly deteriorates on Lalonde data set. Overall, the equal noise setting (1.0 : 1.0), utilized throughout this paper, is not always the best, but a generally safe choice.

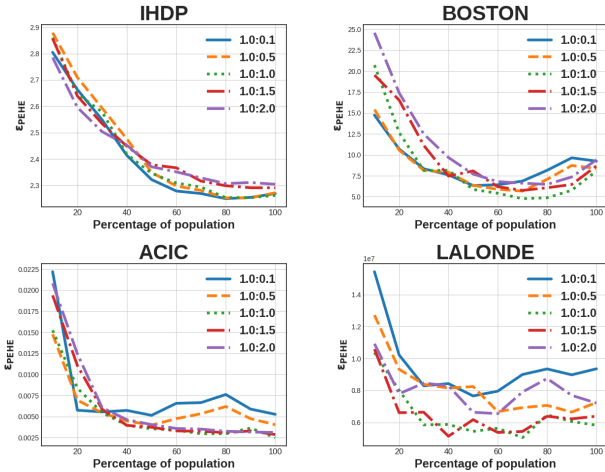


Figure 5:  $\epsilon_{PEHE}$  for different  $\sigma_\epsilon^0 : \sigma_\epsilon^1$ .

## Measuring Computation Time

Here we present the time to sample the whole train data set of Boston data. We compute the mean and standard deviation of the computation time by iterating 10 times. As a result, Leverage shows nearly the same computation time as Naive, while ABC3 shows a time comparable to Mackay. ACE requires nearly twice as much time than ABC3. However, most policies require less than 1 second to sample the whole data set. The result shows that ABC3 is a feasible policy with reasonable computation time.

Naive	Leverage	Mackay	ACE	ABC3
0.2567	0.2517	0.7595	1.6506	0.8871
(0.0321)	(0.0212)	(0.3803)	(0.7961)	(0.0373)

Table 1: Mean and standard deviation of the computation time (in second) on Boston data set.

## Empirical Validation of Assumption

In Theorem 4.5, we introduced the following assumption:  $\epsilon^*(I_n) \leq 2\delta^*(I_n), \forall I_n$  where  $M = \frac{1}{N^2} \sum_{i,j=1}^N k(x_i, x_j)$ ,  $\delta^*(I_n) = \frac{1}{n} \sum_{i \in I_n} \int_{\mathcal{X}} k(x_i, x) d\mathbb{P}(x)$  and  $\epsilon^*(I_n) = M - \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$ . To verify the empirical satisfaction of the assumption, we compute and plot  $\epsilon^*(I_n)$  and  $2\delta^*(I_n)$  for data sets used in our paper. As computing all  $\epsilon^*(I_n)$  and  $\delta^*(I_n)$  for all  $I_n$  is computationally infeasible, we compute the value of the leading principal submatrix by randomly permuting 100 times and present points with a minimum value of  $2\delta^*(I_n) - \epsilon^*(I_n)$ . The result, shown in Figure 6, supports the empirical satisfaction of the assumption.

## 6 Discussion & Limitation

ABC3 algorithm utilizes Gaussian process at its heart, hence the improvements pertaining to Gaussian process also happen in ABC3. For example, as multiple researchers attempted to extend the Gaussian Process to large data sets,

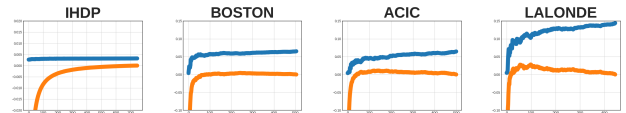


Figure 6: Empirical validation of the assumption ( $\epsilon^*(I_n) \leq 2\delta^*(I_n)$ ) from Theorem 4.5. The blue line is for  $2\delta^*(I_n)$ , the orange line is for  $\epsilon^*(I_n)$ , and the  $x$ -axis is for  $n$ .

e.g. Hensman, Fusi, and Lawrence (2013) and Wang et al. (2019), ABC3 can be extended to be more scalable. In Appendix E, we showcase a possible direction of extending ABC3 with sample-and-optimize approach, which outperforms the Naive policy on a large Weather data set with much better efficiency. This demonstrates the potential of ABC3, which goes beyond its original design principle to maximize the learning efficiency of the expensive and limited data set for causal inference.

Practitioners may consider using ABC3 just as an intelligent sampling policy, in conjunction to external regressors other than Gaussian process. The performance of the causal effect estimation when ABC3 is used with a different regressor, like a neural network or a random forest, is reported in Appendix F. According to the result, the best choice of regressor may depend not only on the sampling process design but also on the data set itself, which suggests another future research direction on designing optimal regressor given the data set and its sampling algorithm.

Lastly, ABC3 may be used as a selective data set cleanser that learns to avoid sampling potentially toxic observation points. As shown in Figure 1, for Boston and ACIC data sets, ABC3 shows the best performance when observing only a portion of data set. The result may imply the existence of *toxic* data points for CATE estimation, and ABC3 successfully avoids sampling those data points unless forced to sample all. We believe that more detailed analysis of this behavior deserves a separate future research as well.

## 7 Conclusion

We present ABC3, an active learning based sampling policy for causal inference. ABC3 minimizes the expected error of CATE estimation from a Bayesian perspective, without violating the key assumptions in causal inference. Using maximum mean discrepancy, we prove that ABC3 minimizes the upper bound of imbalance between observed treatment and control groups. Moreover, ABC3 theoretically minimizes the upper bound of type 1 error probability. ABC3 empirically outperforms other active learning policies, and its theoretical properties as well as empirical robustness are also validated to give additional support to the general applicability of ABC3. We expect ABC3 and its extensions will deepen theoretical insights and general applicability of active learning on causal inference tasks.

## Acknowledgements

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea Institute for Advancement

of Technology (KIAT) through the Materials and Parts Industry Technology Development Base Construction Project (P0022334) and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1G1A1102828).

## References

- Addanki, R.; Arbour, D.; Mai, T.; Musco, C. N.; and Rao, A. 2022. Sample Constrained Treatment Effect Estimation. In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.
- Antognini, A. B.; and Zagoraïou, M. 2011. The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika*, 98(3): 519–535.
- Atkinson, A. C. 2014. Selecting a Biased-Coin Design. *Statistical Science*, 29(1): 144 – 163.
- Brooks-Gunn, J.; Liaw, F.-r.; and Klebanov, P. K. 1992. Effects of early intervention on cognitive function of low birth weight preterm infants. *The Journal of pediatrics*, 120(3): 350–359.
- Cohn, D.; Ghahramani, Z.; and Jordan, M. 1994. Active Learning with Statistical Models. In Tesauro, G.; Touretzky, D.; and Leen, T., eds., *Advances in Neural Information Processing Systems*, volume 7. MIT Press.
- Dai, J.; Gradu, P.; and Harshaw, C. 2023. CLIP-OGD: An Experimental Design for Adaptive Neyman Allocation in Sequential Experiments. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 32235–32269. Curran Associates, Inc.
- Deng, K.; Pineau, J.; and Murphy, S. 2011. Active learning for personalizing treatment. In *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, 32–39.
- Efron, B. 1971. Forcing a sequential experiment to be balanced. *Biometrika*, 58(3): 403–417.
- Fisher, R. A. 1970. Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution*, 66–70. Springer.
- Ghadiri, M.; Arbour, D.; Mai, T.; Musco, C. N.; and Rao, A. 2023. Finite Population Regression Adjustment and Non-asymptotic Guarantees for Treatment Effect Estimation. In *Thirty-seventh Conference on Neural Information Processing Systems*.
- Gramacy, R. B. 2020. *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Boca Raton, Florida: Chapman Hall/CRC. <http://bobby.gramacy.com/surrogates/>.
- Gretton, A.; Borgwardt, K. M.; Rasch, M. J.; Schölkopf, B.; and Smola, A. 2012. A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13(25): 723–773.
- Gruber, S.; Lefebvre, G.; Schuster, T.; ; and Piché, A. 2019. Atlantic Causal Inference Conference Data Challenge, 2019. <https://sites.google.com/view/acic2019datachallenge/>. Accessed: 2024-07-18.
- Ham, D. W.; Bojinov, I.; Lindon, M.; and Tingley, M. 2023. Design-Based Confidence Sequences: A General Approach to Risk Mitigation in Online Experimentation. arXiv:2210.08639.
- Harrison, D.; and Rubinfeld, D. L. 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1): 81–102.
- Harshaw, C.; Sävje, F.; Spielman, D.; and Zhang, P. 2023. Balancing Covariates in Randomized Experiments with the Gram-Schmidt Walk Design. arXiv:1911.03071.
- Hensman, J.; Fusi, N.; and Lawrence, N. D. 2013. Gaussian processes for Big data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence, UAI'13*, 282–290. Arlington, Virginia, USA: AUAI Press.
- Hill, J. L. 2011. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, 20(1): 217–240.
- Jesson, A.; Tigas, P.; van Amersfoort, J.; Kirsch, A.; Shalit, U.; and Gal, Y. 2021. Causal-BALD: Deep Bayesian Active Learning of Outcomes to Infer Treatment-Effects from Observational Data. In Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*.
- LaLonde, R. J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, 76(4): 604–620.
- MacKay, D. J. C. 1992. Information-Based Objective Functions for Active Data Selection. *Neural Computation*, 4(4): 590–604.
- Rasmussen, C. E.; and Williams, C. K. I. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.
- Rubin, D. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66.
- Seo, S.; Wallat, M.; Graepel, T.; and Obermayer, K. 2000. Gaussian process regression: active data selection and test point rejection. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, 241–246 vol.3.
- Settles, B. 2009. Active learning literature survey.
- Shalit, U.; Johansson, F. D.; and Sontag, D. 2017. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, 3076–3085. PMLR.
- Song, D.; Mak, S.; and Wu, C. F. J. 2023. ACE: Active Learning for Causal Inference with Expensive Experiments. arXiv:2306.07480.
- Sundin, I.; Schulam, P.; Siivola, E.; Vehtari, A.; Saria, S.; and Kaski, S. 2019. Active learning for decision-making from imbalanced observational data. In *International conference on machine learning*, 6046–6055. PMLR.
- Toth, C.; Lorch, L.; Knoll, C.; Krause, A.; Pernkopf, F.; Peharz, R.; and von Kügelgen, J. 2022. Active Bayesian Causal Inference. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural*



*Information Processing Systems*, volume 35, 16261–16275. Curran Associates, Inc.

Wang, K.; Pleiss, G.; Gardner, J.; Tyree, S.; Weinberger, K. Q.; and Wilson, A. G. 2019. Exact Gaussian Processes on a Million Data Points. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Zhao, J.; and Zhou, Z. 2024. Pigeonhole Design: Balancing Sequential Experiments from an Online Matching Perspective. arXiv:2201.12936.

Zhu, H.; Zhang, S.; Su, Y.; Zhao, Z.; and Chen, N. 2024. Integrating Active Learning in Causal Inference with Interference: A Novel Approach in Online Experiments. arXiv:2402.12710.

# Appendices

This is a technical appendix for the manuscript: **ABC3: Active Bayesian Causal Inference with Cohn Criteria in Randomized Experiments.**

## A Proofs

### A.1 Proof on Theorem 4.1.

To prove the theorem, we need a lemma.

**Lemma A.1.** *For a random variable  $X$  and two filtration  $\mathcal{F}' \subset \mathcal{F}$ , if  $\mathbb{V}[X|\mathcal{F}]$  is a known constant in  $\mathcal{F}'$ , then  $\mathbb{V}[\mathbb{E}[X|\mathcal{F}]|\mathcal{F}'] = \mathbb{V}[X|\mathcal{F}'] - \mathbb{V}[X|\mathcal{F}]$*

*Proof.*

$$\begin{aligned}
\mathbb{V}[X|\mathcal{F}] &= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}] \\
&= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'] + \mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}] \\
&= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'])^2|\mathcal{F}] + 2\mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'])(\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}]|\mathcal{F}] \\
&\quad + \mathbb{E}[(\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}] \\
&= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'])^2|\mathcal{F}] + 2(\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}'])(\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}]) \\
&\quad + (\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}])^2 \\
&= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'])^2|\mathcal{F}] - (\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}])^2
\end{aligned}$$

where the fourth equality holds as  $\mathbb{E}[X|\mathcal{F}']$  and  $\mathbb{E}[X|\mathcal{F}]$  are already known at  $\mathcal{F}$ . By taking an expectation with respect to  $\mathcal{F}'$ , as  $\mathbb{V}[X|\mathcal{F}]$  is a known constant in  $\mathcal{F}'$ ,

$$\begin{aligned}
\mathbb{V}[X|\mathcal{F}] &= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}'])^2|\mathcal{F}'] - \mathbb{E}[(\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}'] \\
&= \mathbb{V}[X|\mathcal{F}'] - \mathbb{V}[\mathbb{E}[X|\mathcal{F}]|\mathcal{F}']
\end{aligned}$$

by the tower property.  $\square$

**Theorem A.2.** *Assume  $|k(x, x')| < \infty$  and  $|y_i^a| < \infty$  for all  $x, x' \in \mathcal{X}$  and  $a, i$ , as a result,  $\epsilon_{PEHE}^\Omega(C\hat{A}TE_t) < \infty, \forall t$ . Let our estimator  $C\hat{A}TE_t(x) = \hat{y}_t^1(x) - \hat{y}_t^0(x)$ , where  $\hat{y}_t^a(x) = \mathbb{E}_t[Y^a(x)]$  is a mean posterior distribution of gaussian process  $Y^a$  trained on data set  $D_t^a$ . Assume two Gaussian processes  $Y^1$  and  $Y^0$  are independent. Then*

$$\begin{aligned}
&\arg \min_{x_{t+1}, a_{t+1}} \mathbb{E}_{t+1} \left[ \epsilon_{PEHE}^\Omega(C\hat{A}TE_{t+1}) \right] \\
&= \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1}[Y^1(x)] + \mathbb{V}_{t+1}[Y^0(x)] d\mathbb{P}(x)
\end{aligned}$$

*Proof.* We can write our target quantity,

$$\begin{aligned} & \mathbb{E}_{t+1} \left[ \epsilon_{PEHE}^\Omega(C\hat{A}TE_{t+1}) \right] \\ &= \mathbb{E}_{t+1} \left[ \int_{\mathcal{X}} (C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x))^2 d\mathbb{P}(x) \right] \end{aligned}$$

Note

$$\begin{aligned} & \mathbb{E}_{t+1} \left[ \int_{\mathcal{X}} (C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x))^2 d\mathbb{P}(x) \right] \\ &= \int_{\mathcal{X}} \mathbb{E}_{t+1} \left[ (C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x))^2 \right] d\mathbb{P}(x) \text{ by Fubini} \\ &= \int_{\mathcal{X}} \mathbb{E}_{t+1} \left[ C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x) \right]^2 + \mathbb{V}_{t+1} \left[ C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x) \right] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{V}_{t+1} \left[ C\hat{A}TE_{t+1}(x) - C\hat{A}TE_\Omega(x) \right] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{V}_{t+1} \left[ C\hat{A}TE_\Omega(x) \right] d\mathbb{P}(x) \end{aligned}$$

The fourth line holds because  $\mathbb{E}_{t+1} \left[ C\hat{A}TE_\Omega(x) \right] = \mathbb{E}_{t+1} \left[ C\hat{A}TE_{t+1}(x) \right]$  by the tower property. We can use Fubini Theorem in the second line as we assumed  $\epsilon_{PEHE}^\Omega(C\hat{A}TE_t) < \infty, \forall t$ . Also,  $\mathbb{V}_{t+1} \left[ C\hat{A}TE_{t+1}(x) \right] = 0$  as  $C\hat{A}TE_{t+1}$  is fixed at  $t+1$ .

By the property of the Gaussian process, we can compute  $\mathbb{V}_\Omega [Y^1(x) - Y^0(x)]$ , i.e. it is a known constant at  $t+1$ . Then by applying Lemma A.1.,

$$\begin{aligned} & \int_{\mathcal{X}} \mathbb{V}_{t+1} \left[ C\hat{A}TE_\Omega(x) \right] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{V}_{t+1} \left[ \mathbb{E}_\Omega [Y^1(x) - Y^0(x)] \right] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x) - Y^0(x)] - \mathbb{V}_\Omega [Y^1(x) - Y^0(x)] d\mathbb{P}(x) \end{aligned}$$

As our choice on  $x_{t+1}$  and  $a_{t+1}$  does not affect  $\mathbb{V}_\Omega [Y^1(x) - Y^0(x)]$ , we obtain

$$\begin{aligned} & \arg \min_{x_{t+1}, a_{t+1}} \mathbb{E}_{t+1} \left[ \epsilon_{PEHE}^\Omega(C\hat{A}TE_{t+1}) \right] \\ &= \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x) - Y^0(x)] d\mathbb{P}(x) \\ &= \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) \end{aligned}$$

by the independence assumption between  $Y^1$  and  $Y^0$ . □

## A.2 Proof on Proposition 4.2.

To prove the proposition, we need a lemma.

**Lemma A.3.**

$$\begin{aligned} & \arg \min_{x_{t+1}, a} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) \\ &= \arg \max_{x_{t+1}, a} \int_{\mathcal{X}} \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \end{aligned}$$

*Proof.*  $x_{t+1} = \tilde{x}, a_{t+1} = a$  holds, if

$$= \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^a(x)] + \mathbb{V}_t [Y^{|1-a|}(x)] d\mathbb{P}(x) \quad (1)$$

$$\leq \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^{|1-a|}(x)] + \mathbb{V}_t [Y^a(x)] d\mathbb{P}(x) \quad (2)$$

and

$$= \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^a(x)] + \mathbb{V}_t [Y^{|1-a|}(x)] d\mathbb{P}(x) \quad (3)$$

$$\leq \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^a(x)] + \mathbb{V}_t [Y^{|1-a|}(x)] d\mathbb{P}(x) \quad (4)$$

for all  $\bar{x}$ , where  $\mathbb{V}_{t+1} = \mathbb{V}[\cdot | \mathcal{F}_{t+1}]$ ,  $\mathbb{V}_{t+1} = \mathbb{V}[\cdot | \mathcal{F}_{t+1}]$ ,  $\mathcal{F}_{t+1} = \mathcal{F}_t \cup \{\tilde{x}\}$ ,  $\mathcal{F}_{t+1} = \mathcal{F}_t \cup \{\bar{x}\}$ , i.e. filtration assuming the future observation  $x_{t+1} = \tilde{x}$  and  $\bar{x}$ . Proposition holds as

$$\begin{aligned} (2) & \iff \int_{\mathcal{X}} \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \\ & \geq \int_{\mathcal{X}} \mathbb{V}_t [Y^{|1-a|}(x)] - \mathbb{V}_{t+1} [Y^{|1-a|}(x)] d\mathbb{P}(x) \end{aligned}$$

and

$$\begin{aligned} (4) & \iff \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \leq \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \\ & \iff \int_{\mathcal{X}} -\mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \geq \int_{\mathcal{X}} -\mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \\ & \iff \int_{\mathcal{X}} \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \\ & \geq \int_{\mathcal{X}} \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] d\mathbb{P}(x) \end{aligned}$$

□

Now we can prove the proposition

**Proposition A.4.**

$$\begin{aligned} & \arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) = \\ & \arg \max_{x_{t+1}, a} \frac{\int_{\mathcal{X}} \left[ (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x) \right]^2 d\mathbb{P}(x)}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{k}}_{t+1}^a} \end{aligned}$$

where  $\tilde{\mathbf{k}}_{t+1}^a = [k(x_i, x_{t+1})]_{i \in I_\epsilon^a}$

*Proof.* As shown in Lemma A.3.,  $\arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x) = \arg \max_{x_{t+1}, a} \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)]$ .

Note,

$$\begin{aligned} & \mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] \\ &= (\mathbf{k}_{t+1,*}^a(x))^T [\mathbb{K}_{t+1}^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t+1,*}^a(x) - (\mathbf{k}_{t,*}^a(x))^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) \end{aligned}$$

and

$$\begin{aligned} \mathbb{K}_{t+1}^a &= \begin{bmatrix} \mathbb{K}_t^a & \tilde{\mathbf{k}}_{t+1}^a \\ (\tilde{\mathbf{k}}_{t+1}^a)^T & k(x_{t+1}, x_{t+1}) \end{bmatrix} \\ \mathbf{k}_{t+1,*}^a(x) &= [\mathbf{k}_{t,*}^a(x) \quad k(x_{t+1}, x)] \end{aligned}$$

By utilizing the bordering method (Faddeev and Faddeeva, 1981), we can compute the inverse matrix.

$$(\mathbb{K}_{t+1}^a + \sigma_\epsilon^2 \mathbf{I})^{-1} = \begin{bmatrix} (i) & -\frac{(\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1} \tilde{\mathbf{k}}_{t+1}^a}{(ii)} \\ -\frac{(\tilde{\mathbf{k}}_{t+1}^a)^T (\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1}}{(ii)} & \frac{1}{(ii)} \end{bmatrix}$$

where

$$\begin{aligned} (i) &= (\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1} + \frac{(\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1} \tilde{\mathbf{k}}_{t+1}^a (\tilde{\mathbf{k}}_{t+1}^a)^T (\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1}}{(ii)} \\ (ii) &= k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T (\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I})^{-1} \tilde{\mathbf{k}}_{t+1}^a \end{aligned}$$

By multiplying  $\mathbf{k}_{t+1,*}^a$  to both sides, we obtain

$$\begin{aligned} & (\mathbf{k}_{t+1,*}^a(x))^T [\mathbb{K}_{t+1}^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t+1,*}^a(x) \\ &= \frac{\left[ (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x) \right]^2}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{k}}_{t+1}^a} \\ &+ (\mathbf{k}_{t,*}^a(x))^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) \end{aligned}$$

So  $\mathbb{V}_t [Y^a(x)] - \mathbb{V}_{t+1} [Y^a(x)] = \frac{[(\tilde{\mathbf{K}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x)]^2}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{K}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{K}}_{t+1}^a}$  so the proposition holds.  $\square$

### A.3 Proof on Theorem 4.5

**Theorem A.5.** Assume  $0 < |k(x, x')| < \infty$  and  $|y_i^a| < \infty$  for all  $x, x' \in \mathcal{X}$  and  $a, i$ . Let  $\lambda^*$  be a maximum eigenvalue of  $\mathbb{K}_\Omega$ , i.e. covariance matrix of whole covariates. Let  $M = \frac{1}{N^2} \sum_{i,j=1}^N k(x_i, x_j)$ . Define functions  $\delta^*(I_n) = \frac{1}{n} \sum_{i \in I_n} \int_{\mathcal{X}} k(x_i, x) d\mathbb{P}(x)$ , and  $\epsilon^*(I_n) = M - \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$ , where  $I_n \subset \{1, \dots, N\}$  is an  $n$ -element index set. Assume  $\epsilon^*(I_n) \leq 2\delta^*(I_n), \forall I_n$ . Then

$$MMD(\mathbb{P}_t^1, \mathbb{P}_t^0, \mathcal{F})^2 \leq 4 \frac{\lambda^*}{|I_t^1|} + 4 \frac{\lambda^*}{|I_t^0|} + 2 \int_{\mathcal{X}} \mathbb{V}_t [Y^1(x)] + \mathbb{V}_t [Y^0(x)] d\mathbb{P}(x)$$

*Proof.* Note,  $MMD(\mathbb{P}_t^1, \mathbb{P}_t^0, \mathcal{F})^2 \leq 2MMD(\mathbb{P}_t^1, \mathbb{P}, \mathcal{F})^2 + 2MMD(\mathbb{P}_t^0, \mathbb{P}, \mathcal{F})^2$ , as

$$\begin{aligned} & MMD(\mathbb{P}_t^1, \mathbb{P}_t^0, \mathcal{F})^2 \\ & \leq [MMD(\mathbb{P}_t^1, \mathbb{P}, \mathcal{F}) + MMD(\mathbb{P}_t^0, \mathbb{P}, \mathcal{F})]^2 \text{ by Triangle Inequality} \\ & = MMD(\mathbb{P}_t^1, \mathbb{P}, \mathcal{F})^2 + MMD(\mathbb{P}_t^0, \mathbb{P}, \mathcal{F})^2 + 2MMD(\mathbb{P}_t^1, \mathbb{P}, \mathcal{F})MMD(\mathbb{P}_t^0, \mathbb{P}, \mathcal{F}) \\ & \leq 2MMD(\mathbb{P}_t^1, \mathbb{P}, \mathcal{F})^2 + 2MMD(\mathbb{P}_t^0, \mathbb{P}, \mathcal{F})^2 \end{aligned}$$

We focus on one  $MMD(\mathbb{P}_t^a, \mathbb{P}, \mathcal{F})$  as we can extend the arguments to  $MMD(\mathbb{P}_t^{|1-a|}, \mathbb{P}, \mathcal{F})$ . For notational convenience, we omit  $a$  if trivial.

Define a weighted empirical distribution  $\tilde{\mathbb{P}}_t = \sum_{i \in I_t} \omega_{t,i} \delta_{x_i}$ , where the weight vector  $\omega_t = (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t$ ,  $\tilde{\mathbf{k}}_t = [\int k(x_i, x) d\mathbb{P}(x)]_{i \in I_t}$ . As

$$MMD(\mathbb{P}_t, \mathbb{P}, \mathcal{F})^2 = \|\mu_{\mathbb{P}_t} - \mu_{\mathbb{P}}\|^2 \quad (5)$$

$$= \bar{\mathbf{1}}^T \mathbb{K}_t \bar{\mathbf{1}} + \int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') d\mathbb{P}(x) d\mathbb{P}(x') - 2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} \quad (6)$$

$$= \bar{\mathbf{1}}^T \mathbb{K}_t \bar{\mathbf{1}} \quad (7)$$

$$+ \int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') d\mathbb{P}(x) d\mathbb{P}(x') - \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t \quad (8)$$

$$+ \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t - 2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} \quad (9)$$

where  $\bar{\mathbf{1}} = \frac{1}{|I_t|} \mathbf{1}_{|I_t|}$  is a vector computing the mean. As  $\mathbb{K}_t$  is a symmetric and positive-definite matrix, we obtain,

$$(7) \leq \lambda_{max}(\mathbb{K}_t) \bar{\mathbf{1}}^T \bar{\mathbf{1}} = \frac{1}{|I_t|} \lambda_{max}(\mathbb{K}_t) \leq \frac{1}{|I_t|} \lambda_{max}(\mathbb{K}_\Omega) = \frac{\lambda^*}{|I_t|} \quad (10)$$

where  $\lambda_{max}$  is a maximum eigenvalue of a matrix.

From the Bayesian quadrature literature, it is known that (8) =  $\mathbb{V}_t [\int Y(x)d\mathbb{P}(x)]$  (See Section 6.2. of Kanagawa et al. (2018)). Since

$$\mathbb{V}_t \left[ \int Y(x)d\mathbb{P}(x) \right] = \mathbb{E}_t \left[ \left[ \int Y(x)d\mathbb{P}(x) - \mathbb{E}_t \left[ \int Y(x)d\mathbb{P}(x) \right] \right]^2 \right] \quad (11)$$

$$= \mathbb{E}_t \left[ \left[ \int Y(x) - \mathbb{E}_t [Y(x)] d\mathbb{P}(x) \right]^2 \right] \text{ by Fubini} \quad (12)$$

$$\leq \mathbb{E}_t \left[ \int [Y(x) - \mathbb{E}_t [Y(x)]]^2 d\mathbb{P}(x) \right] \text{ by Jensen's inequality} \quad (13)$$

$$= \int \mathbb{E}_t \left[ [Y(x) - \mathbb{E}_t [Y(x)]]^2 \right] d\mathbb{P}(x) \text{ by Fubini} \quad (14)$$

$$= \int \mathbb{V}_t [Y(x)] d\mathbb{P}(x) \quad (15)$$

We can apply the Fubini theorem to (12) because,

$$\begin{aligned} \int \mathbb{E}_t [Y(x)] d\mathbb{P}(x) &= \int \mathbf{k}_{t,*}(x)^T (\mathbb{K}_t)^{-1} \mathbf{y}_t d\mathbb{P}(x) \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{k}_{t,*}(x_i)^T (\mathbb{K}_t)^{-1} \mathbf{y}_t < \infty \end{aligned}$$

by assumptions. Likewise for (14),

$$\begin{aligned} \int \mathbb{E}_t \left[ [Y(x) - \mathbb{E}_t [Y(x)]]^2 \right] d\mathbb{P}(x) &= \int \mathbb{V}_t [Y(x)] d\mathbb{P}(x) \\ &= \int k(x, x) - \mathbf{k}_{t,*}(x)^T (\mathbb{K}_t)^{-1} \mathbf{k}_{t,*}(x) d\mathbb{P}(x) \\ &= \frac{1}{N} \sum_{i=1}^N k(x_i, x_i) - \mathbf{k}_{t,*}(x_i)^T (\mathbb{K}_t)^{-1} \mathbf{k}_{t,*}(x_i) < \infty \end{aligned}$$

So

$$(8) \leq \int \mathbb{V}_t [Y(x)] d\mathbb{P}(x) \quad (16)$$

As  $MMD(\mathbb{P}, \tilde{\mathbb{P}}_t, \mathcal{F})^2 = \int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') d\mathbb{P}(x) d\mathbb{P}(x') - \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t \geq 0$ , we have  $M = \frac{1}{N^2} \sum_{i,j=1}^N k(x_i, x_j) = \int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') d\mathbb{P}(x) d\mathbb{P}(x') \geq \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t$ . By definition,  $2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} = 2\delta^*(I_t)$  and  $\bar{\mathbf{1}}^T \mathbb{K}_t \bar{\mathbf{1}} = M - \epsilon^*(I_t)$ . By our assumption,

$$\epsilon^*(I_t) \leq 2\delta^*(I_t) \iff M - \epsilon^*(I_t) \geq M - 2\delta^*(I_t) \quad (17)$$

$$\iff \bar{\mathbf{1}}^T \mathbb{K}_t \bar{\mathbf{1}} \geq M - 2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} \quad (18)$$

$$\implies \bar{\mathbf{1}}^T \mathbb{K}_t \bar{\mathbf{1}} \geq \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t - 2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} \quad (19)$$

$$\implies \frac{\lambda^*}{|I_t|} \geq \tilde{\mathbf{k}}_t^T (\mathbb{K}_t)^{-1} \tilde{\mathbf{k}}_t - 2\tilde{\mathbf{k}}_t^T \bar{\mathbf{1}} \text{ by (10)} \quad (20)$$

By combining the results from (10), (16), and (20), we obtain

$$MMD(\mathbb{P}_t^a, \mathbb{P}, \mathcal{F})^2 \leq 2 \frac{\lambda^*}{|I_t^a|} + \int \mathbb{V}_t [Y^a(x)] d\mathbb{P}(x)$$

and as a result,

$$MMD(\mathbb{P}_t^1, \mathbb{P}_t^0, \mathcal{F})^2 \leq 4 \frac{\lambda^*}{|I_t^1|} + 4 \frac{\lambda^*}{|I_t^0|} + 2 \int \mathbb{V}_t [Y^1(x)] + \mathbb{V}_t [Y^0(x)] d\mathbb{P}(x)$$

□

#### A.4 Proof on Theorem 4.7.

**Theorem A.6.** *Under Fisher's Sharp null hypothesis,  $\arg \min_{x_{t+1}, a_{t+1}} \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] d\mathbb{P}(x)$  also minimizes the upper bound of the  $\int_{\mathcal{X}} \mathbb{P}_{t+1} [\text{Type 1 Error}(x)] d\mathbb{P}(x)$*

*Proof.* Note,

$$\mathbb{P}_{t+1} [\text{Type 1 Error}(x)] = \mathbb{P}_{t+1} [ |Y^1(x) - Y^0(x)| > \alpha ]$$

By the Markov inequality,

$$\begin{aligned} & \int_{\mathcal{X}} \mathbb{P}_{t+1} [ |Y^1(x) - Y^0(x)| > \alpha ] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{P}_{t+1} [ (Y^1(x) - Y^0(x))^2 > \alpha^2 ] d\mathbb{P}(x) \\ &\leq \int_{\mathcal{X}} \frac{\mathbb{E}_{t+1} [(Y^1(x) - Y^0(x))^2]}{\alpha^2} d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \frac{\mathbb{V}_{t+1} [(Y^1(x) - Y^0(x))] + \mathbb{E}_{t+1} [(Y^1(x) - Y^0(x))]^2}{\alpha^2} d\mathbb{P}(x) \\ &\propto \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + [Y^0(x)] + C\hat{ATE}_{t+1}(x)^2 d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} \mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)] + (C\hat{ATE}_{t+1}(x) - C\hat{ATE}_{\Omega}(x))^2 d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} [\mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)]] + \mathbb{E}_{t+1} [(C\hat{ATE}_{t+1}(x) - C\hat{ATE}_{\Omega}(x))^2] d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} [\mathbb{V}_{t+1} [Y^1(x)] + \mathbb{V}_{t+1} [Y^0(x)]] d\mathbb{P}(x) + \mathbb{E}_{t+1} [\epsilon_{PEHE}^{\Omega}(C\hat{ATE}_{t+1})] \end{aligned}$$

where the last three lines hold as  $C\hat{ATE}_{\Omega}(x) = 0, \forall x$  under the null hypothesis, and  $C\hat{ATE}_{t+1}(x) - C\hat{ATE}_{\Omega}(x)$  is known at  $t + 1$ . The last line holds by applying the Fubini Theorem with our assumption that  $\epsilon_{PEHE}^{\Omega}(C\hat{ATE}_t) < \infty, \forall t$ . As shown in Theorem 4.1., both terms are our minimization target. □



## B Analysis on the Assumption of Theorem 4.5.

Recall the key assumption in Theorem 4.5. was  $\epsilon^*(I_n) \leq 2\delta^*(I_n), \forall I_n$  where  $M = \frac{1}{N^2} \sum_{i,j=1}^N k(x_i, x_j)$ ,  $\delta^*(I_n) = \frac{1}{n} \sum_{i \in I_n} \int_{\mathcal{X}} k(x_i, x) d\mathbb{P}(x)$  and  $\epsilon^*(I_n) = M - \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$ . The assumption is equivalent to  $M - 2\delta^*(I_n) \leq \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$ . Assume the kernel is a non-negative valued kernel with maximum value 1, which is achieved when  $x = x'$  (e.g. RBF and Matern kernels, used in our paper).

Then  $\inf_{I_n} \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$  is lower bounded by  $\frac{1}{n}$ . The lower bound is achieved when the covariates are (roughly speaking) extremely and absolutely isolated in the kernel space, i.e.  $k(x_i, x_j) = 0, \forall i \neq j$ . The gap between the line and  $\inf_{I_n} \frac{1}{n^2} \sum_{i,j \in I_n} k(x_i, x_j)$  represents the absolute uniformness of the covariates. Meanwhile,  $M - 2\inf_{I_n} \delta^*(I_n)$  is lower bounded by the straight line connecting  $(0, M)$  and  $(N, -M)$ . The lower bound is achieved when the covariates are (again, roughly speaking) extremely uniformly distributed, i.e.  $\int k(x_i, x) d\mathbb{P}(x) = \int k(x_j, x) d\mathbb{P}(x), \forall i, j$ . The gap between the line and  $M - 2\inf_{I_n} \delta^*(I_n)$  represents the relative isolatedness of the covariates.

So the assumption may not hold if the covariates are absolutely and relatively isolated. However, as shown in Section 5.5, this assumption holds for general data sets with a large margin. Figure 1 visualizes the condition.

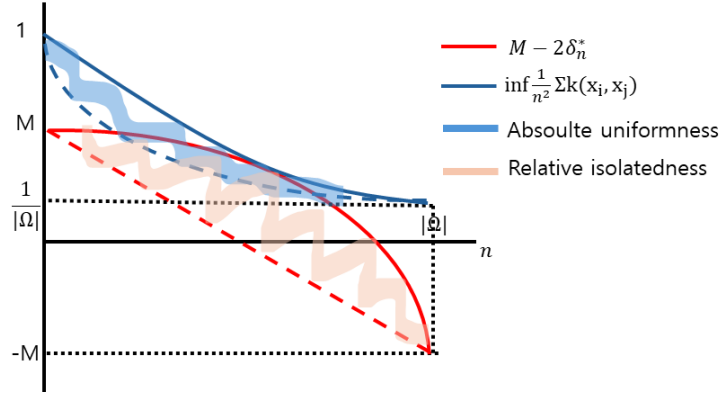


Figure 1: Intuitive visualization of the assumption used in Theorem 4.5.

## C Data Set & Experimental Setting

We present more detailed explanations of utilized data sets.

**IHDP** (Brooks-Gunn, Liaw, and Klebanov (1992) and Hill (2011)): This study measures the cognitive development of low-birth-weight and premature infants after intensive high-quality child care and home visits. It consists of 747 samples with 25 covariates describing the children and parents with simulated potential outcomes.

**Boston** (Harrison and Rubinfeld, 1978): This data set measures the median house prices in the Boston area. It consists of 506 samples with 13 covariates. Following

Addanki et al. (2022), we set  $Y^0 = Y^1$  to simulate the null hypothesis circumstance (i.e.  $CATE(x) = 0, \forall x$ ).

**ACIC** (Gruber et al., 2019): We utilize the first test data set of the Atlantic Causal Inference Conference 2019 Data Challenge data set. It consists of 500 samples with 22 covariates. We report the results for the other 7 test data sets on Appendix G.

**Lalonde** (LaLonde, 1986): This data set measures the effectiveness of a job training program on each individual’s earnings. It consists of 506 samples with 10 covariates. As in Boston, we simulate the null hypothesis circumstance.

For the experiment, we use a machine with AMD Ryzen 9 5900X 12-Core Processor CPU and NVIDIA RTX 3090 GPU.

## D Standard Deviation of Measured $\epsilon_{PEHE}$

We present the standard deviation of measured  $\epsilon_{PEHE}$ .

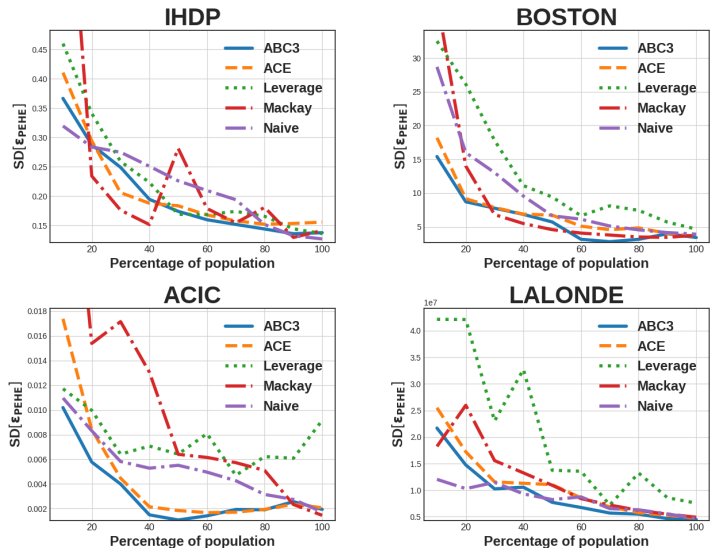


Figure 2: Standard deviation of  $\epsilon_{PEHE}$ .

## E Extending ABC3 to Large Data Set

Though the majority of randomized experimental data is relatively small by nature, practitioners may want to apply ABC3 to the huge number of covariates. However, computing our target quantity (Equation (1) in the main paper) for whole unseen covariates is computationally infeasible.

To bypass the problem, we showcase the extension of ABC3, which utilizes sampling and optimization techniques. It samples  $N$  items among whole subjects and ob-

served outcomes and maximizes the sampling version of our target quantity with optimization techniques. We use L-BFGS as our optimizer, and uniform sampling. We present the policy in greater detail in Algorithm 1.

To verify the empirical applicability, we use a large **Weather** data set (Budincsevity, 2016), consisting of 100k weather information in Szeged, where we can not apply original ABC3 and other active learning policies because variance computation is computationally infeasible. We should predict visibility with other weather-related covariates, e.g. temperature, humidity, or wind speed, utilizing only 0.1~1% of data. As it is not a randomized experimental data set, we set  $Y^1 = Y^0$ . We run 10 experiments with different train-test splits with a ratio of 9:1. We compare ABC3 with **Naive** and **Sample** policies, where the latter does not perform optimization, but computes our target quantity within a set of sampled unseen covariates. The result is in Figure 2.

---

**Algorithm 1** Extension of ABC3

---

**Input:** Current time step  $t$ , sampling number  $n$ , whole covariates set  $X_\Omega$ , covariates distribution  $\mathbb{P}$ , previous observations  $X_t^1$  and  $X_t^0$ , maximizing optimizer  $\mathcal{O}$ , tolerance  $\epsilon$ , kernel  $k$ , noise parameter  $\sigma_\epsilon$

**Output:**  $x_{t+1}, a_{t+1}$

- 1: Set an initial value  $x_0^\mathcal{O} = \text{feature-wise mean}(X_\Omega \setminus (X_t^1 \cup X_t^0))$
  - 2: Set an initial tolerance  $t_0 = \infty, i = 0$
  - 3: **while**  $t_i > \epsilon$  **do**
  - 4:   Sample  $n$  elements  $X_n \subset X_\Omega$  and  $\{I_t^a\}_n \subset I_t^a$  for  $a \in \{0, 1\}$
  - 5:   Compute  $\tilde{\mathbf{k}}_{t+1}^a, \mathbb{K}_t^a, \mathbf{k}_{t,*}^a$  assuming  $x_{t+1} = x_i^\mathcal{O}$  and  $I_t^a = \{I_t^a\}_n$  for  $a \in \{0, 1\}$
  - 6:    $x_{i+1}^\mathcal{O} = \mathcal{O}_{x_{t+1}} \left[ \max_a \frac{\int_{X_n} [(\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x)]^2 d\mathbb{P}(x)}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{k}}_{t+1}^a} \right]$
  - 7:    $t_{i+1} = \|x_{i+1}^\mathcal{O} - x_i^\mathcal{O}\|$
  - 8:    $i = i + 1$
  - 9: **end while**
  - 10:  $x_{t+1} = \arg \max_{x \in X_\Omega \setminus (X_t^1 \cup X_t^0)} k(x, x_i^\mathcal{O})$
  - 11:  $a_{t+1} = \arg \max_a \left[ \frac{\int_{X_n} [(\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{k}_{t,*}^a(x) - k(x_{t+1}, x)]^2 d\mathbb{P}(x)}{k(x_{t+1}, x_{t+1}) + \sigma_\epsilon^2 - (\tilde{\mathbf{k}}_{t+1}^a)^T [\mathbb{K}_t^a + \sigma_\epsilon^2 \mathbf{I}]^{-1} \tilde{\mathbf{k}}_{t+1}^a} \right]$
  - 12: **return**  $x_{t+1}, a_{t+1}$
- 

ABC3 substantially outperforms other policies despite small sample sizes. Especially, the gap between ABC3 and Sample policy shows the importance of the optimization step. Meanwhile, the performances of Sample and Naive policies are indistinguishable. The result shows the extensibility of ABC3 to large data sets when equipped with suitable techniques.

## F Plugging-in Other Regressor

In our paper, we draw ABC3 by assuming our data follows the Gaussian process. However, we can flexibly change the regressors trained on observed data sets, while

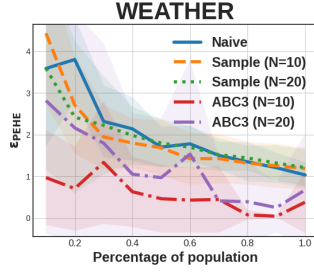


Figure 3: Mean and standard deviation of  $\epsilon_{PEHE}$  on large Weather data set.  $N$  is the number of utilized samples.

maintaining ABC3 policy. It would be beneficial if we achieve lower error by utilizing more sophisticated regressors, like neural networks. We present  $\epsilon_{PEHE}$  for different regressors. We test Gaussian process (GP), random forest (RF, Breiman (2001)), kernel support vector machine (SVM, Cristianini and Scholkopf (2002)), and 3 layered neural network with hidden size 100 (NN). The result is in Figure 3.

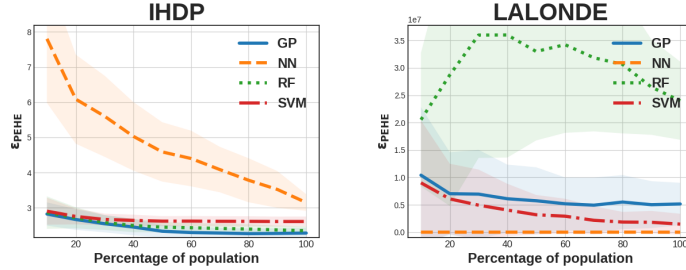


Figure 4:  $\epsilon_{PEHE}$  for different regressors.

Regressors show different trends depending on data sets. For IHDP, NN significantly deteriorates compared to other models. SVM underperforms GP and RF for IHDP. However, for Lalonde, NN significantly outperforms the others. Also, SVM outperforms GP, while RF shows the worst performance.

The result shows that we can obtain better estimation with a more appropriate regressor, while it can also significantly deteriorate. We recommend practitioners to check the suitability of regressors before applying them.

## G Results for ACIC Data Sets

### References

Addanki, R.; Arbour, D.; Mai, T.; Musco, C. N.; and Rao, A. 2022. Sample Constrained Treatment Effect Estimation. In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K.,

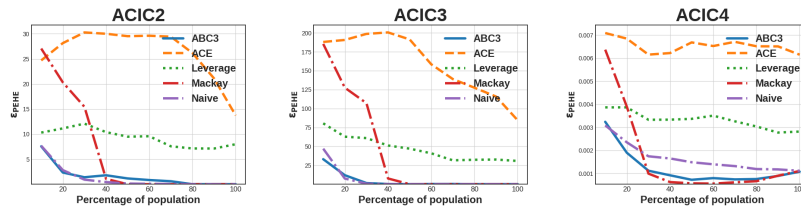


Figure 5:  $\epsilon_{PEHE}$  for all ACIC test data sets (cont.).

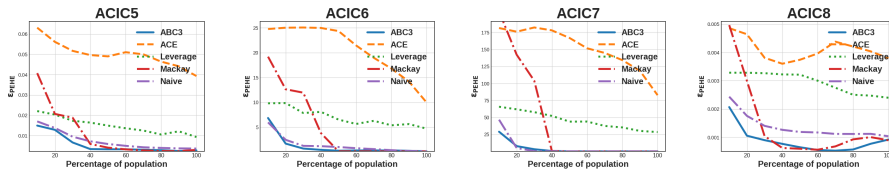


Figure 6:  $\epsilon_{PEHE}$  for all ACIC test data sets.

eds., *Advances in Neural Information Processing Systems*.

Breiman, L. 2001. Random forests. *Machine learning*, 45: 5–32.

Brooks-Gunn, J.; Liaw, F.-r.; and Klebanov, P. K. 1992. Effects of early intervention on cognitive function of low birth weight preterm infants. *The Journal of pediatrics*, 120(3): 350–359.

Budincsevy, N. 2016. Weather in Szeged 2006-2016. <https://www.kaggle.com/datasets/budincsevy/szeged-weather/data>. Accessed: 2024-08-08.

Cristianini, N.; and Scholkopf, B. 2002. Support vector machines and kernel methods: the new generation of learning machines. *Ai Magazine*, 23(3): 31–31.

Faddeev, D. K.; and Faddeeva, V. N. 1981. Computational methods of linear algebra. *Journal of Soviet Mathematics*, 15(5): 531–650.

Gruber, S.; Lefebvre, G.; Schuster, T.; ; and Piché, A. 2019. Atlantic Causal Inference Conference Data Challenge, 2019. <https://sites.google.com/view/acic2019datachallenge/>. Accessed: 2024-07-18.

Harrison, D.; and Rubinfeld, D. L. 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1): 81–102.

Hill, J. L. 2011. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, 20(1): 217–240.

Kanagawa, M.; Hennig, P.; Sejdinovic, D.; and Sriperumbudur, B. K. 2018. Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences. arXiv:1807.02582.

LaLonde, R. J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, 76(4): 604–620.