Electronic Journal of Statistics
Vol. 16 (2022) 3523–3560
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2025

Augmented direct learning for conditional average treatment effect estimation with double robustness

Haomiao Meng and Xingye Qiao

*Department of Mathematical Sciences,
Binghamton University, State University of New York,
Binghamton, NY 13902-6000, USA*
e-mail: hmeng3@binghamton.edu*;* qiao@math.binghamton.edu

Abstract: Inferring the heterogeneous treatment effect is a fundamental problem in many applications. In this paper, we focus on estimating the Conditional Average Treatment Effect (CATE), that is, the difference in the conditional mean outcome between treatments given covariates. Traditionally, Q-Learning based approaches estimate each conditional mean outcome. However, they are subject to model misspecification. Recently, flexible one-step methods to directly learn (D-Learning) the CATE without outcome model specifications have been proposed. However, they require a specification of the propensity score. We propose robust direct learning (RD-Learning), to augment D-learning, leading to doubly robust estimators of the treatment effect. The consistency for our CATE estimator is guaranteed if either the main effect model or the propensity score model is correctly specified. The framework can be used in both the binary and the multi-arm settings and is general enough to allow different function spaces and incorporate different generic learning algorithms. We conduct a thorough theoretical analysis of the prediction error of our CATE estimator using statistical learning theory under both linear and non-linear settings. The effectiveness of our proposed method is demonstrated by simulation studies and a real data example about an AIDS Clinical Trials study.

Keywords and phrases: Heterogeneous treatment effects, doubly robust estimator, multi-arm treatments, angle-based approach, statistical learning theory.

Received September 2021.

1. Introduction

Inferring the heterogeneous treatment effect is a fundamental problem in the sciences and commercial applications. Examples include studies on the effect of certain advertising or marketing efforts on consumer behavior [5], research on the effectiveness of public policies [47], and "A/B tests" in the context of tech companies for product development [44]. In particular, it can be useful in personalized medicine: based on many biomarkers, how can we determine which patients can potentially benefit from a treatment [36]?

Under the potential outcome framework [37, 18], we are interested in the comparison between the observed outcome and the counterfactual outcome we would have observed under a different regime or treatment. Assuming that there are two treatment arms ($\{+1, -1\}$), we want to estimate the difference in the conditional mean outcome between the two treatments, given the individual's pre-treatment covariates. This problem is typically known as Conditional Average Treatment Effect (CATE) estimation. CATE is also closely associated with the optimal individualized treatment rule (ITR). The latter maximizes the mean of a (clinical) outcome in a population of interest.

Traditional approaches for estimating CATE and the optimal ITR include Q-Learning ("Q" denoting "quality") [54, 31, 24] and A-Learning ("A" denoting "advantage") [25, 32]. Q-Learning estimates CATE through modeling the conditional mean outcome and taking their differences, and A-Learning estimates CATE by modeling the interaction between treatment and predictors based on a pre-estimated propensity score. A-Learning is known to be robust against the misspecification of the baseline mean function, given the propensity score model is correctly specified.

Recently, there is a growing literature on using machine learning for CATE or ITR estimation, including regression trees [1, 43], random forests [49], boosting [28], neural nets [19], Bayesian machine learning [9, 16, 44, 14], and combinations of the above [23]. Besides those under the Q-Learning framework, many of these methods can be categorized to *modified outcome methods* and *modified covariate methods*, following the categorization of [21]. *Modified outcome methods* were proposed in the Ph.D. thesis of James Signorovitch, Harvard University [41] in the randomized experimental setting. In the observational data setting, the inverse probability weighted (IPW) estimator can be used to modify the outcome. Similar modified outcome approaches have been proposed which allows directly using off-the-shelf machine learning algorithms for CATE or ITR estimation [3, 11, 55]. A drawback of the IPW estimator is that its performance hinges upon accurate estimation of the propensity score. The doubly robust (DR) augmented inverse probability weighted (AIPW) approach [33, 2] was formulated by [56]. It requires an estimation of the treatment propensity score and the conditional mean outcome given treatment and covariates. The double robustness of this method is well studied: as long as the model for either the conditional mean outcome or the propensity score is correctly specified, the estimator is consistent. See more work on double robustness in [20, 6, 57, 63, 13, 62]. Many of these methods focus on estimating optimal ITRs instead of CATE.

The *modified covariate method* was introduced by [45] for the experimental setting and was later generalized to the observational setting by [8]. [30] proposed a variant of the modified covariate method for estimating the optimal ITR. Modified covariate methods do not require to specify any model of the main effect or the conditional mean outcome function and they directly estimate the CATE. We refer to them as the D-Learning ("D" for "direct"). While D-Learning models the treatment effect directly, it relies on an accurate estimate of the propensity score in the observational setting. [45] and [8] both described the possibility to increase the efficiency of their estimators. This efficiency augmentation variant replaces the outcome by the residual of the outcome less the conditional mean outcome function. Though such efficiency augmentation has been shown to work well in certain scenarios, a double robustness property has not yet been discovered among *modified covariate methods*. We are not aware of any further theoretical analyses of the statistical properties of this approach.

The first contribution of this paper is to propose an augmented form of the modified-covariate (D-Learning) method with a double robustness property for CATE estimation. Specifically, the consistency for our CATE estimation is guaranteed if either the main effect model or the propensity score model is correctly specified. In contrast, to achieve the double robustness in the AIPW literature, the outcome model for each of the k arms and the propensity score model need to be pre-estimated. The second goal of this paper is to generalize our method to the multi-arm case. [30, 29] discussed the D-Learning in the multi-arm setting, but their method mainly focused on ITR estimation (instead of CATE) and did not have a double robustness property. Thirdly, we provide a theoretical analysis of the convergence rate for the prediction error under the general augmented D-Learning framework (including D-learning as a special case), and justify the benefit of the augmentation under the two misspecification scenarios. As a byproduct, we propose an efficient estimator for the main effect under a special setting with known propensity score. Since our method can be viewed as the robustified version of the D-Learning, we call it RD-Learning.

One related result to our work can be found in [27] and [40], which we refer to as the R-Learning, and offer an optimization problem in the form of A-Learning ("R" refers to Robinson's transformation in [34], "residual", and "robust"). They modify both the outcome and the covariates, and offer some robustness protection. However, neither enjoys the double robustness property. When this paper was being written, we noticed a similar work done parallelly by [39] which generalized [51] to continuous treatments. These methods were motivated by [32] which was doubly robust against the treatment-free outcome model (not the main effect model) and the propensity score model. However, as noted in [51], their estimation equations were similar, but not identical, to those in [32].

The rest of the paper is organized as follows. In Section 2, we introduce some notations and preliminaries. We present the RD-Learning method in Section 3. Theoretical properties are studied in Section 4. In Section 5, we conduct simulation studies to validate the proposed method, followed by a real data example about an AIDS clinical trial in Section 6. Section 7 concludes the paper. All technical proofs are provided in the Appendix.

2. Notations and preliminaries

In the two-treatment arm setting, a patient, with pre-treatment covariates $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$, is assigned to treatment $A \in \mathcal{A} = \{1, -1\}$. Let $Y^*(j) \in \mathbb{R}$ be the potential outcome the patient would have got by receiving treatment $j \in \mathcal{A}$. The observed outcome is denoted by $Y = Y^*(A)$. Let $p_j(\boldsymbol{x}) = \mathbb{P}(A = j \mid \boldsymbol{X} = \boldsymbol{x})$ be the propensity score for treatment j. We assume the unconfoundedness [35] and common support.

Assumption 1. *For all $j \in \mathcal{A}$, $Y^*(j) \perp A \mid \boldsymbol{X}$ and $p_j(\boldsymbol{x}) \geq c$ for some $c \in (0,1)$.*

Let P be the distribution of the triplet $(\boldsymbol{X}, A, Y)$. The goal is to estimate the Conditional Average Treatment Effect (CATE), defined as

$$\mathbb{E}(Y^*(1) - Y^*(-1) \mid \boldsymbol{X} = \boldsymbol{x}),$$

based on an i.i.d. training sample $\{(\boldsymbol{x}_i, a_i, y_i)\}_{i=1}^n$ randomly drawn from P. Let the conditional mean outcome for treatment j be $\mu_j(\boldsymbol{x}) \triangleq \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}, A = j)$. For continuous outcomes, a common model to characterize the interaction between the treatment and covariates is

$$Y = m(\boldsymbol{X}) + A\delta(\boldsymbol{X}) + \epsilon, \quad \text{where } \mathbb{E}(\epsilon) = 0,\ \mathrm{Var}(\epsilon) = \sigma^2 < \infty, \tag{1}$$

in which the main effect function $m(\boldsymbol{x}) \triangleq [\mu_1(\boldsymbol{x}) + \mu_{-1}(\boldsymbol{x})]/2$, and the treatment effect $\delta(\boldsymbol{x}) \triangleq [\mu_1(\boldsymbol{x}) - \mu_{-1}(\boldsymbol{x})]/2$. Thus, to estimate CATE is equivalent to estimate $\delta(\boldsymbol{x})$. In this paper, we simply refer to $\delta(\boldsymbol{x})$ as the treatment effect.

The Q-Learning [31] approach to estimating $\delta(\boldsymbol{x})$ is to conduct regressions for $\mu_j(\boldsymbol{x})$ for each j. Q-Learning may be vulnerable to model misspecification of $m(\boldsymbol{x})$ and $\delta(\boldsymbol{x})$, especially when the function space is small.

[45] proposed a method without specifying the model for $m(\boldsymbol{x})$ under the experimental setting with $p_1(\boldsymbol{x}) = 1/2$. Based on the observation that

$$\mathbb{E}(AY \mid \boldsymbol{X} = \boldsymbol{x}) = \delta(\boldsymbol{x}),$$

one can estimate $\delta(\boldsymbol{x})$ directly by regressing the modified outcome AY on $\boldsymbol{X}$ using least square, or equivalently by regressing Y on the modified covariate $A\boldsymbol{X}$ (by noting that $A^2 \equiv 1$). This method was later generalized to observational studies [8, 30]. Specifically, using a linear model, $\delta(\boldsymbol{x})$ can be estimated by $\boldsymbol{x}\hat{\boldsymbol{\beta}}$ where

$$\hat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n \frac{1}{p_{a_i}(\boldsymbol{x}_i)} (a_i y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2. \tag{2}$$

One advantage of this approach is that it avoids misspecification of the main effect $m(\boldsymbol{x})$ by estimating $\delta(\boldsymbol{x})$ directly. Hence, it is named D-Learning [30]. However, existing consistency results for D-Learning assume that the propensity score $p_j(\boldsymbol{x})$ is known or at least correctly specified, which may be challenging in observational studies.

The AIPW estimator is a well-studied approach that can address the misspecification issue of the propensity score. It was first proposed to estimate the (unconditional) average treatment effect, $\Delta \triangleq \mathbb{E}(Y^*(1) - Y^*(0))$ using $\hat{\Delta} \triangleq \hat{\psi}_1 - \hat{\psi}_{-1}$, where

$$\hat{\psi}_j = \frac{1}{n} \sum_{i=1}^n \frac{y_i \mathbb{1}[a_i = j]}{\hat{p}_j(\boldsymbol{x}_i)} - \frac{\mathbb{1}[a_i = j] - \hat{p}_j(\boldsymbol{x}_i)}{\hat{p}_j(\boldsymbol{x}_i)} \hat{\mu}_j(\boldsymbol{x}_i) \text{ for } j \in \{1, -1\},$$

where $\hat{p}_j(\boldsymbol{x})$ is an estimator for $p_j(\boldsymbol{x})$ and $\hat{\mu}_j(\boldsymbol{x})$ is an estimator for $\mu_j(\boldsymbol{x})$. The AIPW estimator has a double robustness property in that $\hat{\Delta}$ is consistent if either $p_j(\boldsymbol{x})$ or $\mu_j(\boldsymbol{x})$ is correctly specified for each j. The main product of this paper is a D-Learning method for CATE estimation with a similar doubly robust property.

3. RD-learning

We first introduce our proposed doubly robust D-Learning (RD-Learning) approach in the binary case. We then generalize it to the multi-arm setting. Lastly, we propose a direct method to estimate the main effect model.

3.1. RD-learning in the binary case

Given a training sample $\{\boldsymbol{x}_i, a_i, y_i\}_{i=1}^n$, the RD-Learning method is based on an estimator for the propensity score $p_1(\boldsymbol{x})$, denoted by $\hat{p}_1(\boldsymbol{x})$, and an estimator for the main effect $m(\boldsymbol{x})$, denoted by $\hat{m}(\boldsymbol{x})$. If we consider linear modeling for the treatment effect, i.e., $\delta(\boldsymbol{x}) = \mathbf{x}^T\boldsymbol{\beta}$ with $\mathbf{x}_i \triangleq (1, \boldsymbol{x}_i^T)^T$, then RD-Learning estimator for $\boldsymbol{\beta}$ is obtained by solving

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\hat{p}_{a_i}(\boldsymbol{x}_i)}(y_i - \hat{m}(\boldsymbol{x}_i) - a_i\mathbf{x}_i^T\boldsymbol{\beta})^2, \tag{3}$$

The RD-Learning CATE estimator is then $\hat{\delta}(\boldsymbol{x}) = \mathbf{x}^T\hat{\boldsymbol{\beta}}$.

Comparing (3) with (2), the RD-Learning is an augmented version of D-Learning by replacing the outcome y_i in (2) with the residual $y_i - \hat{m}(\boldsymbol{x})$. In the literature, similar procedures in which the outcome is replaced by a certain residual have been proposed to improve the efficiency of the estimation. For example, R-Learning methods [40, 27] replaced the outcome y_i in the A-Learning framework [25, 32] by the residual $y_i - \hat{\Phi}(\boldsymbol{x}_i)$, where $\hat{\Phi}(\boldsymbol{x})$ is an estimator for the conditional mean outcome function $\Phi(\boldsymbol{x}) \triangleq \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})$. In the ITR literature, residual weighted learning [64, RWL] replaced the outcome y_i in the outcome weighted learning [61, OWL] by residual $y_i - \hat{m}(\boldsymbol{x}_i)$ with respect to the main effect estimator $\hat{m}(\boldsymbol{x}_i)$. In general, these procedures reduce the variance of the estimator. Moreover, it has been shown that in these works [40, 27, 64], the CATE estimators are still consistent even when $\hat{\Phi}(\boldsymbol{x})$ or $\hat{m}(\boldsymbol{x})$ is mis-specified, so long as the propensity score $p_1(\boldsymbol{x})$ is known or can be consistently estimated. In other words, they are robust against misspecification for $\Phi(\boldsymbol{x})$ or $m(\boldsymbol{x})$.

Although residual based approaches have been studied in the aforementioned work, it has not been thoroughly studied for modified-covariate (D-Learning) methods. Specifically, while it is commonly recognized that replacing the outcome by the residual can improve efficiency [45], it was unclear which residual should be used. In RD-Learning (3), we propose to use the residual with respect to the main effect instead of the conditional mean outcome $m(\boldsymbol{x})$. As will be shown below, the proposed RD-Learning estimator is not only robust to misspecification for $m(\boldsymbol{x})$, but also has an "double robustness" property similar to that of the AIPW estimator [56]. This is described in the following theorem.

Theorem 1. *Assume that Assumption 1 holds under model (1). Let $\tilde{p}_1(\boldsymbol{x})$ be a working model for the propensity score $p_1(\boldsymbol{x})$ with $0 < \tilde{p}_1(\boldsymbol{x}) < 1$ and $\tilde{m}(\boldsymbol{x})$ be a working model for the main effect $m(\boldsymbol{x})$. Then we have*

$$\delta \in \underset{f \in \{\mathcal{X} \to \mathbb{R}\}}{\operatorname{argmin}} \, \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{X})}(Y - \tilde{m}(\boldsymbol{X}) - Af(\boldsymbol{X}))^2\right]$$

if either $\tilde{p}_1(\boldsymbol{x}) = p_1(\boldsymbol{x})$ or $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$ almost surely.

Theorem 1 also holds when, the functions $\tilde{p}_1(\boldsymbol{x})$ and $\tilde{m}(\boldsymbol{x})$ are replaced by the limiting functions of estimators $\hat{p}(\boldsymbol{x})$ and $\hat{m}(\boldsymbol{x})$. This suggests that the empirical version of the above minimizer $\hat{\delta}(\boldsymbol{x})$ (whose definitions are given in (3) and in Section 3.2) will be consistent with $\delta(\boldsymbol{x})$ if either $\hat{p}_1(\boldsymbol{x})$ or $\hat{m}(\boldsymbol{x})$ is consistent. Compared to the sole robustness against the misspecification of $m(\boldsymbol{x})$, this estimator is robust against two types of model misspecification, with respect to both $p_1(\boldsymbol{x})$ and $m(\boldsymbol{x})$. We defer more discussions on theoretical properties of this "double robustness" property to Section 4.

Remark 1. The double robust property of our method is different from previous work. The traditional AIPW-based approaches [33, 2] require the estimation for each outcome model (e.g. $\mu_1(\boldsymbol{x})$ and $\mu_{-1}(\boldsymbol{x})$ in the binary case), while a recent work by [39] requires the specification of the treatment-free model, namely, $\mu_{-1}(\boldsymbol{x})$ only. The double robustness of our proposed method is with respect to the estimation of the **main effect** model, i.e., $m(\boldsymbol{x})$.

Because RD-Learning is essentially a weighted least square, it is well posited to be generalized for high-dimensional data using sparsity. For example, we may solve a LASSO problem,

$$\min_{\substack{\boldsymbol{\beta} \in \mathbb{R}^p \\ \beta_0 \in \mathbb{R}}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{p}_{a_i}(\boldsymbol{x}_i)} \left(y_i - \hat{m}(\boldsymbol{x}_i) - a_i(\boldsymbol{x}_i^T \boldsymbol{\beta} + \beta_0)\right)^2 + \lambda \|\boldsymbol{\beta}\|_1 \tag{4}$$

with the tuning parameter $\lambda > 0$. To adopt a richer model space, we could also consider a non-linear function form for $\delta(\boldsymbol{x})$. For example, we may solve a kernel ridge regression

$$\min_{\substack{\boldsymbol{\beta} \in \mathbb{R}^n \\ \beta_0 \in \mathbb{R}}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{p}_{a_i}(\boldsymbol{x}_i)} \left(y_i - \hat{m}(\boldsymbol{x}_i) - a_i(\boldsymbol{K}_i^T \boldsymbol{\beta} + \beta_0)\right)^2 + \lambda \boldsymbol{\beta}^T \mathbf{K} \boldsymbol{\beta},$$

where $\boldsymbol{K}_i$ is the ith column of the gram matrix $\mathbf{K} = (K(\boldsymbol{x}_i, \boldsymbol{x}_j))_{n \times n}$, with $K(\cdot, \cdot)$ a kernel function. Other non-linear regression models such as generalized additive model and gradient boosting can be also applied to the RD-Learning framework.

Figure 1 compares Q-Learning, D-Learning, and RD-Learning using two toy examples. In each example, the subpopulation treatment effect pattern plots (STEPP), a typical visualization method for exploring the heterogeneity of treatment effects [4], shows the relationship between the estimated CATEs and predictor x_1. Case I has a non-linear main effect and a linear treatment effect, while both the main effect and the treatment effect in Case II have a linear form. The true CATE is $\mu_1(\boldsymbol{x})-\mu_{-1}(\boldsymbol{x}) = -x_1$ in both cases. From Figure 1, it is clear that RD-Learning is a robust method. In particular, compared to Q-Learning, RD-Learning reduces bias in estimating CATE when the main effect tends to be mis-specified (Case I). Compared to D-Learning, RD-Learning reduces variance in both examples.

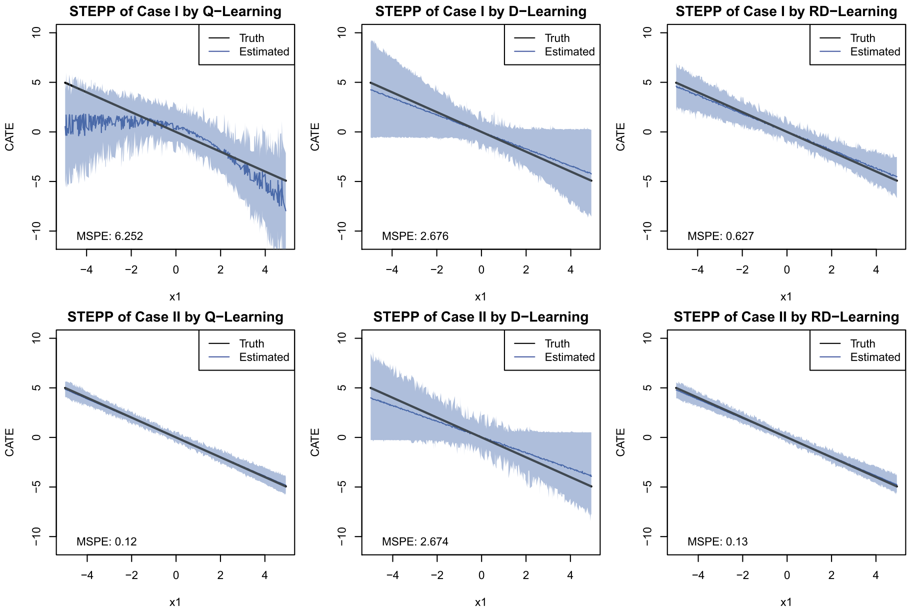

Fig 1. *Subpopulation treatment effect pattern plots (STEPP) by different methods on two simulated data with $\mathcal{X} \subseteq \mathbb{R}^{20}$. Blue regions are 95% confidence region based on 200 replications, and the black line is the true CATE. Case I: $\mu_j(\boldsymbol{x}) = 2\cos(x_1+\pi/4)+(3-j)x_1/2-\tanh(x_1)$ for $j \in \{1,-1\}$ and $p_1(\boldsymbol{x}) = 0.2+0.6\mathbb{1}[x_1<0]$; Case II: $\mu_j(\boldsymbol{x}) = (3-j)x_1/2+x_2$ for $j \in \{1,-1\}$ and $p_1(\boldsymbol{x}) = 1/2$. In both cases RD-Learning has a good performance.*

3.2. RD-learning in the multi-arm case

We now generalize RD-Learning to the multi-arm case with $\mathcal{A} = \{1, \ldots, k\}$. Consider

$$Y = m(\boldsymbol{X}) + \delta_A(\boldsymbol{X}) + \epsilon, \tag{5}$$

$$\text{where } \sum_{j=1}^{k} \delta_j(\boldsymbol{x}) = 0,\ \mathbb{E}(\epsilon) = 0,\ \mathrm{Var}(\epsilon|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}) < \infty.$$

Similar to the binary case, $m(\boldsymbol{x}) \triangleq \sum_{j=1}^{k} \mu_j(\boldsymbol{x})/k$ is the main effect. The jth treatment effect $\delta_j(\cdot)$ measures the difference between the conditional mean outcome of treatment j and the main effect, i.e., $\delta_j(\boldsymbol{x}) = \mu_j(\boldsymbol{x}) - m(\boldsymbol{x})$. The sum-to-zero constraint guarantees the model is identifiable. We allow heteroskedasticity in the error term ϵ to make the model more general. To estimate the treatment effect $\delta_j(\boldsymbol{x})$, we consider the following angle-based approach.

The angle-based approach [59] is a method used in multicategory classification problem and recently it has been used to solve a multi-arm ITR problem [58, 29]. In the angle-based framework, we represent k arms by k vertices of a $(k-1)$-dimensional simplex, denoted by $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_k$: $\boldsymbol{W}_1 = (k-1)^{-1/2}\mathbf{1}_{k-1}$ and $\boldsymbol{W}_j = -(1+k^{1/2})(k-1)^{-3/2}\mathbf{1}_{k-1} + [k/(k-1)]^{1/2}\boldsymbol{e}_{j-1}$ for $2 \le j \le k$, where $\mathbf{1}_{k-1}$ is a $(k-1)$-dimensional vector with all elements equal to 1 and $\boldsymbol{e}_{j-1} \in \mathbb{R}^{k-1}$ is a vector with the $(j-1)$th element 1 and 0 elsewhere. It is easy to check that $\|\boldsymbol{W}_j\| = 1$ and the angle $\angle(\boldsymbol{W}_i, \boldsymbol{W}_j)$ is the same for all $i \neq j$. The angle-based approach uses a $(k-1)$-dimensional vector-valued function $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_{k-1}(\boldsymbol{x}))^T$ as the decision function. Moreover, for any $\boldsymbol{f} \in \mathbb{R}^{k-1}$, $\sum_{j=1}^{k}\langle \boldsymbol{W}_j, \boldsymbol{f}\rangle = 0$; so the common sum-of-zero constraint in the multicategory classification problem is automatically satisfied. This will reduce the computational cost in the optimization problem. In an ITR problem, by computing the angles between $\boldsymbol{f}(\boldsymbol{x})$ and these vertices $\boldsymbol{W}_j$'s, the optimal treatment for $\boldsymbol{x}$ is chosen to be $\mathrm{argmin}_{j\in\mathcal{A}} \angle(\boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x}))$.

In a multi-arm CATE problem with model (5), there is also a sum-to-zero constraint for treatment effects $\{\delta_j(\boldsymbol{x})\}_{j=1}^{k}$, i.e., $\sum_{j=1}^{k} \delta_j(\boldsymbol{x}) = 0$. Then we may replace $\delta_j(\boldsymbol{x})$ in model (5) by $\langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x})\rangle$, where $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{k-1}$ such that $\langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x})\rangle = \delta_j(\boldsymbol{x})$ for all $j \in \mathcal{A}$. By design of the angle-based approach, the sum-to-zero constraint is implicitly satisfied. This allows us to estimate the treatment effect $\delta_j(\boldsymbol{x}) = \langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x})\rangle$ as follows.

Theorem 2. *Let $\tilde{p}_j(\boldsymbol{x}) > 0$ be a working model for $p_j(\boldsymbol{x})$ and $\tilde{m}(\boldsymbol{x})$ for $m(\boldsymbol{x})$. Define*

$$\boldsymbol{f}^* \in \underset{\boldsymbol{f}\in\{\mathcal{X}\to\mathbb{R}^{k-1}\}}{\mathrm{argmin}}\ \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{X})}(Y - \tilde{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle)^2\right].$$

Under Assumption 1 and model (5), if either $\tilde{p}_j(\boldsymbol{x}) = p_j(\boldsymbol{x})$ or $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x})$ holds for $\boldsymbol{x} \in \mathcal{X}$ almost surely and all $j \in \mathcal{A}$, then $\delta_j(\boldsymbol{x}) = \langle \boldsymbol{W}_j, \boldsymbol{f}^*(\boldsymbol{x})\rangle$ almost surely.

In light of Theorem 2, the angle-based RD-Learning[1] is obtained by solving

$$\min_{\boldsymbol{f}\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\hat{p}_{a_i}(\boldsymbol{x}_i)}(y_i - \hat{m}(\boldsymbol{x}_i) - \langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle)^2, \tag{6}$$

[1] Angle-based D-Learning has been studied in [29] with a different formulation.

where $\mathcal{F}$ is a function space of interest. We may consider a linear space, $\mathcal{F} = \{\boldsymbol{f} = (f_1, \dots, f_{k-1})^T; f_j(\boldsymbol{x}) = \mathbf{x}^T \boldsymbol{\beta}_j, j = 1, \dots, k-1\}$; or we may consider Reproducing Kernel Hilbert Space (RKHS) with kernel function $K(\cdot, \cdot)$, $\mathcal{F} = \{\boldsymbol{f} = (f_1, \dots, f_{k-1})^T; f_j(\boldsymbol{x}) = \sum_{i=1}^n K(\boldsymbol{x}_i, \boldsymbol{x})\beta_{ij} + \beta_{0j}, j = 1, \dots, k-1\}$. A regularization term on $\boldsymbol{f}$ can be also added to prevent overfitting. Denote the solution to (6) by $\hat{\boldsymbol{f}}$. Then the angle-based RD-Learning estimator for the jth treatment effect is given by

$$\hat{\delta}_j(\boldsymbol{x}) = \langle \boldsymbol{W}_j, \hat{\boldsymbol{f}}(\boldsymbol{x}) \rangle. \tag{7}$$

The binary RD-Learning in Section 3.1 is a special case of the angle-based RD-Learning. This is because $W_1 = 1$ and $W_2 = -1$ when $k = 2$; hence $\langle W_a, f(\boldsymbol{x}) \rangle = f(\boldsymbol{x})$ for $a = 1$ and $\langle W_a, f(\boldsymbol{x}) \rangle = -f(\boldsymbol{x})$ for $a = 2$, and (6) reduces to (3) for the linear function space.

3.3. A direct method in estimating the main effect

The RD-Learning framework we proposed is a two-step procedure. The discussion so far focuses on the second step, i.e., estimating the treatment effect $\delta_j(\boldsymbol{x})$ provided that the main effect and the propensity score have been estimated in the first step. In practice how to estimate the main effect $m(\boldsymbol{x})$ is also important. As will be shown in Section 4, an accurate $\hat{m}(\boldsymbol{x})$ estimation can reduces the variance of $\hat{\delta}_j(\boldsymbol{x})$. Moreover, the problem of main effect estimation has its own interest in many applications. For example, in biomedical studies, it can help researchers to identify prognostic biomarkers [22].

Here we propose a direct method to estimate the main effect given the propensity score,

$$\hat{m} = \underset{g \in \mathcal{G}}{\operatorname{argmin}} \sum_{i=1}^n \frac{1}{p_{a_i}(\boldsymbol{x}_i)} (y_i - g(\boldsymbol{x}_i))^2, \tag{8}$$

where $\mathcal{G}$ is an appropriate function space. This estimator is motivated by the important observation that under model (5),

$$\begin{aligned} \mathbb{E}\left[\frac{(Y - g(\boldsymbol{x}))^2}{p_A(\boldsymbol{x})} \middle| \boldsymbol{X} = \boldsymbol{x} \right] &= \mathbb{E}\left[\sum_{j=1}^k (Y^*(j) - g(\boldsymbol{x}))^2 \middle| \boldsymbol{X} = \boldsymbol{x} \right] \\ &= \sum_{j=1}^k (\mu_j(\boldsymbol{x}) - g(\boldsymbol{x}))^2 + k\sigma^2(\boldsymbol{x}), \end{aligned}$$

and the fact that $m(\boldsymbol{x}) = k^{-1} \sum_{j=1}^k \mu_j(\boldsymbol{x}) = \operatorname{argmin}_{g(\boldsymbol{x}) \in \mathbb{R}} \sum_{j=1}^k (\mu_j(\boldsymbol{x}) - g(\boldsymbol{x}))^2$.

Theorem 3. *Under Assumption 1, we have*

$$m \in \underset{g \in \{\mathcal{X} \to \mathbb{R}\}}{\operatorname{argmin}} \mathbb{E}\left[\frac{1}{p_A(\boldsymbol{X})} (Y - g(\boldsymbol{X}))^2 \right].$$

Theorem 3 implies that this estimator is consistent if $m \in \mathcal{G}$. Compared to the naive approach of estimating each $\mu_j(\boldsymbol{x})$ separately then taking the average to estimate $m(\cdot)$, (8) uses all the data units all at once to estimate the main effect. Due to its weighted least square form, it can be easily generalized to other regression methods, such as LASSO, kernel ridge regression, generalized additive model, gradient boosting, and so on.

A limitation of the proposed method to estimate the main effect is worth mentioning. Similar to D-Learning, this method is sensitive to the weight $p_{a_i}^{-1}(\boldsymbol{x}_i)$. The variance of the estimator will be large if the propensity score $p_{a_i}(\boldsymbol{x}_i)$ is close to zero. Furthermore, in an observational study where we have to estimate $p_{a_i}(\boldsymbol{x}_i)$, the accuracy will be low if there is a large bias to the estimation of $p_{a_i}(\boldsymbol{x}_i)$. While we find our direct method to be promising (see our simulation in Section 5), in practice, we suggest the user to try several different methods and compare when estimating the main effect.

4. Theoretical analysis of RD-learning

In this section, we study the theoretical property of $\hat{\delta}_j(\boldsymbol{x})$ in (7). Note that it suffices to study the angle-based RD-Learning only since the binary RD-Learning is a special case of the angle-based RD-Learning. Denote $\hat{\boldsymbol{\delta}} = (\hat{\delta}_1, \ldots, \hat{\delta}_k)^T$. The goal of our theoretical study is to obtain the convergence rate for the prediction error (PE) of $\hat{\boldsymbol{\delta}}$, defined by

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) = \|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|_2^2 = \mathbb{E}\sum_{j=1}^{k}\left(\hat{\delta}_j(\boldsymbol{X}) - \delta_j(\boldsymbol{X})\right)^2,$$

where the expectation is with respect to $\boldsymbol{X}$. Note that since $\hat{\boldsymbol{\delta}}$ depends on the training data, $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ is a random quantity. We consider linear models and non-linear models separately. Before we present the main results, we make two additional assumptions for the two estimators $\hat{p}_j(\boldsymbol{x})$ and $\hat{m}(\boldsymbol{x})$.

Assumption 2. $\|\hat{p}_j^{-1}(\boldsymbol{x}) - p_j^{-1}(\boldsymbol{x})\|_\infty \le r_p$ *with constant* $r_p > 0$.

Assumption 3. $\|\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x})\|_\infty \le r_m$ *and* $|Y - \hat{m}(\boldsymbol{X})| \le C_m$ *with* $r_m > 0$ *and* $C_m > 0$.

Assumptions 2 and 3 state that the estimation error for $\hat{p}_j^{-1}(\boldsymbol{x})$ and $\hat{m}(\boldsymbol{x})$ are bounded with r_p and r_m characterizing the accuracy for both estimators. Recall that $\tilde{p}_j(\boldsymbol{x})$ and $\tilde{m}(\boldsymbol{x})$ are the limiting functions of $\hat{p}_j(\boldsymbol{x})$ and $\hat{m}(\boldsymbol{x})$ in Theorem 2. The case of $\tilde{p}_j(\boldsymbol{x}) = p_j(\boldsymbol{x})$ corresponds to $r_p \ll r_m$; the case of $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x})$ corresponds to $r_m \ll r_p$.

Remark 2. In the literature, many L^2 type error bounds of machine learning and statistical nonparametric methods have been established. The L^∞ error bound in Assumption 2 and 3 can be relaxed to L^2 error bounds (though in Assumption 2 we will need an additional L^∞ bound on $\hat{p}_j^{-1}(\boldsymbol{x})$). Along with Assumption 1, one can show that $\|\hat{p}_j^{-1}(\boldsymbol{x}) - p_j^{-1}(\boldsymbol{x})\|_\infty < \infty$ which is denoted as r_p

in the subsequent theorems. The inequality $|Y - \hat{m}(\boldsymbol{X})| \leq C_m$ in Assumption 3 is a hard bound that is assumed to hold everywhere. It may be weakened to be (1) $E|Y - \hat{m}(\boldsymbol{X})|^2 \leq C_m$ and (2) $|y_i - \hat{m}(\boldsymbol{x}_i)| \leq C_m$ for $i = 1, \ldots, n$. We have chosen the current form for brevity of the presentation.

4.1. Linear function space

We consider a linear function space $\mathcal{F}$ with a bounded L^1 norm:

$$\mathcal{F} = \mathcal{F}(p, s) \triangleq \{\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T; f_j(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}_j + \beta_{0j}, \\ j = 1, \ldots, k-1, \sum_{j=1}^{k-1} \|\boldsymbol{\beta}_j\|_1 \leq s\}.$$

We bound each covariate in $[-1, 1]$ for simplicity.

Assumption 4. $\boldsymbol{X} \in \mathcal{X} = [-1, 1]^p$.

The result still holds if we bound each covariate in $[-B, B]$ for any large number $B > 0$.

Theorem 4. *Denote p_n be the data dimension. Let $\mathcal{F} = \mathcal{F}(p_n, s_n)$ and $\tau_n = (n^{-1}\log p_n)^{1/2} \to 0$ as $n \to \infty$. Under Assumptions 1 to 4, we have*

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(\max\left\{(C_m + s_n)^2\tau_n \log \tau_n^{-1}, \min\{r_1, r_2\}, d_n\right\}\right),$$

almost surely, given estimators $\hat{p}_j(\cdot)$ and $\hat{m}(\cdot)$, where $r_1 = (C_m + s_n)^2 r_p$, $r_2 = (1 + r_p)r_m^2$, and $d_n = \inf_{\boldsymbol{f}\in\mathcal{F}(p_n,s_n)} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

Remark 3. Theorem 4 claims that $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ is determined by three terms. The first term is the estimation error similar to the excess risk in the classification literature. As n grows, the term will vanish for fixed s_n, while for fixed n and p_n, it increases as $s_n \to \infty$ indicating a more complicated function space. The second term is determined by the accuracy of the two preliminary estimators $\hat{p}_j(\cdot)$ and $\hat{m}(\cdot)$. Specifically, r_1 describes the error from $\hat{p}_j(\boldsymbol{x})$ while r_2 describes the error from $\hat{m}(\boldsymbol{x})$. This term is small as long as either r_p or r_m is small. Hence, this term reflects the "double robustness" property of the proposed estimator. The third term d_n is the approximation error of the function space $\mathcal{F}(p_n, s_n)$, and it will decrease as s_n increases in general. The choice of s_n represents a trade-off between the three terms.

Remark 4. By Theorem 4, RD-Learning improves D-Learning in the following two aspects. Firstly, the second term in the upper bound of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ offers an additional way to decrease the error. Note that D-Learning is a special case of RD-Learning with $\hat{m} \equiv 0$, which means r_2 is a large number. Therefore, for D-Learning to work well, r_1 must be small. In contrast, RD-Learning offers a good CATE as long as either r_1 or r_2 is small. Secondly, the estimator of RD-Learning has a smaller variance than that of D-Learning. This is because by replacing

y_i in D-Learning with $y_i - \hat{m}(\boldsymbol{x}_i)$, the upper bound C_m for $|Y - \hat{m}(\boldsymbol{X})|$ in Assumption 3 also becomes smaller in general, which further reduces the first term in $\mathrm{PE}(\hat{\boldsymbol{\delta}})$. This explains the narrower confidence bands of RD-Learning shown in Figure 1.

Theorem 4 is a general statement for the convergence rate of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$. It neither makes assumptions on the magnitude of r_p and r_m, nor assumes the true treatment effect falls in a particular function space $\mathcal{F}$. If we assume one of r_p and r_m is zero, the second term can be ignored. For example, in clinical trial $p_j(\boldsymbol{x})$ is known so $\hat{p}_j(\boldsymbol{x}) = p_j(\boldsymbol{x})$ and $r_p = 0$. If we further assume $\delta_j(\boldsymbol{x})$ to be a linear function that only depends on finitely many covariates for each j, the third term can be also eliminated. This is because in that case, there exists a finite p^* and s^* such that the true population minimizer $\boldsymbol{f}^*$ belongs to the $\mathcal{F}(p_n, s_n)$ as long as the function space we consider is large enough so that $p_n \geq p^*$ and $s_n \geq s^*$. In this case, the third term $d_n = 0$ for sufficient large n. The result is given in Corollary 1.

Corollary 1. *Let $\mathcal{F} = \mathcal{F}(p_n, s_n)$ and $\tau_n = (n^{-1}\log p_n)^{1/2} \to 0$ as $n \to \infty$. Suppose $\delta_j(\cdot)$ depends on finitely many covariates for each $j \in \mathcal{A}$. Under Assumptions 1 to 4, if either $r_p = 0$ or $r_m = 0$ holds, then $\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}(\tau_n \log \tau_n^{-1})$ almost surely.*

From Corollary 1, we first observe that the convergence of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ requires that p_n increases with the order at most $\exp(n)$. Secondly, since $\mathcal{O}(\log x) < \mathcal{O}(x^t)$ for all $t > 0$, $\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}(\tau_n^{1-t})$ for any small positive t. This implies that the upper bound of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ is almost $\mathcal{O}(\tau_n) = \mathcal{O}\left((n^{-1}\log p_n)^{1/2}\right)$. Furthermore, when p_n is a fixed number, i.e., $p_n = \mathcal{O}(1)$, the rate is almost $\mathcal{O}(n^{-1/2})$. These results are coincident with most of the classical LASSO theory.

4.2. Reproducing kernel Hilbert space

We consider $\mathcal{F}$ to be a Reproducing Kernel Hilbert Space (RKHS) to demonstrate the results for non-linear learning. There is a vast literature on RKHS. See [42] and [17] for more details. Let $\mathcal{H}_K$ be a RKHS with kernel function $K(\cdot,\cdot)$. By the Mercer's theorem, K has an eigen-expansion $K(\boldsymbol{x},\boldsymbol{x}') = \sum_{i=1}^{\infty}\gamma_i\phi_i(\boldsymbol{x})\phi_i(\boldsymbol{x}')$ with $\gamma_i \geq 0$ and $\sum_{i=1}^{\infty}\gamma_i^2 < \infty$. Any function in $\mathcal{H}_K$ can be written as $f(\boldsymbol{x}) = \sum_{i=1}^{\infty} c_i\phi_i(\boldsymbol{x})$ under the constraint that $\|f\|_{\mathcal{H}_K}^2 = \sum_{i=1}^{\infty} c_i^2/\gamma_i < \infty$. Define the function space $\mathcal{F}$ as

$$\mathcal{F} = \mathcal{F}(s) \triangleq \{\boldsymbol{f} = (f_1,\ldots,f_{k-1})^T; f_j = f_j' + b_j, \\ j = 1,\ldots,k-1, \sum_{j=1}^{k-1}\|f_j'\|_{\mathcal{H}_K}^2 \leq s^2\}.$$

Note that as in the linear case, the penalty term does not include the intercept term b_j. Rewrite the solution to (6) under such $\mathcal{F}$ as $\hat{\boldsymbol{f}} = \hat{\boldsymbol{f}}' + \hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{f}}' = (\hat{f}_1',\ldots,\hat{f}_k')^T$ with $f_j' \in \mathcal{H}_K$. By the representer theorem [50], $\hat{f}_j'$ can be represented by $\hat{f}'_j(\boldsymbol{x}) = \sum_{i=1}^n K(\boldsymbol{x}_i, \boldsymbol{x})\hat{\beta}_{ij}$, and the penalty term is written as $\|\hat{f}'_j\|^2_{\mathcal{H}_K} = \sum_{i=1}^n \sum_{l=1}^n K(\boldsymbol{x}_i, \boldsymbol{x}_l)\hat{\beta}_{ij}\hat{\beta}_{lj}$.

The separability of the RKHS is commonly assumed in many papers concerning RKHS.

Assumption 5. *The RKHS $\mathcal{H}_K$ is separable and $\sup_{\boldsymbol{x}} K(\boldsymbol{x}, \boldsymbol{x}) = B < \infty$.*

A bounded kernel ensures that $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ does not explode. It naturally holds for popular kernels like the Gaussian radial basis kernel, where $B = 1$. In general, it requires $\mathcal{X}$ to be covered by a compact set and the dimension of $\boldsymbol{X}$ to be finite.

Theorem 5. *Let $\mathcal{F} = \mathcal{F}(s_n)$. Under Assumptions 1, 2, 3, and 5, we have*

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(\max\left\{(C_m + Bs_n)^2 n^{-1/2}\log n, \min\{r_1, r_2\}, d_n\right\}\right),$$

almost surely, given estimators $\hat{p}_j(\cdot)$ and $\hat{m}(\cdot)$, where $r_1 = (C_m + Bs_n)^2 r_p$, $r_2 = (1 + r_p)r_m^2$, and $d_n = \inf_{\boldsymbol{f}\in\mathcal{F}(s_n)} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

Remark 5. Similar to Theorem 4, there is a trade-off between the estimation error, the approximation error, and the error from $\hat{p}_j$ and $\hat{m}$ for kernel learning. s_n is the tuning parameter to balance these three terms. The result also shows that compared to D-Learning, RD-Learning still enjoys a better convergence rate through a smaller r_m and C_m.

Theorem 5 can be simplified in some special cases. Firstly, the second term can be ignored when r_p or r_m is negligible. Secondly, by assuming the approximation error $d_n \leq \mathcal{O}(s_n^{-q})$ for some $q > 0$, which is standard in the RKHS literature, we have a neat convergence rate by appropriately choosing s_n, shown in Corollary 2.

Corollary 2. *Let $\mathcal{F} = \mathcal{F}(s_n)$. Suppose $d_n = \inf_{\boldsymbol{f}\in\mathcal{F}(s_n)} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2 \leq \mathcal{O}(s_n^{-q})$ for some $q > 0$. Under Assumption 1, 2, 3, and 5, if either $r_p = 0$ or $r_m = 0$ holds, then by choosing $s_n = \mathcal{O}\left((n^{1/2}\log^{-1} n)^{\frac{1}{q+2}}\right)$, we have*

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(n^{-\frac{q}{2q+5}}\right)$$

almost surely.

Given Corollary 2, the convergence rate of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ approaches to $\mathcal{O}\left(n^{-1/2}\right)$ for sufficiently large q, corresponding to the case that $\boldsymbol{f}^*$ is well approximated by a function in $\mathcal{F}(s_n)$.

5. Simulation studies

We compare the proposed RD-Learning method with four popular competing methods, Q-Learning [31], R-Learning [40, 27, R-Learning], causal forests [49], and D-Learning [8, 30]. Except for Q-Learning and D-Learning, all methods are preceded by first estimating either the main effect function $m(\boldsymbol{x})$ or the conditional mean outcome function $\Phi(\boldsymbol{x})$. We fix the dimension to be 100, where X_1, X_2 and X_3 are i.i.d. from $N(0,3)$, and $X_4,\ldots,X_{100}$ are i.i.d. from Uniform$(0,1)$. For each setting, we let the number of observations to be $n=50$, 100, 150, and 200. The parameters in each case are tuned through cross validation. The prediction error is reported based on a testing set of size 400.

Case I: It is a two-arm design, with $\mu_1(\boldsymbol{x})=2\cos(x_1+\pi/4)+x_1-\tanh(x_2)$ and $\mu_{-1}(\boldsymbol{x})=2\cos(x_1+\pi/4)+2x_1-\tanh(x_2)$. The treatment assignment depends on $\boldsymbol{x}$. Specifically, $p_1(\boldsymbol{x})=0.2+0.6\mathbb{1}[x_1<0]$. Since $\mu_1(\cdot)$, $\mu_{-1}(\cdot)$ and $m(\cdot)$ are non-linear functions of $\boldsymbol{x}$, we consider kernel machines in estimating $\mu_j(\boldsymbol{x})$ in Q-Learning, and in estimating $p_j(\boldsymbol{x})$, $\mu_j(\boldsymbol{x})$, or $\Phi(\boldsymbol{x})$ in the preliminary step of causal forests, R-Learning, and RD-Learning. On the other hand, because the treatment effect is linear, we use linear models with an L_1 penalty in D-Learning as well as in the main procedure of R-Learning and RD-Learning.[2]

Case II: This is an example to test the robustness against misspecification of the main effect. In this case, we have $\mu_1(\boldsymbol{x})=\tanh(x_1)-4/(1+\exp(x_2-x_1))+3$ and $\mu_{-1}(\boldsymbol{x})=\tanh(x_1)+4/(1+\exp(x_2-x_1))$. It is a randomized design with $p_1(\boldsymbol{x})=1/5$. Both the main effect and the treatment effect are non-linear. Hence we are supposed to use non-linear function spaces for all the methods. However, to test the robustness of the proposed RD-Learning method, we deliberately use the (wrong) linear model with an L_1 penalty to estimate the main effect in the first step and use kernel ridge regression in the second step. For comparison purposes, we adopt the same function spaces (linear and kernel) in all the other two-step procedures, and use kernel ridge regression in Q-Learning and D-Learning.

Case III: This is an example to test the robustness against misspecification of the propensity score. In this example, $\mu_1(\boldsymbol{x})=x_1-x_2+x_3$ and $\mu_{-1}(\boldsymbol{x})=2x_1-x_2$. The propensity score is defined as $p_1(\boldsymbol{x})=2/(2+\exp(x_1))$. In this case, we use linear models with an L_1 penalty in all methods and both steps. To test the robustness of RD-Learning, we deliberately use a wrong propensity score $\hat{p}_1(\boldsymbol{x})=1/2$. For comparison, we let $\hat{p}_1(\boldsymbol{x})=1/2$ in other methods.

Case IV: This is a three-arm case, with $\mu_1(\boldsymbol{x})=(x_1^2+x_2^2+x_3^2)/3+x_1-x_2$, $\mu_2(\boldsymbol{x})=(x_1^2+x_2^2+x_3^2)/3+x_2-x_3$, and $\mu_3(\boldsymbol{x})=(x_1^2+x_2^2+x_3^2)/3+x_3-x_1$. The propensity scores are $(p_1(\boldsymbol{x}),p_2(\boldsymbol{x}),p_3(\boldsymbol{x}))=(1/2,1/4,1/4)$ for $x_1\geq x_2$ and $x_1\geq x_3$, $(1/4,1/2,1/4)$, for $x_2>x_1$ and $x_2\geq x_3$, and $(1/4,1/4,1/2)$, for $x_3>x_2$ and $x_3>x_1$. This setting is similar to Case I in the sense that it has a non-linear main effect and a linear treatment effect. We use the same function space as in Case I. We do not report the results by causal forests and R-Learning because their current implementations for the multi-arm case report the estimators for the contrasts $\mu_j(\boldsymbol{x})-\mu_1(\boldsymbol{x})$ instead of $\delta_j(\boldsymbol{x})$.

Estimation of the treatment effect

From Figure 2, we first see that the RD-Learning method has the smallest prediction error in most scenarios. Secondly, Q-Learning and D-Learning

[2] By default, the second step of causal forests uses a regression tree which is by nature non-linear.

typically have a larger variance than those two-step procedures. This is consistent with the well known intuition (see also Theorems 4 and 5) that by replacing y_i with $y_i - \hat{m}(\boldsymbol{x}_i)$ (or $y_i - \hat{\Phi}(\boldsymbol{x}_i)$), the variance of estimators can be reduced. Thirdly, we see that RD-Learning is indeed "doubly robust" against misspecification of the main effect (in Case II) and the propensity score (in Case III).

Recall that Case II is an example where we deliberately use a wrong function space for the main effect. Since R-Learning is robust against this kind of misspecification, it has a better performance than Q-Learning. However, in Case III where we deliberately use a wrong propensity score, R-Learning has a much worse performance than Q-Learning since it relies on a correctly-specified $p_j(\cdot)$. But RD-Learning is as good as, and in many cases much better, than these two in both settings.

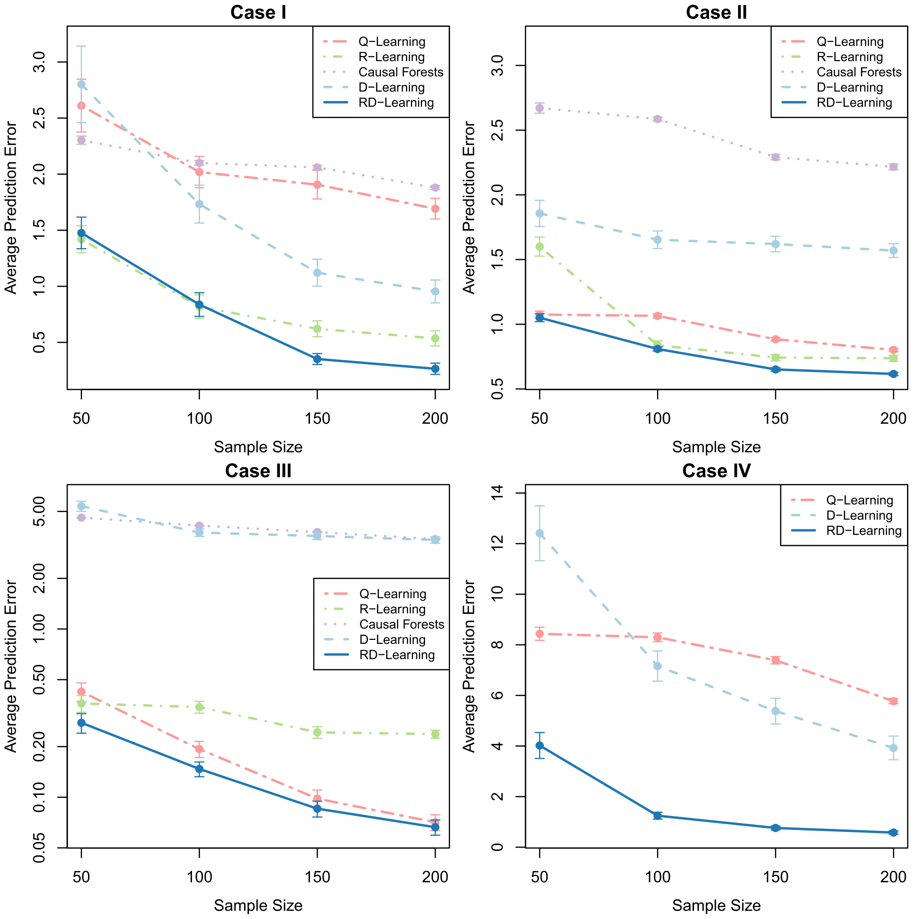

FIG 2. *The average prediction error of $\hat{\delta}_1$ based on 200 replications with standard error by different methods. In all cases RD-Learning has the best performance.*

Estimation of the main effect

We also report the performance for the main effect estimation using the proposed direct method in Section 3.3 and the Q-Learning method which estimates each $\mu_j(\cdot)$ and takes the average. Figure 3 shows the result based on the same simulation data in Case I and Case IV. We observe that by using all the data at once with propensity score as the weight, the proposed method has a better performance compared to the Q-Learning method.

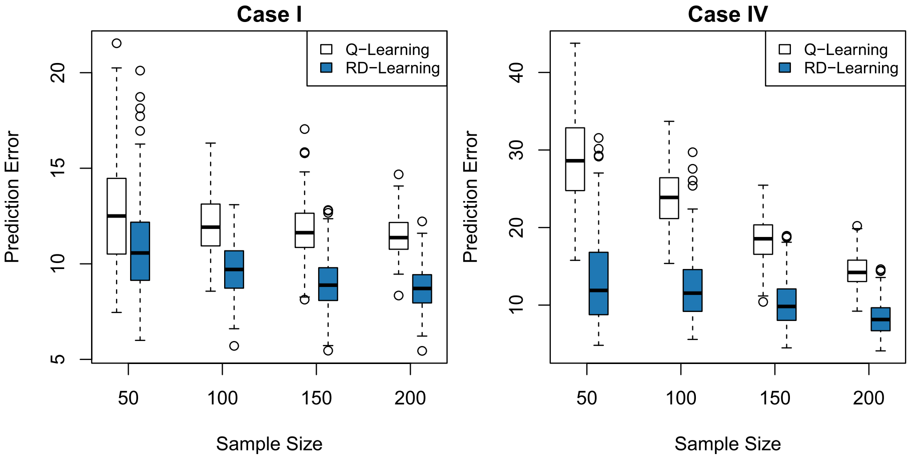

FIG 3. *Boxplots for the prediction error of $\hat{m}$ based on 200 replications in Case I (left) and Case IV (right). The proposed method has a smaller error than Q-Learning in estimating the main effect.*

6. Real data analysis

We apply RD-Learning on a real dataset from the AIDS Clinical Trials Group Study 175 [15, ACTG175]. The dataset includes 2,139 HIV-1 infected subjects. They were randomly assigned with equal probabilities to one of the four treatments: zidovudine (ZDV) only, ZDV with didanosine (ddI), ZDV with zalcitabine (ddC), and ddI only. The endpoint (outcome) we consider is the change of the CD4 cell count (per cubic millimeter) at 20 ± 5 weeks from the baseline. Note that a decrease in the number of CD4 cell count usually implies a progression to AIDS. In other words, a larger value indicates a better outcome.

To apply the RD-Learning method, we first estimate the main effect using the direct estimator proposed in Section 3.3 based on the 18 variables that were measured prior to the initiation of the study. Specifically, we use the generalized additive model (GAM) to solve the weighted least square problem (8). The best GAM model is selected using stepwise AIC.

For the main procedure, we follow the analysis of [12] and [29] and consider only 12 variables measured at baseline as the covariates for each subject. Five of the 12 covariates are continuous: age (years), weight (kilogram), Karnofsky score (on a scale of 0-100), CD4 cell counts (per cubic millimeter), and CD8 cell counts (per cubic millimeter). The rest seven are binary: hemophilia (0=no, 1=yes), homosexual activity (0=no, 1=yes), history of intravenous drug use (0=no, 1=yes), race (0=white, 1=non-white), gender (0=female, 1=male), antiretroviral history (0=naive, 1=experienced), and symptomatic indicator (0=asymptomatic, 1=symptomatic). We use RD-Learning with linear $\boldsymbol{f}$ (6) to estimate the coefficients.

We compare the performance of RD-Learning with that of Q-Learning and D-Learning through 5-fold cross validation. However, since it is a real dataset in which the true treatment effect is not observed, instead of evaluating the prediction error, we first derive the estimated optimal ITR of each method by $\hat{d}(\boldsymbol{x}_i) = \text{argmax}_j \, \hat{\delta}_j(\boldsymbol{x}_i)$. Then we calculate the empirical expected outcome under the obtained ITR $\hat{d}$, defined as

$$V(\hat{d}) = \frac{\sum_{i=1}^n \left(y_i \mathbb{1}\left[a_i = \hat{d}(\boldsymbol{x}_i) \right] / p_{a_i}(\boldsymbol{x}_i) \right)}{\sum_{i=1}^n \left(\mathbb{1}\left[a_i = \hat{d}(\boldsymbol{x}_i) \right] / p_{a_i}(\boldsymbol{x}_i) \right)}$$

[26, 61]. Note that in this application $V(\hat{d})$ measures the average increase in CD4 cell counts (per cubic millimeter) by taking the recommended treatment. Larger value $V(\hat{d})$ is preferred. Finally, we replicate the procedure for 400 times.

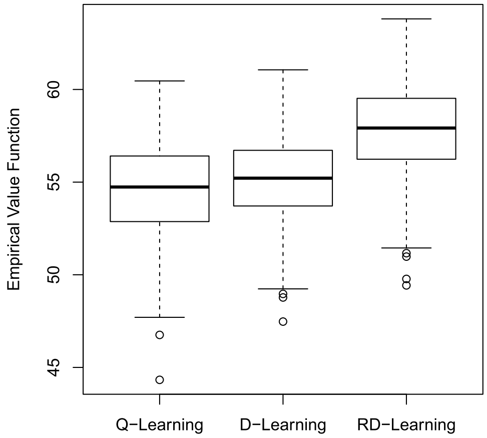

Fig 4. *5-fold cross validation scores of $V(\hat{d})$ based on 400 replications by different methods for the ACTG175 data. RD-Learning has the highest empirical value on average.*

The boxplot of $V(\hat{d})$ is shown in Figure 4. We see that RD-Learning yields the largest value, and $V(\hat{d})$ of D-Learning is slightly higher than that of Q-Learning. This implies that patients would benefit more by following the recommended treatment that is based on the treatment effect estimated by RD-Learning.

To visualize the heterogeneity in treatment effects, in Figure 5, we project the data on two important biomarkers "age" and "CD4 baseline" and mark each point according to its optimal treatment assignment estimated by RD-Learning. We see that the treatment ZDV is inferior to the other three treatments. This result is consistent with previous findings [15, 12, 29]. Furthermore, for the majority of the patients, ZDV with ddI is the best treatment. ZDV with ddC is most effective on young patients (age < 25), and ddI alone is better than the others for patients who have more CD4 cells (CD4 counts > 500 per cubic millimeter) at baseline.

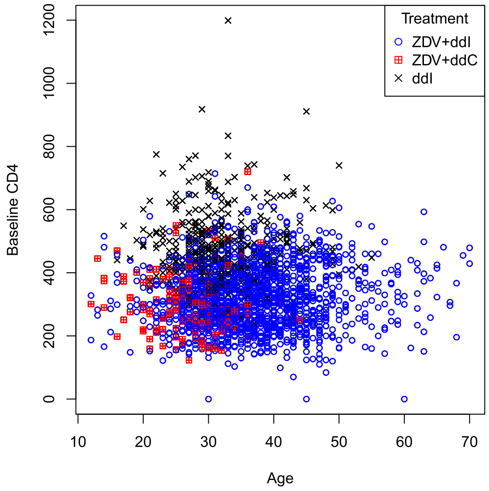

Fig 5. *ACTG175 data projected on "age" and "CD4 baseline", with the best treatment based on the estimated treatment effect by the RD-Learning marked by different colors and symbols.*

7. Conclusion

In this work, we propose RD-Learning to estimate CATE under both two-arm and multi-arm settings. The estimator is consistent if either the model for the main effect or the model for the propensity score is correctly specified. The proposed framework is flexible enough to incorporate existing generic procedures such as LASSO, kernel ridge regression, and generalized additive model. We also propose a direct estimation approach for the main effect when the propensity scores are known.

There are a few possible future research directions based on this work. Firstly, by modifying the quadratic loss function, the framework can be extended to other types of outcome, such as binary outcome and survival outcome. Secondly, one may want to improve our two-step procedure to a one-step method based on (6), i.e., estimating $p_j(\boldsymbol{x})$, $m(\boldsymbol{x})$, and $\delta_j(\boldsymbol{x})$ simultaneously. Such CATE estimator would still enjoy a doubly-robust property while the convergence rate of PE($\hat{\boldsymbol{\delta}}$) may be different from the proposed method in this paper. Thirdly, statistical inference based on RD-Learning can be investigated, so that in addition to doubly robust estimators, we may also have doubly robust confident regions. Finally the method can applied to dynamic treatment regime so that a sequence of treatment effects and treatment rules can be estimated robustly and sequentially.

Appendix A: Preliminaries on angle-based approach

In this section, we provide some basics for angle-based approach. One can refer to [59] for more details.

Recall that for a $(k-1)$-dimension simplex, we denote the vertex set as $\{\boldsymbol{W}_1,\dots,\boldsymbol{W}_k\}$. Specifically,

$$\boldsymbol{W}_j = \begin{cases} (k-1)^{-1/2}\mathbf{1}_{k-1}, & j=1 \\ -(1+k^{1/2})(k-1)^{-3/2}\mathbf{1}_{k-1} + [k/(k-1)]^{1/2}\,\boldsymbol{e}_{j-1}, & 2\le j\le k, \end{cases}$$

where $\mathbf{1}_{k-1}$ is a $(k-1)$-dimensional vector with all 1 and $\boldsymbol{e}_{j-1}\in\mathbb{R}^{k-1}$ is a vector with the $(j-1)$th element 1 and 0 elsewhere. By definition, it enjoys some nice properties.

Lemma 1. *A $(k-1)$-dimension simplex ($k\ge 2$) with vertex set $\{\boldsymbol{W}_1,\dots,\boldsymbol{W}_k\}$ has the following properties:*

1) $\|\boldsymbol{W}_j\|=1$ for any $j=1,\dots,k$.
2) $\langle\boldsymbol{W}_i,\boldsymbol{W}_j\rangle=\langle\boldsymbol{W}_{i'},\boldsymbol{W}_{j'}\rangle$ for any $i\ne j$ and $i'\ne j'$.
3) For any $\boldsymbol{f}\in\mathbb{R}^{k-1}$, $\sum_{j=1}^k\langle\boldsymbol{W}_j,\boldsymbol{f}\rangle=0$.
4) For any $\boldsymbol{f}\in\mathbb{R}^{k-1}$, $\sum_{j=1}^k\langle\boldsymbol{W}_j,\boldsymbol{f}\rangle^2=\frac{k}{k-1}\|\boldsymbol{f}\|^2$.
5) Any $\boldsymbol{f}\in\mathbb{R}^{k-1}$ can be uniquely represented by $\langle\boldsymbol{W}_j,\boldsymbol{f}\rangle$ for $j=1,\dots,k$. Specifically,

$$\boldsymbol{f}=\frac{k-1}{k}\sum_{j=1}^k\langle\boldsymbol{W}_j,\boldsymbol{f}\rangle\boldsymbol{W}_j.$$

Lemma 1 can be proved from the definition of $\boldsymbol{W}_j$ directly. From the last property in Lemma 1, for any $\boldsymbol{\delta}=(\delta_1,\dots,\delta_k)^T\in\mathbb{R}^k$ with $\sum_{j=1}^k\delta_j=0$, we can construct a function $\boldsymbol{f}\in\mathbb{R}^{k-1}$ by letting $\boldsymbol{f}=k^{-1}(k-1)\sum_{j=1}^k\delta_j\boldsymbol{W}_j$. On the other hand, for any $\boldsymbol{f}\in\mathbb{R}^{k-1}$, we can map it to a $\boldsymbol{\delta}\in\mathbb{R}^k$ by letting $\boldsymbol{\delta}=(\langle\boldsymbol{W}_1,\boldsymbol{f}\rangle,\dots,\langle\boldsymbol{W}_k,\boldsymbol{f}\rangle)^T$ and $\sum_{j=1}^k\langle\boldsymbol{W}_j,\boldsymbol{f}\rangle=0$. So there is a one-to-one map from $\mathbb{R}^{k-1}$ to the hyperplane $\{\boldsymbol{\delta}=(\delta_1,\dots,\delta_k)^T;\sum_{j=1}^k\delta_j=0\}$ in $\mathbb{R}^k$. This implies that one can use the $(k-1)$-dimension function $\boldsymbol{f}$ to represent **any** k-dimension treatment effect $\boldsymbol{\delta}$, which is also the intuition to generalize RD-Learning from the binary case (3) to the multi-arm case (6).

Appendix B: Fisher consistency

In this section, we prove Theorem 1, Theorem 2, and Theorem 3.

Proof of Theorem 1. Let $\ell(\boldsymbol{X}, \boldsymbol{f}) = \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{X})}(Y - \tilde{m}(\boldsymbol{X}) - Af(\boldsymbol{X}))^2 \;\middle|\; \boldsymbol{X}\right]$ and $L(f) = \mathbb{E}\ell(\boldsymbol{X}, f)$. Denote $f^* \in \operatorname{argmin}_{f\in\mathbb{R}} L(f)$ and $f^+ \in \operatorname{argmin}_{f\in\mathbb{R}} \ell(\boldsymbol{X}, f)$. From the definition of f^*,

$$L(f^*) \le L(f^+) = \mathbb{E}\ell(\boldsymbol{X}, f^+).$$

On the other hand, from the definition of f^+,

$$L(f^*) = \mathbb{E}\ell(\boldsymbol{X}, f^*) \ge \mathbb{E}\ell(\boldsymbol{X}, f^+) = L(f^+).$$

The two equations imply $f^* \in \operatorname{argmin}_{f\in\mathbb{R}} \ell(\boldsymbol{X}, f)$ and $f^+ \in \operatorname{argmin}_{f\in\mathbb{R}} L(f)$. So it suffices to show $f^+ = \delta$.

Under model (1),

$$\begin{aligned}
\ell(\boldsymbol{x}, f) &= \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{x})}(m - \tilde{m}(\boldsymbol{x}) + A\delta(\boldsymbol{x}) - Af(\boldsymbol{x}) + \epsilon)^2 \;\middle|\; \boldsymbol{X} = \boldsymbol{x}\right] \\
&= \frac{p_1(\boldsymbol{x})}{\tilde{p}_1(\boldsymbol{x})}\left[(m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x}) + \delta(\boldsymbol{x}) - f(\boldsymbol{x}))^2 + \sigma^2\right] \\
&\quad + \frac{1 - p_1(\boldsymbol{x})}{1 - \tilde{p}_1(\boldsymbol{x})}\left[(m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x}) - \delta(\boldsymbol{x}) + f(\boldsymbol{x}))^2 + \sigma^2\right].
\end{aligned}$$

Note that $\ell(\boldsymbol{x}, f)$ is convex in f. So it is sufficient to consider f such that $\partial\ell(\boldsymbol{x}, f)/\partial f = 0$. There are two situations.

Case I: When $\tilde{p}_1(\boldsymbol{x}) = p_1(\boldsymbol{x})$ for almost all $\boldsymbol{x} \in \mathcal{X}$,

$$\frac{\partial\ell(\boldsymbol{x}, f)}{\partial f} = 2(f(\boldsymbol{x}) - \delta(\boldsymbol{x})).$$

Let $\partial\ell(\boldsymbol{x}, f)/\partial f = 0$ then the conclusion holds.

Case II: When $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x})$ for almost all $\boldsymbol{x} \in \mathcal{X}$,

$$\frac{\partial\ell(\boldsymbol{x}, f)}{\partial f} = 2\left(\frac{p_1(\boldsymbol{x})}{\tilde{p}_1(\boldsymbol{x})} + \frac{1 - p_1(\boldsymbol{x})}{1 - \tilde{p}_1(\boldsymbol{x})}\right)\Big(f(\boldsymbol{x}) - \delta(\boldsymbol{x})\Big).$$

Check that $p_1(\boldsymbol{x})/\tilde{p}_1(\boldsymbol{x}) + (1 - p_1(\boldsymbol{x}))/(1 - \tilde{p}_1(\boldsymbol{x})) > 0$. So the conclusion still holds by letting $\partial\ell(\boldsymbol{x}, f)/\partial f = 0$. □

Proof of Theorem 2. Denote $L(\boldsymbol{f}) = \mathbb{E}\ell(\boldsymbol{X}, \boldsymbol{f})$ where

$$\ell(\boldsymbol{X}, \boldsymbol{f}) = \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{X})}(Y - \tilde{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle)^2 \;\middle|\; \boldsymbol{X}\right].$$

By the similar argument in the proof of Theorem 1, it suffices to consider $\boldsymbol{f}^*$ such that $\boldsymbol{f}^*(\boldsymbol{x}) \in \operatorname{argmin}_{\boldsymbol{f}\in\mathbb{R}^{k-1}} \ell(\boldsymbol{x}, \boldsymbol{f})$. The minimizer of $L(\boldsymbol{f})$ only differs from such $\boldsymbol{f}^*$ on a set of measure zero.

By the last property in Lemma 1, there is a one-to-one map between the hyperplane $\{(u_1, \dots, u_k); \sum_{j=1}^k u_j = 0\}$ and $\mathbb{R}^{k-1}$. So $\min_{\boldsymbol{f} \in \mathbb{R}^{k-1}} \ell(\boldsymbol{x}, \boldsymbol{f})$ is equivalent to

$$\min_{\{u_j\}} \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{x})}(Y - \tilde{m}(\boldsymbol{x}) - u_A)^2 \;\Big|\; \boldsymbol{X} = \boldsymbol{x}\right], \quad \text{s.t.} \sum_{j=1}^k u_j = 0. \tag{9}$$

Denote the solution of (9) by $\{u_j^*\}$. From the last property of Lemma 1, we may assign $\boldsymbol{f}^*(\boldsymbol{x})$ with value $\frac{k-1}{k} \sum_{j=1}^k u_j^* \boldsymbol{W}_j$. Check that $\langle \boldsymbol{W}_j, \boldsymbol{f}^*(\boldsymbol{x}) \rangle = u_j^*$ for each j. So it remains to show $u_j^* = \delta_j(\boldsymbol{x})$.

We solve (9) by the method of Lagrange multipliers. Let

$$\mathcal{L} = \mathbb{E}\left[\frac{1}{\tilde{p}_A(\boldsymbol{x})}(Y - \tilde{m}(\boldsymbol{x}) - u_A)^2 \;\Big|\; \boldsymbol{X} = \boldsymbol{x}\right] + \lambda \sum_{j=1}^k u_j.$$

Under model (5), we have

$$\frac{\partial \mathcal{L}}{\partial u_j} = -2\frac{p_j(\boldsymbol{x})}{\tilde{p}_j(\boldsymbol{x})}(m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - u_j) + \lambda = 0, \quad \text{for } j = 1, \dots, k.$$

There are two cases.

Case I: When $\tilde{p}_j(\boldsymbol{x}) = p_j(\boldsymbol{x})$ almost everywhere for each $j \in \mathcal{A}$, $\lambda = 2(m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - u_j)$. By summing over j and multiplying k^{-1} on both sides, we can solve

$$\lambda = \frac{2}{k}\sum_{j=1}^k (m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - u_j(\boldsymbol{x})) = 2(m(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x})).$$

By plugging it back to the first order condition we have $u_j^* = \delta_j(\boldsymbol{x})$.

Case II: When $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x})$ almost everywhere, $\lambda = 2\frac{p_j(\boldsymbol{x})}{\hat{p}_j(\boldsymbol{x})}(\delta_j(\boldsymbol{x}) - u_j)$. By rearranging the term and summing over j, we have

$$\frac{\lambda}{2}\sum_{j=1}^k \frac{\tilde{p}_j(\boldsymbol{x})}{p_j(\boldsymbol{x})} = \sum_{j=1}^k (\delta_j(\boldsymbol{x}) - u_j) = 0.$$

Since $\tilde{p}_j(\boldsymbol{x}) > 0$ and $p_j(\boldsymbol{x}) > 0$, $\sum_{j=1}^k \tilde{p}_j(\boldsymbol{x})/p_j(\boldsymbol{x}) > 0$ implying $\lambda = 0$. By plugging it back to the first order condition we still have $u_j^* = \delta_j(\boldsymbol{x})$. □

Proof of Theorem 3. Denote $\ell(\boldsymbol{X}, g) = \mathbb{E}\left[\frac{1}{p_A(\boldsymbol{X})}(Y - g(\boldsymbol{X}))^2 \;\big|\; \boldsymbol{X}\right]$. It suffices to show $m(\boldsymbol{x}) \in \operatorname{argmin}_{g \in \mathbb{R}} \ell(\boldsymbol{x}, g)$.

Under model (5), check that

$$\ell(\boldsymbol{x}, g) = \mathbb{E}\left[\sum_{j=1}^k (m(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - g(\boldsymbol{x}) + \epsilon)^2 \;\Big|\; \boldsymbol{X} = \boldsymbol{x}\right]$$

$$= \sum_{j=1}^{k} (m(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - g(\boldsymbol{x}))^2 + k\sigma^2(\boldsymbol{x}).$$

Let $\partial \ell(\boldsymbol{x}, g)/\partial g = 0$, we have

$$\frac{\partial \ell(\boldsymbol{x}, g)}{\partial g} = -2\sum_{j=1}^{k} (m(\boldsymbol{x}) + \delta_j(\boldsymbol{x}) - g(\boldsymbol{x})) = 2k(g(\boldsymbol{x}) - m(\boldsymbol{x})) = 0$$

implying that $m(\boldsymbol{x})$ is the solution. □

Appendix C: Upper bounds for prediction error

In this section we provide the proofs for Theorem 4 and Theorem 5, where the goal is to find an upper bound for the prediction error of $\hat{\boldsymbol{\delta}}$, $\mathrm{PE}(\hat{\boldsymbol{\delta}})$, in both linear and kernel settings. To avoid confusion, we use $\|\boldsymbol{f}(\boldsymbol{x})\|^2 = \sum_{j=1}^{k-1} f_j^2(\boldsymbol{x})$ denote the squared ℓ^2 norm of $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_{k-1}(\boldsymbol{x}))^T$. Note that $\|\boldsymbol{f}(\boldsymbol{x})\|$ is a one-dimension function of $\boldsymbol{x}$. We let $\|\boldsymbol{f}\|_2^2 = \mathbb{E}\|\boldsymbol{f}(\boldsymbol{X})\|^2$ denote the squared L^2 norm of $\|\boldsymbol{f}(\boldsymbol{x})\|$, and $\|\boldsymbol{f}\|_\infty$ denote the L^∞ norm of $\|\boldsymbol{f}(\boldsymbol{x})\|$. From the definition, $\mathrm{PE}(\boldsymbol{f}) = \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

To begin with, we denote

$$L_1(\boldsymbol{f}) = \mathbb{E}\left[\frac{1}{p_A(\boldsymbol{X})} \left(Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle\right)^2\right], \quad \text{and}$$
$$L_2(\boldsymbol{f}) = \mathbb{E}\left[\frac{1}{\hat{p}_A(\boldsymbol{X})} \left(Y - m(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle\right)^2\right].$$

By Theorem 2, it can be checked that $\boldsymbol{f}^* = \operatorname{argmin}_{\boldsymbol{f}} L_1(\boldsymbol{f}) = \operatorname{argmin}_{\boldsymbol{f}} L_2(\boldsymbol{f})$. Given any function $\boldsymbol{f}^0 \in \mathbb{R}^{k-1}$, we are interested in $L_1(\boldsymbol{f}^0) - L_1(\boldsymbol{f}^*)$ and $L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*)$. In statistical learning theory, they are called the excess risks of $\boldsymbol{f}^0$. The following lemma gives the relationship between $\mathrm{PE}(\boldsymbol{f}^0)$ and the excess risk of $\boldsymbol{f}^0$.

Lemma 2. *Suppose the model (5) holds. Under Assumption 1 and 2, for any $\boldsymbol{f}^0 \in \mathbb{R}^{k-1}$ we have*

$$\mathrm{PE}(\boldsymbol{\delta}^0) = \frac{k-1}{k}\|\boldsymbol{f}^0 - \boldsymbol{f}^*\|_2^2 = L_1(\boldsymbol{f}^0) - L_1(\boldsymbol{f}^*) = \mathcal{O}\left(L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*)\right),$$

where $\boldsymbol{\delta}^0 = (\delta_1^0, \ldots, \delta_k^0)^T = (\langle \boldsymbol{W}_1, \boldsymbol{f}^0\rangle, \ldots, \langle \boldsymbol{W}_k, \boldsymbol{f}^0\rangle)^T$.

Proof of Lemma 2. Firstly, for the excess risk $L_1(\boldsymbol{f}^0) - L_1(\boldsymbol{f}^*)$, check that

$$L_1(\boldsymbol{f}^0) - L_1(\boldsymbol{f}^*)$$
$$= \mathbb{E}\left[\frac{1}{p_A(\boldsymbol{X})} \left(2(Y - \hat{m}(\boldsymbol{X})) - \langle \boldsymbol{W}_A, \boldsymbol{f}^0(\boldsymbol{X}) + \boldsymbol{f}^*(\boldsymbol{X})\rangle\right) \langle \boldsymbol{W}_A, \boldsymbol{f}^*(\boldsymbol{X}) - \boldsymbol{f}^0(\boldsymbol{X})\rangle\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{j=1}^{k}\big(2(r(\boldsymbol{X})+\delta_j(\boldsymbol{X}))-(\delta_j^0(\boldsymbol{X})+\delta_j(\boldsymbol{X}))\big)\big(\delta_j(\boldsymbol{X})-\delta_j^0(\boldsymbol{X})\big)\right]\\
&= \mathbb{E}\left[2r(\boldsymbol{X})\sum_{j=1}^{k}\big(\delta_j(\boldsymbol{X})-\delta_j^0(\boldsymbol{X})\big)+\sum_{j=1}^{k}\big(\delta_j(\boldsymbol{X})-\delta_j^0(\boldsymbol{X})\big)^2\right]\\
&= \mathrm{PE}(\boldsymbol{\delta}^0),
\end{aligned}
$$

where $r(\boldsymbol{X}) = m(\boldsymbol{X}) - \hat{m}(\boldsymbol{X})$.

Secondly, for the excess risk $L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*)$, check that

$$
\begin{aligned}
&L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*)\\
&= \mathbb{E}\left[\frac{1}{\hat{p}_A(\boldsymbol{X})}\big(2(Y-m(\boldsymbol{X}))-\langle \boldsymbol{W}_A, \boldsymbol{f}^0(\boldsymbol{X})+\boldsymbol{f}^*(\boldsymbol{X})\rangle\big)\langle \boldsymbol{W}_A, \boldsymbol{f}^*(\boldsymbol{X})-\boldsymbol{f}^0(\boldsymbol{X})\rangle\right]\\
&= \mathbb{E}\left[\sum_{j=1}^{k}\frac{p_j(\boldsymbol{X})}{\hat{p}_j(\boldsymbol{X})}\big(2\delta_j(\boldsymbol{X})-(\delta_j^0(\boldsymbol{X})+\delta_j(\boldsymbol{X}))\big)\big(\delta_j(\boldsymbol{X})-\delta_j^0(\boldsymbol{X})\big)\right]\\
&= \mathbb{E}\left[\sum_{j=1}^{k}\frac{p_j(\boldsymbol{X})}{\hat{p}_j(\boldsymbol{X})}\big(\delta_j^0(\boldsymbol{X})-\delta_j(\boldsymbol{X})\big)^2\right].
\end{aligned}
$$

From Assumption 1, $p_j(\boldsymbol{X}) \geq c$ implies that $c \leq p_j(\boldsymbol{X}) \leq 1-c$. For Assumption 2 to be true, there exists a $c_0 > 1$ such that $\hat{p}_j^{-1}(\boldsymbol{X}) \leq c_0$, which implies that $(c_0-1)^{-1}c_0 \leq \hat{p}_j^{-1}(\boldsymbol{X}) \leq c_0$. So for each $j \in \mathcal{A}$, we have $(c_0-1)^{-1}cc_0 \leq p_j(\boldsymbol{X})/\hat{p}_j(\boldsymbol{X}) \leq (1-c)c_0$. Therefore,

$$
\frac{cc_0}{c_0-1}\mathrm{PE}(\boldsymbol{\delta}^0) \leq L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*) \leq (1-c)c_0\mathrm{PE}(\boldsymbol{\delta}^0) \tag{10}
$$

indicating that $\mathrm{PE}(\boldsymbol{\delta}^0) = \mathcal{O}\big(L_2(\boldsymbol{f}^0) - L_2(\boldsymbol{f}^*)\big)$.

Finally, by the property 4 of Lemma 1, we have

$$
\begin{aligned}
\frac{k}{k-1}\|\boldsymbol{f}^0-\boldsymbol{f}^*\|_2^2 &= \mathbb{E}\sum_{j=1}^{k}\langle \boldsymbol{W}_j, \boldsymbol{f}^0(\boldsymbol{X})-\boldsymbol{f}^*(\boldsymbol{X})\rangle^2\\
&= \mathbb{E}\sum_{j=1}^{k}\big(\delta_j^0(\boldsymbol{X})-\delta_j(\boldsymbol{X})\big)^2 = \mathrm{PE}(\boldsymbol{\delta}^0)
\end{aligned}
$$

and the proof is complete. □

Lemma 2 implies that the order of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ is the same as the excess risk $L_1(\hat{\boldsymbol{f}}) - L_1(\boldsymbol{f}^*)$ and $L_2(\hat{\boldsymbol{f}}) - L_2(\boldsymbol{f}^*)$. This provides us an innovate way to derive the upper bound for $\mathrm{PE}(\hat{\boldsymbol{\delta}})$. That is, we derive the upper bounds for $L_1(\hat{\boldsymbol{f}}) - L_1(\boldsymbol{f}^*)$ and $L_2(\hat{\boldsymbol{f}}) - L_2(\boldsymbol{f}^*)$ respectively, and take the smaller one.

Let

$$
L(\boldsymbol{f}) = \mathbb{E}\left[\frac{1}{\hat{p}_A(\boldsymbol{X})}\big(Y-\hat{m}(\boldsymbol{X})-\langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle\big)^2\right].
$$

Lemma 3 is a general statement for the upper bound of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ when one considers the optimization problem (6) under a certain function space $\mathcal{F}$.

Lemma 3. *Let $\mathcal{F}$ be a function space with bounded L^2 norm in $\mathbb{R}^{k-1}$. Under Assumption 1 to 3,*

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(\max\left\{L(\hat{\boldsymbol{f}}) - \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}), \min\{r_1, r_2\}, \inf_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2\right\}\right),$$

almost surely under P, where $r_1 = \left(C_m + \sup_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f}\|_\infty\right)^2 r_p$, $r_2 = k(1+r_p)r_m^2$.

Proof of Lemma 3. Let $\tilde{\boldsymbol{f}}_1 \in \operatorname{argmin}_{\boldsymbol{f}\in\mathcal{F}} L_1(\boldsymbol{f})$. By Lemma 2, we have

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) = L_1(\hat{\boldsymbol{f}}) - L_1(\boldsymbol{f}^*) = \left(L_1(\hat{\boldsymbol{f}}) - L_1(\tilde{\boldsymbol{f}}_1)\right) + \left(L_1(\tilde{\boldsymbol{f}}_1) - L_1(\boldsymbol{f}^*)\right).$$

Check that

$$\begin{aligned}&\inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}) - L(\hat{\boldsymbol{f}}) + \left(L(\hat{\boldsymbol{f}}) - L_1(\hat{\boldsymbol{f}})\right) + \left(L_1(\hat{\boldsymbol{f}}) - L_1(\tilde{\boldsymbol{f}}_1)\right) + \left(L_1(\tilde{\boldsymbol{f}}_1) - L(\tilde{\boldsymbol{f}}_1)\right) \\ &= \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}) - L(\tilde{\boldsymbol{f}}_1) \leq 0.\end{aligned}$$

So one can bound the first term of $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ by

$$\begin{aligned}L_1(\hat{\boldsymbol{f}}) - L_1(\tilde{\boldsymbol{f}}_1) &\leq L(\hat{\boldsymbol{f}}) - \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}) + \left(L(\tilde{\boldsymbol{f}}_1) - L_1(\tilde{\boldsymbol{f}}_1)\right) - \left(L(\hat{\boldsymbol{f}}) - L_1(\hat{\boldsymbol{f}})\right) \\ &\leq L(\hat{\boldsymbol{f}}) - \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}) + 2\sup_{\boldsymbol{f}\in\mathcal{F}} |L(\boldsymbol{f}) - L_1(\boldsymbol{f})|.\end{aligned}$$

This implies that

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(\max\left\{L(\hat{\boldsymbol{f}}) - \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}), \sup_{\boldsymbol{f}\in\mathcal{F}} |L(\boldsymbol{f}) - L_1(\boldsymbol{f})|, L_1(\tilde{\boldsymbol{f}}_1) - L_1(\boldsymbol{f}^*)\right\}\right).$$

Similarly, by denoting $\tilde{\boldsymbol{f}}_2 \in \operatorname{argmin}_{\boldsymbol{f}\in\mathcal{F}} L_2(\boldsymbol{f})$ and applying Lemma 2 on L_2, we have

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \leq \mathcal{O}\left(\max\left\{L(\hat{\boldsymbol{f}}) - \inf_{\boldsymbol{f}\in\mathcal{F}} L(\boldsymbol{f}), \sup_{\boldsymbol{f}\in\mathcal{F}} |L(\boldsymbol{f}) - L_2(\boldsymbol{f})|, L_2(\tilde{\boldsymbol{f}}_2) - L_2(\boldsymbol{f}^*)\right\}\right).$$

The rest of the proof is split into two parts. In part I, we derive an upper bound for the second term in each scenario. In part II, we show that the third term is of the same order of $\inf_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

Part I: For any $\boldsymbol{f} \in \mathcal{F}$, check that

$$\begin{aligned}L(\boldsymbol{f}) - L_1(\boldsymbol{f}) &= \mathbb{E}\left[\left(\frac{1}{\hat{p}_A(\boldsymbol{X})} - \frac{1}{p_A(\boldsymbol{X})}\right)(Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle)^2\right] \\ &\leq r_p\left(C_m + \|\boldsymbol{f}\|_\infty\right)^2.\end{aligned}$$

So $\sup_{\boldsymbol{f}\in\mathcal{F}} |L(\boldsymbol{f}) - L_2(\boldsymbol{f})| \leq \left(C_m + \sup_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f}\|_\infty\right)^2 r_p \triangleq r_1$.

On the other hand, by letting $r(\boldsymbol{x}) = m(\boldsymbol{x}) - \hat{m}(\boldsymbol{x})$,

$$\begin{aligned}
&L(\boldsymbol{f}) - L_2(\boldsymbol{f}) \\
=&\ \mathbb{E}\left[\frac{1}{\hat{p}_A(\boldsymbol{X})}\left((Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle)^2 - (Y - m(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle)^2\right)\right] \\
=&\ \mathbb{E}\left[\frac{p_A(\boldsymbol{X})}{\hat{p}_A(\boldsymbol{X})}\frac{1}{p_A(\boldsymbol{X})}(r(\boldsymbol{X}) + 2\delta_A(\boldsymbol{X}) - 2\langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle) r(\boldsymbol{X})\right] \\
\leq&\ \left\|\frac{p_j(\boldsymbol{x})}{\hat{p}_j(\boldsymbol{x})}\right\|_\infty \mathbb{E}\left[\sum_{j=1}^k r^2(\boldsymbol{X}) + 2r(\boldsymbol{X})\sum_{j=1}^k(\delta_j(\boldsymbol{X}) - \langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{X})\rangle)\right] \\
\leq&\ k r_m^2 \|p_j(\boldsymbol{x})/\hat{p}_j(\boldsymbol{x})\|_\infty,
\end{aligned}$$

where the last equality follows from the property 3 of Lemma 1 and the fact that $\sum_{j=1}^k \delta_j(\boldsymbol{x}) = 0$. By Assumption 1, $c \leq p_j(\boldsymbol{x}) \leq 1 - c$. Hence by Assumption 2 we have

$$\begin{aligned}
(1-c)^{-1}\|p_j(\boldsymbol{x})/\hat{p}_j(\boldsymbol{x}) - 1\|_\infty &\leq \|(p_j(\boldsymbol{x})/\hat{p}_j(\boldsymbol{x}) - 1)p_j^{-1}(\boldsymbol{x})\|_\infty \\
&= \|\hat{p}_j^{-1}(\boldsymbol{x}) - p_j^{-1}(\boldsymbol{x})\|_\infty \leq r_p
\end{aligned}$$

implying $\|p_j(\boldsymbol{x})/\hat{p}_j(\boldsymbol{x})\|_\infty \leq 1 + (1-c)r_p \leq 1 + r_p$. So $\sup_{\boldsymbol{f}\in\mathcal{F}} |L(\boldsymbol{f}) - L_2(\boldsymbol{f})| \leq k(1 + r_p)r_m^2 \triangleq r_2$.

Part II: By Lemma 2 one can check that

$$\begin{aligned}
L_1(\tilde{\boldsymbol{f}}_1) - L_1(\boldsymbol{f}^*) = \inf_{\boldsymbol{f}\in\mathcal{F}} L_1(\boldsymbol{f}) - L_1(\boldsymbol{f}^*) &= \inf_{\boldsymbol{f}\in\mathcal{F}}(L_1(\boldsymbol{f}) - L_1(\boldsymbol{f}^*)) \\
&= \inf_{\boldsymbol{f}\in\mathcal{F}} \frac{k-1}{k}\|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2.
\end{aligned}$$

This implies that (1) $\tilde{\boldsymbol{f}}_1 \in \operatorname{argmin}_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$; (2) $L_1(\tilde{\boldsymbol{f}}_1) - L_1(\boldsymbol{f}^*)$ has the same order of $\inf_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

On the other hand, by Lemma 2 and the inequality (10),

$$\begin{aligned}
&L_2(\tilde{\boldsymbol{f}}_2) - L_2(\boldsymbol{f}^*) \leq L_2(\tilde{\boldsymbol{f}}_1) - L_2(\boldsymbol{f}^*) \leq \frac{(1-k)(1-c)c_0}{k}\|\tilde{\boldsymbol{f}}_1 - \boldsymbol{f}^*\|_2^2, \quad \text{and} \\
&L_2(\tilde{\boldsymbol{f}}_2) - L_2(\boldsymbol{f}^*) \geq \frac{(1-k)cc_0}{k(c_0-1)}\|\tilde{\boldsymbol{f}}_2 - \boldsymbol{f}^*\|_2^2 \geq \frac{(1-k)cc_0}{k(c_0-1)}\|\tilde{\boldsymbol{f}}_1 - \boldsymbol{f}^*\|_2^2,
\end{aligned}$$

where $c_0 > 1$ is a constant such that $\hat{p}_j^{-1}(\boldsymbol{x}) \leq c_0$. Since $\|\tilde{\boldsymbol{f}}_1 - \boldsymbol{f}^*\|_2^2 = \inf_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$, we conclude that $L_2(\tilde{\boldsymbol{f}}_1) - L_2(\boldsymbol{f}^*)$ also has the same order of $\inf_{\boldsymbol{f}\in\mathcal{F}} \|\boldsymbol{f} - \boldsymbol{f}^*\|_2^2$.

Finally, by collecting the two inequalities for $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ derived under L_1 and L_2, we have the desired result by picking the smaller bound. □

Note that the upper bound for $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ in Lemma 3 involves a supreme for $\|\boldsymbol{f}\|_\infty$ for $\boldsymbol{f} \in \mathcal{F}$. However, unlike in most of the classical learning theories where the constant term is omitted [7, 10, 42, 52], we include an unpenalized constant term in the function space when developing the theory. This makes our conclusion more general while bringing additional challenges to bound $\|\boldsymbol{f}\|_\infty$. The following Lemma is useful to address this issue by providing a bound for the constant term.

Lemma 4. *Let*

$$\mathcal{F} = \{\boldsymbol{f} \in \mathbb{R}^{k-1}; \boldsymbol{f} = \boldsymbol{f}' + \boldsymbol{b}, \boldsymbol{f}' \in \mathcal{F}', \boldsymbol{b} \in \mathbb{R}^{k-1}\},$$

where $\mathcal{F}'$ is a collection of functions with bounded L^2 norm in $\mathbb{R}^{k-1}$. Write $\hat{\boldsymbol{f}}$ as $\hat{\boldsymbol{f}} = \hat{\boldsymbol{f}}' + \hat{\boldsymbol{b}}$ where $\hat{\boldsymbol{b}}$ is the constant term, under Assumption 2 and 3 we have

$$\|\hat{\boldsymbol{b}}\| \le \sqrt{2k(k-1)}\left(c_0 - 1\right)\left(C_m + \sup_{\boldsymbol{f}' \in \mathcal{F}'} \|\boldsymbol{f}'\|_\infty\right).$$

Proof of Lemma 4. By the definition of $\hat{\boldsymbol{f}}$, $\hat{\boldsymbol{b}}$ is also the solution to

$$\min_{\boldsymbol{b} \in \mathbb{R}^{k-1}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{p}_{a_i}(\boldsymbol{x}_i)} \left(y_i - \hat{m}(\boldsymbol{x}_i) - \langle \boldsymbol{W}_{a_i}, \hat{\boldsymbol{f}}'(\boldsymbol{x}_i) + \boldsymbol{b} \rangle\right)^2.$$

If we let $u_j = \langle \boldsymbol{W}_j, \boldsymbol{b} \rangle$, the optimization problem above is equivalent to

$$\sum_{j=1}^{k} \sum_{i \in I_j} \frac{1}{\hat{p}_j(\boldsymbol{x}_i)} (r_i - u_j)^2, \quad \text{s.t.} \sum_{j=1}^{k} u_j = 0, \tag{11}$$

where $I_j = \{i; a_i = j\}$ and $r_i = y_i - \hat{m}(\boldsymbol{x}_i) - \langle \boldsymbol{W}_{a_i}, \hat{\boldsymbol{f}}'(\boldsymbol{x}_i) \rangle$. By Assumption 3, we have $|r_i| \le C_m + \|\boldsymbol{f}'\|_\infty \le C_m + \sup_{\boldsymbol{f}' \in \mathcal{F}'} \|\boldsymbol{f}'\|_\infty$ for all i.

By solving (11) by the method of Lagrange multipliers, the estimator $\hat{u}_j$ can be written as

$$\hat{u}_j = \theta_j - \frac{\nu_j}{\bar{\nu}} \bar{\theta},$$

where $\nu_j = \left(\sum_{i \in I_j} \hat{p}_j^{-1}(\boldsymbol{x}_i)\right)^{-1}$, $\theta_j = \nu_j \sum_{i \in I_j} r_i \hat{p}_j^{-1}(\boldsymbol{x}_i)$, and $\bar{\nu}$ and $\bar{\theta}$ are the averages of the sequence $\{\nu_j\}$ and $\{\theta_j\}$. According to the property 4 of Lemma 1, finding an upper bound for $\|\hat{\boldsymbol{b}}\|^2$ is equivalent to finding an upper bound for $\sum_{j=1}^{k} \langle \boldsymbol{W}_j, \hat{\boldsymbol{b}} \rangle^2 = \sum_{j=1}^{k} \hat{u}_j^2$.

By rearranging the terms in $\hat{\mu}_j$, we have

$$\sum_{j=1}^{k} \hat{u}_j^2 = \sum_{j=1}^{k} \left(\theta_j - \bar{\theta} - \left(\frac{\nu_j}{\bar{\nu}} - 1\right)\bar{\theta}\right)^2 \le 2\sum_{j=1}^{k} (\theta_j - \bar{\theta})^2 + 2\bar{\theta}^2 \sum_{j=1}^{k} \left(\frac{\nu_j}{\bar{\nu}} - 1\right)^2.$$

Notice that the last term $\sum_{j=1}^{k} \left(\nu_j/\bar{\nu} - 1\right)^2$ is the variance of the sequence $\{\nu_j/\bar{\nu}\}$ multiplying by k. Since for any nonnegative sequence of length k with mean 1, the maximum variance is achieved when one of them equals to k and others equal to 0. This implies that $\sum_{j=1}^{k} (\nu_j/\bar{\nu} - 1)^2 \le (k-1)^2 + (k-1)(0-1)^2 = k(k-1)$. So

$$\sum_{j=1}^{k} \hat{u}_j^2 \le 2\sum_{j=1}^{k}(\theta_j - \bar{\theta})^2 + 2k(k-1)\bar{\theta}^2 \le 2\sum_{j=1}^{k}\theta_j^2 + 2k(k-1)\cdot\frac{1}{k}\sum_{j=1}^{k}\theta_j^2 = 2k\sum_{j=1}^{k}\theta_j^2.$$

Let $c_0 > 1$ such that $\hat{p}_j(\boldsymbol{x}) \ge c_0^{-1}$, we have $(c_0-1)^{-1}c_0 \le \hat{p}_j^{-1}(\boldsymbol{x}) \le c_0$. So we have $|\theta_j| \le \nu_j \sum_{i\in I_j} c_0|r_j| \le (c_0-1)\left(C_m + \sup_{\boldsymbol{f}'\in\mathcal{F}'}\|\boldsymbol{f}'\|_\infty\right)$ and $\nu_j^{-1} \ge |I_j|(c_0-1)^{-1}c_0$, where $|I_j|$ denotes the cardinality of I_j. By applying Lemma 1, finally we have

$$\begin{aligned}\|\hat{\boldsymbol{b}}\|^2 = \frac{k-1}{k}\sum_{j=1}^{k}\hat{u}_j^2 &\le 2(k-1)\sum_{j=1}^{k}\theta_j^2\\ &\le 2k\left(k-1\right)\left(c_0-1\right)^2\left(C_m + \sup_{\boldsymbol{f}'\in\mathcal{F}'}\|\boldsymbol{f}'\|_\infty\right)^2\end{aligned}$$

and the proof is complete. □

C.1. Prediction error for linear learning

The goal of this section is to prove Theorem 4. From Lemma 2, it suffices to bound the excess risk $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(p,s)})$, where $\boldsymbol{f}^{(p,s)} \in \operatorname{argmin}_{\boldsymbol{f}\in\mathcal{F}(p,s)} L(\boldsymbol{f})$.

Before the proof, we introduce some additional notations. For each $\boldsymbol{f} \in \mathcal{F}(p,s)$, we write it as $\boldsymbol{f} = \boldsymbol{f}' + \boldsymbol{b}$, where $\boldsymbol{b}$ is the constant term. By Assumption 4, since each covariate is bounded in $[-1,1]$, we have $\|\boldsymbol{f}'(\boldsymbol{x})\| \le \|\boldsymbol{f}'(\boldsymbol{x})\|_1 = \sum_{j=1}^{k-1}|\boldsymbol{x}^T\boldsymbol{\beta}_j| \le \sum_{j=1}^{k-1}\sum_{l=1}^{p}|\beta_{jl}x_l| \le s$ for all $\boldsymbol{x}$ and any $\boldsymbol{f}'$, which implies $\sup\|\boldsymbol{f}'\|_\infty \le s$. Then by applying Lemma 4, $\|\hat{\boldsymbol{b}}\| \le \sqrt{2k(k-1)}(c_0-1)(C_m+s)$, where c_0 is an upper bound for $\hat{p}_j^{-1}(\boldsymbol{x})$. If we let

$$\mathcal{F}^b(p,s) = \mathcal{F}(p,s) \cap \{\boldsymbol{f};\|\boldsymbol{b}\| \le \sqrt{2k(k-1)}(c_0-1)(C_m+s)\},$$

then $\hat{\boldsymbol{f}} \in \mathcal{F}^b(p,s)$. Hence it suffices to consider the new function space $\mathcal{F}^b(p,s)$. Check that for any $\boldsymbol{f} \in \mathcal{F}^b(p,s)$, $\|\boldsymbol{f}(\boldsymbol{x})\| \le \|\boldsymbol{f}'(\boldsymbol{x})\| + \|\boldsymbol{b}\| \le s + \sqrt{2k(k-1)}(c_0-1)(C_m+s)$ for all $\boldsymbol{x} \in \mathcal{X}$. So

$$\sup_{\boldsymbol{f}\in\mathcal{F}^b(p,s)}\|\boldsymbol{f}\|_\infty \le s + \sqrt{2k(k-1)}(c_0-1)(C_m+s).$$

Next, we let $g_{\boldsymbol{f}}(\boldsymbol{Z}) = \frac{1}{\hat{p}_A(\boldsymbol{X})}\left(Y - \hat{m}(\boldsymbol{X}) - \langle\boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle\right)^2$, where the random vector $\boldsymbol{Z} = (\boldsymbol{X}, A, Y)$. From Assumption 3, for any $\boldsymbol{f} \in \mathcal{F}^b(p,s)$ we have

$$|g_{\boldsymbol{f}}(\boldsymbol{Z})| \le c_0(C_m + \|\boldsymbol{f}\|_\infty)^2 \le c_0\left((c_0-1)\sqrt{2k(k-1)}+1\right)^2\left(C_m+s\right)^2 \triangleq c_0 s_0^2.$$

Notice that here we use $s_0 = \Big((c_0-1)\sqrt{2k(k-1)}+1\Big)\Big(C_m+s\Big)$ to denote an upper bound for $|Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle|$. Then we consider the following function space

$$\mathcal{H} = \{\left(2c_0 s_0^2\right)^{-1}\left(g_{\boldsymbol{f}} - g_{\boldsymbol{f}^{(p,s)}}\right); \boldsymbol{f} \in \mathcal{F}^b(p,s)\}.$$

Check that $\boldsymbol{f}^{(p,s)} \in \mathcal{F}^b(p,s)$. And for any $h_{\boldsymbol{f}} \in \mathcal{H}$, we have $|h_{\boldsymbol{f}}| = \left(2c_0 s_0^2\right)^{-1}|g_{\boldsymbol{f}} - g_{\boldsymbol{f}^{(p,s)}}| \leq 1$.

Finally, for $h_{\boldsymbol{f}} \in \mathcal{H}$, we define the empirical process $h_{\boldsymbol{f}} \to P_n h_{\boldsymbol{f}} - Ph_{\boldsymbol{f}}$, where $Ph_{\boldsymbol{f}} = \int h_{\boldsymbol{f}} dP$ and $P_n h_{\boldsymbol{f}} = n^{-1}\sum_{i=1}^n h_{\boldsymbol{f}}(\boldsymbol{Z}_i)$ with $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ i.i.d. from distribution P. Denote the entropy of $\mathcal{H}$ under distribution Q as $H(\epsilon, \mathcal{H}, L_2(Q))$ and let $H(\epsilon, \mathcal{H}) = \sup_Q H(\epsilon, \mathcal{H}, L_2(Q))$.

The following Lemma gives the tail probability of the excess risk $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(p,s)})$, from which we can derive its convergence rate.

Lemma 5. *Let $0 \leq \theta \leq 1$. Assume there exists an $M > 0$ such that*

$$\epsilon_0 \log_2 \frac{16\sqrt{6}\epsilon_0}{\theta M} + \epsilon_0 \leq \frac{\theta}{32\sqrt{2}}, \tag{12}$$

where $\epsilon_0 > 0$ is such that the entropy of $\mathcal{H}$

$$H(\epsilon_0, \mathcal{H}) \leq \frac{\theta}{2} n M^2. \tag{13}$$

Then for arbitrary $\boldsymbol{f}^0 \in \mathcal{F}$,

$$P\left(L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^0) \geq 16 c_0 s_0^2 M\right) \leq 6\left(1 - \frac{1}{4nM^2}\right)^{-1} \exp\left(-2(1-\theta)nM^2\right).$$

Proof of Lemma 5. Firstly, by the definition of $\mathcal{H}$, we observe that

$$Ph_{\hat{\boldsymbol{f}}} - Ph_{\boldsymbol{f}^0} = \left(2c_0 s_0^2\right)^{-1}\left(L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^0)\right).$$

Secondly, from the definition of $\hat{\boldsymbol{f}}$, $P_n h_{\hat{\boldsymbol{f}}} - P_n h_{\boldsymbol{f}^0} \leq 0$, so

$$\left(P_n h_{\hat{\boldsymbol{f}}} - Ph_{\hat{\boldsymbol{f}}}\right) + \left(Ph_{\hat{\boldsymbol{f}}} - Ph_{\boldsymbol{f}^0}\right) + \left(Ph_{\boldsymbol{f}^0} - P_n h_{\boldsymbol{f}^0}\right) = P_n h_{\hat{\boldsymbol{f}}} - P_n h_{\boldsymbol{f}^0} \leq 0.$$

By rearranging the terms of the left hand side, we have

$$Ph_{\hat{\boldsymbol{f}}} - Ph_{\boldsymbol{f}^0} \leq \left(P_n h_{\boldsymbol{f}^0} - Ph_{\boldsymbol{f}^0}\right) - \left(P_n h_{\hat{\boldsymbol{f}}} - Ph_{\hat{\boldsymbol{f}}}\right) \leq 2\sup_{h\in\mathcal{H}}|P_n h - Ph|.$$

Therefore,

$$\begin{aligned} P\left(L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^0) \geq 16c_0 s_0^2 M\right) &= P\left(Ph_{\hat{\boldsymbol{f}}} - Ph_{\boldsymbol{f}^0} \geq 8M\right) \\ &\leq P^*\left(\sup_{h\in\mathcal{H}}|P_n h - Ph| \geq 4M\right), \end{aligned}$$

where P^* is the outer probability.

Finally, check that for any $h \in \mathcal{H}$,

$$P(h - Ph)^2 = Ph^2 - (Ph)^2 \leq Ph^2 \leq 1.$$

Then the result follows from Lemma 6 by taking $\sigma^2 = 1$. □

Lemma 6 ([53])**.** *Let $\mathcal{H}$ be a collection of functions such that $P(h - Ph)^2 \leq \sigma^2$ holds for all $h \in \mathcal{H}$. Given n and $0 < \theta < 1$, assume $M > 0$ and ϵ_0 satisfy (12) and (13). Then*

$$P^*\left(\sup_{h\in\mathcal{H}} |P_n h - Ph| \geq 4M\right) \leq 6\left(1 - \frac{\sigma^2}{4nM^2}\right)^{-1} \exp\left(-2(1-\theta)nM^2\right),$$

where P^* denotes the outer probability.

Note that Lemma 6 is essentially the same as Lemma 4 in [53], while we rewrite its condition to make it more general. For the detailed proof, one can refer to Lemma 4 from [53] and Lemma 2.14.18 from [48].

With the help of Lemma 5, we can prove Theorem 4 by choosing an appropriate M. However, from condition (13), we notice that M is closely related to the entropy of the function space $\mathcal{H}$. So before the final step, we still need to find an upper bound for $H(\epsilon, \mathcal{H})$.

Lemma 7. *For any $\epsilon > 0$, $H(\epsilon, \mathcal{H}) \leq \frac{8(k-1)^2}{\epsilon^2} \log\left(e + e\frac{p+1}{2(k-1)}\epsilon^2\right)$.*

Proof of Lemma 7. Take $\boldsymbol{f}_1, \boldsymbol{f}_2 \in \mathcal{F}^b(p, s)$. For any distribution Q, check that

$$\begin{aligned}
&\|h_{\boldsymbol{f}_1} - h_{\boldsymbol{f}_2}\|_{Q,2}^2 \\
&= \left(2c_0 s_0^2\right)^{-2} \|g_{\boldsymbol{f}_1} - g_{\boldsymbol{f}_2}\|_{Q,2}^2 \\
&= \frac{1}{4c_0^2 s_0^4}\mathbb{E}\Big(\frac{1}{p_A(\boldsymbol{X})}\left(2(Y - \hat{m}(\boldsymbol{X})) - \langle \boldsymbol{W}_A, \boldsymbol{f}_1(\boldsymbol{X}) + \boldsymbol{f}_2(\boldsymbol{X})\rangle\right) \\
&\qquad\qquad \times \langle \boldsymbol{W}_A, \boldsymbol{f}_1(\boldsymbol{X}) - \boldsymbol{f}_2(\boldsymbol{X})\rangle\Big)^2 \\
&\leq \frac{c_0^2}{4c_0^2 s_0^4}\mathbb{E}\left(|2(Y - \hat{m}(\boldsymbol{X})) - \langle \boldsymbol{W}_A, \boldsymbol{f}_1(\boldsymbol{X}) + \boldsymbol{f}_2(\boldsymbol{X})\rangle| \cdot \|\boldsymbol{f}_1(\boldsymbol{X}) - \boldsymbol{f}_2(\boldsymbol{X})\|\right)^2 \\
&\leq s_0^{-2}\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{Q,2}^2.
\end{aligned}$$

This implies that an $s_0\epsilon$-net on $\mathcal{F}^b(p, s)$ introduces an ϵ-net on $\mathcal{H}$ and the remaining work is to find the entropy number for $\mathcal{F}^b(p, s)$. However, since $\mathcal{F}^b(p, s)$ is a collection of $(k-1)$-dimensional functions while most of the learning theories only give the entropy number for a one-dimensional function space, we use the following "measure changing trick" to address this problem.

Let $\tilde{Q}$ be the distribution of $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_{k-1}) = (\delta_1\mathbf{X}_1, \ldots, \delta_{k-1}\mathbf{X}_{k-1})$, where $\mathbf{X}_j = (1, \boldsymbol{X}_j^T)^T \in \mathbb{R}^{p+1}$ with $\boldsymbol{X}_j$ i.i.d. from distribution Q, and the random vector $(\delta_1, \ldots, \delta_{k-1})$ has a joint distribution $P\left((\delta_1, \ldots, \delta_{k-1})^T = \boldsymbol{e}_j\right) = (k-1)^{-1}$ for $j = 1, \ldots, k-1$. For any $\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T \in \mathcal{F}^b(p, s)$ with

$f_j(\mathbf{x}) = \mathbf{x}^T\tilde{\boldsymbol{\beta}}_j$, where $\tilde{\boldsymbol{\beta}}_j = (\beta_{0j}, \boldsymbol{\beta}_j^T)^T$, we check that the **one-dimensional** function $\tilde{f}(\tilde{\boldsymbol{x}}) = \sum_{j=1}^{k-1} f_j(\tilde{\boldsymbol{x}}_j) = \sum_{j=1}^{k-1} \tilde{\boldsymbol{x}}_j^T \tilde{\boldsymbol{\beta}}_j$ is $\tilde{Q}$-measurable and $\|\tilde{f}\|_{\tilde{Q},2}^2 = (k-1)^{-1}\|\boldsymbol{f}\|_{Q,2}^2$. This implies that for any $\boldsymbol{f}_1, \boldsymbol{f}_2 \in \mathcal{F}^b(p,s)$,

$$\|h_{\boldsymbol{f}_1} - h_{\boldsymbol{f}_2}\|_{Q,2} \leq s_0^{-1}\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{Q,2} = s_0^{-1}\sqrt{k-1}\|\tilde{f}_1 - \tilde{f}_2\|_{\tilde{Q},2} \tag{14}$$

On the other hand, by Cauchy—Schwarz inequality,

$$\begin{aligned}\sum_{j=1}^{k-1}\|\tilde{\boldsymbol{\beta}}_j\|_1 = \sum_{j=1}^{k-1}\|\boldsymbol{\beta}_j\|_1 + \|\boldsymbol{b}\|_1 &\leq s + \sqrt{k-1}\|\boldsymbol{b}\| \\ &\leq s + 2\sqrt{k}(k-1)(c_0+1)(C_m+s) \leq \sqrt{k-1}s_0.\end{aligned}$$

Thus it suffices to consider the entropy for the one-dimensional function space

$$\tilde{\mathcal{F}} = \{\tilde{f}; \tilde{f}(\tilde{\boldsymbol{x}}) = \sum_{j=1}^{k-1}\sum_{l=0}^{p}\beta_{lj}\tilde{x}_{jl}, \sum_{l,j}|\beta_{lj}| \leq \sqrt{k-1}s_0\},$$

where $\tilde{x}_{jl}$ is the lth element of $\tilde{\boldsymbol{x}}_j$. By Assumption 4, it can be verified that $|\tilde{x}_{jl}| \leq 1$.

Observe that for any $\tilde{f} \in \tilde{\mathcal{F}}$, $\tilde{f}(\tilde{\boldsymbol{x}}) = \sum_{j=1}^{k-1}\sum_{l=0}^{p}\frac{\beta_{lj}}{\sqrt{k-1}s_0}\left(\text{sign}(\beta_{lj})\tilde{f}_{jl}(\tilde{\boldsymbol{x}})\right)$, where $\tilde{f}_{jl}(\tilde{\boldsymbol{x}}) = \sqrt{k-1}s_0\tilde{x}_{jl}$. In fact, $\mathcal{J} = \{\pm\tilde{f}_{jl}\}$ forms a bisis for $\tilde{\mathcal{F}}$ with $\text{diam}\mathcal{J} \leq 2\sqrt{k-1}s_0$. Check that $\sum_{j=1}^{k-1}\sum_{l=0}^{p}\frac{|\beta_{lj}|}{\sqrt{k-1}s_0} \leq 1$. So $\tilde{\mathcal{F}}$ is a convex hull of $\mathcal{J}$, i.e., $\tilde{\mathcal{F}} = \text{conv}\mathcal{J}$. By applying Lemma 2.6.11 of [48], the covering number of $\text{conv}\mathcal{J}$, $N\left(\epsilon\text{diam}\mathcal{J}, \text{conv}\mathcal{J}, L_2(\tilde{Q})\right) \leq \left(e + 2(k-1)(p+1)\epsilon^2 e\right)^{2/\epsilon^2}$.

Finally, by (14), we have

$$\begin{aligned}N\left(2(k-1)\epsilon, \mathcal{H}, L_2(Q)\right) &\leq N\left(2\sqrt{k-1}s_0\epsilon, \tilde{\mathcal{F}}, L_2(\tilde{Q})\right) \\ &\leq \left(e + 2(k-1)(p+1)\epsilon^2 e\right)^{2/\epsilon^2}.\end{aligned}$$

Since it is true for any Q, we conclude that $N(\epsilon, \mathcal{H}) \leq \left(e + e\frac{p+1}{2(k-1)}\epsilon^2\right)^{8(k-1)^2/\epsilon^2}$ and the result follows by taking the logarithm. □

By combining the results from Lemma 5 and 7, we can prove Theorem 4.

Proof of Theorem 4. We start from Lemma 5 to find an upper bound for the estimation error (the first quantity). Take M in Lemma 5 as $M = 4\tau_n \log\tau_n^{-1}$, where $\tau_n = (n^{-1}\log p_n)^{1/2} \to 0$ as $n \to \infty$. Check that $nM^2 \to \infty$ so the tail probability goes to 0. Thus we only need to check condition (12) and (13) before we draw the conclusion that $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(p_n,s_n)}) \leq \mathcal{O}_p(16c_0s_0^2M)$.

We first check condition (13). Let $\epsilon_0 = \mathcal{O}\left(\log^{-1}\tau_n^{-1}\right)$. By Lemma 7, we have $H(\epsilon_0, \mathcal{H}) \leq \frac{8(k-1)^2}{\epsilon_0^2}\log\left(e + e\frac{p+1}{2(k-1)}\epsilon_0^2\right)$. So by choosing such ϵ_0, for the left hand side of (13),

$$H(\epsilon_0, \mathcal{H}) \leq \mathcal{O}\left(\log^2 \tau_n^{-1} \log\left(p_n \epsilon_0^2\right)\right) \leq \mathcal{O}\left(\log p_n \log^2 \tau_n^{-1}\right).$$

For the right hand side of (13), check that $nM^2 = 4 \log p_n \log^2 \tau_n^{-1}$. Therefore condition (13) holds.

Next we check condition (12). Notice that

$$\mathcal{O}\left(\log_2 \frac{\epsilon_0}{M}\right) \leq \mathcal{O}(\log M^{-1}) \leq \mathcal{O}(\log \tau_n^{-1}) = \mathcal{O}(\epsilon_0^{-1}).$$

Thus the left hand side of (12) is smaller than $\mathcal{O}(1)$ so condition (12) also holds.

With condition (12) and (13) verified, we let $\theta = 1/2$ and check that

$$nM^2 = 4 \log p_n \log^2 \tau_n^{-1} \geq 4 \log p_n \log \tau_n^{-1} = 2 \log p_n \log \frac{n}{\log p_n} \geq 2 \log n$$

for sufficiently large n. This implies that $\exp(-nM^2) \leq \exp(-2 \log n) = n^{-2}$. So we can apply Borel-Cantelli Lemma and have an even stronger statement that $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(p_n, s_n)}) \leq \mathcal{O}(16 c_0 s_0^2 M) = \mathcal{O}((C_m + s_n)^2 \tau_n \log \tau_n^{-1})$.

Finally, check that $r_1 = \left(C_m + \sup_{\boldsymbol{f} \in \mathcal{F}^b(p_n, s_n)} \|\boldsymbol{f}\|_\infty\right)^2 r_p \leq s_0 r_p = \mathcal{O}((C_m + s_n)^2 r_p)$ and the result follows by applying Lemma 3. □

Proof of Corollary 1. If either $r_p = 0$ or $r_p = 0$, the second term in $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ vanishes. Since the true treatment effect $\delta_j(\cdot)$ depends on finite many covariates, there exists a finite p^* and s^* such that $\boldsymbol{f}^* \in \mathcal{F}(p, s)$ when $p \geq p^*$ and $s \geq s^*$. Then by taking $s_n = s^*$ we have $d_n = 0$ and the result follows from Theorem 4. □

C.2. Prediction error for RKHS learning

The goal of this section is to prove Theorem 5, where the function space is $\mathcal{F} = \mathcal{F}(s)$. The schema for the proof of kernel learning is similar to that of linear learning. That is, we first find an upper bound for the estimation error of $\hat{\boldsymbol{f}}$ given by $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(s)})$, where $\boldsymbol{f}^{(s)} \in \operatorname{argmin}_{\boldsymbol{f} \in \mathcal{F}(s)} L(\boldsymbol{f})$, then apply Lemma 3 to finish the proof.

Take any $\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T \in \mathcal{F}(s)$ where $f_j = f_j' + b_j$ with $f_j' \in \mathcal{H}_K$ and $b_j \in \mathbb{R}$. From RKHS theory [50, 38, 46], $\langle K(\boldsymbol{x}, \cdot), f_j' \rangle_{\mathcal{H}_K} = f_j'(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{X}$. So

$$\|\boldsymbol{f}'(\boldsymbol{x})\|^2 = \sum_{j=1}^{k-1} f_j'^2(\boldsymbol{x}) = \sum_{j=1}^{k-1} \langle K(\boldsymbol{x}, \cdot), f_j'(\boldsymbol{x}) \rangle_{\mathcal{H}_K}^2 \leq \sum_{j=1}^{k-1} |K(\boldsymbol{x}, \cdot)|^2 \cdot \|f_j'\|_{\mathcal{H}_K}^2 \leq B^2 s^2$$

implying $\sup \|\boldsymbol{f}'\|_\infty \leq Bs$. On the other hand, by Lemma 4 and denoting $\hat{\boldsymbol{b}}$ as the intercept term of $\hat{\boldsymbol{f}}$, we have $\|\hat{\boldsymbol{b}}\| \leq \sqrt{2k(k-1)}(c_0 - 1)(C_m + Bs)$, where

c_0 is an upper bound for $\hat{p}_j^{-1}(\boldsymbol{x})$. So instead of $\mathcal{F}(s)$, it suffices to consider the following function space:

$$\mathcal{F}^b(s) = \mathcal{F}(s) \cap \{\boldsymbol{f}; \|\boldsymbol{b}\| \leq \sqrt{2k(k-1)}(c_0 - 1)(C_m + Bs)\}.$$

Next, as in linear learning, let $g_{\boldsymbol{f}}(\boldsymbol{Z}) = \frac{1}{\hat{p}_A(\boldsymbol{X})}\left(Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle\right)^2$. It can be checked that $|g_{\boldsymbol{f}}| \leq c_0 s_0^2$ holds for any $\boldsymbol{f} \in \mathcal{F}^b(s)$, where

$$s_0 = \left((c_0 - 1)\sqrt{2k(k-1)} + 1\right)\left(C_m + Bs\right)$$

is an upper bound for $|Y - \hat{m}(\boldsymbol{X}) - \langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X})\rangle|$. Let

$$\mathcal{H} = \{\left(2c_0 s_0^2\right)^{-1}\left(g_{\boldsymbol{f}} - g_{\boldsymbol{f}^{(s)}}\right); \boldsymbol{f} \in \mathcal{F}^b(s)\}.$$

It can be checked that $\boldsymbol{f}^{(s)} \in \mathcal{F}^b(s)$ so $|h_{\boldsymbol{f}}| \leq 1$ for any $h_{\boldsymbol{f}} \in \mathcal{H}$.

Finally, let $T_{\boldsymbol{Z}}$ be the empirical distribution of the training set $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$. Denote the entropy of $\mathcal{H}$ under $T_{\boldsymbol{Z}}$ as $H(\epsilon, \mathcal{H}, L_2(T_{\boldsymbol{Z}}))$ and let $H(\epsilon, \mathcal{H}) = \sup_{T_{\boldsymbol{Z}}} H(\epsilon, \mathcal{H}, L_2(T_{\boldsymbol{Z}}))$.

Lemma 8 gives an upper bound for $H(\epsilon, \mathcal{H})$.

Lemma 8. *For sufficient small $\epsilon > 0$, there is a constant $C_0 > 0$, such that $H(\epsilon, \mathcal{H}) \leq C_0 \epsilon^{-2}$.*

Proof of Lemma 8. By following the similar argument in the proof of Lemma 7, we observe that for any $\boldsymbol{f}_1, \boldsymbol{f}_2 \in \mathcal{F}^b(s)$,

$$\|h_{\boldsymbol{f}_1} - h_{\boldsymbol{f}_2}\|_{T_{\boldsymbol{Z}},2}^2 \leq s_0^{-2}\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{T_{\boldsymbol{Z}},2}^2. \tag{15}$$

So the rest of the work is to find the entropy number for $\mathcal{F}^b(s)$. Since $\mathcal{F}^b(s)$ is a $(k-1)$-dimensional function, it is difficult to compute its entropy directly. To address this issue, we start with the entropy of one-dimensional RKHS.

Let $\mathcal{F}_1 = \cdots = \mathcal{F}_{k-1} = \{f; f = f' + b, f' \in \mathcal{H}_K, |f'| + |b| \leq s_0\}$. Notice that this is a function space with penalized constant term. The covering number of such function space is given in [60], i.e.,

$$\sup_{T_{\boldsymbol{Z}}} N(2s_0\epsilon, \mathcal{F}_1, L_2(T_{\boldsymbol{Z}})) \leq \frac{5\exp(C'\epsilon^{-2})}{\epsilon}$$

for sufficient small $\epsilon > 0$ and some $C' > 0$. One can verify that $\mathcal{F}^b(s) \subseteq \mathcal{F}_1 \times \cdots \times \mathcal{F}_{k-1} \triangleq \mathcal{F}^{k-1}$. Next we will cover $\mathcal{F}^{k-1}$ by constructing a coverage for each $\mathcal{F}_j$.

Let $\mathcal{G}_j$ be a minimal $(k-1)^{-1/2}\epsilon$-net for $\mathcal{F}_j$. That is, for any $f_j \in \mathcal{F}_j$, there exists a $g_j \in \mathcal{G}_j$, such that $\|f_j - g_j\|_{T_{\boldsymbol{Z}},2} \leq (k-1)^{-1/2}\epsilon$. Take an arbitrary $\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T \in \mathcal{F}^{k-1}$. Then there is a $\boldsymbol{g} = (g_1, \ldots, g_{k-1})^T$ where $g_j \in \mathcal{G}_j$, such that

$$\|\boldsymbol{g} - \boldsymbol{f}\|_{T_{\boldsymbol{Z}},2}^2 = \sum_{j=1}^{k-1}\|f_j - g_j\|_{T_{\boldsymbol{Z}},2}^2 \leq \epsilon^2.$$

By construction we have $\boldsymbol{g} \in \mathcal{F}^{k-1}$. So we can cover $\mathcal{F}^{k-1}$ by these ϵ-nets centered at such $\boldsymbol{g}$. This implies that

$$N\left(\epsilon, \mathcal{F}^{k-1}, L_2(T_{\boldsymbol{Z}})\right) \leq N\left((k-1)^{-1/2}\epsilon, \mathcal{F}_1, L_2(T_{\boldsymbol{Z}})\right)^{k-1}.$$

Combining the results with (15), we have

$$\begin{aligned} N\left(2(k-1)^{1/2}\epsilon, \mathcal{H}, L_2(T_{\boldsymbol{Z}})\right) &\leq N\left(2s_0(k-1)^{1/2}\epsilon, \mathcal{F}^b(s), L_2(T_{\boldsymbol{Z}})\right) \\ &\leq N\left(2s_0\epsilon, \mathcal{F}_1, L_2(T_{\boldsymbol{Z}})\right)^{k-1}. \end{aligned}$$

Note that the inequality holds for any $T_{\boldsymbol{Z}}$. Therefore

$$\sup_{T_{\boldsymbol{Z}}} N\left(2(k-1)^{1/2}\epsilon, \mathcal{H}, L_2(T_{\boldsymbol{Z}})\right) \leq \left(\frac{5}{\epsilon}\right)^{k-1} \exp\left(\frac{(k-1)C'}{\epsilon^2}\right).$$

By taking the logarithm,

$$H(\epsilon, \mathcal{H}) \leq \frac{k-1}{2} \log \frac{100(k-1)}{\epsilon^2} + \frac{4(k-1)^2 C'}{\epsilon^2}.$$

Because $\mathcal{O}(\log \epsilon^{-2}) < \mathcal{O}(\epsilon^{-2})$, there is a constant $C_0 > 0$, such that $H(\epsilon, \mathcal{H}) \leq C_0\epsilon^{-2}$ holds for sufficient small $\epsilon > 0$. □

Now we can prove Theorem 5 using Lemma 3, Lemma 5, and Lemma 8. Note that though we state Lemma 5 in linear learning, after replacing each counterpart, the proof is the same and the conclusion is still valid.

Proof of Theorem 5. We start with the upper bound of the estimation error given by $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(s)})$. By Lemma 5, $L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(s)}) = \mathcal{O}_p(16c_0s_0^2M)$. Then it remains to choose an appropriate M that satisfies condition (12) and condition (13).

Let $M = 2n^{-1/2}\log n$ and take $\epsilon_0 = \mathcal{O}(\log^{-1} n)$. By Lemma 8, $H(\epsilon_0, \mathcal{H}) \leq \mathcal{O}(\log^2 n)$. Since $nM^2 = 2\log^2 n$, condition (13) holds. For condition (12), check that

$$\mathcal{O}\left(\log_2 \frac{\epsilon_0}{M}\right) \leq \mathcal{O}(\log M^{-1}) \leq \mathcal{O}(\log n) = \mathcal{O}(\epsilon_0^{-1}).$$

So the left hand side of (12) is smaller than $\mathcal{O}(1)$ thus condition (12) holds.

With condition (12) and (12) verified, let $\theta = 1/2$ and check that $nM^2 = 2\log^2 n \geq 2\log n$ for sufficient large n. So $\exp(-nM^2) \leq \exp(-2\log n) = n^{-2}$. By applying Borel-Cantelli Lemma we state that

$$L(\hat{\boldsymbol{f}}) - L(\boldsymbol{f}^{(s_n)}) \leq \mathcal{O}\left((C_m + Bs_n)^2 n^{-1/2}\log n\right).$$

Finally, check that $r_1 = \left(C_m + \sup_{\boldsymbol{f}\in\mathcal{F}^b(s_n)} \|\boldsymbol{f}\|_\infty\right)^2 r_p \leq s_0 r_p = \mathcal{O}((C_m + Bs_n)^2 r_p)$ and the result follows by applying Lemma 3. □

Proof of Corollary 2. If either $r_p = 0$ or $r_m = 0$, the second term in $\mathrm{PE}(\hat{\boldsymbol{\delta}})$ vanishes. If we further assume $d_n \le \mathcal{O}(s_n^{-q})$ for some $q > 0$, we have

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \le \mathcal{O}\left(\max\left\{s_n^2 n^{-1/2}\log n, s_n^{-q}\right\}\right).$$

Note that s_n is the tuning parameter to balance the estimation error and the approximation error, and the optimal s_n is chosen such that these two are the same. In this case, we let

$$s_n = \mathcal{O}\left((n^{1/2}\log^{-1} n)^{\frac{1}{q+2}}\right).$$

Then $\mathrm{PE}(\hat{\boldsymbol{\delta}}) \le \mathcal{O}\left((n^{-1/2}\log n)^{\frac{q}{q+2}}\right)$. Notice that $\mathcal{O}(\log n) < \mathcal{O}(n^t)$ for any $t > 0$. For better displaying the result we may let $t = (4q+10)^{-1}$. Check that

$$\mathrm{PE}(\hat{\boldsymbol{\delta}}) \le \mathcal{O}\left(n^{-\frac{q}{2(q+2)}+\frac{q}{q+2}t}\right) = \mathcal{O}\left(n^{-\frac{q}{2q+5}}\right)$$

and the proof is complete. □

References

[1] Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* **113** 7353–7360. MR3531135

[2] Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–973. MR2216189

[3] Beygelzimer, A. and Langford, J. (2009). The offset tree for learning with partial labels. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* 129–138.

[4] Bonetti, M. and Gelber, R. D. (2004). Patterns of treatment effects in subsets of patients in clinical trials. *Biostatistics* **5** 465–481.

[5] Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugaly, E., Ray, D., Simard, P. and Snelson, E. (2013). Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research* **14** 3207–3260. MR3144461

[6] Cao, W., Tsiatis, A. A. and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96** 723–734. MR2538768

[7] Chatterjee, S. (2013). Assumptionless consistency of the lasso. arXiv preprint arXiv: 1303.5817.

[8] Chen, S., Tian, L., Cai, T. and Yu, M. (2017). A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics* **73** 1199–1209. MR3744534

[9] Chipman, H. A., George, E. I. and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics* **4** 266–298. MR2758172

[10] Dalalyan, A. S., Hebiri, M., Lederer, J. et al. (2017). On the prediction performance of the lasso. *Bernoulli* **23** 552–581. MR3556784

[11] Dudík, M., Langford, J. and Li, L. (2011). Doubly robust policy evaluation and learning. arXiv preprint arXiv: 1103.4601.

[12] Fan, C., Lu, W., Song, R. and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 1565–1582. MR3731676

[13] Fan, J., Imai, K., Liu, H., Ning, Y. and Yang, X. (2016). Improving covariate balancing propensity score: A doubly robust and efficient approach Technical Report, Technical report, Princeton Univ.

[14] Hahn, P. R., Murray, J. S. and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. *Bayesian Analysis*. MR4154846

[15] Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M. et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* **335** 1081–1090.

[16] Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics* **20** 217–240. MR2816546

[17] Hofmann, T., Schölkopf, B. and Smola, A. J. (2008). Kernel methods in machine learning. *The annals of statistics* 1171–1220. MR2418654

[18] Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press. MR3309951

[19] Johansson, F., Shalit, U. and Sontag, D. (2016). Learning representations for counterfactual inference. In *International conference on machine learning* 3020–3029.

[20] Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science* **22** 523–539. MR2420458

[21] Knaus, M. C., Lechner, M. and Strittmatter, A. (2020). Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal*. utaa014. MR4228632

[22] Kosorok, M. R. and Laber, E. B. (2019). Precision medicine. *Annual review of statistics and its application* **6** 263–286. MR3939521

[23] Künzel, S. R., Sekhon, J. S., Bickel, P. J. and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences* **116** 4156–4165.

[24] Moodie, E. E., Dean, N. and Sun, Y. R. (2014). Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences* **6** 223–243.

[25] Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 331–355. MR1983752

[26] Murphy, S. A., van der Laan, M. J., Robins, J. M. and Group, C. P. P. R. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* **96** 1410–1423. MR1946586

[27] Nie, X. and Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. arXiv preprint arXiv: 1712.04912. MR4259133

[28] Powers, S., Qian, J., Jung, K., Schuler, A., Shah, N. H., Hastie, T. and Tibshirani, R. (2018). Some methods for heterogeneous treatment effect estimation in high dimensions. *Statistics in medicine* **37** 1767–1787. MR3799840

[29] Qi, Z., Liu, D., Fu, H. and Liu, Y. (2019). Multi-Armed Angle-Based Direct Learning for Estimating Optimal Individualized Treatment Rules With Various Outcomes. *Journal of the American Statistical Association* 1–33. MR4107672

[30] Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics* **12** 3601–3638. MR3870507

[31] Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics* **39** 1180. MR2816351

[32] Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics* 189–326. Springer. MR2129402

[33] Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association* **89** 846–866. MR1294730

[34] Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society* 931–954. MR0951762

[35] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

[36] Royston, P. and Sauerbrei, W. (2008). Interactions between treatment and continuous covariates: a step toward individualizing therapy.

[37] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66** 688.

[38] Scholkopf, B. and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

[39] Schulz, J. and Moodie, E. E. (2021). Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association* **116** 256–268. MR4227692

[40] Shi, C., Song, R. and Lu, W. (2016). Robust learning for optimal treatment decision with NP-dimensionality. *Electronic journal of statistics* **10** 2894. MR3557316

[41] Signorovitch, J. E. (2007). Identifying informative biological markers in high-dimensional genomic data and clinical trials, PhD thesis, Harvard University.

[42] Steinwart, I. and Scovel, C. (2007). Fast rates for support vector machines using Gaussian kernels. *The Annals of Statistics* **35** 575–607. MR2336860

[43] Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M. and Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research* **10**.

[44] Taddy, M., Gardner, M., Chen, L. and Draper, D. (2016). A nonparametric bayesian analysis of heterogenous treatment effects in digital experimentation. *Journal of Business & Economic Statistics* **34** 661–672. MR3548002

[45] Tian, L., Alizadeh, A. A., Gentles, A. J. and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **109** 1517–1532. MR3293607

[46] Trevor, H., Robert, T. and JH, F. (2009). The elements of statistical learning: data mining, inference, and prediction. MR2722294

[47] Turney, K. and Wildeman, C. (2015). Detrimental for some? Heterogeneous effects of maternal incarceration on child wellbeing. *Criminology & Public Policy* **14** 125–156.

[48] Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer. MR1385671

[49] Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* **113** 1228–1242. MR3862353

[50] Wahba, G. (1990). *Spline models for observational data* **59**. Siam. MR1045442

[51] Wallace, M. P. and Moodie, E. E. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71** 636–644. MR3402599

[52] Wang, B. and Zou, H. (2018). Another look at distance-weighted discrimination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 177–198. MR3744717

[53] Wang, L. and Shen, X. (2007). On L1-norm multiclass support vector machines: methodology and theory. *Journal of the American Statistical Association* **102** 583–594. MR2370855

[54] Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning* **8** 279–292.

[55] Weisberg, H. I. and Pontes, V. P. (2015). Post hoc subgroups in clinical trials: Anathema or analytics? *Clinical trials* **12** 357–364.

[56] Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007

[57] Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445

[58] Zhang, C., Chen, J., Fu, H., He, X., Zhao, Y. and Liu, Y. (2018). Multicategory Outcome Weighted Margin-based Learning for Estimating Individualized Treatment Rules. *Statistica Sinica*. MR4260747

[59] Zhang, C. and Liu, Y. (2014). Multicategory angle-based large-margin

classification. *Biometrika* **101** 625–640. MR3254905

[60] Zhang, C., Liu, Y. and Wu, Y. (2016). On quantile regression in reproducing kernel Hilbert spaces with the data sparsity constraint. *The Journal of Machine Learning Research* **17** 1374–1418. MR3491134

[61] Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* **107** 1106–1118. MR3010898

[62] Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S. and Sands, B. E. (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Research* **20** 1–23. MR3948088

[63] Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M. and Kosorok, M. R. (2014). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* **102** 151–168. MR3335102

[64] Zhou, X., Mayer-Hamblett, N., Khan, U. and Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association* **112** 169–187. MR3646564