Published in final edited form as:
J Am Stat Assoc. 2021 ; 116(534): 659–674. doi:10.1080/01621459.2020.1796359.

Learning Optimal Distributionally Robust Individualized Treatment Rules

Weibin Mo,
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Zhengling Qi,
Department of Decision Sciences, George Washington University, Washington, D.C. 20052, USA

Yufeng Liu
Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Science, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC 27599, USA

Abstract

Recent development in the data-driven decision science has seen great advances in individualized decision making. Given data with individual covariates, treatment assignments and outcomes, policy makers best individualized treatment rule (ITR) that maximizes the expected outcome, known as the value function. Many existing methods assume that the training and testing distributions are the same. However, the estimated optimal ITR may have poor generalizability when the training and testing distributions are not identical. In this paper, we consider the problem of finding an optimal ITR from a restricted ITR class where there is some unknown covariate changes between the training and testing distributions. We propose a novel distributionally robust ITR (DR-ITR) framework that maximizes the worst-case value function across the values under a set of underlying distributions that are "close" to the training distribution. The resulting DR-ITR can guarantee the performance among all such distributions reasonably well. We further propose a calibrating procedure that tunes the DR-ITR adaptively to a small amount of calibration data from a target population. In this way, the calibrated DR-ITR can be shown to enjoy better generalizability than the standard ITR based on our numerical studies.

Keywords

Covariate shifts; Distributionally robust optimization; Generalizability; Individualized treatment rules

Yufeng Liu is Professor, yfliu@email.unc.edu.
Weibin Mo and Zhengling Qi are co-first authors for the paper.

1 Introduction

Data-driven individualized decision making problems are commonly seen in practice and have been studied intensively in the literature. In disease management, the physician may decide whether to introduce or switch a therapy for a patient based on his/her characteristics in order to achieve a better clinical outcome (Bertsimas et al., 2017). In public policy making, a policy that allocates the resource based on the characteristics of the targets can improve the overall resource allocation efficiency (Kube et al., 2019). In a context-based recommender system, the use of the contextual information such as time, location and social connection can increase the effectiveness of the recommendation process (Aggarwal, 2016). One common goal of these problems is to find the optimal *individualized treatment rule (ITR)* mapping from the individual characteristics or contextual information to the treatment assignment, that maximizes the expected outcome, known as the *value function* (Manski, 2004; Qian and Murphy, 2011).

One approach for estimating an optimal ITR is to first estimate the conditional mean outcome, known as the *Q-function*, given the individual characteristics and the treatment assignment, and then induce the ITR that prescribes the treatment by maximizing the estimated Q-function (Qian and Murphy, 2011). In the binary treatment case, such an approach can be reformulated as estimating the *conditional treatment effect (CTE)* as the difference of the conditional mean outcomes under two candidate treatments (Zhao et al., 2017; Chen et al., 2017; Qi et al., 2020). Another approach is to directly estimate the value function using the *inverse-probability weighted estimator (IPWE)*, and then search for the ITR that maximizes the corresponding value function (Zhao et al., 2012; Kitagawa and Tetenov, 2018; Liu et al., 2018; Zhang et al., 2019). Since there are potential model misspecification issues of these approaches, the *augmented IPWE (AIPWE)* of the value function combines the estimates of the Q-function and the treatment propensity score. AIPWE is *doubly robust* in the sense that the consistency of the value function estimate is guaranteed as long as either the Q-function model or the propensity score model is correctly specified (Dudík et al., 2011; Zhang et al., 2012b; Athey and Wager, 2017; Zhao et al., 2019a). While the doubly robust property can protect against the violation of the model assumptions, one key assumption behind is that the training and testing distributions should be identical.

When the training and testing distributions are different, an estimated optimal ITR may not generalize well on the testing data (Zhao et al., 2019b). Similar phenomenon for causal inference in randomized controlled trials (RCTs) has also been pointed out by Muller (2014); Gatsonis and Morton (2017). Specifically, due to the inclusion and exclusion criteria of an RCT, the training sample can be unrepresentative of the testing population we are interested in. Therefore, the corresponding casual evidence may not be broadly applicable or relevant for the real-world practice. In causal inference literature, it is common to regard the training data as a selected sample from the pooled population of training and testing. The selection bias can be adjusted by reweighing or stratifying the training data according to the relationship between training and testing (O’Muircheartaigh and Hedges, 2014; Buchanan et al., 2018). However, it requires strong assumptions on completely measuring the selection confounders and correctly specifying the selection model, and thus can only work well on a prespecified testing population. There are many other practical scenarios where the difference between the training and testing distributions is unknown. One example is that the training data can be confounded by some unidentified effects such as batch effects, which may cause potential covariate shifts (Luo et al., 2010). Another possibility is that the testing distribution may evolve over time (Hand, 2006). There is also a widely studied scenario that multiple datasets are aggregated to perform combined analysis (Alyass et al., 2015; Shi et al., 2018; Li et al., 2020). Aggregating data from various sources can benefit from sharing common information, transferring knowledge from different but related samples, and maintaining certain privacy. However, due to the heterogeneity among data sources, standard approaches of finding pooled optimal ITRs may not generalize well on all these sources. One way of handling the heterogeneity is to formulate it as a problem of distributional changes, where we train on the mixture of subpopulations while testing on one of the subpopulations (Duchi et al., 2019). In all these applications, an optimal ITR that is robust to unattended distributional differences is of great interest.

Despite a vast literature in ITR, much less work has been done on the problem when the training and testing distributions are different. Imai and Ratkovic (2013) and Johansson et al. (2018) estimated the CTE function by reweighing the training loss to ensure the estimators generalizable on a prespecified testing distribution. Zhao et al. (2019b) aimed to find an ITR that optimizes the worst-case quality assessment among all testing covariate distributions satisfying some moment conditions. However, since their method only requires some moment conditions, the uncertainty set of the testing distributions can be very large. Recent developments in the distributionally robust optimization (DRO) literature provide the opportunities to quantify the difference between the training and testing distributions more precisely (Ben-Tal et al., 2013; Duchi and Namkoong, 2018; Rahimian and Mehrotra, 2019). Motivated by the DRO literature, we develop a new robust optimal ITR framework in this paper.

In this paper, we consider the problem of finding an optimal ITR from a restricted ITR class, where there is some unknown covariate changes between the training and testing distributions. We propose to use the *distributionally robust ITR (DR-ITR)* that maximizes the defined worst-case value function among value functions under a set of underlying distributions. More specifically, value functions are evaluated under all testing covariate distributions that are “close” to the training distribution, and the worst-case situation takes a minimal one. Our distributionally robust ITR framework is different from the existing doubly robust ITR framwork that uses an AIPWE. In particular, an AIPWE robustifies the model specification assumptions, while our DR-ITR robustifes the underlying distributions. The DR-ITR aims to guarantee reasonable performance across all testing distributions in an uncertainty set around the training distribution by optimizing the worst-case scenarios. In particular, we parameterize the amount of “closeness” by the *distributional robustness-constant (DR-constant)*, where the smallest possible DR-constant corresponds to the *standard ITR* that maximizes the value function under the training distribution. To ensure the performance of the DR-ITR on a specific testing distribution, we fit a class of DR-ITRs for a spectrum of DR-constants at the training stage, and calibrate the DR-constant based on a small amount of the calibrating data from the testing distribution. In this way, the correctly calibrated DR-constant ensures that the DR-ITR performs at least as well as, often much better than, the standard ITR. Using our illustrative example, we show that the standard ITR can have very poor values on many testing distributions, while our calibrated DR-ITRs still maintain relatively good performance. In particular, our proposed calibrating procedures can tune DR-constants based on the small calibrating sample. To solve the worst-case optimization problem, we make use of the difference-of-convex (DC) relaxation of the nonsmooth indicator, and propose two algorithms to solve the related nonconvex optimization problems. We also provide the finite sample regret bound for the proposed DR-ITR.

The rest of this paper is organized as follows. In Section 2, we discuss an illustrative example that the optimality of an ITR can be sensitive to the underlying distribution, and introduce the DR-ITR that can generalize well across all testing distributions considered in this example. Then we propose the DR-ITR framework and the corresponding learning problem. In Section 3, we justify the theoretical guarantees of the finite sample approximations for the learning problem. In Section 4, we evaluate the generalizability of our proposed DR-ITR on two simulation studies: the problem of covariate shifts and the problem of mixture of multiple subgroups. We apply our proposed DR-ITR on the AIDS clinical dataset ACTG 175 and evaluate its generalizability on the subgroup of female patients in Section 5. Some related discussions and extensions are given in Section 6. The implementation details, technical proofs and some additional numerical results are all given in the Supplementary Material.

2 Methodology

In this section, we introduce the value maximization framework in the current literature, and discuss its limitation when the training and testing distributions are different. Then we propose the DR-value function that optimizes the worse-case value function across all distributions within an uncertainty set around the training distribution.

2.1 Maximizing the Value Function

Consider the training data $(\boldsymbol{X}, A, Y) \sim \mathbb{P}$, where $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ denotes the covariates, $A \in \mathcal{A} = \{+1, -1\}$ is the binary treatment assignment, and $Y \in \mathcal{Y} \subseteq \mathbb{R}$ is the observed outcome. We assume that the larger outcome is better. Let $Y(+1)$, $Y(-1)$ be the potential outcomes. Consider a prespecified ITR class $\mathcal{D} \subseteq \{\pm 1\}^{\mathcal{X}}$. For $d \in \mathcal{D}$, denote $Y(d) := Y(1)\mathbb{1}[d(\boldsymbol{X}) = 1] + Y(-1)\mathbb{1}[d(\boldsymbol{X}) = -1]$ as the potential outcome following the treatment assignment prescribed by the ITR *d*. Then the value function under the training distribution $\mathbb{P}$ is defined as

$$\mathcal{V}(d) := \mathbb{E}[Y(d)] .$$

Denote $\pi(a \mid \boldsymbol{x}) := \mathbb{P}(A = a \mid \boldsymbol{X} = \boldsymbol{x})$ as the training propensity score function for treatment assignment. If we assume 1) the *consistency* of the observed outcome $Y = Y(A)$; 2) the *strict overlap* $\pi(\pm 1 \mid \boldsymbol{x}) \geqslant \tau > 0$ for any $\boldsymbol{x} \in \mathcal{X}$; and 3) the *strong ignorability*

$(Y(1+), Y(-1)) \perp\!\!\!\perp A \mid \boldsymbol{X}$ (Rubin, 1974), then we can identify $\mathscr{V}(d)$ in terms of the observed data $(\boldsymbol{X}, A, Y)$ by the IPWE of $\mathbb{E}\left(\frac{\mathbb{1}[d(\boldsymbol{X})=A]}{\pi(A \mid \boldsymbol{X})} Y\right)$.

Instead of targeting the value function directly, we instead consider the CTE function as $C(\boldsymbol{x}) := \mathbb{E}[Y(+1) - Y(-1) \mid \boldsymbol{X} = \boldsymbol{x}]$ under the training distribution $\mathbb{P}$. Note that for an ITR d and all $\boldsymbol{x} \in \mathscr{X}$, the prescribed treatment assignment satisfies $d(\boldsymbol{x}) \in \{\pm 1\}$. Then we have $C(\boldsymbol{x})d(\boldsymbol{x}) = \mathbb{E}[Y(d) - Y(-d) \big| \boldsymbol{X} = \boldsymbol{x}]$. Based on this representation, we define another value function

$$\mathscr{V}_1(d) := \mathbb{E}[C(\boldsymbol{X})d(\boldsymbol{X})] = \mathbb{E}[Y(d) - Y(-d)]\,. \tag{1}$$

Since $Y(d) + Y(-d) \equiv Y(1) + Y(-1)$, it can be observed that $\mathscr{V}_1(d) = 2\left[\mathscr{V}(d) - \frac{\mathbb{E}[Y(+1) + Y(-1)]}{2}\right] = 2[\mathscr{V}(d) - \mathscr{V}(d_{\text{rand}})]$, where $d_{\text{rand}}(\boldsymbol{x}) = +1$ with probability $1/2$ and -1 with probability $1/2$. Therefore, $\mathscr{V}_1(d)$ can be interpreted as the value improvement of the ITR d upon the completely random treatment rule d_{rand}. In terms of the optimal ITR, the resulting rules by optimizing the value functions $\mathscr{V}_1(d)$ and $\mathscr{V}(d)$ over d are equivalent.

By the definition (1), we have $\mathscr{V}_1(d) \leqslant \mathbb{E}[\mid C(\boldsymbol{X}) \mid]$ with equality if $d(\boldsymbol{X}) = \text{sign}[C(\boldsymbol{X})]$ almost surely. Such an ITR is the global optimal ITR when $\mathscr{D}$ consists of all measurable functions from $\mathscr{X}$ to $\{\pm 1\}$. To obtain the global optimal ITR, we can estimate $C(\boldsymbol{X})$ from data using flexible nonparametric techniques, such as the Bayesian additive regression tree (BART) (Hill, 2011), or the casual forest (Wager and Athey, 2018). However, in general, the global optimal ITR $\boldsymbol{x} \mapsto \text{sign}[C(\boldsymbol{x})]$ can take a very complicated functional form, while decision makers may want to have a simpler ITR (Kitagawa and Tetenov, 2018). Then the ITR class $\mathscr{D}$ is often considered as a restricted subset of measurable functions from $\mathscr{X}$ to $\{\pm 1\}$. The following two-step procedure can be implemented to estimate the restricted optimal ITR on $\mathscr{D}$: first we estimate the CTE function $\boldsymbol{x} \mapsto \widehat{C}(\boldsymbol{x})$ using flexible nonparametric techniques; and then we estimate the ITR by solving $\max_{d \in \mathscr{D}} \mathbb{E}_n[\widehat{C}(\boldsymbol{X})d(\boldsymbol{X})]$ on the restricted ITR class $\mathscr{D}$ (Zhang et al., 2012a). Here, $\mathbb{E}_n$ is the empirical average based on the training data.

2.2 Covariate Changes

It can be observed that the value functions defined in Section 2.1 depend on the underlying distribution. Suppose we are interested in a testing distribution $\mathbb{P}_{\text{test}}$ that may be different from the training distribution $\mathbb{P}$ to some extent. Then ITRs estimated by most existing methods may not be able to perform well on our target population. In order to address this problem, we first make the following assumption on the potential difference between $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$.

Assumption 1 (Covariate Changes). For every training distribution $\mathbb{P}$ and testing distribution $\mathbb{P}_{\text{test}}$ considered in this paper, we assume the followings:

I. $\mathbb{P}_{\text{test}} \ll \mathbb{P}$;

II. There exists $w: \mathscr{X} \rightarrow \mathbb{R}_+$ such that $\mathbb{E}_{\mathbb{P}}w(\boldsymbol{X}) = 1$, and $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P} = w(\boldsymbol{X})$.

Assumption 1 (I) requires that the support of the testing distribution cannot go beyond the training distribution. Assumption 1 (II) is mathematically equivalent to assuming that the differences between $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ only appear in the covariate distributions. The treatment-response relationship conditional on covariates remains unchanged across training and testing distributions. Specifically, let $p_{\boldsymbol{X}}(\boldsymbol{x})p_{Y \mid \boldsymbol{X}}(y(1), y(-1) \mid \boldsymbol{x})$ and $q_{\boldsymbol{X}}(\boldsymbol{x})q_{Y \mid \boldsymbol{X}}(y(1), y(-1) \mid \boldsymbol{x})$ be the training and testing densities of the data $(\boldsymbol{X}, Y(1), Y(-1))$. Then the density ratio $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}$ becomes

$$\frac{\mathrm{d}\mathbb{P}_{\text{test}}}{\mathrm{d}\mathbb{P}} = \frac{q_{\boldsymbol{X}}(\boldsymbol{X})}{p_{\boldsymbol{X}}(\boldsymbol{X})} \times \frac{q_{Y \mid \boldsymbol{X}}(Y(1), TY(-1) \big| \boldsymbol{X})}{p_{Y \mid \boldsymbol{X}}(Y(1), TY(-1) \big| \boldsymbol{X})}.$$

If $q_{Y \mid \boldsymbol{X}}(Y(1), Y(-1) \mid \boldsymbol{X}) = p_{Y \mid \boldsymbol{X}}(Y(1), Y(-1) \mid \boldsymbol{X})$, *i.e.*, the conditional distributions $(Y(1), Y(-1)) \mid \boldsymbol{X}$ are identical under $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$, then $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P} = q_{\boldsymbol{X}}(\boldsymbol{X})/p_{\boldsymbol{X}}(\boldsymbol{X})$, which is the weighting function $w(\boldsymbol{X})$ in Assumption 1 (II).

The assumption of covariate changes is commonly seen in the setting of randomized trial. Consider the training and testing populations together as a pooled population with finite subjects. For each subject $i \in \{1, 2, \ldots, N\}$, let $S_i \in \{0, 1\}$ be a selection random variable such that $S_i = 1$ if i is a training sample point, and $S_i = 0$ if i is a testing sample point. Let the distributions of $(\boldsymbol{X}_i, Y_i(1), Y_i(-1)) \mid (S_i = 1)$ and $(\boldsymbol{X}_i Y_i(1), Y_i(-1)) \mid (S_i = 0)$ be the training distribution $\mathbb{P}$ and the testing distribution $\mathbb{P}_{\text{test}}$ respectively. Denote $\overline{\mathbb{P}}$ as the joint distribution of $(\boldsymbol{X}_i Y_i(1), Y_i(-1), S_i)$. Then conditions in Assumption 1 can correspond to the following (Hotz et al., 2005; Stuart et al., 2011):

- (Overlapping Support) $0 < \overline{\mathbb{P}}(S_i = 1 \mid \boldsymbol{X}_i) < 1$;
- (Selection Unconfoundedness) $S_i \perp\!\!\!\perp (Y_i(1), Y_i(-1)) \mid \boldsymbol{X}_i$.

In particular, under this finite population setting, the overlapping support condition is equivalent to that $\mathbb{P}_{\text{test}} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{P}_{\text{test}}$, and the selection unconfoundedness condition is equivalent to Assumption 1 (II). Such a correspondence can bring more intuitive implications of Assumption 1 under the randomized trial setting. Specifically, the overlapping support requires the chances of each subject being selected into the training and testing populations to be both positive. The selection unconfoundnedness requires that the selection mechanism is independent of the potential outcomes given the covariates. Both conditions can be satisfied by a successful trial design (Pearl and Bareinboim, 2014). The phenomenon of covariate changes between $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ can exist if $\overline{\mathbb{P}}(S_i = 1 \mid \boldsymbol{X}_i) \neq \overline{\mathbb{P}}(S_i = 0 \mid \boldsymbol{X}_i)$ with a positive probability. This can be often the case if the subject needs to satisfy certain requirements before enrolling a trial.

As a consequence from Assumption 1, the CTE function $C(\boldsymbol{X}) = \mathbb{E}_{\mathbb{P}}[Y(1) - Y(-1) \mid \boldsymbol{X}] = \mathbb{E}_{\text{test}}[Y(1) - Y(-1) \mid \boldsymbol{X}]$ remains unchanged under $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$.

Then it can be convenient to consider the value functions $\mathscr{V}_1(d) = \mathbb{E}_{\mathbb{P}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ and $\mathscr{V}_{1,\,\text{test}}(d) = \mathbb{E}_{\text{test}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ defined in (1). When the testing value function $\mathscr{V}_{1,\,\text{test}}(d)$ is of interest, maximizing the training value function $\mathscr{V}_1(d)$ may not be optimal. Alternatively, we can rewrite the testing value function $\mathscr{V}_{1,\,\text{test}}(d) = \mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})]$ where $w(\boldsymbol{X}) = \mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}$. Then based on the training data from $\mathbb{P}$, we can maximize $\mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})]$ that targets the correct objective. It amounts to determine the weighting function w that captures the differences between $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$.

Remark 1. Notice that for any weighting function $w: \mathscr{X} \to \mathbb{R}_+$, we have $\mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})] \leqslant \mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})|C(\boldsymbol{X})|]$ with equality if $d(\boldsymbol{X}) = \text{sign}[C(\boldsymbol{X})]$. That is, if $\mathscr{D}$ consists of all measurable functions from $\mathscr{X}$ to $\{\pm 1\}$, then the global optimal ITR is *not* sensitive to any covariate changes in the testing distribution. However, the problem of covariate changes induces a challenge if $\mathscr{D}$ is a restricted ITR class.

Remark 2. Our methodology only relies on the fact that $C(\boldsymbol{X})$ remains unchanged under $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$. Therefore, it can be possible to relax Assumption 1 to allowing distributional changes in $(Y(1), Y(-1))|\boldsymbol{X}$, while assuming that the CTE function $C(\cdot)$ remains identical across $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$. Furthermore, our methodology can also be meaningful if the testing CTE function can be different from training, but the optimal treatment assignment remains unchanged. We will discuss this extension in Remark 5.

2.3 An Illustrative Example

In this section, we begin with an example as in Figure 1 that the optimality of an ITR depends on the underlying distribution. There are two underlying bivariate normal distributions of means $(0, 0)^\top$; (training) and $(1.47, 1.69)^\top$; (testing) respectively. We obtain the standard ITR by maximizing the value function $\mathscr{V}_1(d)$ under the training distribution over the linear ITR class. We also obtain the DR-ITR by maximizing the DR-value function $\mathscr{V}_c^k(d)$ to be introduced in Section 2.4 over the linear ITR class. Then the DR-ITR is compared with the standard ITR through the value functions $\mathscr{V}_1$ under the training distribution and $\mathscr{V}_{1,\,\text{test}}$ under the testing distribution as in Table 1. Since the values can be comparable only through the same value function but not across different value functions, we further define the criteria *relative regret* of an ITR as $[\text{value(LB-ITR)} - \text{value(ITR)}]/|\text{value(LB-ITR)}|$, where “value” can be $\mathscr{V}_1$ or $\mathscr{V}_{1,\,\text{test}}$, and the LB-ITR maximizes the corresponding value function over the linear ITR class. In this sense, value(LB-ITR) is the best achievable value among the linear ITR class for the corresponding value function, and becomes the benchmark reference for the relative regret criteria.

Two facts can be concluded from Table 1: 1) the optimality of an ITR can be different across different distributions; and 2) maximizing the training value function may have poor testing performance when covariate changes exist. In Table 1, even though the standard ITR is optimal under the training distribution, it can be far from optimal (94.49% off in terms of relative regret) under the testing distribution. In contrast, the DR-ITR may not enjoy high training value, but can have much better testing performance (only 9.16% off in terms of relative regret).

Remark 3. Figure 1 also illustrates how the covariate changes affect the optimality of ITRs. Specifically, we can divide the covariate domain into two types of subdomains, annotated in blue and red, on which the DR-ITR and standard ITR have different treatment assignments. On the blue subdomain, the standard ITR assignment shares the same sign with the CTE function, while the DR-ITR does not. In this case, the standard ITR outperforms the DR-ITR with the difference of value $|C(\boldsymbol{X})|$ at the individual level. The case reverses on the red subdomain on which the DR-ITR outperforms the standard ITR. The overall difference of values integrates the individual difference with respect to the training or testing density.

The overall outperformance of the DR-ITR under the testing distribution can be explained from the following three perspectives: 1) the 95% confidence ellipsoid of the training domain only covers a small area of the red subdomain, while that of the testing domain covers a much larger area; 2) the distance of the red subdomain from the testing centroid is much closer than its distance from the training centroid. Then the red subdomain concentrates higher testing density than training; and 3) the individual value differences $|C(\boldsymbol{X})|'s$ are generally larger on the red subdomain intersected with the testing domain than that intersected with the training domain. Therefore, the DR-ITR performs much better than the standard ITR on the testing distribution.

2.4 Maximizing the Distributionally Robust Value (DR-Value) Function

We begin to introduce our DR-ITR that can show strong generalizability as in Figure 1. As discussed in Section 1, our goal in this paper is not to find an ITR that is generalizable on a specific testing distribution, but rather, to find an ITR that guarantees reasonable performance across an uncertain set of testing distributions. We first define the k-th *power uncertainty set* in two equivalent ways under Assumption 1:

$$\mathscr{P}_c^k(\mathbb{P}) := \left\{\mathbb{Q} \ll \mathbb{P} \,\middle|\, \|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^k(\mathbb{P})} \leqslant c\right\} \tag{2}$$

$$= \left\{\mathbb{Q} \ll \mathbb{P} \,\middle|\, w: \mathcal{X} \to \mathbb{R}_+, \mathbb{E}_{\mathbb{P}} w(\boldsymbol{X}) = 1, \mathbb{E}_{\mathbb{P}} w(\boldsymbol{X})^k \le c^k, \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = w(\boldsymbol{X})\right\}. \tag{3}$$

The set $\mathscr{P}_c^k(\mathbb{P})$ consists of the probability distributions $\mathbb{Q}$ such that the $L^k(\mathbb{P})$-norm of the density ratio $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ is bounded above by the DR-constant c. The definition (3) highlights that the density ratio is a weighting function w of $\boldsymbol{X}$, and the distribution $\mathbb{Q}$ in $\mathscr{P}_c^k(\mathbb{P})$ can be characterized by the weighting function w satisfying the conditions in (3). Here the DR-constant $c \geqslant 1$ controls the degree of the distributional robustness that measures how "close" $\mathbb{Q}$ is from $\mathbb{P}$. In particular, $c = 1$ reduces the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ to the singleton $\{\mathbb{P}\}$. The power order $1 < k \leqslant +\infty$ parametrizes the measurement of the distance of $\mathbb{Q}$ from $\mathbb{P}$. In particular, the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ increases in c as k is fixed, and decreases in k as c is fixed. The latter one is due to the Lyapunov's inequality: $\|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^k(\mathbb{P})} \leqslant \|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^{k'}(\mathbb{P})}$ whenever $1 < k \leqslant k' \leqslant +\infty$. In the Supplementary Material, we will discuss the explicit form of $\mathscr{P}_c^k(\mathbb{P})$ in the context of specific parametric families of distributions, and how it depends on the DR-constant c and the power k. One important conclusion from Example S.2 in the Supplementary Material for the mean-shifted p-dimensional normal distribution is that $\mathscr{N}_p(\mu, \boldsymbol{I}_p) \in \mathscr{P}_c^k(\mathscr{N}_p(0_p, \boldsymbol{I}_p))$ if and only if $\|\mu\|_2^2 \leqslant \frac{2\log c}{k-1}$.

With the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$, we propose to robustly maximize the following worst-case value function among the values under $\mathbb{Q} \in \mathscr{P}_c^k(\mathbb{P})$:

$$\mathscr{V}_c^k(d) := \inf_{\mathbb{Q} \in \mathscr{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[C(\boldsymbol{X})d(\boldsymbol{X})], \tag{4}$$

which we term as the *DR-value function*. In particular, $c = 1$ reduces the DR-value function $\mathscr{V}_1^k(d)$ to the standard value function $\mathscr{V}_1(d) = \mathbb{E}_{\mathbb{P}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ in the definition (1).

Remark 4 (Optimality). The “optimality” of the DR-ITR is with respect to the DR-value function $\mathscr{V}_c^k$, which highlights its difference from the traditional “optimal” ITR with respect to the standard value function $\mathscr{V}_1$.

In the example in Section 2.3, the standard ITR maximizes the value function under the training distribution over the linear ITR class, while the DR-ITR maximizes the DR-value function $\mathscr{V}_c^k(d)$ of $k = 2$ and $c = 20$ over the linear ITR class. In particular, the randomness of $\mathbb{P}$ comes from the training covariate distribution $\mathscr{N}_2(0_2, \boldsymbol{I}_2)$. Such a choice of $\mathscr{P}_c^k(\mathbb{P})$ contains the mean-shifted normal distributions $\mathscr{N}_2(\mu, \boldsymbol{I}_2)$ for all $\mu \in \{(\mu_1, \mu_2)^\top : \mu_1^2 + \mu_2^2 \leqslant 4\log 5\}$. In Figure 2a, we enumerate such mean-shifted normal distributions as the testing distributions, and evaluate the *relative improvement* of the DR-ITR over the standard ITR as the difference of their relative regrets. Among all testing distributions, the relative improvements of the DR-ITR span from −37.4% to 85.3%, suggesting that the potential of improvement can be large. Besides the DR-constant $c = 20$, we also consider the case $c = 2.71, 6.57, 10.31$ in the Supplementary Material. As c increases, the range of relative improvements becomes wider. The increase in the relative improvement upper bound is in general much larger than the decrease in the lower bound.

Based on these observations, the DR-constant c should be carefully chosen. On one hand, as can be seen from Figure 2a, the DR-ITR for a fixed DR-constant c may or may not improve over the standard ITR on a specific testing distribution within $\mathscr{P}_c^k(\mathbb{P})$. When the DR-constant c can be tuned adaptive to the specific testing distribution, then the DR-ITR can perform at least as well as the standard ITR. On the other hand, we may not even have any prior information on c to ensure that the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ contains the testing distribution of interest. Both cases ask for additional information to calibrate the choice of c so that the DR-ITR performs well on a specific testing distribution. Suppose we are able to obtain a small size of calibrating sample from the testing distribution. We propose the following training-calibrating procedure to choose c: 1) at the training stage, we estimate

DR-ITRs $\{\hat{d}_c\}_{c \in \mathscr{C}}$ where c is the DR-constant to compute $\hat{d}_c$, and C is a set of candidate DR-constants; 2) we obtain a calibrating sample from the testing distribution, on which we estimate the testing values of $\{\hat{d}_c\}_{c \in \mathscr{C}}$; 3) we select the $\hat{c}$ that maximizes the value of $\hat{d}_c$ among $c \in \mathscr{C}$.

In order to estimate the value function under the testing distribution, we consider the following two possible calibration scenarios: 1) the calibrating sample is a randomized controlled trial (RCT) dataset $(\boldsymbol{X}, A, Y)$ from the testing distribution; and 2) the calibrating sample only consists of the covariates $\boldsymbol{X}$ from the testing distribution. Scenario 1 will be more ideal than Scenario 2 since we have the testing information of both the treatment and the outcome. We can evaluate an ITR d using the IPWE $\widehat{\mathscr{V}}_{\text{calib}}^{\text{IPWE}}(d) = \mathbb{E}_{n_{\text{calib}}}\{\mathbb{1}[d(\boldsymbol{X}) = A]Y/\pi_{\text{calib}}(A \mid \boldsymbol{X})\}$, where $\mathbb{E}_{n_{\text{calib}}}$ is the empirical average over the calibrating sample, $\boldsymbol{\pi}_{\text{calib}}$ is the corresponding propensity score function, and $\boldsymbol{\pi}_{\text{calib}}$ is known or estimable from the calibrating data. We call the corresponding calibrate DR-ITR as *RCT-DR-ITR*. In Scenario 2, we do not have the treatment-response information from the testing distribution. We can instead use the value function estimate $\widehat{\mathscr{V}}_{\text{calib}}^{\text{CTE}}(d) = \mathbb{E}_{n_{\text{calib}}}[\hat{C}_n(\boldsymbol{X})d(\boldsymbol{X})]$ to evaluate d, where $\hat{C}_n(\boldsymbol{X})$ is estimated at the training stage. However, the CTE estimate $\hat{C}_n(\cdot)$ may also suffer from a potential generalizability problem on the testing distribution. Practitioners need to be careful of the generalizability of the CTE estimate when performing the calibration. We call the corresponding DR-ITR as *CTE-DR-ITR*.

RCT-DR-ITR and CTE-DR-ITR are different in their use of information for calibration. Specifically, the RCT-DR-ITR makes use of $(\boldsymbol{X}, A, Y)$ from the testing distribution, while the CTE-DR-ITR only makes use of $\boldsymbol{X}$ from the testing distribution, and the underlying CTE function $C(\boldsymbol{X})$. In practice, $C(\boldsymbol{X})$ is estimated from training data. It requires Assumption 1 to generalize the CTE estimate $\hat{C}_n(\boldsymbol{X})$ from training to testing. If Assumption 1 holds, then CTE-DR-ITR can have better performance than RCT-DR-ITR, since CTE-DR-ITR captures less variance from calibrated data. If Assumption 1 is violated, which will be illustrated in Section 4.2, then CTE-DR-ITR can have poorer performance than RCT-DR-ITR, since the testing value function estimate of CTE-DR-ITR can be biased.

In Figure 2b, we generate a calibrating RCT sample from $\mathbb{P}_{\text{test}}$ of size 50. It shows that across the mean-shifted testing distributions, the relative improvements of the calibrated DR-ITRs range from −1.70% to 82.4%. It suggests that the small sample size 50 is sufficient for a reasonably good calibration, with the positive relative improvements being maintained.

Remark 5 (Extending Covariate Changes). Consider the case that Assumption 1 is violated. Let C_{test} be the testing CTE function that can be different from the training CTE function C. We use the notations $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ to refer to the training and testing covariate distributions. Assume that $\text{sign}[C_{\text{test}}(\boldsymbol{X})] = \text{sign}[C(\boldsymbol{X})]$ almost surely. Then we can still represent the value function under the testing distribution as follows:

$$\mathbb{E}_{\text{test}}[C_{\text{test}}(\boldsymbol{X})d(\boldsymbol{X})] = \mathbb{E}_{\mathbb{P}}\left\{\frac{\mathrm{d}\mathbb{P}_{\text{test}}}{\mathrm{d}\mathbb{P}}\frac{C_{\text{test}}(\boldsymbol{X})}{C(\boldsymbol{X})}\mathbb{1}[C(\boldsymbol{X}) \neq 0] \times C(\boldsymbol{X})d(\boldsymbol{X})\right\}.$$

The definition of the DR-value function (4) can be robust with respect to the change of $(\mathbb{P}_{\text{test}}, C_{\text{test}})$ from $(\mathbb{P}, C)$, such that $w(\boldsymbol{X}) := (\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}) \times [C_{\text{test}}(\boldsymbol{X})/C(\boldsymbol{X})]\mathbb{1}[C(\boldsymbol{X}) \neq 0]$ satisfies $\mathbb{E}_{\mathbb{P}}w(\boldsymbol{X}) = 1$ and $\mathbb{E}_{\mathbb{P}}w(\boldsymbol{X})^k \leqslant c^k$.

Remark 6. The calibration procedure ensures that among the DR-ITRs of various DR-constants, the best one is chosen to maximize the testing value function. In this sense, the calibrated DR-ITR can have potential of improving the generalizability from training to testing. However, if the testing distribution is very far from the training distribution, one cannot expect that an ITR estimated by any method from the training data can perform well on the test data, even though our proposed method may be able to protect against such a distributional change to some extent. Therefore, in practice, we suggest to use our method when training and testing distributions are relatively close.

2.5 Distributionally Robust Expectation

In this section, we first discuss the rationale of considering the L^k-norm of the density ratio as the measurement of distributional distance. We show that the k-th power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ is equivalent to the distributional ball induced by the ϕ-divergence (Pardo, 2005) for some specific divergence ϕ. Then we derive the dual form of the worst-case expectation over $\mathscr{P}_c^k(\mathbb{P})$, which provides a more tractable optimization problem.

2.5.1 Equivalence to the Divergence-Based Distributional Ball—

As a generalization of the conventional likelihood-based framework which corresponds to the Kullback-Leibler (KL) divergence, the framework of general ϕ-divergence between distributions has been well studied in the context of parameter estimation and hypothesis testing (Pardo, 2005). The ϕ-divergence between two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ is defined as follows:

$$D_{\phi}(\mathbb{Q}\|\mathbb{P}) := \int \phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)\mathrm{d}\mathbb{P} = \mathbb{E}_{\mathbb{P}}\phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right); \quad \phi \in \Phi,$$

where Φ is a class of convex functions on $\mathbb{R}$ that satisfies the regularity conditions: $\phi(w) = +\infty$ for $w > 0$, $\phi(1) = \phi'(1) = 0$, and $\lim_{w \to 0_+} w\phi(p/w) = \lim_{w \to +\infty} \phi(w)/w$ for $p > 0$. The definition with various choices of ϕ's includes the empirical likelihood $\phi_{\text{EL}}(w) = -\log w + w - 1$, the KL divergence $\phi_{\text{KL}}(w) = w\log w - w + 1$, and the χ^2-divergence $\phi_{\chi}2(w) = \frac{1}{2}(w-1)^2$. There is another important special case that relates to the power uncertainty set of $k = +\infty$. Consider the optimization indicator for $c \geqslant 1 : \phi_{\infty,c} = 0$ if $u \in [0, c]$ and $+\infty$ otherwise, for which $D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0$ if $\|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^{\infty}(\mathbb{P})} \leqslant c$, and $+\infty$ otherwise. Then $D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0$ if and only if $\mathscr{P}_c^{\infty}(\mathbb{P})$.

Although D_ϕ is not a proper metric between probability distributions since it is asymmetric, we can still define a D_ϕ-distributional ball as $\mathscr{P}_\rho^\phi(\mathbb{P}) := \{\mathbb{Q} \ll \mathbb{P}: D_\phi(\mathbb{Q}\|\mathbb{P}) \leqslant \rho\}$, where $\mathbb{P}$ is the center and $\rho \geqslant 0$ is the radius. Then for any $\rho \geqslant 0$, the $D_{\phi_{\infty,c}}$-distributional ball $\mathscr{P}_\rho^{\phi_{\infty,c}}(\mathbb{P}) \equiv \{\mathbb{Q} \ll \mathbb{P}: D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0\}$, which coincides with the power uncertainty set $\mathscr{P}_c^\infty(\mathbb{P})$ defined in (2) for $k = \infty$. Such an equivalence can be extended to all finite $k \in (1, +\infty)$ when a Cressie-Read (CR) family (Cressie and Read, 1984) of divergence functions $\Phi_{\text{CR}} \subseteq \Phi$ is taken into consideration. For $k > 1$, the corresponding $\phi_k \in \Phi_{\text{CR}}$ is defined as

$$\phi_k(w) := \frac{w^k - kw + k - 1}{k(k-1)};\ \ w \geqslant 0.$$

Here, ϕ_k effectively measures the probability-distributional distance by the k-th moment of the density ratio, since $D_{\phi_k}(\mathbb{Q}\|\mathbb{P}) = \frac{1}{k(k-1)}[\mathbb{E}_\mathbb{P}(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P})^k - 1]$ as long as $\mathbb{Q}$ is a probability distribution. Then it can be inferred that the D_{ϕ_k}-distributional ball $\mathscr{P}_\rho^{\phi_k}(\mathbb{P})$ is actually equivalent to the power uncertainty set $\mathscr{P}_{c_k}^k(\mathbb{P})$ in (2). Here, there is a one-to-one correspondence between the DR-constant c and the radius ρ of the D_{ϕ_k}-distributional ball with $c_k(\rho) := [k(k-1)\rho + 1]^{1/k}$. We conclude the case $k = +\infty$ and $1 < k < +\infty$ with the following:

$$\mathscr{P}_\rho^{\phi_{\infty,c}}(\mathbb{P}) = \mathscr{P}_c^\infty(\mathbb{P});\quad \mathscr{P}_\rho^{\phi_k}(\mathbb{P}) = \mathscr{P}_{c_k(\rho)}^k(\mathbb{P});\quad \rho \geqslant 0. \tag{5}$$

2.5.2 Dual Representation—

We begin with a general result on the dual representation of the ϕ-divergence-based distributionally robust expectation. We state the following lemma and refer readers to Duchi and Namkoong (2018, Proposition 1).

Lemma 1. *Fix a random variable Z on $\mathbb{R}$ with distribution $\mathbb{P}$. Let $\phi \in \Phi$ be a legitimate divergence function. Define the convex conjugate of ϕ as*

$$\phi^*(\boldsymbol{x}^*) := \sup_{\boldsymbol{x} \in \mathbb{R}} \{\langle \boldsymbol{x}^*, \boldsymbol{x}\rangle - \phi(\boldsymbol{x})\};\quad \boldsymbol{x}^* \in \mathbb{R}.$$

Then for $\rho > 0$,

$$\sup_{\mathbb{Q} \in \mathscr{P}_\rho^\phi(\mathbb{P})} \mathbb{E}_\mathbb{Q} Z = \inf_{\substack{\lambda \geqslant 0 \\ \eta \in \mathbb{R}}} \left\{\mathbb{E}_\mathbb{P}\left[\lambda\phi^*\left(\frac{Z-\eta}{\lambda}\right)\right] + \lambda\rho + \eta\right\}. \tag{6}$$

Let $c \geqslant 1$. Lemma 1 can be directly applied to the optimization indicator: $\phi_{\infty,c}(u) := 0$ if $u \in [0, c]$ and $+\infty$ otherwise, whose convex conjugate is given by $\phi_{\infty,c}^*(u) = c\max\{u, 0\}$. Then λ in (6) attains the infimum at $\lambda = 0$, so that

$$\sup_{\mathbb{Q} \in \mathscr{P}_{\rho}^{\phi_{\infty}, c}(\mathbb{P})} \mathbb{E}_{\mathbb{Q}} Z = \inf_{\eta \in \mathbb{R}} \left\{ c \mathbb{E}_{\mathbb{P}} (Z - \eta)_{+} + \eta \right\}. \tag{7}$$

In particular, the right hand side of (7) is solved by the $(1 - 1/c)$-*value-at-risk* $\mathrm{VaR}_{1 - 1/c}$ in finance, or equivalently, the $(1 - 1/c)$-quantile of Z under the center distribution $\mathbb{P}$. The right hand side of (7) itself is defined as the $(1 - 1/c)$-*conditional value-at-risk* $\mathrm{CVaR}_{1 - 1/c}$ (Rockafellar and Uryasev, 2000). Next, we apply Lemma 1 to the k-th power divergence ϕ_k to derive the dual problem of the worst-case expectation over $\mathscr{P}_{c}^{k}(\mathbb{P})$.

Lemma 2. *Let* Φ_{CR} *be the Cressie-Read family of divergence functions,* k*,* $k^{*} \in (1, +\infty)$ *be conjugate numbers, i.e.,* $\frac{1}{k} + \frac{1}{k^{*}} = 1$*, and* $\phi_k \in \Phi_{\mathrm{CR}}$*. Then we have following conclusions*:

I. *The convex conjugate of* ϕ_k *is given by*

$$\phi_k^{*}(z) = \frac{1}{k}\left\{ [(k - 1)z + 1]_{+}^{k^{*}} - 1 \right\}.$$

II. *Fix a probability measure* $\mathbb{P}$ *and a random variable* Z *on* $\mathbb{R}$*. Then for* $\rho \geqslant 0$,

$$\sup_{\mathbb{Q} \in \mathscr{P}_{\rho}^{\phi_k}(\mathbb{P})} \mathbb{E}_{\mathbb{Q}} Z = \inf_{\eta \in \mathbb{R}} \left\{ c_k(\rho) [\mathbb{E}_{\mathbb{P}} (Z - \eta)_{+}^{k^{*}}]^{1/k^{*}} + \eta \right\}, \tag{8}$$

where $c_k(\rho) = [k(k - 1)\rho + 1]^{1/k}$.

Note that the right hand side of (8) and its optimizer η are both coherent risk measures as the higher-order generalizations of the CVaR and VaR (Krokhmal, 2007).

Using the equivalence in (5), the worst-case expectation over the power uncertainty set $\mathscr{P}_{\rho}^{\phi_k}(\mathbb{P})$ for $k \in (1, \infty]$ and $k^{*} = \frac{k}{k - 1}$ (in particular,$k = \infty \Leftrightarrow k^{*} = 1$) unifies (7) and (8) as follows:

$$\sup_{\mathbb{Q} \in \mathscr{P}_{c}^{k}(\mathbb{P})} \mathbb{E}_{\mathbb{Q}} Z = \inf_{\eta \in \mathbb{R}} \left\{ c [\mathbb{E}_{\mathbb{P}} (Z - \eta)_{+}^{k^{*}}]^{1/k^{*}} + \eta \right\}; \quad c \geqslant 1. \tag{9}$$

By inspecting the dual problem (9), the right hand side is computationally more tractable than the left hand side, since instead of optimizing over an infinite-dimensional probability measure $\mathbb{Q}$, we only need to optimize over a univariate variable η.

In order to apply the duality result to the DR-ITR problem, we negate the DR-value maximization to a risk minimization problem. Denote the *risk function* under the training distribution $\mathbb{P}$ as $\mathscr{R}_1(d) := \mathscr{V}_1(d) = \mathbb{E}_{\mathbb{P}}\{C(\boldsymbol{X})[-d(\boldsymbol{X})]\}$. Then for $k \in (1, \infty]$ and $c \geqslant 1$, the *DR-risk function* is defined as

$$\mathscr{R}_c^k(d) := \sup_{\mathbb{Q} \in \mathscr{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}\{C(\boldsymbol{X})[-d(\boldsymbol{X})]\}.$$

Using the fact $Z = -C(\boldsymbol{X})d(\boldsymbol{X}) = C(\boldsymbol{X})\mathbb{1}[d(\boldsymbol{X}) = -\mathbb{1}] + [-C(\boldsymbol{X})]1[d(\boldsymbol{X}) = 1]$, the dual representation (9) can be expressed in the following particular form (10).

Corollary 3 (Dual Representation of the DR-Risk Function). *Let* $k \in (1, +\infty]$, $k^* = \frac{k}{k-1}$ *if* $k < +\infty$ *and* $k^* = 1$ *if* $k = +\infty$, $c \geqslant 1$. *Then the DR-risk function* $\mathscr{R}_c^k$ *has the following dual representation:*

$$\mathscr{R}_c^k(d) = \inf_{\eta \in \mathbb{R}} \left\{c\left[\mathbb{E}\left([C(\boldsymbol{X}) - \eta]_+^{k^*} 1[d(\boldsymbol{X}) = -1] + [-C(\boldsymbol{X}) - \eta]_+^{k^*} 1[d(\boldsymbol{X}) = 1]\right)\right]^{1/k^*} + \eta\right\}. \tag{10}$$

2.6 Implementation

In this section, we introduce the implementation of DR-risk minimization based on the empirical data. We cast the learning problem as finding a decision function $f: \mathscr{X} \to \mathbb{R}$ that induces an ITR based on its sign: $d(\boldsymbol{x}) = \text{sign}[f(\boldsymbol{x})]$. The ITR class $\mathscr{D}$ can correspond to a prespecified decision function class $\mathscr{F}$. The DR-risk function as a functional of the decision function becomes $\mathscr{R}_c^k(f) = \sup_{\mathbb{Q} \in \mathscr{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}\{C(\boldsymbol{X})\text{sign}[-f(\boldsymbol{X})]\}$. However, directly optimizing the risk $\mathscr{R}_c^k(f)$ is challenging, since the $\text{sign}(\cdot)$ operation is nonconvex and nonsmooth. We consider a specific difference-of-convex (DC) relaxation of the sign operator.

We propose to relax the indicators in the dual form (10) by the following robust smoothed ramp loss (Zhou et al., 2017): $\psi(u) := (1-u)^2\mathbb{1}(0 \leqslant u \leqslant 1) + [2 - (1+u)^2\mathbb{1}(-1 \leqslant u \leqslant 0) + 2\mathbb{1}(u \leqslant -1)]$. The DC representation is given by $\psi(u) = \psi_+(u) - \psi\_(u)$, where

$\psi_+(u) = (1-u)^2\mathbb{1}(0 \leqslant u \leqslant 1) + (1-2u)\mathbb{1}(u \leqslant 0)$,

$\psi\_(u) = u^2\mathbb{1}(-1 \leqslant u \leqslant 0) + (-1-2u)\mathbb{1}(u \leqslant -1)$. The advantages of using the symmetric nonconvex loss can be: 1) to protect from outliers in $\boldsymbol{X}$ and improve generalizability (Shen et al., 2003; Wu and Liu, 2007), and 2) to equally indicate $f(\boldsymbol{X}) < 0$ and $f(\boldsymbol{X}) > 0$. We would like to point out that $\mathbb{1}[f(\boldsymbol{X}) < 0] + \mathbb{1}[f(\boldsymbol{X}) > 0] \equiv 1$ will be preserved to $\frac{\psi[f(\boldsymbol{X})]}{2} + \frac{\psi[-f(\boldsymbol{X})]}{2} \equiv 1$ in this surrogate loss. Then we define the DR-ψ-risk function as

$$\mathscr{R}_{c,\psi}^k(f) := \inf_{\eta \in \mathbb{R}} \left\{c\left[\mathbb{E}\left([C(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[f(\boldsymbol{X})]}{2} + [-C(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[-f(\boldsymbol{X})]}{2}\right)\right]^{1/k^*} + \eta\right\}. \tag{11}$$

Algebraically, we can invert (11) to its primal representation $\mathscr{R}^k_{c,\psi}(f) = \sup_{\mathbb{Q} \in \mathscr{P}^k_c(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[C(\boldsymbol{X})\zeta_\psi(f)]$ by introducing a sign random variable $\zeta_\psi(f) \in \left\{\pm 1\right\}$ with $\mathbb{P}(\zeta_\psi(f) = \pm 1 \mid \boldsymbol{X}) : = \frac{\psi[\pm f(\boldsymbol{X})]}{2}$. That is, given the covariate $\boldsymbol{X}$, the original deterministic sign $\mathrm{sign}[-f(\boldsymbol{X})]$ is relaxed to the random sign $\zeta_\psi(f)$ with ± 1 probability $\frac{\psi[\pm f(\boldsymbol{X})]}{2}$. In particular, if $f(\boldsymbol{X}) > 0$, then $\mathrm{sign}[-f(\boldsymbol{X})] = -1$ is a hard sign while $\zeta_\psi(f)$ is a soft sign with $\mathbb{P}(\zeta_\psi(f) = -1 \mid \boldsymbol{X}) = \frac{\psi[-f(\boldsymbol{X})]}{2} > \frac{\psi[f(\boldsymbol{X})]}{2} = \mathbb{P}(\zeta_\psi(f) = 1 \mid \boldsymbol{X})$. When $c = 1$, the DR-risk function reduces to the risk function under the training distribution, and the DC relaxation here is equivalent to the relaxation in Zhou et al. (2017).

The DR-ψ-risk function provides the learning objective based on the empirical data. In particular, the population expectation $\mathbb{E}$ is replaced by the empirical average $\mathbb{E}_n$, and the CTE function $C(\cdot)$ is replaced by a plug-in estimate $\widehat{C}_n(\cdot)$. The corresponding empirical objective is minimized over the decision function f and the auxiliary variables (η, λ) jointly:

$$\min_{f \in \mathscr{F}, \eta \in \mathbb{R}} \left\{ c\left[\mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[f(\boldsymbol{X})]}{2} + [-\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[-f(\boldsymbol{X})]}{2}\right)\right]^{1/k^*} + \eta \right\}$$

$$= \min_{f \in \mathscr{F}, \eta \in \mathbb{R}, \lambda \geqslant 0} \left\{ \frac{c}{k^* \lambda^{k^*-1}} \mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[f(\boldsymbol{X})]}{2} + [-\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[-f(\boldsymbol{X})]}{2}\right) + \frac{c\lambda}{k} + \eta \right\}.$$

The objective function is a summation of multiple products of DC functions. For $k < +\infty$, we consider a block successive upper-bound minimization algorithm (Razaviyayn et al., 2013) to alternatively minimize the convex upper bounds over the decision function f and the auxiliary variables (η, λ) respectively. For $k = +\infty$, it requires a further probabilistic enhancement to break ties at argmin and ensure the convergence to stationarity (Qi et al., 2019a,b). The implementation details are given in the Supplementary Material.

3 Theoretical Properties

In this section, we justify the validity of the DC relaxation and the empirical substitution. First of all, we introduce the following joint stochastic objectives:

$$\ell^k_c(f, \eta, \lambda; \widetilde{C}) : = \frac{c}{k^* \lambda^{k^* - \mathbb{1}}}\left([\widetilde{C}(\boldsymbol{X}) - \eta]_+^{k^*} \mathbb{1}[f(\boldsymbol{X}) < 0] + [-\widetilde{C}(\boldsymbol{X}) - \eta]_+^{k^*} 1[f(\boldsymbol{X}) > 0]\right) + \frac{c\lambda}{k} + \eta;$$

$$\ell^k_{c,\psi}(f, \eta, \lambda; \widetilde{C}) : = \frac{c}{k^* \lambda^{k^* - 1}}\left([\widetilde{C}(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[f(\boldsymbol{X})]}{2} + [-\widetilde{C}(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[-f(\boldsymbol{X})]}{2}\right) + \frac{c\lambda}{k} + \eta.$$

Here, $\widetilde{C}$ can be the plug-in estimate $\widetilde{C}_n$ or the underlying true CTE C. Denote $\mathscr{L}^k_c(f, \eta, \lambda) : = \mathbb{E}\ell^k_c(f, \eta, \lambda; C)$, $\mathscr{L}^k_{c,\psi}(f, \eta, \lambda) : = \mathbb{E}\ell^k_{c,\psi}(f, \eta, \lambda; C)$. Then by Corollary 3, we have

$\mathscr{R}_c^k(f) = \inf_{\eta \in \mathbb{R}, \lambda \geqslant 0}\mathscr{L}_c^k(f, \eta, \lambda)$, $\mathscr{R}_{c,\psi}^k(f) = \inf_{\eta \in \mathbb{R}, \lambda \geqslant 0}\mathscr{L}_{c,\psi}^k(f, \eta, \lambda)$. In the following proposition, we show the validity of the DC re-laxation.

Proposition 4 (Fisher Consistency and Excess Risk). *Suppose* $\mathscr{R}_c^k$, $\mathscr{R}_{c,\psi}^k$, $\mathscr{L}_c^k$ *and* $\mathscr{L}_{c,\psi}^k$ *are defined as above. Fix* $k \in (1, +\infty]$, $k^* = \frac{k}{k-1}$, $c \geqslant 1$, $\eta \in \mathbb{R}$, $\lambda > 0$. *Then the following results hold:*

I. *(Fisher Consistency)*

$$\underset{f:\mathscr{X} \to [-1,1]}{\operatorname{argmin}} \mathscr{L}_{c,\psi}^k(f, \eta, \lambda) = \underset{f:\mathscr{X} \to \{\pm 1\}}{\operatorname{argmin}} \mathscr{L}_c^k(f, \eta, \lambda), \quad \min_{f:\mathscr{X} \to [-1,1]} \mathscr{L}_{c,\psi}^k(f, \eta, \lambda) = \min_{f:\mathscr{X} \to \{\pm 1\}} \mathscr{L}_c^k(f, \eta, \lambda);$$

II. *(Excess Risk) Denote* $\mathscr{L}_c^{k,*}(\eta, \lambda) := \min_{f \in \mathscr{X} \to \{\pm 1\}}\mathscr{L}_c^k(f, \eta, \lambda)$. *Then for* $f:\mathscr{X} \to \mathbb{R}$, *we have*

$$\mathscr{L}_c^k(f, \eta, \lambda) - \mathscr{L}_c^{k,*}(\eta, \lambda) \leqslant 2[\mathscr{L}_{c,\psi}^k(f, \eta, \lambda) - \mathscr{L}_c^{k,*}(\eta, \lambda)].$$

Denote $\mathscr{R}_c^{k,*} := \inf_{\eta \in \mathbb{R}, \lambda \geqslant 0}\mathscr{L}_c^{k,*}(\eta, \lambda)$. *Then for* $f:\mathscr{X} \to \mathbb{R}$, *we have*

$$\mathscr{L}_c^k(f, \eta, \lambda) - \mathscr{R}_c^{k,*} \leqslant 2[\mathscr{L}_{c,\psi}^k(f, \eta, \lambda) - \mathscr{R}_c^{k,*}], \quad \mathscr{R}_c^k(f) - \mathscr{R}_c^{k,*} \leqslant 2[\mathscr{R}_{c,\psi}^k(f) - \mathscr{R}_c^{k,*}].$$

Suppose $\mathscr{F}$ is a functional class on $\mathscr{X}$ with norm $\| \cdot \|_{\mathscr{F}}$ that characterizes the complexity of function. Motivated by Steinwart and Scovel (2007, (6)), we define for $\gamma \geqslant 0$ the constrained version of the approximation error

$$\mathscr{A}_c^k(\gamma) := \inf_{f \in \mathscr{F}}\left\{\mathscr{R}_{c,\psi}^k(f) : \|f\|_{\mathscr{F}} \leqslant \gamma\right\} - \mathscr{R}_c^{k,*}.$$

Similarly to that in Steinwart and Scovel (2007), $\mathscr{A}_c^k(\gamma)$ with the appropriately chosen tuning parameter γ can trade off the learnability and the approximatability of $\mathscr{F}$ towards the population Bayes rule $\operatorname{argmin}_{f:\mathscr{X} \to \{\pm 1\}}\mathscr{R}_c^k(f)$. Specifically, as γ increases, the population approximation error ("bias") $\mathscr{A}_c^k(\gamma)$ decreases with γ, while the empirical complexity ("variance") increases with γ. The trade-off will be stated more explicitly in the following Assumption 5.

Next, we make the following assumptions to show the regret bound for the empirical minimization of the ψ-risk $\mathbb{E}_n\ell_{c,\psi}^k(f, \eta, \lambda; \widehat{C}_n)$. Without loss of generality, we restrict to consider the functional class $\mathscr{F}$ as the Reproducing Kernel Hilbert Space (RKHS) with the Gaussian radial basis function kernels, where $\| \cdot \|_{\mathscr{F}}$ is the RKHS-norm. General results can be established by adopting the covering number argument as in Zhao et al. (2019a, Theorem 3.1).

Assumption 2 (Boundedness). There exists $M < +\infty$ such that $|C(\boldsymbol{X})| \leqslant M$ almost surely.

Assumption 3 (Diffuse Property). The distribution of $C(\boldsymbol{X})$ has a uniformly bounded density with respect to the Lebesgue measure.

Assumption 4 (Convergence of the Plug-in CTE). For the CTE estimate $\widehat{C}_n(\boldsymbol{X})$, we assume that $\|\widehat{C}_n - C\|_\infty := \sup_{\boldsymbol{x} \in \mathscr{X}} \left|\widehat{C}_n(\boldsymbol{x}) - C(\boldsymbol{x})\right| \xrightarrow{\mathbb{P}} 0$.

Assumption 5 (Approximation Error Rate). There exists $\beta \in (0, 1]$ and $K_{\mathscr{A}} < +\infty$ such that for all small enough $\gamma > 0$, we have $\mathscr{A}_c^k(\gamma) \leqslant K_{\mathscr{A}}\gamma^{-\beta}$.

As a remark, we note that Assumption 2 can hold if the difference of potential outcomes $Y(1) - Y(-1)$ is uniformly bounded, or $\mathscr{X}$ is compact and $\boldsymbol{x} \mapsto C(\boldsymbol{x})$ is continuous. Assumption 3 holds if $\boldsymbol{X}$ has a diffuse distribution, *i.e.*, $\boldsymbol{X}$ doesn't contain points with positive mass; and $\boldsymbol{x} \mapsto C(\boldsymbol{x})$ is injective. Assumption 3 is the key assumption to bound λ away from 0. This assumption will not be necessary if $k = +\infty$ and $k^* = 1$. Assumption 4 can be met if $\mathscr{X}$ is compact and $\widehat{C}_n$ is a random forest estimate (Wager and Walther, 2015). Following Steinwart and Scovel (2007, Theorem 2.7), Assumption 5 can be shown valid if the Tsybakov's noise assumption on the population margin is met and the kernel bandwidth parameter is chosen appropriately. In the following proposition, we establish the regret bound.

Proposition 5 (Regret Bound). *Suppose* $\mathscr{R}_c^k$, $\mathscr{R}_{c,\psi}^k$, *and* $\mathscr{L}_c^k$, $\mathscr{L}_{c,\psi}^k$ *are defined as above. Fix* $k \in (1, +\infty]$, $k^* = \frac{k}{k-1}$, $c > 1$. *Assume that Assumptions 2–5 hold. Let*

$$(\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) \in \underset{f \in \mathscr{F}, \eta \in \mathbb{R}, \lambda \geqslant 0}{\operatorname{argmin}} \left\{\mathbb{E}_n \ell_{c,\psi}^k(f, \eta, \lambda; \widehat{C}_n) : \|f\|_{\mathscr{F}} \leqslant \gamma_n\right\},$$

with the tuning parameter γ_n *satisfying* $\gamma_n = \mathscr{O}(n^{-\frac{1}{2\beta+1}})$ *as* $n \to \infty$. *Then there exists constants* $K_0 = K_0(c, M) < +\infty$ *and* $K_1 = K_1(c, M) < +\infty$ *such that for* $0 < \delta < 1$, *with probability at least* $1 - \delta$, *we have*

$$\mathscr{R}_c^k(\widehat{f}_n) - \mathscr{R}_c^{k,*} \leqslant \mathscr{L}_c^k(\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) - \mathscr{R}_c^{k,*} \leqslant K_0\sqrt{\log(2/\delta)}\,n^{-\frac{\beta}{2\beta+1}} + K_1\|\widehat{C}_n - C\|_\infty.$$

In particular, there exists K_{01}, K_{02}, K_{11}, $K_{12} < +\infty$ *not depending on c, M, such that*

$$K_0(c, M) = \begin{cases} K_{01} \dfrac{c^{\frac{(k^*+1)(2k^*-1)}{k^*-1} + \frac{1}{2}}}{(c-1)^{k^*+1/2}} M^{k^*+1/2}, & k < +\infty; \\ K_{02} c M^{3/2} & k = +\infty; \end{cases} \quad K_1(c, M)$$
$$= \begin{cases} K_{11} \dfrac{c^{2k^*+1}}{(c-1)^{k^*-1}} M^{k^*-1} & k < +\infty; \\ K_{12} c, & k = +\infty\,. \end{cases}$$

In Proposition 5, it can be of theoretical interest to understand how the regret bound depends on the DR-constant c and the power order k. Specifically, as $c \to +\infty$, η approaches to the essential supremum of $[C(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[f(\boldsymbol{X})]}{2} + [-C(\boldsymbol{X}) - \eta]_+^{k^*} \frac{\psi[-f(\boldsymbol{X})]}{2}$ (Krokhmal, 2007, Example 2.3). Then λ vanishes to 0 so that $1/\lambda$ tends to $+\infty$. Since the Lipschitz constant of $\ell_{c,\psi}^{k}(f, \eta, \lambda)$ with respect to λ scales with $1/\lambda^{k^*}$, the universal constants K_0 and K_1 grow to $+\infty$ as well.

Another important fact is that the conjugate number k^* of k appears in the polynomial orders of c and M respectively in the universal constants K_0 and K_1. In particular, for a large conjugate order k^*, the universal constants K_0 and K_1 increase with the DR-constant c and the CTE bound M more rapidly. In order to achieve a tighter finite sample regret bound, a smaller k^* and hence a larger k is preferred. Such a phenomenon complements the fact that the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ decreases in k. Specifically, as the power order k increases, its conjugate order k^* decreases, and the regret bound in Proposition 5 becomes tighter. On the contrary, the power uncertainty set $\mathscr{P}_c^k(\mathbb{P})$ gets smaller, and the worst-case objective is less distributionally robust. Therefore, the power order k trades off between the distributional robustness in terms of the size of $\mathscr{P}_c^k(\mathbb{P})$, and the finite sample regret bound.

4 Simulation Studies

In this section, we carry out two simulation studies to evaluate the generalizability of the DR-ITR on the testing distributions that are different from the training distribution. The first simulation considers the covariat shifts. The second simulation considers the mixture of subgroups.

4.1 Covariate Shifts

In this section, we extend the motivating example in Section 2.3 to a more practical simulation setting. Consider the training data generating process: n = 1,000, p =10, $\boldsymbol{X} \sim \mathscr{N}_p(0_p, \mathbf{I}_p)$, $A \mid \boldsymbol{X} \sim$Bernoulli(1/2) and $Y(\boldsymbol{X}, A) = m(\boldsymbol{X}) + (A - 1/2)C(\boldsymbol{X}) + \mathscr{N}(0, 1)$, where $m(\boldsymbol{x}) = 1 + \frac{1}{p}\sum_{j=1}^{p} x_j$, $C(\boldsymbol{x}) = x_2 - (x_1^3 - 2x_1)$.

At the training stage, we first obtain a CTE function estimate $\widehat{C}_n$ by fitting a casual forest (Wager and Athey, 2018) on the training data. Then we obtain the out-of-bag prediction at the training covariates $\widehat{C}_n(\boldsymbol{X})$. Next we fit the standard ITR by empirically minimizing

$\mathbb{E}_n\{\hat{C}_n(\boldsymbol{X})(\psi[f(\boldsymbol{X})] - 1)\}$ as the ψ-relaxation of the empirical risk function $\mathbb{E}_n\{\hat{C}_n(\boldsymbol{X})\text{sign}(-f(\boldsymbol{X}))]\}$, over the linear function class $\mathscr{F}_\gamma: = \{f(\boldsymbol{x}) = b + \boldsymbol{\beta}^\top \boldsymbol{x}: b \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p, \|\boldsymbol{\beta}\|_2 \leqslant \gamma\}$. The tuning parameter $\gamma \geqslant 0$ is determined by 10-fold cross-validation among {0.1, 0.5, 1, 2, 4}. Finally, we fit the DR-ITRs for $k = 2$ and $c \in \mathscr{C} = \{1.19, 1.38, \cdots, 20\}$ from the function class $\mathscr{F}_\gamma$, where γ is the same as that of the standard ITR.

We consider the mean-shifted testing distribution $\boldsymbol{X} \sim \mathscr{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ for various covariate centroids $\boldsymbol{\mu}$'s. In order to calibrate the DR-constant c for every fixed $\boldsymbol{\mu}$, we generate a calibrating dataset of size $n_{\text{calib}} = 50$ from the testing distribution. The following two scenarios for the calibrating data are considered here: 1) a randomized controlled trial (RCT) dataset $(\boldsymbol{X}, A, Y)$ is generated, with $\boldsymbol{X} \sim \mathscr{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ and (A, Y) as before; and 2) only the covariate vector $\boldsymbol{X} \sim \mathscr{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ is generated. In Scenario 1, we use the IPWE of the calibrating value function $\widehat{\mathscr{V}}_{\text{calib}}^{\text{IPWE}}(\hat{f}_c): = \mathbb{E}_{n_{\text{calib}}}\{Y\mathbb{1}[(2A-1)\hat{f}_c(\boldsymbol{X}) > 0]/(1/2)\}$ to evaulate the DR-constant c, while in Scenario 2, we use the CTE-based calibrating value function $\widehat{\mathscr{V}}_{\text{calib}}^{\text{CTE}}(\hat{f}_c): = \mathbb{E}_{n_{\text{calib}}}\{\hat{C}_n(\boldsymbol{X})\text{sign}[\hat{f}_c(\boldsymbol{X})]\}$ instead. Here, the estimated CTE function $\hat{C}_n$ is obtained from the training stage.

For comparison, we consider the following: 1) the LB-ITR that maximizes the value function under the testing distribution; 2) the ℓ_1-penalized least-square (ℓ_1-PLS) (Qian and Murphy, 2011) of $Q(\boldsymbol{X}, A) = \mathbb{E}(Y \mid \boldsymbol{X}, A)$ on $(1, \boldsymbol{X}, A, A\boldsymbol{X})$ and the corresponding estimated ITR $\hat{d}(\boldsymbol{x}) \in \text{argmin}_{a \in \{\pm 1\}} \hat{Q}_n(\boldsymbol{x}, a)$; 3) the standard ITR; 4) the RCT-DR-ITR for the calibrating Scenario 1; and 5) the CTE-DR-ITR for the calibrating Scenario 2. We compare the testing values $\mathbb{E}_{n_{\text{test}}}[C(\boldsymbol{X})\hat{d}(\boldsymbol{X})]$ based on an independent testing dataset of size $n_{\text{test}} = 100,000$ for every testing distribution. The testing values across different testing distributions are not comparable. For a specific testing distribution, the LB-ITR can be a benchmark to be compared to, since its testing value is the best achievable in theory among the linear ITR class. The training-calibrating-testing procedure is replicated for 500 times. The testing values (standard errors) for $n_{\text{calib}} = 50$ are reported in Table 2.

When the testing distribution is the same as training $(\mu_1, \mu_2) = (0, 0)$, the calibration procedures for the DR-ITRs are expected to choose $c = 1$, which corresponds to the standard ITR. With the finite calibrating sample, some DR-constant c greater than 1 can be possibly chosen, leading to smaller testing values for the DR-ITRs in Table 2. In particular, the testing value of the CTE-DR-ITR is higher than that of the RCT-DR-ITR, and is closer to the testing value of the standard ITR in this case. The reason is that, the RCT-based calibrating value function estimate $\widehat{\mathscr{V}}_{\text{calib}}^{\text{IPWE}}$ depends on $(\boldsymbol{X}, A, Y)$ in the calibrating data, while the CTE-based one $\widehat{\mathscr{V}}_{\text{calib}}^{\text{CTE}}$ depends on $\boldsymbol{X}$ only. As a consequence, the CTE-based calibration can be more accurate than the RCT-based one.

When $(\mu_1, \mu_2) \neq (0, 0)$, the testing distribution is different from training, and the performance of the standard ITR deteriorates while the DR-ITRs still maintain reasonably good performance. The phenomenon is more evident when $\mu_1, \mu_2 \in \{1.469, 1.958\}$. In particular at $(\mu_1, \mu_2) = (1.958, 1.958)$, the value of the standard ITR can be as low as 17% of the best achievable value among the linear ITR class, while the DR-ITRs can maintain more than 90%. In fact, such a phenomenon is general. In Figure 3a, we further enumerate the testing covariate centroid $\mu = (\mu_1, \mu_2, 0, \dots, 0)^\top$ for $\mu_1, \mu_2 \in [-2.448, 2.448]$ and compute the relative regrets of the standard ITR and the RCT-DR-ITR. Across all mean-shifted testing distributions, the relative regrets of the standard ITRs can be as high as 108%, in which case the standard ITR value is negative, and hence even worse than the completely random treatment rule d_{rand}. On the contrary, the relative regrets for the RCT-DR-ITR ($n_{\text{calib}} = 50$) shown in Figure 3b are at most 24% across all testing centroids. This suggests that the RCT-DR-ITR maintains relatively good performance on all such testing distributions, while the standard ITR fails. Figure 4 further shows that the DR-ITR provides substantial testing value improvements over the standard ITR. This demonstrates that the small sample size $n_{\text{calib}} = 50$ is sufficient for calibrating the DR-ITR with significant testing improvement.

From Table 2, it can be also observed that ℓ_1-PLS can have better performance than the standard ITR when training and testing distributions are different. The reason is that, the objective of ℓ_1-PLS does not target the value function under the training distribution directly, but rather, the mean squared error of the linear approximation to $Q(\boldsymbol{X}, A)$ under the training distribution. Such a linear approximation can perform well when the testing distribution is not far from the training distribution. However, in the case $\mu_1, \mu_2 \in \{1.469, 1.958\}$ in the sense that the testing distribution deviates more from the training one, the DR-ITRs enjoy notably higher testing values than ℓ_1-PLS.

In the Supplementary Material, we provide more detailed results for other comparisons including the relative regrets/improvements on all mean-shifted covariate domains of all centroids, the misclassification rates on all mean-shifted covariate domains of all centroids, the comparison with some other methods in relative regrets and misclassification rates, and the case of $k \in \{1.25, 1.5, 2, 3, \infty\}$. In particular, the misclassification rates inform similar conclusions as the relative regrets/improvements. If we increase the calibrating sample size from 50 to 100, then the testing values of DR-ITRs can be further improved. We also find that among our simulation scenarios, the testing values of the DR-ITR are not very sensitive to difference choices of k.

4.2 Performance on the Mixture of Subgroups

In this section, we consider a population that consists of two subgroups, with each following a distinct CTE function. We aim to find an ITR that can generalize well on different mixtures of subgroups.

We modify the simulation setup in Section 4.1 as follows: $\boldsymbol{X} \mid \xi \sim \xi \mathcal{N}_p(\boldsymbol{\mu}_1, \mathbf{I}_p) + (1 - \xi)\mathcal{N}_p(\boldsymbol{\mu}_0, \mathbf{I}_p)$, where $\xi \sim \textbf{Bernoulli}(p_{\text{mix}})$ is the unobservable mixture/subgroup indicator with subgroup 1 probability p_{mix} and subgroup 0 probability $1-p_{\text{mix}}$, and the subgroup means $\boldsymbol{\mu}_1 = (-1/2, 1/2, 0, \dots, 0)^\top$ and $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$. We consider the CTE function $C(\boldsymbol{x}; \xi) := (2\xi - 1)\beta_0 + \beta_1 x_1 + \beta_2 x_2$ that is linear in the covariate vector, but with a subgroup-dependent intercept $(2\xi - 1)\beta_0$, and $(\beta_0, \beta_1, \beta_2) := (-3/2, -2, 1)$. The unconditional CTE function is nonlinear:

$$C(\boldsymbol{x}) := \mathbb{E}[C(\boldsymbol{x}; \xi) \mid \boldsymbol{X} = \boldsymbol{x}] = \frac{p_{\text{mix}}\exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_1\|_2^2/2) - (1 - p_{\text{mix}})\exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_0\|_2^2/2)}{p_{\text{mix}}\exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_1\|_2^2/2) - (1 - p_{\text{mix}})\exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_0\|_2^2/2)}\beta_0 + \beta_1 x_1 + \beta_2 x_2 .$$

In particular, the unconditional CTE function $C(\boldsymbol{x})$ depends on the subgroup 1 probability p_{mix}. The distributional changes are due to the subgroup 1 probability. Specifically, the training subgroup 1 probability is 0.75, while the testing subgroup 1 probability varies in {0.1, 0.25, 0.5, 0.75, 0.9}. Since the training and testing CTE functions can be different, Assumption 1 cannot be fully met. Therefore, our proposed DR-ITR can be robust to such distributional changes only to some extent.

We consider the same training-calibrating-testing procedure as that in Section 4.1, except that the DR-constant c ranges in $\{1.18, 1.27, \cdots, 10\}$. The testing values of the ITRs are reported in Table 3. When the training and testing distributions are the same at $p_{\text{mix}} = 0.75$, all ITRs have similar testing performance. The standard ITRs have higher testing values than the DR-ITRs in this case. When the testing p_{mix} becomes smaller, the DR-ITRs show better testing performance than the standard ITR. When the testing $p_{\text{mix}} = 0.25$ or 0.1, the RCT-DR-ITR has the highest testing values among all. Since the true testing CTE function changes along with the testing p_{mix}, the corresponding estimate $\widehat{C}_n$ based on the training data can suffer from the generalizability problem. Therefore, the CTE-based calibration performs slightly worse than the RCT-based calibration in this case. However, the CTE-based DR-ITR is superior to the standard ITR, and is comparable to the ℓ_1-PLS. More detailed comparisons and the case $n_{\text{calib}} = 100$ are provided in the Supplementary Material.

5 Application to the ACTG 175 Trial Data

In this section, we evaluate the generalizability of our proposed DR-ITR on a clinical trial dataset from the “AIDS clinical trial group study 175” (Hammer et al., 1996). The goal of this study was to compare four treatment arms among 2,139 randomly assigned subjects with human immunodeficiency virus type 1 (HIV-1), whose CD4 counts were 200–500 cells/mm^3. The four treatments are the zidovudine (ZDV) monotherapy, the didanosine (ddI) monotherapy, the ZDV combined with ddI, and the ZDV combined with zalcitabine (ZAL).

The evidence found from the AIDS trial data can have some generalizability problems. When studying women living with HIV and women at risk for HIV infection in the USA cohort, the Women’s Interagency HIV Study (WIHS) (Bacon et al., 2005) has been considered to be representative. However, it was reported in Gandhi et al. (2005) that 28–68% of the HIV positive women in WIHS were excluded from the eligibility criteria of many ACTG studies. In the ACTG 175 dataset, the number of female patients is only 368 out of 2139. Thus we suspect that the female patients may be underrepresented in this dataset, and the ITR based on the dataset may not generalize well on the women subgroup. In this section, we study the generalizability of DR-ITR when the testing dataset consists of female patients only. Specifically, the training dataset is a subsample from ACTG 175 with original male/female proportion, while the testing dataset is a subsample from the female patients of ACTG 175, and there is no overlap across training and testing. We try to resemble the ideal world that we can have independent testing data from the female population.

We consider the outcome Y as the difference between the early stage (at 20±5 weeks from baseline) CD4 cell counts and the CD4 counts at baseline. We focus on the treatment comparison between the ZDV + ZAL ($A = 1$) and the ddI ($A = -1$), and the corresponding patients from the dataset. In particular, only 180 of them are women. The average treatment effects on the male and female subgroups are −8.97 and −1.39 respectively, which suggests that there is treatment effect discrepancy between these subgroups. We sample the training data from the ACTG 175 dataset in the ZDV + ZAL or ddI arm of sample size 1, 085 × 60% = 651 stratified to the gender. In particular, the training dataset includes 180 × 60% = 108 female patients. The remaining female data (180 − 108 = 72) are used for testing. We only consider female patients in testing. We further sample 50 from the testing female data for calibration, and the remaining (72 − 50 = 22) are the testing dataset. We also consider 12 selected baseline covariates $\boldsymbol{X}$ as was studied in Lu et al. (2013). There are 5 continuous covariates: age (year), weight (kg, coded as `wtkg`), CD4 count (cells/mm^3) at baseline, Karnofsky score (scale of 0–100, coded as `karnof`), CD8 count (cells/mm^3) at baseline. They are centered and scaled before further analysis. In addition, there are 7 binary variables: gender (1 = male, 0 = female), homosexual activity (`homo`, 1 = yes, 0 = no), race (1 = nonwhite, 0 = white), history of intravenous drug use (`drug`, 1 = yes, 0 = no), symptomatic status (`symptom`, 1 = symptomatic, 0 = asymptomatic), antiretroviral history (`str2`, 1 = experienced, 0 = naive) and hemophilia (`hemo`, 1 = yes, 0 = no).

Before fitting ITRs, we estimate the CTE function $C(\boldsymbol{X})$ by the following regress-and-subtract procedure: first we fit two separate random forests by regressing Y on $\boldsymbol{X}$ restricted on $A = 1$ and $A = -1$ respectively; then we subtract two regression models to obtain the CTE function estimate $\widehat{C}_n(\boldsymbol{X})$. We follow the same implementation as in Section 4.1 to fit the standard ITR and DR-ITRs over a constrained linear function class $\mathscr{F}_\gamma := \left\{ f(\boldsymbol{x}) = b + \boldsymbol{\beta}^\top \boldsymbol{x} : b \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p, \|\boldsymbol{\beta}\|_2 \leqslant \gamma \right\}$ on the training data. The testing performance is evaluated by the IPWE of the value function on the testing data. The training-calibrating-testing procedure is repeated for 1,500 times. The testing values are reported in Table 4, where the value can be interpreted as the expected CD4 count improvement from baseline at the early stage (20 ± 5 weeks). In addition to the calibrated DR-ITRs, we also include the value of the *best DR-ITR* that enjoys the highest testing performance among all DR-constants. For comparison, we include the results of residual weighted learning (RWL) (Zhou et al., 2017) with linear kernel. Both RWL and the standard ITR share similar implementation, except that RWL can be shown equivalently using $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$ as a plug-in CTE estimate.

The testing results show that our proposed DR-ITRs can have better values than the standard ITR and RWL. In particular, the improvement of the best DR-ITR is substantial, while the improvements of the calibrated ITRs are not as strong. We plot the testing values of the DR-ITRs against the corresponding DR-constants in Figure 5. It suggests that the testing values generally increase with the DR-constant. In this analysis, the calibrated DR-constants are not close to the optimal DR-constant. As a result, the testing performance of the calibrated DR-ITRs is not as good as the best DR-ITR. One reason for this phenomenon can be that the outcome Y has a heavy tail distribution, as was highlighted in Qi et al. (2019b), so that the value function estimate is highly variable based on the small calibrating sample. Another reason can be that the random forest regress-and-subtract estimate of the CTE function does not generalize well on the testing distribution.

On the overall dataset, we fit the DR-ITRs and report their fitted coefficients in Table 5 for selected DR-constants. To stabilize the randomness from the random forest estimate of the CTE function, we refit the random forest 20 times and average the corresponding DR-ITR coefficients. We find that there are noticeable changes in the coefficients of the intercept and the homosexual activity when the DR-constant gets large. Within the ACTG 175 dataset (ZDV + ZAL or ddI), we find that only 6 female patients have homosexual activity. Four of them are treated with ZDV + ZAL, and the change of their CD4 counts are 123, 34, −11 and 158 respectively. Two of them are treated with ddI, and the change of their CD4 counts are −41, −182. Therefore, the ZDV + ZAL ($A = -1$) may have more benefits compared to the ddI ($A = -1$) on these patients. This helps to explain why the larger coefficients in homosexual activity for the larger DR-constants can be beneficial for the female patients.

6 Discussion

In this paper, we propose a new framework for learning a distributionally robust ITR by maximizing the worst-case value function among values under distributions within the power uncertainty set. We introduce two possible calibration scenarios under which the DR-constant can be tuned adaptively to a small amount of the calibrating data from the target population. In this way, when the training and testing distributions are identical, the calibrated DR-ITRs can achieve similar performance as compared to the standard ITR. When the testing distribution deviates from the training distribution, we show that there are many possible scenarios that the standard ITR generalizes poorly, while the calibrated DR-ITRs maintain relatively good testing performance. Our simulation studies and an application to the ACTG 175 dataset demonstrate the competitive generalizability of our proposed DR-ITR.

The main assumption on the changes of covariates in our DR-ITR framework is equivalent to the selection unconfoundedness assumption in a randomized controlled trial. In practice, there may exist unmeasured selection confounding problems for the trial data, and the distributional changes affect both the covariates and the CTE function. One possible extension is to consider the simultaneous changes of the covariate distribution and the CTE function, and leverage more general robustness measure against these changes.

In our DR-ITR framework, we require an estimate of the CTE function based on the flexible nonparametric techniques. The performance of our DR-ITR can depend on the quality of the CTE function estimate. An alternative strategy is to avoid plugging in a CTE estimate. Instead, the dual representation (10) can be identified from $(\boldsymbol{X}, A, Y)$ directly using a variational representation of $[\pm C(\boldsymbol{X}) - \eta]_+^{k*}$ (Duchi et al., 2019). This can be a possible extension of our framework.

Another possible extension is to consider the problem of high-dimensional covariates. Our current formulation involves an ℓ_2-constraint to control the model complexity. It can be extended to obtain sparse solutions when a ℓ_1-constraint is used instead. Besides the high-dimensional extension, our current theoretical results assume that $C(\boldsymbol{X})$ is uniformly bounded. It will be interesting to relax the assumption, such as sub-Gaussianity. Further investigations along these lines can be pursued.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the editor, the associate editor, and reviewers, whose helpful comments and suggestions led to a much improved presentation.

Funding

The authors were supported in part by NSF grants IIS-1632951, DMS-1821231, and NIH grants R01GM126550 and P01 CA-142538.

References

Aggarwal CC (2016), Recommender systems, Springer.

Alyass A, Turcotte M, and Meyre D (2015), “From big data analysis to personalized medicine for all: challenges and opportunities,” BMC Medical Genomics, 8, 33. [PubMed: 26112054]

Athey S and Wager S (2017), “Efficient policy learning,” arXiv preprint arXiv:1702.02896.

Bacon MC, Von Wyl V, Alden C, Sharp G, Robison E, Hessol N, Gange S, Barranday Y, Holman S, and Weber K (2005), “The Women’s Interagency HIV Study: an observational cohort brings clinical sciences to the bench,” Clin. Diagn. Lab. Immunol, 12, 1013–1019. [PubMed: 16148165]

Ben-Tal A, Den Hertog D, De Waegenaere A, Melenberg B, and Rennen G (2013), “Robust solutions of optimization problems affected by uncertain probabilities,” Management Science, 59, 341–357.

Bertsimas D, Kallus N, Weinstein AM, and Zhuo YD (2017), “Personalized diabetes management using electronic medical records,” Diabetes Care, 40, 210–217. [PubMed: 27920019]

Buchanan AL, Hudgens MG, Cole SR, Mollan KR, Sax PE, Daar ES, Adimora AA, Eron JJ, and Mugavero MJ (2018), “Generalizing evidence from randomized trials using inverse probability of sampling weights,” Journal of the Royal Statistical Society: Series A (Statistics in Society), 181, 1193–1209.

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. [PubMed: 28211943]

Cressie N and Read TR (1984), “Multinomial goodness-of-fit tests,” Journal of the Royal Statistical Society: Series B (Methodological), 46, 440–464.

Duchi JC, Hashimoto T, and Namkoong H (2019), “Distributionally robust losses against mixture covariate shifts,” Under Review.

Duchi JC and Namkoong H (2018), “Learning Models with Uniform Performance via Distributionally Robust Optimization,” arXiv preprint arXiv:1810.08750.

Dudík M, Langford J, and Li L (2011), “Doubly Robust Policy Evaluation and Learning,” in Proceedings of the 28th International Conference on International Conference on Machine Learning, Madison, WI, USA: Omnipress, ICML’11, p. 1097–1104.

Gandhi M, Ameli N, Bacchetti P, Sharp GB, French AL, Young M, Gange SJ, Anastos K, Holman S, and Levine A (2005), “Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols,” Aids, 19, 1885–1896. [PubMed: 16227797]

Gatsonis C and Morton SC (2017), Methods in Comparative Effectiveness Research, CRC Press.

Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, and Niu M (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.

Hand DJ (2006), “Classifier Technology and the Illusion of Progress,” Statistical Science, 21, 1–14. [PubMed: 17906740]

Hill JL (2011), “Bayesian nonparametric modeling for causal inference,” Journal of Computational and Graphical Statistics, 20, 217–240.

Hotz VJ, Imbens GW, and Mortimer JH (2005), “Predicting the efficacy of future training programs using past experiences at other locations,” Journal of Econometrics, 125, 241–270.

Imai K and Ratkovic M (2013), “Estimating treatment effect heterogeneity in randomized program evaluation,” The Annals of Applied Statistics, 7, 443–470.

Johansson FD, Kallus N, Shalit U, and Sontag D (2018), “Learning weighted representations for generalization across designs,” arXiv preprint arXiv:1802.08598.

Kitagawa T and Tetenov A (2018), “Who should be treated? empirical welfare maximization methods for treatment choice,” Econometrica, 86, 591–616.

Krokhmal PA (2007), “Higher moment coherent risk measures,” Quantitative Finance, 7, 373–387.

Kube A, Das S, and Fowler PJ (2019), “Allocating interventions based on predicted outcomes: A case study on homelessness services,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 622–629.

Li S, Cai TT, and Li H (2020), “Transfer Learning for High-dimensional Linear Regression: Prediction, Estimation, and Minimax Optimality,” arXiv preprint arXiv:2006.10593.

Liu Y, Wang Y, Kosorok MR, Zhao Y, and Zeng D (2018), “Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens,” Statistics in Medicine, 37, 3776–3788. [PubMed: 29873099]

Lu W, Zhang HH, and Zeng D (2013), “Variable selection for optimal treatment decision,” Statistical Methods in Medical Research, 22, 493–504. [PubMed: 22116341]

Luo J, Schumacher M, Scherer A, Sanoudou D, Megherbi D, Davison T, Shi T, Tong W, Shi L, and Hong H (2010), “A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data,” The Pharmacogenomics Journal, 10, 278–291. [PubMed: 20676067]

Manski CF (2004), “Statistical treatment rules for heterogeneous populations,” Econometrica, 72, 1221–1246.

Muller S (2014), “Randomised trials for policy: a review of the external validity of treatment effects,” A Southern Africa Labour and Development Research Unit Working Paper Number 127.

O’Muircheartaigh C and Hedges LV (2014), “Generalizing from unrepresentative experiments: a stratified propensity score approach,” Journal of the Royal Statistical Society: Series C: Applied Statistics, 195–210.

Pardo L (2005), Statistical inference based on divergence measures, Chapman and Hall/CRC.

Pearl J and Bareinboim E (2014), “External validity: From do-calculus to transportability across populations,” Statistical Science, 579–595.

Qi Z, Cui Y, Liu Y, and Pang J-S (2019a), “Estimation of Individualized Decision Rules Based on an Optimized Covariate-Dependent Equivalent of Random Outcomes,” SIAM Journal on Optimization, 29, 2337–2362.

Qi Z, Liu D, Fu H, and Liu Y (2020), “Multi-Armed Angle-Based Direct Learning for Estimating Optimal Individualized Treatment Rules With Various Outcomes,” Journal of the American Statistical Association, 115, 678–691.

Qi Z, Pang J-S, and Liu Y (2019b), “Estimating Individualized Decision Rules with Tail Controls,” arXiv preprint arXiv:1903.04367.

Qian M and Murphy SA (2011), “Performance guarantees for individualized treatment rules,” Annals of Statistics, 39, 1180–1210. [PubMed: 21666835]

Rahimian H and Mehrotra S (2019), “Distributionally robust optimization: A review,” arXiv preprint arXiv:1908.05659.

Razaviyayn M, Hong M, and Luo Z-Q (2013), “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” SIAM Journal on Optimization, 23, 1126–1153.

Rockafellar RT and Uryasev S (2000), “Optimization of conditional value-at-risk,” Journal of Risk, 2, 21–42.

Rubin DB (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies,” Journal of Educational Psychology, 66, 688–701.

Shen X, Tseng GC, Zhang X, and Wong WH (2003), “On ψ-learning,” Journal of the American Statistical Association, 98, 724–734.

Shi C, Song R, Lu W, and Fu B (2018), “Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects,” Journal of the Royal Statistical Society. Series B, Statistical methodology, 80, 681. [PubMed: 30555269]

Steinwart I and Scovel C (2007), “Fast rates for support vector machines using Gaussian kernels,” The Annals of Statistics, 35, 575–607.

Stuart EA, Cole SR, Bradshaw CP, and Leaf PJ (2011), “The use of propensity scores to assess the generalizability of results from randomized trials,” Journal of the Royal Statistical Society: Series A (Statistics in Society), 174, 369–386.

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.

Wager S and Walther G (2015), “Adaptive concentration of regression trees, with application to random forests,” arXiv preprint arXiv:1503.06388.

Wu Y and Liu Y (2007), “Robust truncated hinge loss support vector machines,” Journal of the American Statistical Association, 102, 974–983.

Zhang B, Tsiatis AA, Davidian M, Zhang M, and Laber E (2012a), “Estimating optimal treatment regimes from a classification perspective,” Stat, 1, 103–114. [PubMed: 23645940]

Zhang B, Tsiatis AA, Laber EB, and Davidian M (2012b), “A robust method for estimating optimal treatment regimes,” Biometrics, 68, 1010–1018. [PubMed: 22550953]

Zhang C, Chen J, Fu H, He X, Zhao Y, and Liu Y (2019), “Multicategory outcome weighted margin-based learning for estimating individualized treatment rules,” Statistica Sinica.

Zhao Q, Small DS, and Ertefaie A (2017), “Selective inference for effect modification via the lasso,” arXiv preprint arXiv:1705.08020.

Zhao Y, Zeng D, Rush AJ, and Kosorok MR (2012), “Estimating individualized treatment rules using outcome weighted learning,” Journal of the American Statistical Association, 107, 1106–1118. [PubMed: 23630406]

Zhao Y-Q, Laber EB, Ning Y, Saha S, and Sands BE (2019a), “Efficient augmentation and relaxation learning for individualized treatment rules using observational data.” Journal of Machine Learning Research, 20, 1–23.

Zhao Y-Q, Zeng D, Tangen CM, and Leblanc ML (2019b), “Robustifying trial-derived optimal treatment rules for a target population,” Electronic Journal of Statistics, 13, 1717–1743. [PubMed: 31440323]

Zhou X, Mayer-Hamblett N, Khan U, and Kosorok MR (2017), “Residual weighted learning for estimating individualized treatment rules,” Journal of the American Statistical Association, 112, 169–187. [PubMed: 28943682]

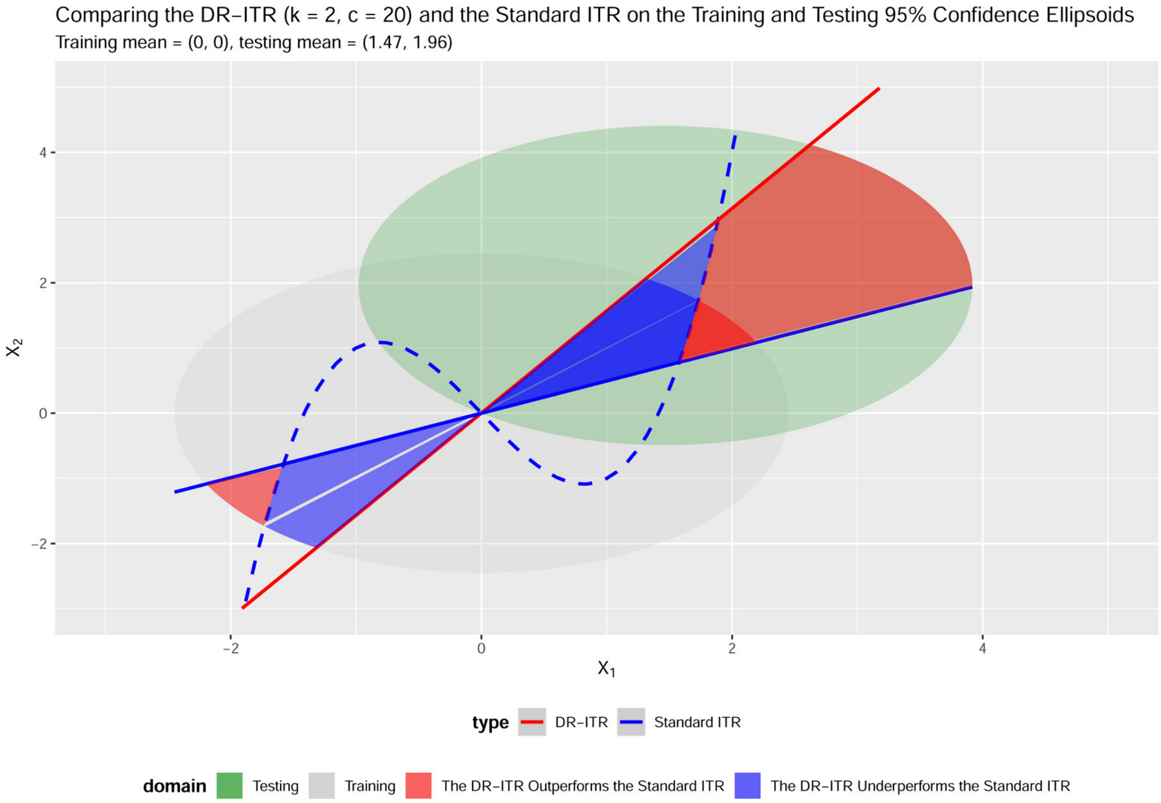
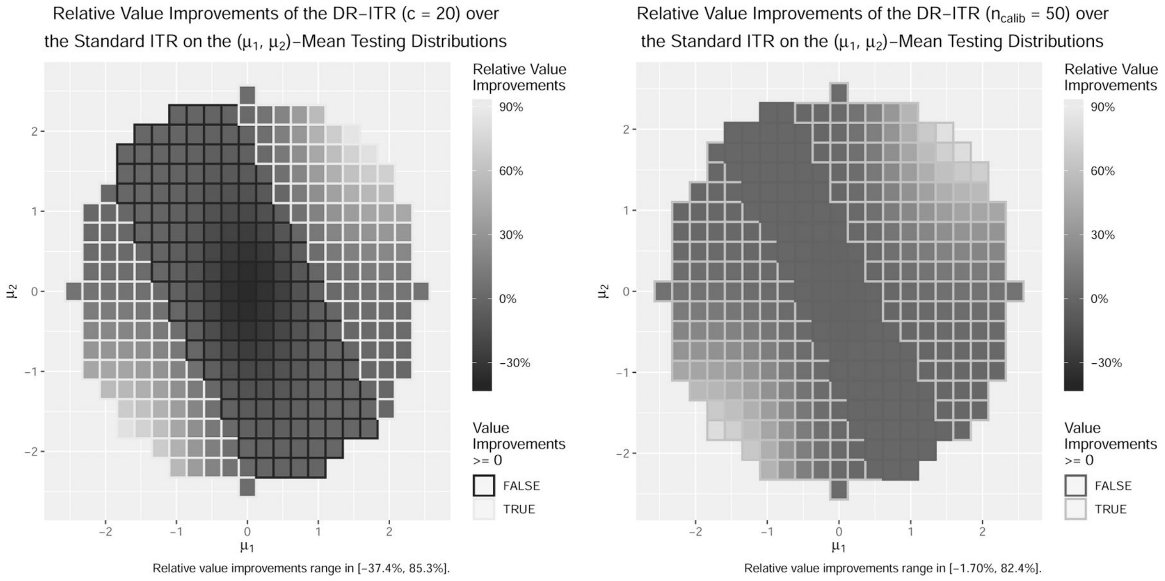


Figure 1:

ITRs and the 95% confidence ellipsoids of the training distribution $(X_1, X_2) \sim \mathcal{N}_2((0,0)^\top, \mathbf{I}_2)$ and the testing distribution $(X_1, X_2) \sim \mathcal{N}_2((1.47, 1.96)^\top, \mathbf{I}_2)$. The blue dashed curve is the underlying CTE boundary $C(X_1, X_2) = X_2 - (X_1^3 - 2X_1) = 0$.



Relative value improvements range in [−37.4%, 85.3%].

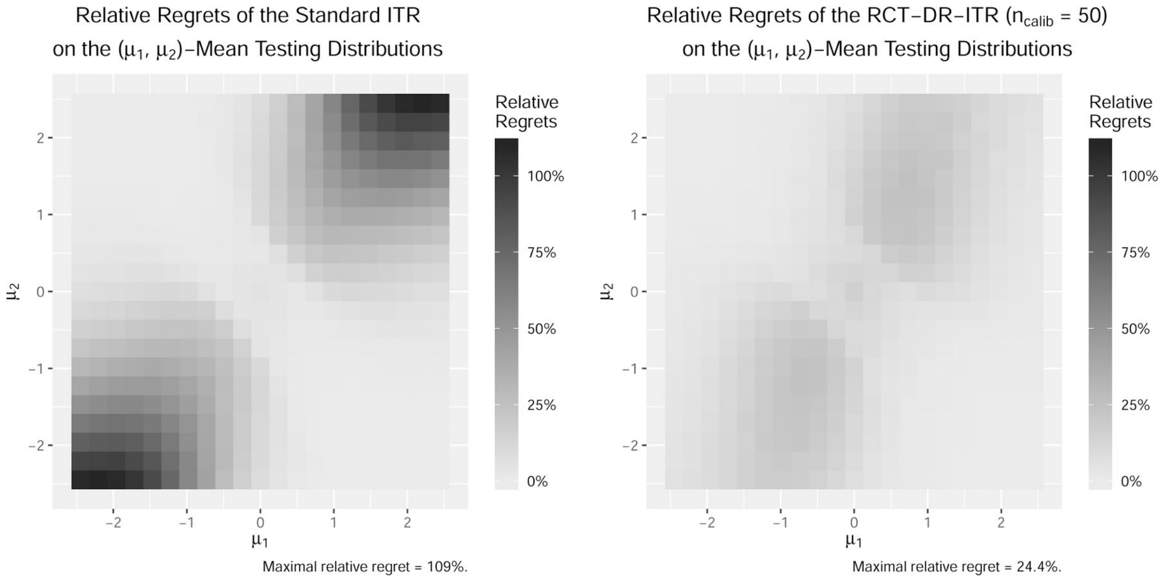
Relative value improvements range in [−1.70%, 82.4%].

(a) $c = 20$

(b) Calibrating c on a size-50 Sample

Figure 2:

Relative improvements of the DR-ITR over the standard ITR as the difference of relative regrets on testing distributions $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{I}_2)$ of $\boldsymbol{\mu} \in \{(\mu_1, \mu_2)^\top \in \mathbb{R}^2 : \mu_1^2 + \mu_2^2 \leqslant 4\log 5\}$ (lighter the better).



(a) Standard ITR

(b) RCT-DR-ITR ($n_{\text{calib}} = 50$)

Figure 3:

Relative Regrets on the Mean-Shifted Covariate Domains (lighter the better).

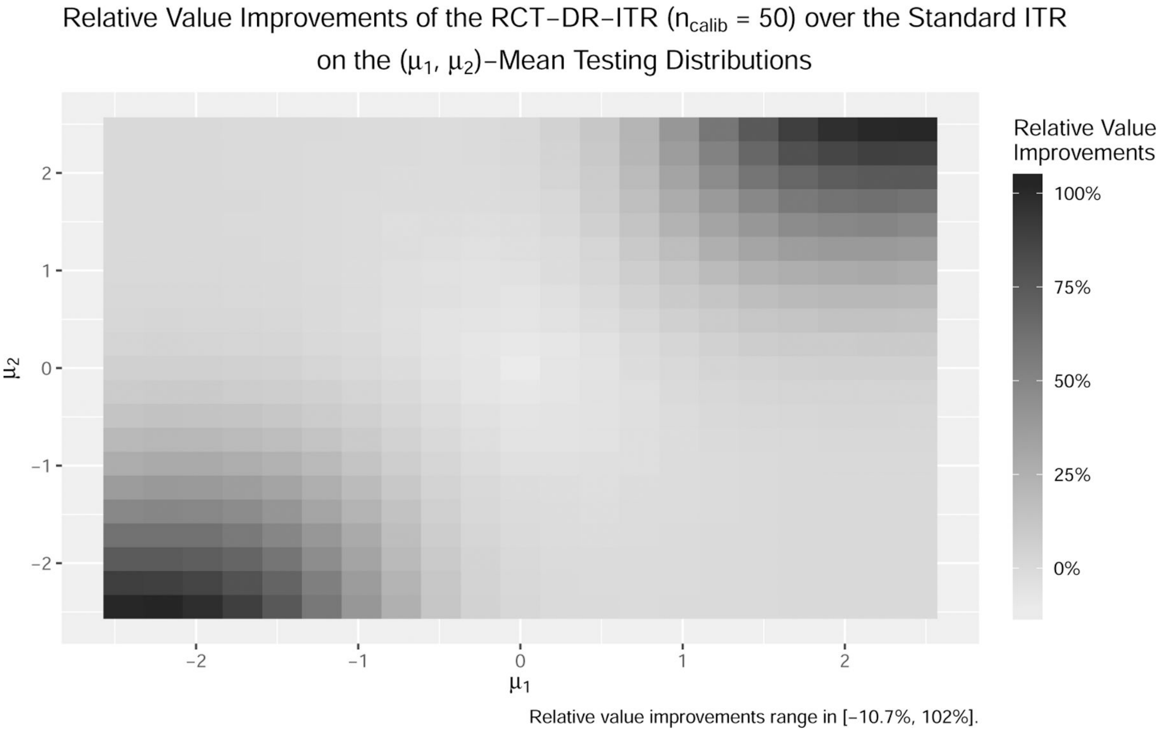


Figure 4:

Relative improvements of the RCT-DR-ITR over the standard ITR as the difference of their relative regrets on the mean-shifted covariate domains (n_{calib} = 50, darker the better).

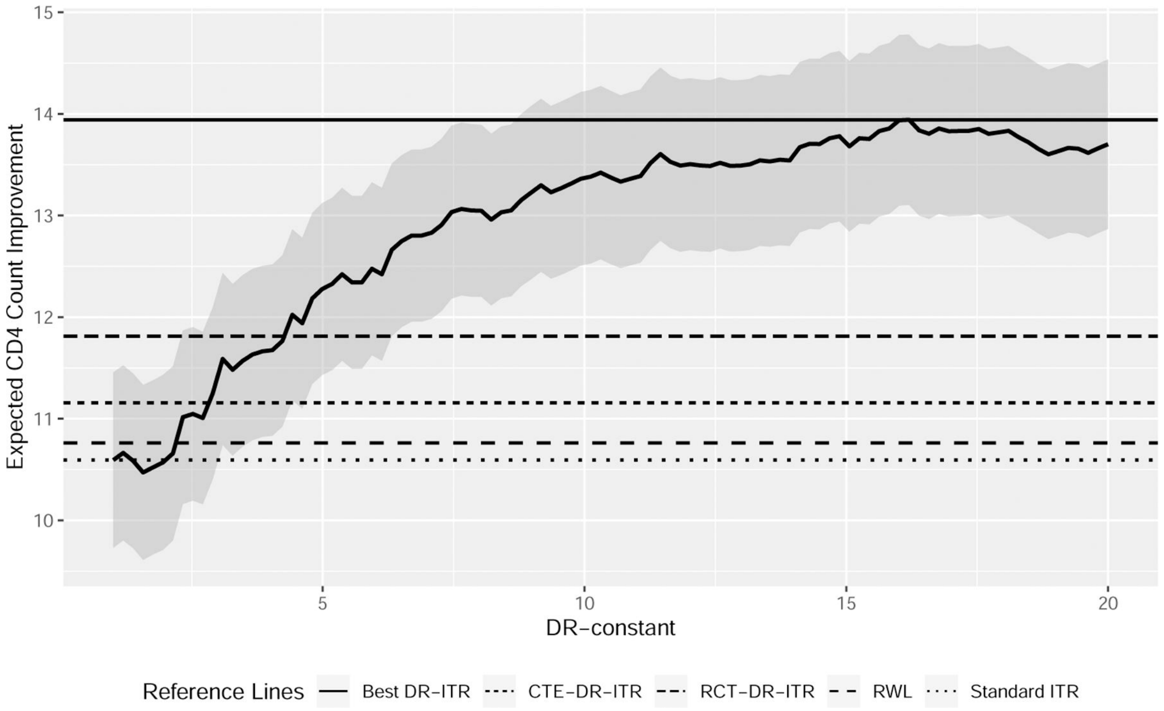


Figure 5:

Expected CD4 Count Improvement (cells/mm^3) from Baseline at the Early Stage (20±5 weeks) of the DR-ITRs of Various DR-Constants on the ACTG 175 Female Patients (higher the better)

Table 1:

Testing Values (Relative Regrets) Comparisons of ITRs

Value \ ITR	DR-ITR	Standard ITR	LB-ITR
Training $\mathscr{V}_1$	0.6253 (37.36%)	0.9982 (0%)	0.9982
Testing $\mathscr{V}_{1,\text{ test}}$	4.8230 (9.16%)	0.2927 (94.49%)	5.3096

[1] DR-ITR maximizes $\mathscr{V}_c^k(d)$ defined in (4) with $k = 2$ and $c = 20$ over the linear ITR class.

[2] Standard ITR maximizes $\mathscr{V}_1(d)$ over the linear ITR class.

[3] LB-ITR maximizes $\mathscr{V}_1(d)$ or $\mathscr{V}_{1,\text{ test}}(d)$ over the linear ITR class.

[4] Values (larger the better) can be comparable within rows but incomparable between rows.

[5] Relative $\text{regret(ITR)} = [\text{value(LB} - \text{ITR)} - \text{value(ITR)}]/|\text{value(LB} - \text{ITR)}|$ (smaller the better)

[6] A size-10,000 sample is generated for fitting DR-ITR and LB-ITRs, and an independent size-100,000 sample is generated for evaluation under $\mathscr{V}_1$ and $\mathscr{V}_{1,\text{ test}}$.

Table 2:

Testing Values (Standard Errors) on the Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

$\mu_2 \diagdown \mu_1$	**Type**	**0**	**0.734**	**1.469**	**1.958**
1.958	LB-ITR	*2.333 (0.00244)*	*2.907 (0.011)*	*5.334 (0.0362)*	*9.27 (0.0154)*
	ℓ_1-PLS	**2.124** (0.0022)	2.235 (0.011)	3.613 (0.0505)	6.32 (0.103)
	Standard ITR	2.089 (0.00158)	1.735 (0.013)	1.348 (0.0595)	1.567 (0.13)
	RCT-DR-ITR	2.085 (0.00444)	2.286 (0.0114)	4.545 (0.0255)	8.371 (0.0451)
	CTE-DR-ITR	2.098 (0.00348)	**2.304** (0.0106)	**4.551** (0.0238)	**8.459** (0.0424)
1.469	LB-ITR	*1.893 (0.00712)*	*2.627 (0.00656)*	*5.28 (0.0213)*	*9.379 (0.0128)*
	ℓ_1-PLS	1.667 (0.00307)	**2.021** (0.0076)	4.095 (0.0342)	7.573 (0.0706)
	Standard ITR	**1.674** (0.00152)	1.645 (0.0127)	2.377 (0.0553)	4.011 (0.119)
	RCT-DR-ITR	1.627 (0.00688)	1.987 (0.00997)	4.484 (0.0192)	8.611 (0.0285)
	CTE-DR-ITR	1.663 (0.00326)	1.997 (0.00992)	**4.55** (0.0163)	**8.686** (0.0269)
0.734	LB-ITR	*1.227 (0.00244)*	*2.144 (0.00609)*	*5.269 (0.00931)*	*9.608 (0.00898)*
	ℓ_1-PLS	1.094 (0.00418)	**1.676** (0.00442)	4.587 (0.0151)	8.8 (0.0314)
	Standard ITR	**1.174** (0.00149)	1.553 (0.00806)	3.739 (0.0379)	7.06 (0.0763)
	RCT-DR-ITR	1.094 (0.00753)	1.651 (0.00675)	4.622 (0.0109)	9.036 (0.015)
	CTE-DR-ITR	1.152 (0.00292)	1.667 (0.00588)	**4.648** (0.0113)	**9.06** (0.0161)
0.000	LB-ITR	*0.9942 (0.00202)*	*1.774 (0.0034)*	*5.232 (0.00559)*	*9.767 (0.0068)*
	ℓ_1-PLS	0.8296 (0.00454)	1.648 (0.0036)	**4.914** (0.00501)	**9.476** (0.0103)
	Standard ITR	**0.9437** (0.00153)	1.679 (0.00336)	4.654 (0.017)	8.895 (0.0342)
	RCT-DR-ITR	0.8374 (0.00821)	1.647 (0.00574)	4.868 (0.00797)	9.444 (0.00841)
	CTE-DR-ITR	0.9206 (0.00272)	**1.688** (0.00289)	4.888 (0.00698)	9.442 (0.00999)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \ldots, 0)^\top$ with μ_1 in column and μ_2 in row is the testing covariate centroid.

[2] Values (larger the better) can be comparable for the same (μ_1, μ_2) but incomparable across different (μ_1, μ_2).

[3] LB-ITR maximizes the testing value function at (μ_1, μ_2) over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.

Table 3:

Testing Values (Standard Errors) on the Mixture of Subgroups ($n_{\text{calib}} = 50$)

	Testing Subgroup 1 Probability				
type	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**
LB-ITR	*1.665 (0.0067)*	*1.537 (0.00618)*	*1.444 (0.00412)*	*1.545 (0.00537)*	*1.679 (0.00585)*
ℓ_1-PLS	1.182 (0.00191)	1.264 (0.0014)	**1.399** (0.000591)	**1.537** (0.000333)	1.624 (0.000781)
Standard ITR	1.143 (0.00434)	1.232 (0.00329)	1.383 (0.0015)	1.535 (0.000543)	**1.632** (0.00142)
RCT-DR-ITR	**1.267** (0.0066)	**1.305** (0.00423)	1.395 (0.00256)	1.52 (0.00212)	1.614 (0.00234)
CTE-DR-ITR	1.16 (0.00409)	1.247 (0.00323)	1.388 (0.00137)	1.534 (0.00055)	1.628 (0.00149)

[1] Testing subgroup 1 probability = 0.75 is the same as the training one.

[2] Values (larger the better) can be comparable for the same subgroup 1 probability but incomparable across different subgroup 1 probabilities

[3] LB-ITR maximizes the testing value function over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.

Table 4:

Expected CD4 Count Improvement (cells/mm^3) from Baseline at the Early Stage (20±5 weeks) and Standard Errors on the ACTG-175 Female Patients (higher the better).

RWL	Standard ITR	Best DR-ITR	RCT-DR-ITR	CTE-DR-ITR
10.7617 (0.8636)	10.593 (0.8627)	**13.9423** (0.8378)	11.8133 (0.8357)	11.1563 (0.8514)

Standard errors are computed based on 1,500 replications.

Table 5:

Linear Coefficients of the DR-ITRs Fitted on the ACTG 175 Dataset

DR-constant	Intercept	age	wtkg	cd40	karnof	cd80	gender	homo	race	drugs	symptom	str2	hemo
1	−0.02	−0.25	0.06	−0.58	−0.06	0.53	−0.16	−0.4	0.16	0.16	0.16	0.16	0.09
4.8	−0.31	−0.23	0.12	−0.67	0.11	0.55	−0.12	−0.21	0.2	0.12	0.1	−0.06	0.09
8.6	−0.43	−0.23	0.11	−0.64	0.16	0.54	−0.11	−0.05	0.12	0.04	0.07	−0.24	0.01
12.4	−0.54	−0.22	0.1	−0.64	0.19	0.51	−0.04	0.01	0.08	0.05	0.04	−0.27	−0.02
16.2	−0.61	−0.23	0.1	−0.64	0.2	0.51	0	0.03	0.06	0.05	0.02	−0.27	−0.02
20	−0.64	−0.24	0.09	−0.63	0.22	0.5	0.01	0.03	0.05	0.07	0.01	−0.26	−0.01

[1] DR-constant = 1 corresponds to the standard ITR; DR-constant = 16.2 has the highest testing value in Figure 5.