Toward Fair and Robust Optimal Treatment Regimes

Kwangho Kim
Harvard Medical School
kkim@hcp.med.harvard.edu

José R. Zubizarreta
Harvard University
zubizarreta@hcp.med.harvard.edu

Abstract

We propose a new framework for robust nonparametric estimation of optimal treatment regimes under flexible fairness constraints. Under standard regularity conditions we show that the resulting estimators possess the double robustness property. We use this framework to characterize the trade-off between fairness and the maximum utility that is achievable by the optimal treatment policy.

1 Introduction

In today's world, an increasing number of decisions that affect people's lives are automatically made by machine learning models. Such decision-making systems are implemented in various settings ranging from financial investment to healthcare policy. Considering the importance of such decisions at an individual and societal level, it is crucial to ensure that the underlying models are not only accurate but fair. In this work, by *fairness* we mean that the models are not biased so that they do not systematically benefit or harm a specific group of people, such as a minority ethnic group. The need to address such algorithmic biases has given rise to an explosion of works studying algorithmic fairness (e.g., see [3] for a review). However, despite the considerable amount of studies in this area, comparatively little attention has been given to fairness in causal inference. In this work, we propose a novel framework for estimating optimal treatment assignments or regimes in a fair and robust manner, leveraging recent developments in counterfactual optimization [20, 21].

1.1 Related Work

Much of the earlier work on estimating optimal treatment regimes involves postulating a parametric model for the outcome regression function [e.g., 4, 10, 30, 37]. More robust approaches based on the idea of doubly robust estimation have also been proposed, for example, in [46, 47]. In recent studies [1, 13, 22], flexible nonparametric approaches are discussed where an optimal policy is deployed from a pre-specified class that can encode problem-specific constraints. However, they do not provide means to incorporate general fairness constraints.

In order to mitigate algorithmic biases where model performance varies over sensitive features, a wide array of fairness criteria have been developed typically by placing restrictions on the joint distribution of model outcomes and sensitive features. Popular fairness criteria include independence (or statistical parity) [3] and separation (or equalized odds) [9]. In some cases such as risk assessment settings [e.g., 7], counterfactual fairness may be of interest where fairness criteria depend on potential (or counterfactual) outcomes with respect to the sensitive feature [e.g., 23, 31] or a decision variable [e.g., 7, 28, 29]. Some proposed constraint-based frameworks to flexibly incorporate such fairness criteria in classification [e.g., 28, 45], but it is not clear how to extend these frameworks to enable the design of fair optimal treatment regimes. It is also well known that there exists a fairness-accuracy tradeoff, because in some cases the most accurate models under consideration do not satisfy a chosen fairness criterion [e.g., 27, 29, 36]. However, the tradeoff between fairness and treatment utility, if any, has never been formally explored.

NeurIPS 2022Workshop on Causal Machine Learning for Real-World Impact (CML4Impact 2022).

Interestingly, little work has been done at the intersection of these two areas. Most results in the algorithmic fairness literature are not directly applicable to optimal treatment regimes where our objective function involves a particular form of counterfactual functionals. A few important exceptions include [32] which integrates algorithmic fairness and policy learning using tools from mediation analysis, and [44] which proposes an estimator for the Pareto optimal policy that minimizes unfairness through a mixed-integer quadratic programming.

1.2 Contribution

Our method builds on a promising literature at the intersection of algorithmic fairness, causal inference, and stochastic optimization, bridging the gap between algorithmic fairness and optimal treatment regimes. At this intersection, our contribution is twofold. First, we propose a robust estimator of optimal treatment regimes under general fairness constraints. We cast our estimator as a convex quadratic program that can be readily solved with off-the-shelf solvers. We show that the resulting estimators are doubly robust under standard regularity conditions. Our proposed approach contributes to [32] in terms of robustness, and to [44] in terms of ease of implementation and interpretability. Second, by analyzing the regret bound, we characterize the trade-off between the maximum possible benefit and fairness. This will be useful for understanding, for example, how a desired level of fairness requires a utility compromise.

2 Setup and Framework

2.1 Optimal Treatment Regimes

Suppose that we have access to an i.i.d. sample $(Z_1, ..., Z_n)$ of n tuples $Z = (Y, A, S, X) \sim \mathbb{P}$ for some distribution $\mathbb{P}$, outcome $Y \in \mathbb{R}$, binary intervention $A \in \{0, 1\}$, sensitive feature $S \in \{0, 1\}$, and additional covariates $X \in \mathcal{X} \subset \mathbb{R}^{d_x}$ for some compact subset $\mathcal{X}$. Throughout we assume larger values of Y are preferred. We let $W = (S, X) \in \mathcal{W}$ represent the measured pre-intervention variables and let Y^a denote the potential outcome that would have been observed (possibly contrary to fact) under treatment or intervention $A = a$. A policy maker has to choose a treatment policy or a treatment regime[1] that is a function $g : \mathcal{W} \to \{0, 1\}$ to determine whether individuals with covariates W will be assigned to the treatment 0 or 1. For an arbitrary treatment regime g, we define the *welfare* or *utility* function for which the treatment regime $g \in \mathcal{G}$ is applied to the population $\mathbb{P}$ by

$$\mathcal{U}(g) = \mathbb{E}\left\{Y^1 g(W) + Y^0 (1 - g(W))\right\}.$$

Throughout we assume the standard causal assumptions of *consistency*, *no unmeasured confounding*, and *positivity* [e.g., 11, Chapter 12]. Under these assumptions, it is straightforward to show that the optimal treatment regime leading to the largest value of $\mathcal{U}(g)$ is given by

$$g^*(W) = \mathbb{1}\left\{\mu_1(W) > \mu_0(W)\right\}, \tag{1}$$

where $\mu_a(W) = \mathbb{E}[Y \mid W, A = a], \forall a \in \{0, 1\}$; i.e., the optimal regime assigns the treatment that yields the larger mean outcome conditional on the individual characteristics.[2]

2.2 Simple Motivating Example

Sometimes, efficient estimation of g^* in (1) alone can result in unfair treatment policies. Consider the following simple data-generating process

$$A \sim Bernoulli(0.5), \quad X \sim Unif[-1, 1]$$
$$\mathbb{P}(S = 1) = expit(7.5X), \quad \mu_a(W) = AX,$$

where $expit$ and $\text{Unif}(l, u)$ denote the inverse logit function and the uniform distribution over the interval $[l, u]$. Then the optimal treatment regime is $\mathbb{1}(X > 0)$. However, when we generate 100 samples, as can be seen in Figure 1, a serious fairness problem is observed; under the optimal treatment regime only less than 7% of individuals with $S = 0$ are treated, while more than 95% of individuals in the untreated group are $S = 0$.

[1] In this work, we use the terms "treatment policy" and "treatment regime" interchangeably to refer to any mapping from the pre-treatment variables to the treatment.

[2] Here, the strict inequality follows from the convention [see, e.g., 46].

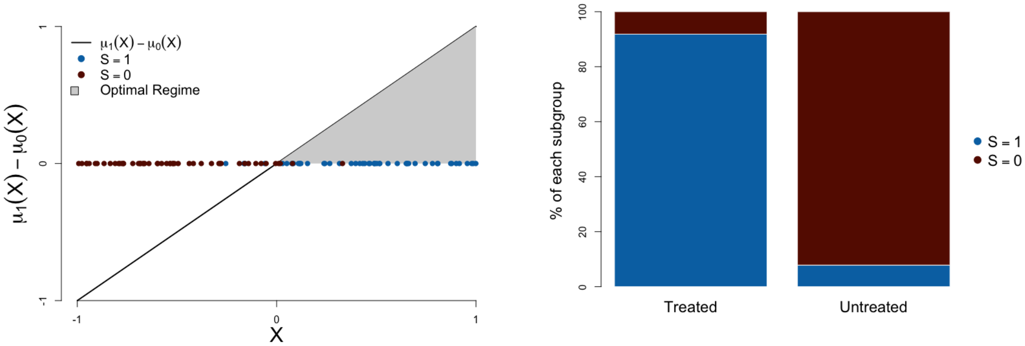

Figure 1: When the optimal treatment regime is applied, only less than 7% of individuals with $S = 0$ are treated while more than 95% of individuals in the untreated group are $S = 0$.

Here, group $S = 0$ is discriminated by the estimated optimal treatment regime that is designed to result in the greatest benefit overall in the population. In data-driven decision-making, this kind of algorithmic bias can lead to critical issues in the real world as illustrated in the following examples.

- Stop-and-Frisk: if A represents the policing practice of stop-and-frisk program, the established optimal treatment regime could be used as a recipe for discriminatory practice of stop-and-frisk toward specific ethnic groups.
- Medical Resource Allocation: if A represents access to medical treatment or health care resources, many recent studies advocate not only cost-effectiveness but also other ethical values for rationing limited health resources [e.g., 8, 36].

2.3 Proposed Framework

In this section, we lay out a framework for estimating optimal treatment regimes where we can minimize algorithmic unfairness below a particular level. Our strategy is to estimate each outcome regression function μ_a satisfying desired fairness criteria, and then plug back into the formula (1) so that the same fairness criteria are also satisfied in the optimal regime.

Specifically, we aim to estimate a functional approximation of μ_a, defined by a projection onto a finite-dimensional parametric model subject to fairness constraints. Our target parameter can be reformulated as the following constrained stochastic optimization problem

$$\begin{aligned} \underset{\beta \in \mathcal{B}}{\text{minimize}} \quad & \mathcal{L}_{\text{MSE}}\left(Y^a, \beta^\top \boldsymbol{b}(W)\right) := \mathbb{E}\left\{\left(Y^a - \beta^\top \boldsymbol{b}(W)\right)^2\right\} \\ \text{subject to} \quad & \beta \in \mathcal{C}_{\text{fair}} := \left\{\beta \;\middle|\; \left|\mathbb{E}\left\{g_j(Y^a, W)\beta^\top \boldsymbol{b}(W)\right\}\right| \le \delta_j,\, j \in J\right\}, \\ & \beta \in \mathcal{C}_{\text{lin}} \end{aligned} \tag{P_{μ_a}}$$

for some $\delta_j \ge 0$ and $J = \{1, ..., m\}$. δ_j is a prespecified tolerance for the maximum acceptable level of unfairness. The solution of the above program corresponds to the coefficients of the estimated best-fitting function of μ_a on the finite-dimensional model space spanned by the basis functions $\boldsymbol{b}(W) = [b_1(W), ..., b_k(W)]^\top$ subject to m fairness constraints in $\mathcal{C}_{\text{fair}}$. $\mathcal{C}_{\text{lin}}$ is a set of other deterministic linear constraints which could be used for penalization or incorporating prior information. This can be generalized to the nonlinear constraints at the expense of stronger regularity conditions [20]. Note that we do not assume anything about the true functional relationship between Y^a and W. This form of aggregated estimators are widely used in nonparametric regression [e.g., 12, 42].

Following [28], we use the canonical form of *fairness function* $g_j : \mathcal{W} \times \mathcal{Y} \to \mathbb{R}$ to accommodate a broad range of fairness measures. For example, the criterion of *independence* that requires our model to be independent of the sensitive feature can be applied by letting

$$g_j(Y^a, W) = \frac{1 - S}{\mathbb{E}(1 - S)} - \frac{S}{\mathbb{E}(S)},$$

which leads to $\left|\mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S=0\right\} - \mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S=1\right\}\right| \le \delta_j$. We refer to [28, Section 3] for more examples.

Similar projection approaches have also been used in causal inference [e.g., 19, 34, 39]. There are several reasons why the above projection approach is preferred in our setting. First, as will be seen shortly, the coefficients β may be estimated with flexible nonparametric methods while achieving the property of double robustness and tractable inference, and so does the target parameter g^*. It also provides interpretability; it allows practitioners to understand and audit the resulting optimal regimes according to the specified level of unfairness. Further, one may flexibly incorporate not only various fairness constraints but also other practical constraints into estimation. Finally, the optimal solution of (P_{μ_a}) can be readily estimated by solving the convex quadratic program that approximates (P_{μ_a}), which will be described in the following section.

Remark 1. *Another notable feature of our framework is that as in [20] we can consider a general setting where only a subset of covariates $V \subseteq W$ can be used for predicting the counterfactual outcome Y^a. This allows for runtime confounding, where some factors used by decision-makers are recorded in the training data (used to construct nuisance estimates) but are not available for prediction (see [6] and references therein).*

Notation. Here we briefly introduce some notation used throughout this paper. For any fixed vector v, we let $\|v\|_q$ denote the L_q-norm. Let $\mathbb{P}_n$ denote the empirical measure over $(Z_1, ..., Z_n)$. Given a sample operator h (e.g., an estimated function), we let $\mathbb{P}$ denote the conditional expectation over a new independent observation Z, as in $\mathbb{P}(h) = \mathbb{P}\{h(Z)\} = \int h(z)d\mathbb{P}(z)$ [3]. Then we use $\|h\|_{q,\mathbb{P}}$ to denote the $L_q(\mathbb{P})$ norm of h defined by $\|h\|_{q,\mathbb{P}} = \left[\int |h(z)|^q d\mathbb{P}(z)\right]^{\frac{1}{q}}$. Lastly, we let $\lesssim$ denote less than or equal to up to a nonnegative constant.

3 Estimation and Inference

(P_{μ_a}) is not directly solvable so we need to find an *approximating program* of the “true" program (P_{μ_a}). A complication arises since standard approaches to stochastic programming such as *stochastic approximation* (SA) and *sample average approximation* (SAA) [e.g., 33, 40] are infeasible in our setting, because i) the relevant sample moments and stochastic (sub)gradients depend on unobserved counterfactuals, and ii) these approaches cannot incorporate efficient semiparametric estimators with cross-fitting [5, 35]. We therefore build our estimators on the recent developments by [20, 21] where counterfactual components are estimated more flexibly.

For convenience, define the following:

$$\pi_a(X) = \mathbb{P}[A = a \mid X],$$
$$\varphi_a(Z;\eta) = \frac{\mathbb{1}(A=a)}{\pi_a(X)}\left\{Y - \mu_A(X)\right\} + \mu_a(X).$$

φ_a is the uncentered efficient influence function for the parameter $\mathbb{E}\left\{\mathbb{E}[Y \mid X, A=a]\right\}$ with a set of the nuisance components defined by $\eta = \{\pi_a(X), \mu_a(X)\}$ [15].

First, we provide influence-function-based semiparametric estimators for each component of (P_{μ_a}). Following [5, 16, 38, 48], we propose to use *sample splitting* (or *cross fitting*) to allow for arbitrarily complex nuisance estimators $\widehat{\eta}$. Specifically, we split the data into K disjoint groups, each with size n/K approximately, by drawing variables $(B_1, ..., B_n)$ independent of the data, with $B_i = b$ indicating that subject i was split into group $b \in \{1, ..., K\}$. Then the semiparametric estimators for $\mathcal{L}_{\text{MSE}}$ and each element in $\mathcal{C}_{\text{fair}}$ based on the efficient influence function and sample splitting are given by

$$\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left\{\left(\varphi_a(Z;\widehat{\eta}_{-b}) - \beta^\top \boldsymbol{b}(W)\right)^2\right\} \equiv \mathbb{P}_n\left\{\left(\varphi_a(Z;\widehat{\eta}_{-K}) - \beta^\top \boldsymbol{b}(W)\right)^2\right\},$$
$$\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left\{g_j(\varphi_a(Z;\widehat{\eta}_{-b}), W)\beta^\top \boldsymbol{b}(W)\right\} \equiv \mathbb{P}_n\left\{g_j(\varphi_a(Z;\widehat{\eta}_{-K}), W)\beta^\top \boldsymbol{b}(W)\right\},$$

[3] When h is a fixed operator, $\mathbb{P}$ and $\mathbb{E}$ are used interchangeably.

where we let $\mathbb{P}_n^b$ denote empirical averages only over the set of units $\{i : B_i = b\}$ in group b and let $\widehat{\eta}_{-B_K}$ denote the nuisance estimator constructed only using those units $\{i : B_i \neq b\}$. Under weak regularity conditions, these sample-splitting-based semiparametric estimators attain the efficiency bound with the double robustness property, and thus allow us to employ flexible machine learning estimation methods while achieving the $\sqrt{n}$-rate of convergence and valid inference [15][4]. Consequently, our approximating program can be found as the following convex quadratic program (QP)

$$\begin{aligned} &\underset{\beta\in\mathcal{B}}{\text{minimize}} \quad \mathbb{P}_n\left\{\left(\varphi_a(Z;\widehat{\eta}_{-b}) - \beta^\top \boldsymbol{b}(W)\right)^2\right\} \\ &\text{subject to} \quad \beta\in\widehat{\mathcal{C}}_{\text{fair}},\ \beta\in\mathcal{C}_{\text{lin}}, \end{aligned} \tag{$\widehat{\mathsf{P}}_{\mu_a}$}$$

where $\widehat{\mathcal{C}}_{\text{fair}} := \left\{\beta \;\middle|\; \left|\mathbb{P}_n\left\{g_j(\varphi_a(Z;\widehat{\eta}_{-b}), W)\beta^\top \boldsymbol{b}(W)\right\}\right| \leq \delta_j,\ j\in J\right\}$. ($\widehat{\mathsf{P}}_{\mu_a}$) can be readily solved using off-the-shelf QP solvers. Next, we introduce the following assumptions for our counterfactual component estimators.

(A1) $\mathbb{P}(\widehat{\pi}_a \in [\epsilon, 1-\epsilon]) = 1$ for some $\epsilon > 0$

(A2) $\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(1)$ or $\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(1)$

(A3) $\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}}\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(n^{-\frac{1}{2}})$

Assumptions (A1) - (A3) are commonly used in semiparametric estimation in the causal inference literature [14]. In the following theorem, we provide the large-sample properties of our proposed estimator.

Theorem 3.1. *Let β^* and $\widehat{\beta}$ denote the optimal solutions to (P_{μ_a}) and ($\widehat{\mathsf{P}}_{\mu_a}$), respectively. If Assumptions (A1) and (A2) hold, then*

$$\|\widehat{\beta} - \beta^*\|_2 = O_{\mathbb{P}}\left(\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}}\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} \vee n^{-\frac{1}{2}}\right).$$

If we additionally assume (A3), uniqueness of β^*, and that the Linear Independence Constraint Qualification (LICQ) and Strict Complementarity (SC) hold at β^*, then $\sqrt{n}(\widehat{\beta} - \beta^*)$ converges in distribution to a zero-mean normal random variable. Further, $\widehat{\beta}$ is efficient, meaning that there exist no other regular asymptotically linear estimators that are asymptotically unbiased and have smaller variance.

The above result immediately follows by Theorems 3.1 and 3.2 of [21], and gives conditions under which $\widehat{\beta}$ is $\sqrt{n}$-consistent and asymptotically normal. Thus, asymptotically valid confidence intervals and hypothesis tests can be constructed via the bootstrap. LICQ and SC are regularity conditions commonly found in the optimization literature [e.g., 40, 41]; see Appendix A for the formal definitions. The uniqueness of β^* simply requires that our basis functions are never perfectly collinear.

Once we obtain $\widehat{\beta}_1$ and $\widehat{\beta}_0$ through ($\widehat{\mathsf{P}}_{\mu_a}$), our proposed estimator for g^* is given by

$$\widehat{g}(W) = \mathbb{1}\left\{\widehat{\beta}_1^\top \boldsymbol{b}(W) > \widehat{\beta}_0^\top \boldsymbol{b}(W)\right\}. \tag{2}$$

Following the convention in the literature, we evaluate the performance of the above estimated treatment regime $\widehat{g}$ in terms of the utility loss or *regret* relative to the maximum obtainable utility $\mathcal{U}(g^*)$, i.e., $\mathcal{U}(g^*) - \mathcal{U}(\widehat{g})$, as will be analyzed in the following section in detail.

4 Regret Bounds and Fairness-Welfare Tradeoff

Here, we analyze the regret upper bounds and discuss its implication in incorporating fairness into optimal treatment regimes. To derive the upper bounds we require a margin condition, which restricts the probability that the two outcome regression functions get too close to each other in the neighborhood of $\mu_1 = \mu_0$.

[4]If one is willing to rely on appropriate empirical process conditions (e.g., Donsker-type or low entropy conditions [43]), then η can be estimated on the same sample without sample splitting. However this would limit the flexibility of the nuisance estimators.

Definition 4.1 (Margin Condition)**.** *For some $\alpha > 0$ and for all t, we have that*

$$\mathbb{P}(|\mu_1(W) - \mu_0(W)| \leq t) \lesssim t^{\alpha}. \tag{3}$$

The above margin condition is analogous to that used in [17, 22, 25, 26] as well as other problems involving estimation of non-smooth parameters such as classification [2], clustering [24].

In the next lemma, adapting the comparison lemmas in [2], we give two useful inequalities between the regrets and the general L_q risks of the corresponding outcome regression estimators proposed in the previous section under the margin condition.

Lemma 4.1. *Assume that the margin condition* (3) *holds with the margin exponent $\alpha > 0$, and let $\Delta \equiv \Delta(W) = \mu_1(W) - \mu_0(W)$ and $\widehat{\Delta} \equiv \widehat{\Delta}(W) = \widehat{\beta}_1^\top \boldsymbol{b}(W) - \widehat{\beta}_0^\top \boldsymbol{b}(W)$. Then we have*

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\| \widehat{\Delta} - \Delta \right\|_{\infty,\mathbb{P}}^{\alpha+1}.$$

Further, for any $1 \leq q < \infty$, we have

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\| \widehat{\Delta} - \Delta \right\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}}.$$

Based on the above lemma, the next theorem gives the upper bounds of the utility regret for our proposed estimator $\widehat{g}$ in (2). Our results are asymptotic in the sample size n.

Theorem 4.1. *Assume (A1) and (A2) and that the margin condition* (3) *holds with the margin exponent $0 < \alpha < \infty$. Also let*

$$\beta_a^* = \underset{\beta \in \mathcal{B}}{\arg\min} \quad \mathbb{E}\left\{ \left(Y^a - \beta^\top \boldsymbol{b}(W)\right)^2 \right\}, \tag{4}$$

and define the remainder terms

$$R_{1,n} = O_{\mathbb{P}}\left(\|\widehat{\pi}(W) - \pi(W)\|_{2,\mathbb{P}} \max_a \|\widehat{\mu}_a(W) - \mu_a(W)\|_{2,\mathbb{P}} \vee n^{-\frac{1}{2}} \right),$$

$$R_2 = O\left(\sum_{a,j} \lambda_j \left\| g_j(Y^a, W)\boldsymbol{b}(W) \right\|_{2,\mathbb{P}} \right),$$

where $\lambda_j \geq 0$ is the Lagrange multiplier associated with the j-th fairness constraint in (P_{μ_a}). *Then we have*

$$\text{(i)}\ \mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \max_a \left\| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right\|_{\infty,\mathbb{P}}^{1+\alpha} + R_{1,n}^{1+\alpha} + R_2^{1+\alpha},$$

$$\text{(ii)}\ \Pr\{\widehat{g}(W) \neq g^*(W)\} \lesssim \max_a \left\| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right\|_{\infty,\mathbb{P}}^{\alpha} + R_{1,n}^{\alpha} + R_2^{\alpha},$$

$$\text{(iii)}\ \mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \max_a \left\| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}} + R_{1,n}^{\frac{q(1+\alpha)}{q+\alpha}} + R_2^{\frac{q(1+\alpha)}{q+\alpha}}, \forall 1 \leq q < \infty.$$

A sketch of the proof is given in Appendix C. In (ii), $\Pr\{\widehat{g}(W) \neq g^*(W)\}$ indicates a probability that $\widehat{g}$ differs from the true optimal treatment policy g^* over a new observation. Theorem 4.1 shows that the utility regret depends on both the nuisance estimation accuracy and the level of fairness which we would like to attain.

Specifically, each bound listed in Theorem 4.1 consists of three terms. The first term is an unavoidable modeling error minimized through least square estimation, which will vanish if $\mu_a(\cdot)$ lies in the function space spanned by the basis functions $\boldsymbol{b}(\cdot)$. From 4, one may further bound this modeling error to obtain a more interpretable form by noticing that

$$\left| \mu_a(W) - {\beta_a^*}^\top \boldsymbol{b}(W) \right| \leq \sqrt{\min_{\beta} \mathbb{E}\left[\left\{ Y^a - \beta^\top \boldsymbol{b}(W) \right\}^2 \mid W \right]}.$$

The second term, $R_{1,n}$, is a doubly robust second-order term that will be small if either π or μ_a are estimated accurately. In nonparametric modeling, the condition $\|\widehat{\pi} - \pi\|_{2,\mathbb{P}} \|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} = O_{\mathbb{P}}(n^{-\frac{1}{2}})$

substantially lowers the bar for the nuisance estimator convergence rate, which allows much more flexible methods to be employed while still achieving $\sqrt{n}$ rates; for example, it suffices that these nuisance functions are estimated consistently at $n^{\frac{1}{4}}$ rates.

The third term, R_2, has particularly important implications. It measures the imbalances in covariate distributions with respect to the sensitive feature, which is closely related to the level of unfairness in the optimal treatment policy; the larger the imbalances, the more likely the estimated optimal policies are unfair. If we use small values of the tolerance level δ_j so that the optimum β^* is constrained by the j-th fairness constraint, then the corresponding Lagrange multiplier, λ_j, is positive. On the contrary, if we loosen the standard by using large values of δ_j so that the j-th fairness constraint does not constrain β^*, λ_j is set to zero. Therefore, our attempts toward making optimal treatment policies more fair may lead to an additional welfare loss (regret) relative to the universally maximum feasible utility $\mathcal{U}(g^*)$. In other words, there is a tradeoff between fairness in the optimal treatment regime and the maximum utility.

In short, Theorem 4.1 implies that although the proposed approach has considerably reduced the burden on nuisance estimation, regardless how accurately we estimate the nuisance components there is a price that comes with imposing fairness constraints for the optimal treatment regime to achieve the desired fairness level.

5 Discussion

We propose a new framework for fair and robust estimation of optimal treatment regimes. Our method is easily implementable and allows practitioners to flexibly incorporate various fairness constraints to meet the desired level of fairness. This affords new opportunities to leverage the recent development in algorithmic fairness for optimal treatment regimes.

There are two important messages in our regret bound analysis. First, the proposed estimator is robust against model misspecification and allow to use more flexible nonparametric methods while still achieving $\sqrt{n}$ convergence rates to the maximum utility. Second, there is a tradeoff between fairness and the maximum utility, which is independent of accuracy of the nuisance estimation.

References

[1] Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021.

[2] Jean-Yves Audibert and Alexandre B Tsybakov. Fast learning rates for plug-in classifiers. *The Annals of statistics*, 35(2):608–633, 2007.

[3] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019. `http://www.fairmlbook.org`.

[4] Jason Brinkley, Anastasios Tsiatis, and Kevin J Anstrom. A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics*, 66(2):512–522, 2010.

[5] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–65, May 2017.

[6] Amanda Coston, Edward Kennedy, and Alexandra Chouldechova. Counterfactual predictions under runtime confounding. In *Advances in neural information processing systems*, volume 33, pages 4150–4162, 2020.

[7] Amanda Coston, Alan Mishler, Edward H Kennedy, and Alexandra Chouldechova. Counterfactual risk assessments, evaluation, and fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 582–593, 2020.

[8] Ezekiel J Emanuel, Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, and James P Phillips. Fair allocation of scarce medical resources in the time of covid-19, 2020.

[9] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29, 2016.

[10] Robin Henderson, Phil Ansell, and Deyadeen Alshibani. Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–1201, 2010.

[11] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[12] Anatoli Juditsky and Arkadii Nemirovski. Functional aggregation for nonparametric regression. *The Annals of Statistics*, 28(3):681–712, 2000.

[13] Nathan Kallus. Balanced policy evaluation and learning. *Advances in neural information processing systems*, 31, 2018.

[14] Edward H Kennedy. Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research*, pages 141–167. Springer, 2016.

[15] Edward H Kennedy. Semiparametric theory. *arXiv preprint arXiv:1709.06418*, 2017.

[16] Edward H Kennedy. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020.

[17] Edward H Kennedy, Steve Harris, and Luke J Keele. Survivor-complier effects in the presence of selection on treatment, with application to a study of prompt icu admission. *Journal of the American Statistical Association*, 114(525):93–104, 2019.

[18] Edward H Kennedy, Sivaraman Balakrishnan, and Max G'Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.

[19] Edward H Kennedy, Sivaraman Balakrishnan, and Larry Wasserman. Semiparametric counterfactual density estimation. *arXiv preprint arXiv:2102.12034*, 2021.

[20] Kwangho Kim, Edward Kennedy, and José Ramon Zubizarreta. Doubly robust counterfactual classification. *Advances in Neural Information Processing Systems*, 36, 2022.

[21] Kwangho Kim, Alan Mishler, and José Ramon Zubizarreta. Counterfactual mean-variance optimization. *arXiv preprint arXiv:2209.09538*, 2022.

[22] Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

[23] Matt J Kusner, Joshua R Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. *arXiv preprint arXiv:1703.06856*, 2017.

[24] Clément Levrard. Quantization/clustering: when and why does k-means work? *Journal de la société française de statistique*, 159(1):1–26, 2018.

[25] Alexander R Luedtke and Mark J van der Laan. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016.

[26] Alexander R Luedtke and Mark J van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713, 2016.

[27] Aditya Krishna Menon and Robert C Williamson. The cost of fairness in binary classification. In *Conference on Fairness, Accountability and Transparency*, pages 107–118. PMLR, 2018.

[28] Alan Mishler and Edward Kennedy. Fade: Fair double ensemble learning for observable and counterfactual outcomes. *arXiv preprint arXiv:2109.00173*, 2021.

[29] Alan Mishler, Edward H Kennedy, and Alexandra Chouldechova. Fairness in risk assessment instruments: Post-processing to achieve counterfactual equalized odds. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 386–400, 2021.

[30] Susan A Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

[31] Razieh Nabi and Ilya Shpitser. Fair inference on outcomes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[32] Razieh Nabi, Daniel Malinsky, and Ilya Shpitser. Learning optimal fair policies. In *International Conference on Machine Learning*, pages 4674–4682. PMLR, 2019.

[33] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4): 1574–1609, 2009.

[34] Romain Neugebauer and Mark van der Laan. Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434, 2007.

[35] Whitney K Newey and James R Robins. Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv preprint arXiv:1801.09138*, 2018.

[36] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

[37] Liliana Orellana, Andrea Rotnitzky, and James M Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *The international journal of biostatistics*, 6(2), 2010.

[38] James Robins, Lingling Li, Eric Tchetgen, and Aad van der Vaart. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, pages 335–421. Institute of Mathematical Statistics, 2008.

[39] Vira Semenova and Victor Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2021.

[40] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

[41] Georg Still. Lectures on parametric optimization: An introduction. *Optimization Online*, 2018.

[42] Alexandre B Tsybakov. Optimal rates of aggregation. In *Learning theory and kernel machines*, pages 303–313. Springer, 2003.

[43] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[44] Davide Viviano and Jelena Bradic. Fair policy targeting. *arXiv preprint arXiv:2005.12395*, 2020.

[45] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P Gummadi. Fairness constraints: A flexible approach for fair classification. *The Journal of Machine Learning Research*, 20(1):2737–2778, 2019.

[46] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

[47] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3): 681–694, 2013.

[48] Wenjing Zheng and Mark J Van Der Laan. Asymptotic theory for cross-valiyeard targeted maximum likelihood estimation. *Working Paper 273*, 2010.

APPENDIX

A Formal Definitions of the Regularity Conditions

First, for a feasible point $\bar{\beta} \in \mathcal{C}_{\text{fair}}$ we define the active index set.

Definition A.1 (Active set)**.** *For $\bar{\beta} \in \mathcal{C}_{fair}$, we define the active index set J_0 by*

$$J_0(\bar{\beta}) = \{1 \leq j \leq m \mid g_j(\bar{\beta}) = 0\}.$$

In what follows, we define LICQ and SC with respect to (P_{μ_a}).

Definition A.2 (LICQ)**.** *Linear independence constraint qualification (LICQ) is satisfied at $\bar{\beta} \in \mathcal{S}$ if the vectors $\nabla_\beta g_j(\bar{\beta})$, $j \in J_0(\bar{\beta})$ are linearly independent.*

Definition A.3 (SC)**.** *Let $L(\beta, \gamma)$ be the Lagrangian. Strict Complementarity (SC) is satisfied at $\bar{\beta} \in \mathcal{S}$ if, with multipliers $\bar{\gamma}_j \geq 0$, $j \in J_0(\bar{\beta})$, the Karush-Kuhn-Tucker (KKT) condition*

$$\nabla_\beta L(\bar{\beta}, \bar{\gamma}) \coloneqq \nabla_\beta \mathcal{L}(\bar{\beta}) + \sum_{j \in J_0(\bar{\beta})} \bar{\gamma}_j \nabla_\beta g_j(\bar{\beta}) = 0,$$

is satisfied such that $\bar{\gamma}_j > 0, \forall j \in J_0(\bar{\beta})$.

LICQ is arguably one of the most widely-used constraint qualifications that admit the first-order necessary conditions. SC means that if the j-th inequality constraint is active then the corresponding dual variable is strictly positive, so exactly one of them is zero for each $1 \leq j \leq m$. SC is widely used in the optimization literature, particularly in the context of parametric optimization [e.g., 40, 41].

B Proof of Lemma 4.1

Proof. The proof mimics the proofs of Lemma 5.1 and Lemma 5.2 in [2]. To show the first inequality, note that

$$\begin{aligned}
\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) &= \mathbb{P}\left[\Delta\left(\mathbb{1}\{\Delta > 0\} - \mathbb{1}\left\{\widehat{\Delta} > 0\right\}\right)\right] \\
&\leq \mathbb{P}\left[|\Delta|\left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right)\right] \\
&\leq \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}} \mathbb{P}\left\{|\Delta| \leq \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}}\right\} \\
&\lesssim \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}}^{\alpha+1},
\end{aligned}$$

where the first inequality follows by Lemma 1 of [18] and the last by the margin condition.

Next, for any $t > 0$ we have

$$\begin{aligned}
\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) &\leq \mathbb{P}\left[|\Delta|\left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right)\mathbb{1}\{|\Delta| \leq t\}\right] + \mathbb{P}\left[|\Delta|\left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right)\mathbb{1}\{|\Delta| > t\}\right] \\
&\leq \mathbb{P}\left[\left|\widehat{\Delta} - \Delta\right|\mathbb{1}\{|\Delta| \leq t\}\right] + \mathbb{P}\left[\left|\widehat{\Delta} - \Delta\right|\mathbb{1}\left\{\left|\widehat{\Delta} - \Delta\right| > t\right\}\right] \\
&\leq \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} \Pr\{|\Delta| \leq t\}^{\frac{q-1}{q}} + \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} \left(\frac{\mathbb{P}|\widehat{\Delta} - \Delta|^q}{t^q}\right)^{\frac{q-1}{q}} \\
&\lesssim \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} t^{\frac{q-1}{q}} + \frac{\left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^q}{t^{q-1}},
\end{aligned}$$

where the third inequality follows by the Hölder and Markov inequalities and the last by the margin condition. Now, the RHS in the last display is minimized when $t = O\left(\left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^{\frac{q}{q+\alpha}}\right)$, yielding

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}}.$$

□

C Proof of Theorem 4.1 (Sketchy)

Proof. It suffices to show the results in the part (i). By the first inequality in Lemma 4.1, we have

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \max_a \left\| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right\|_{\infty,\mathbb{P}}^{\alpha+1}$$

Recall that β_a^* denotes an optimal solution to the following unconstrained optimization problem

$$\beta_a^* = \underset{\beta \in \mathcal{B}}{\text{minimize}} \quad \mathbb{E}\left\{ \left(Y^a - \beta^\top \boldsymbol{b}(W) \right)^2 \right\},$$

and let $\widetilde{\beta}_a$ be an optimal solution to (P_{μ_a}). Then $\forall a$, by the triangle and Cauchy–Schwarz inequalities,

$$\left| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right| \leq \left| \mu_a(W) - \beta_a^{*\top} \boldsymbol{b}(W) \right| + \|\boldsymbol{b}(W)\|_2 \left\{ \|\beta_a^* - \widetilde{\beta}_a\|_2 + \|\widetilde{\beta}_a - \widehat{\beta}_a\|_2 \right\}.$$

Next, one may show that

$$\|\beta_a^* - \widetilde{\beta}_a\|_2 \lesssim \sum_j \lambda_j \left\| g_j(Y^a, W) \boldsymbol{b}(W) \right\|_{2,\mathbb{P}}$$

by noticing that (4) and (P_{μ_a}) can be viewed as the same form of a parametrized program (after writing (4) as a Lagrange dual form) and then applying the stability results [41, Chapter 6]. Further, by Theorem 3.1, it follows that

$$\|\widetilde{\beta}_a - \widehat{\beta}_a\|_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}} \vee \|\widehat{\pi}(W) - \pi(W)\|_{2,\mathbb{P}} \max_a \|\widehat{\mu}_a(W) - \mu_a(W)\|_{2,\mathbb{P}} \right).$$

Since $0 < \alpha < \infty$, we obtain the desired result by putting the pieces together due to Minkowski's inequality.

Then the part (ii) immediately follows by the fact that

$$\begin{aligned}
\Pr\left\{ \widehat{g}(W) \neq g^*(W) \right\} &= \mathbb{P}\left\{ |\widehat{g}(W) - g^*(W)| \right\} \\
&\leq \mathbb{P}\left[\mathbb{1}\left\{ |\mu_1(W) - \mu_0(W)| \leq \sum_a \left| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right| \right\} \right] \\
&\lesssim \max_a \left\| \mu_a(W) - \widehat{\beta}_a^\top \boldsymbol{b}(W) \right\|_{\infty,\mathbb{P}}^{\alpha}.
\end{aligned}$$

$\square$