On Robustness of Individualized Decision Rules

Zhengling Qi, Jong-Shi Pang, and Yufeng Liu

Department of Decision Sciences, George Washington University, Washington, District of Columbia; Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California; Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

ABSTRACT
With the emergence of precision medicine, estimating optimal individualized decision rules (IDRs) has attracted tremendous attention in many scientific areas. Most existing literature has focused on finding optimal IDRs that can maximize the expected outcome for each individual. Motivated by complex individualized decision making procedures and the popular conditional value at risk (CVaR) measure, we propose a new robust criterion to estimate optimal IDRs in order to control the average lower tail of the individuals’ outcomes. In addition to improving the individualized expected outcome, our proposed criterion takes risks into consideration, and thus the resulting IDRs can prevent adverse events. The optimal IDR under our criterion can be interpreted as the decision rule that maximizes the “worst-case” scenario of the individualized outcome when the underlying distribution is perturbed within a constrained set. An efficient non-convex optimization algorithm is proposed with convergence guarantees. We investigate theoretical properties for our estimated optimal IDRs under the proposed criterion such as consistency and finite sample error bounds. Simulation studies and a real data application are used to further demonstrate the robust performance of our methods. Several extensions of the proposed method are also discussed. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2020
Accepted January 2022

KEYWORDS
Conditional value at risk; Individualized decision rules; Nonconvex optimization; Robustness; Tail controls

1. Introduction

Decision making is a long standing research problem in many scientific areas, ranging from engineering, management science, to statistics. In the era of big data, the traditional “one fits all” decision rules are no longer ideal in many applications due to data heterogeneity. A decision rule that works for certain individuals may not necessarily work for others. Motivated by this, it is desirable to make individualized decision rules (IDRs) that map from individual characteristics into available decision assignments. Developing effective IDRs has a wide range of applications. For example, a credit card company hopes to send a special offer for each targeted customer tailoring to their personal needs. An epidemiologist needs to decide whether to deliver a vaccine plan to a specific region in order to prevent the spread of diseases. In medical applications, IDRs can be developed for better prevention and treatment methods that are tailored to each individual patient. Developing optimal IDRs is one of the main goals of precision medicine, also known as personalized medicine.

Most existing methods in the literature, from the data analytic perspective, are focused on estimating the optimal IDR that can maximize the expected outcome or minimize the expected loss for each individual (Qian and Murphy 2011; Manski 2004). For example, we may want to learn an IDR \( d \) that maps an individual’s covariate \( X \in \mathcal{X} \) into a binary decision space \( A = \{1, -1\} \), that is, \( d : \mathcal{X} \to \mathcal{A} \), in order to maximize the expectation of \( R(d) \) or a utility function \( u \) of \( R(d) \). Here \( R(d) \) is the random outcome under IDR \( d \), which will be formally defined in Section 2. Then the problem can be mathematically formulated as the following optimization problem:

\[
\max_{d \in \mathcal{D}_0} E[R(d)],
\]

where \( \mathcal{D}_0 \) is a class of all IDRs. If some utility function \( u \) is used, then one could replace the objective function in (1) with \( E[u(R(d))] \). The equivalent form of Problem (1) is

\[
\max_{t \in \mathbb{R}, d \in \mathcal{D}_0} t \quad \text{subject to} \quad t \leq E[R(d)],
\]

by using the hypographical representation (Rockafellar and Wets (2009)).

According to (2), it may be reasonable to restrict the expected outcome larger than some threshold when stochastic ups and downs of outcome \( R(d) \) can be safely averaged out. Then by implementing the IDR obtained by (2), we can guarantee on average the outcome will be at least as good as some threshold \( t \). However, in practice, practitioners usually want to have a safe margin to protect against undesired outcomes, especially when the lower tails of the outcome are more important. For instance, in reliability engineering (Rockafellar and Royset 2010), people...
are often interested in controlling the failure probability of some designed systems or structures such as buildings and bridges, instead of expected failure time. Here the selection of designs can be interpreted as IDRs, which are based on environmental conditions, existing materials, etc. In precision medicine, sometimes the gain in the expected outcome may be very little by comparing two treatments while the tail of the potential outcome distribution is of direct interest (Wang et al. 2018). Suppose two drugs are used to improve CD4 T cell amount of AIDS patients. Since the normal range of CD4 cells is 500–1500, then in practice the event of a certain subject \( \{ R_i(d) = 600 \} \) may be treated as good as the event of \( \{ R_i(d) = 800 \} \). In contrast, the event \( \{ R_i(d) = 50 \} \) may be considered much worse than \( \{ R_i(d) = 400 \} \). Hence, only using (2) to search for the best IDR may not be sufficient when the tails are more important or the variability of the outcomes needs to be controlled. In this case, one may consider using the truncated mean with a fixed cutoff as a criterion for better decision making, that is, \( \mathbb{E} [ \min (R(d), L) ] \), where \( L \) is a fixed cutoff. However, in general, the cutoff \( L \) is unknown and can be difficult to determine. Even though \( L \) can be pre-determined in some ideal case, when the scale of \( R(d) \) is changed or there is some measurement error or batch effect in the observed \( R_i(d) \), the final \( L \) needs to be properly modified, which could still be hard to choose. In addition, for many applications especially medical studies, normal ranges are usually determined by quantiles of healthy people. Therefore, it may be more reasonable to consider the expected outcome lower than a certain quantile when evaluating IDRs.

In this article, motivated by the conditional value at risk (CVaR) used extensively in finance and risk management (Rockafellar and Uryasev 2000), we propose a new criterion that considers average lower tails of outcome to evaluate IDRs. The resulting IDR under our proposed criterion can optimize the outcome of each individual and provide a safe margin against adverse events jointly. Our work is closely related to the recent article by Qi et al. (2019a), where they established a mathematical framework to study IDRs under general risk measures. In this work, we are more specific on the lower tail of outcomes for making decisions and develop a thorough statistical analysis. We propose an efficient optimization algorithm for our specific problem with better convergence guarantee than that of Qi et al. (2019a). In addition, several practical extensions are also discussed.

### 1.1. Related Literature

There is an increasing body of literature in learning optimal IDR under the framework of (1). The literature spans across various fields such as statistics, economics, and machine learning. Along the line of statistics literature, most existing methods can be roughly divided into two categories: model based methods and direct search methods. Q-learning (Watkins 1989; Murphy 2005; Schulte et al. 2014) and A-learning (Murphy 2003; Robins 2004; Shi et al. 2018) are two representative model based methods. Other variants include Fan et al. (2017), Gunter, Zhu, and Murphy (2011), etc. For direct search methods, by viewing IDR problems as a weighted classification problem, Zhao et al. (2012) proposed to use the weighted support vector machine method to estimate the optimal IDR. Following that, various types of machine learning methods were proposed, such as Liu et al. (2018), Zhou et al. (2017), Zhao et al. (2015), Cui, Zhu, and Kosorok (2017), Tao and Wang (2017), Laber and Zhao (2015), Zhang et al. (2015), Chen et al. (2018), and Zhang et al. (2020).

In addition, Tian et al. (2014), Qi and Liu (2018), and Qi et al. (2019b) proposed to use regression methods to directly estimate the optimal IDR. Recently, Wang et al. (2018) proposed to use the quantile of \( R(d) \) as a criterion to search for best IDRs, which is closely related to this article. Their method can help to obtain robust optimal IDRs by controlling the lower quantile. However, it can be unstable when the potential outcome distribution is discrete. More importantly, as we mentioned in the example of CD4 T cells, lower tails of the outcome should be treated differently: if both \( \{ R_i(d) = 50 \} \) and \( \{ R_i(d) = 400 \} \) are below the 25% quantile, the first event should be considered much worse than the latter one. We will discuss this further in Section 2. Other quantile-based methods include Linn, Laber, and Stefanski (2017), Xiao, Zhang, and Lu (2019), and Fang, Wang, and Wang (2021).

In the econometrics literature, Manski (2004) provided a comprehensive regret analysis on the estimation of optimal IDRs with some connection to statistical decision theory (Savage 1951). Later on, exact finite sample regret analysis was established by Stoye (2009) and Tetenov (2012). Under smooth parametric and semiparametric settings, Hirano and Porter (2009) investigated asymptotic optimality and large sample properties of optimal IDRs. Other related work includes Chamberlain 2011, Bhattacharya and Dupas 2012, and Kasy 2016. Recently, Kitagawa and Tetenov (2018) and Athey and Wager (2021) established rate-optimal regret bounds for learning optimal IDRs. All of these existing developments are based on expected outcome or expected utility. Dehejia (2008) studied the risk aversion of treatment effect evaluation by using the mean-variance trade off criterion.

We would like to point out the increasing literature of learning optimal IDRs in the machine learning community, which is often referred as batch learning from bandit feedback, such as Beygelzimer and Langford 2009, Dudík, Langford, and Li 2011, Swaminathan and Joachims 2015, Dudík, Langford, and Li 2011, and Kallus 2018. Finally, in the reinforcement learning literature, CVaR has been used as constraints in Markov decision processes such as Tamar, Glassner, and Mannor (2015) and Chow et al. (2017).

### 1.2. Main Contributions and Outline

The main contributions of this article can be summarized as follows. We leverage the CVaR criterion and propose a robust criterion to directly estimate the optimal IDR that can improve the expected outcome while simultaneously controlling the average lower tails of the outcome. We discuss several important properties of our criterion and its practical usage by providing safety protection for implementing optimal IDRs. An efficient non-convex optimization algorithm is proposed to compute the solutions with a convergence guarantee. We establish several important theoretical properties of our estimator under the proposed criterion similar to the regret analysis in Zhou et al.
mating the propensity score using various methods such as
studies, the proposed method can be applied as well by esti-
function mapping from the covariates space.

We also let

Assumption 2.2. \( R(a) \perp A | X \) for any \( a \in A \), where \( \perp \) represents independence.

Assumption 2.3. \( \pi(a|X) \geq c \) almost surely for any \( a \in A \) and some positive constant \( c \).

For simplicity, we assume the random outcome \( R \) has a bounded support. This assumption can be relaxed by the high-
order moment condition. See Athey and Wager (2021) for more
details. Without loss of generality, we assume that the larger
indicates the better condition an individual is in. Throughout
this article, we make the following three standard assumptions.

Assumption 2.1. \( R = R(A) \).

Assumption 2.2. \( R(a) \perp A | X \) for any \( a \in A \), where \( \perp \) represents independence.

Assumption 2.3. \( \pi(a|X) \geq c \) almost surely for any \( a \in A \) and some positive constant \( c \).

Despite much progress toward developing optimal IDRs in
the intersection of statistics, econometrics, and machine learn-
ing, only focusing on obtaining the largest expected outcome for
each individual can be too restrictive and sometimes may not
be even safe. For example, doctors may want to know whether a
treatment does the best to improve the worst scenario, in partic-
ular for a patient with high risk. Without such a risk considera-
tion, it may lead to serious consequences, such as exacerbation
or hospitalization in practice. Similar concerns may happen in
the credit card company, where the “best” policy should not
only improve the expected profit, but also reduce the chance of
incurring a heavy loss. This motivates us to control risk exposure
associated with the corresponding decision rules, in addition to
maximizing the expected outcome for each individual.

2.2. Expected Value Function Framework

Before introducing our new criterion and methods, we first
present the value function framework used by the most existing
methods, such as (Manski 2004) and (Qian and Murphy 2011).

The value function of an IDR \( d \) is defined as

\[
V(d) = E[R(d)] = E_X [E[R(d)|X]] = E_X [E[R(d)|X, A = d(X)]] = E_X [E[R(X, A = d(X)) | X]].
\]

so that

\[
\pi(A \perp X | d) = \pi(A \perp X | \tilde{d})
\]

where the first line is based on Assumption 2.2, the first equality in
the second line is based on Assumption 2.1 and the last
equality relies on Assumption 2.3. Based on this value function,
an optimal IDR \( d^* \) under the mean criterion is defined as

\[
d^* \in \arg\max_{d \in D_B} V(d),
\]

which is equivalent to

\[
d^*(X) \in \arg\max_{a \in A} E[R | X, A = a],
\]

almost surely. It is observed that under the value function frame-
work, the optimal IDR is to select the treatment with the largest
expected outcome among all treatments for each individual.

Despite much progress toward developing optimal IDRs in
the intersection of statistics, econometrics, and machine learn-
ing, only focusing on obtaining the largest expected outcome for
each individual can be too restrictive and sometimes may not
be even safe. For example, doctors may want to know whether a
treatment does the best to improve the worst scenario, in partic-
ular for a patient with high risk. Without such a risk considera-
tion, it may lead to serious consequences, such as exacerbation
or hospitalization in practice. Similar concerns may happen in
the credit card company, where the “best” policy should not
only improve the expected profit, but also reduce the chance of
incurring a heavy loss. This motivates us to control risk exposure
associated with the corresponding decision rules, in addition to
maximizing the expected outcome for each individual.

2.3. Conditional Value at Risk

It is natural to consider some robust metrics such as quantiles of
\( R \) given \( X \) and \( A \) to measure the effect of a treatment (Wang et al. 2018). The corresponding optimal IDR \( \tilde{d} \) under the quantile
may be defined as

\[
\tilde{d} \in \arg\max_{d \in D_B} Q_y(R(d)),
\]

where \( Q_y(R(d)) = \inf\{y : P(R(d) \leq y) \geq 1 - \gamma\} \) and \( y \in (0, 1) \). Analogous to Problems (1) and (2), an equivalent form of
Problem (4) can be constructed as follows:

\[
\max_{t \in \mathbb{R}, d \in D_B} t \quad \text{subject to} \quad t \leq Q_y(R(d)).
\]
as we discussed in the aforementioned example, the CD4 T cell \( |R_i(d) = 400| \) below the normal level is considered to be much better than \( |R_i(d) = 50| \), therefore they should be treated differently in practice. Second, \( Q_{1/2} (R(d)) \) is generally not continuous in \( \gamma \), which may cause instability. For instance, if the outcome distribution is discrete and there is a small change in \( \gamma \), the resulting optimal \( \tilde{d} \) may change significantly. Lastly, from the computational perspective, the quantile makes the optimization Problem (4) hard to solve and thus may limit its use in practice.

In order to address the drawbacks of using quantiles, we propose to use the conditional value at risk (CVaR), also known as expected tail loss, which was proposed by Artzner et al. (1999) in risk management. Consider a continuous random variable \( Y \). Then the \( \gamma \)-CVaR of \( Y \) may be defined as

\[
S(F_Y) := \frac{1}{\gamma} E[Y \mathbb{1}(Y \leq Q_{\gamma}(F_Y))],
\]

where \( F_Y \) is the corresponding probability distribution of \( Y \). Based on this definition, the \( \gamma \)-CVaR can be interpreted as a truncated mean lower than \( \gamma \)-quantile of \( Y \). For the general setting, instead of assuming a continuous distribution of \( Y \), \( \gamma \)-CVaR is defined as an optimal value of a concave maximization problem by the celebrated work of Ben-Tal and Teboulle (1997), Rockafellar and Uryasev (2000), which is defined as follows:

\[
S(F_Y) := \sup_{\alpha \in \mathbb{R}} \left\{ \alpha - \frac{1}{\gamma} E[(\alpha - Y)_+] \right\},
\]

where \([t]_+ = \max(0, t)\). The leftmost of the optimal solution set to (7) is \( Q_{\gamma}(F_Y) \) (Rockafellar and Uryasev 2000, Theorem 1). Then one can see that the definition in (7) is equivalent to (6) when the outcome distribution of \( Y \) is continuous. We also remark that \( Y \) is often referred to a loss in the finance literature. Here we call \( Y \) an outcome/reward in order to be consistent with our problem setting.

Finally, we would like to point out that \( \gamma \)-CVaR has several nice properties discussed by Artzner et al. (1999) and it is in general preferable to the quantile measure (Sarykalin, Serraino, and Uryasev 2008). In particular, based on the interpretation of (6), \( \gamma \)-CVaR considers average outcomes lower than \( \gamma \)-quantile, which treats lower tails of outcome differently. This exactly satisfies our purpose. The concave maximization formulation in (7) demonstrates the continuity of \( \gamma \)-CVaR with respect to \( \gamma \), which provides more stability measure compared with the quantile one. Furthermore, \( \gamma \)-CVaR is considered to be more computationally efficient than the quantile criterion because of the concave maximization formulation. Finally, Pflug (2000) and Rockafellar and Uryasev (2002) showed that \( S(F_Y) \leq Q_{\gamma}(F_Y) \), suggesting that \( \gamma \)-CVaR is more conservative than \( \gamma \)-quantile. This also implies that a larger \( \gamma \)-CVaR of a random outcome indicates a larger \( \gamma \)-quantile. Clearly the reverse inequality does not necessarily hold. All these nice properties motivate us to use CVaR in the IDR problems. For the related theoretical discussion about CVaR, we refer to Rockafellar and Uryasev (2002) and the references therein.

### 2.4. A New Robust Criterion for IDR Problems

We borrow the concept of CVaR in order to conduct risk control to obtain a robust optimal IDR that can improve the lower tails of outcomes. Specifically, we define the following decision-rule based \( \gamma \)-CVaR criterion as

\[
M_{0,\gamma}(d) := \sup_{\alpha \in \mathbb{R}} \left\{ \alpha - \frac{1}{\gamma} E[(\alpha - R(d))_+] \right\}.
\]

Note that the difference between the criterion (8) and the original CVaR is to let the outcome \( R \) depend on IDR \( d \). The proposed criterion is not restricted to the continuous outcomes but can also be used in the discrete outcome cases, thus providing a broad application. In addition, if a small value of the outcome is more desirable (i.e., measuring the loss instead of the reward), one can replace \( R(d) \) by \(-R(d)\). All properties preserve. Given the coherent property of CVaR shown in Artzner et al. (1999), we have the following proposition for \( M_{0,\gamma}(d) \).

**Proposition 2.1.** The following properties of \( M_{0,\gamma}(d) \) hold:

(a) If \( R \) is shifted by a constant \( c \), then \( M_{0,\gamma}(d) \) is also shifted by the same constant \( c \);

(b) If \( R \) is multiplied by a positive constant \( c \), then the corresponding \( M_{0,\gamma}(c) \) is also multiplied by the same constant \( c \);

(c) Given two IDRs \( d_1 \) and \( d_2 \), if \( R(d_1) \leq R(d_2) \) almost surely, then

\[
M_{0,\gamma}(d_1) \leq M_{0,\gamma}(d_2);
\]

(d) Given IDR \( d \), \( M_{0,\gamma}(d) \leq \min\{V(d), Q_{\gamma}(R(d))\} \).

In addition, if outcome \( R = c \) almost surely, then \( M_{0,\gamma}(d) = c \).

**Remark 2.1.** Proposition (2.1) justifies the use of \( M_{0,\gamma}(d) \). In particular, (a) and (b) demonstrate that \( M_{0,\gamma}(d) \) is not affected by a constant shift or multiplication. (c) implies that if one IDR is no worse than the other, \( M_{0,\gamma}(d) \) preserves the preferences. The last property indicates that \( M_{0,\gamma}(d) \) is more conservative than the expected outcome and the quantile criterion when evaluating an IDR \( d \).

If the distribution of \( R(d) \) is continuous, then we could rewrite (8) as

\[
M_{0,\gamma}(d) = \frac{E[d|R(d) \leq Q_{\gamma}(R(d))]}{\gamma}.
\]

Note that

\[
P(R(d) < \alpha) = \mathbb{E}[\sum_{a \in A} I(d(X) = a)P(R < \alpha|X, A = a)] = \mathbb{E}[P(R < \alpha|X, A = d(X))].
\]

Then \( Q_{\gamma}(R(d)) \) can be further expressed as

\[
Q_{\gamma}(R(d)) = \inf\{\alpha \mid \mathbb{E}[P(R < \alpha|X, A = d(X))] \geq \gamma\},
\]

which can be interpreted as the average \( \gamma \)-quantile of \( R \) under the decision rule \( d \). Correspondingly \( M_{0,\gamma}(d) \) can be understood as the \( \gamma \)-average CVaR.
According to Proposition 2.1 (d), $M_{0,Y}(d)$ can be regarded as a lower bound of $V(d)$ and $Q_y(R(d))$. Based on this, maximizing $M_{0,Y}(d)$ can thus potentially improve both $V(d)$ and $Q_y(R(d))$. Then the optimal IDR under our proposed robust criterion $M_{0,Y}(d)$ is defined as

$$d_{0,Y} = \arg\max_{d \in D_Y} M_{0,Y}(d).$$

(12)

The optimal IDR with respect to $M_{0,Y}(d)$ is to select a treatment/decision with the largest $\gamma$-average CVaR. Moreover, if we again use hypographical representation similar as those in (2) and (5), then we could formulate (12) as the following constraint optimization problem.

$$\max_{d \in D_Y, t \in R} t$$

subject to $\sup_{a \in R} \{ - \frac{1}{\gamma} E[(\alpha - R(d))_+] \} \geq t$. (13)

Based on the formulation (13), $d_{0,Y}$ can be interpreted as the best IDR with the average lower tail of the outcome being at least as good as some certain threshold. Besides, based on statement (d) in Proposition (2.1), the constraint set in (13) implies $\min(V(d), Q_y(R(d))) \geq t$. Therefore, by using $M_{0,Y}(d)$, the resulting optimal $d_{0,Y}$ can guarantee both $(1 - \gamma) \times 100\%$ of the population’s outcomes and the expected outcome no worse than some threshold $t$.

By a similar derivation as that in (3), we have the following proposition.

**Proposition 2.2.** $M_{0,Y}(d) = \inf_{a \in R} \{ a - \frac{1}{\gamma} E[(\alpha - R(d))_+] \}$. (14)

The definition in (8) and Proposition (2.2) give us an efficient way to compute the optimal IDR $d_{0,Y}$ and $a_0$ via jointly optimizing

$$(d_{0,Y}, a_0) \in \arg\max_{a \in R, d \in D_Y} \left\{ - \frac{1}{\gamma} E^d[(\alpha - R)_+] \right\}. (14)$$

**2.5. Dual Representation**

Note that $M_{0,Y}(d)$ involves concave maximization with respect to $\alpha$. Thus, it would be useful to investigate its dual representation by making use of convex duality theory in optimization (e.g., Rockafellar 1974). To begin with, we first define the following set:

$$\gamma_0^d := \{ V \in L^1(T, \mathcal{F}, P^d) | E^d[V] = 1, 0 \leq V(\omega) \leq \frac{1}{\gamma}, \text{for } \omega \in T \text{ almost surely} \},$$

where $P^d$ is the probability measure under the IDR $d$. We have the following theorem that gives the dual representation of $M_{0,Y}(d)$.

**Theorem 2.1.** $M_{0,Y}(d) = \inf_{\gamma \in \gamma_0^d} E^d[VR]$. (15)

Dual representation of $M_{0,Y}(d)$: According to the dual representation of $M_{0,Y}(d)$ and its proof in the appendix, online supplementary materials, we can define a conditional probability measure $p^V(B) = \int_B VdP^d$ for any measurable set $B \in \mathcal{F}_1$, where $V \in \gamma_0^d$. Then $V = \frac{dp^V}{dP^d}$. Define

$$\zeta_y(u) = \begin{cases} 0 & \text{if } 0 \leq u \leq \frac{1}{\gamma} \\ +\infty & \text{otherwise} \end{cases}.$$

Then we can further rewrite $E^d[VR]$ in $M_{0,Y}(d)$ as

$$E^d[VR] = E^d[R\frac{dp^V}{dP^d}] + E \left[ \zeta_y \left( \frac{dp^V}{dP^d} \right) \right] = E^d[R] + E \left[ \zeta_y \left( \frac{dp^V}{dP^d} \right) \right] = E^d[R] + I_{\zeta_y} \left( \frac{dp^V}{dP^d} \right),$$

where $I_{\zeta_y}(\cdot)$ can be interpreted as the $f$-divergence distance between $p^V$ and $P^d$. Then $M_{0,Y}(d) = \inf_{p^V \ll P^d} E^d[R] + I_{\zeta_y} \left( \frac{dp^V}{dP^d} \right)$, where that $u \ll v$ denotes the probability measure $u$ is absolutely continuous with respect to the probability measure $v$.

Thus, the optimal IDR can also be written as

$$d_{0,Y} \in \arg\max_{d \in D_Y} \left\{ \inf_{p^V \ll P^d} E^d[R] + I_{\zeta_y} \left( \frac{dp^V}{dP^d} \right) \right\},$$

which can be interpreted as choosing an optimal decision rule in terms of the worst expected outcome within the $f$-divergence distance from the original distribution $P^d$.

According to our problem setting, define the density under $p^d$ as $f_0(x)f_1(d(x) = a)f_i(r|x, a)$. Since $P^V \ll p^d$, then the density under $p^V$ should be $v_0(x) = f_0(x)f_1(d(x) = a)v(r|x, a)$ for some conditional probability density $v_0(x)$ and $v_1(r|x, a)$. Then we have

$$\frac{dp^V}{dP^d} = \frac{v_0(x)v_1(r|x, a)}{f_0(x)f_1(d(x) = a)},$$

by the chain rule. Therefore, we can further rewrite $M_{0,Y}(d)$ as

$$M_{0,Y}(d) = \inf_{p^V \ll P^d} \left\{ E^d[R] | p^V \ll p^d, \frac{dp^V}{dP^d} \right\},$$

$$0 \leq \frac{v_0(X)v_1(r|X, a = d(X))}{f_0(X)f_1(r|X, a = d(X))} \leq \frac{1}{\gamma}, \text{almost surely} \right\}. (15)$$

This gives us a natural link to distributionally robust statistical models that can evaluate a decision rule under ambiguity. Maximizing $M_{0,Y}(d)$ over $D_Y$ is equivalent to identifying an optimal IDR, that is, robust to the contamination of both outcome $R$ and covariates $X$ characterized by a probability constraint set. Recently, Mo, Qi, and Liu (2021) investigated distributionally robust ITRs by directly optimizing the expected rewards for the worse ITR within a perturbation set around the training distribution. Their method mainly focuses on potential covariate shifts. In contrast, by using a CVaR-based criterion, our proposed method is robust to distributional changes of both the outcome and covariates.

**3. Statistical Estimation and Optimization**

In this section, we discuss the estimation and optimization procedures for Problem (2.2) given observed data. The optimization
Figure 1. Plot of the smooth surrogate loss function with \( \delta = 1 \)

in (14) can be rewritten as

\[
\max_{\alpha \in \mathbb{R}, d \in \mathcal{D}} E \left[ \frac{(\alpha - (\alpha - R_i)_{+}) \mathbb{I}(A = d(X))}{\pi(A|X)} \right],
\]

where \( \mathcal{D} \) is some prespecified classes of decision rules such as the linear ones. Consider the binary treatment setting and let \( d(X) = \text{sign}(f(X)) \). Suppose we observe independently and identically distributed data \((X_i, A_i, R_i); i = 1, \ldots, n\). Then we can estimate the optimal IDR via the empirical approximation:

\[
\max_{\alpha \in \mathbb{R}, d \in \mathcal{D}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i = \text{sign}(f(X_i))) \left( \frac{\alpha - (\alpha - R_i)_{+}}{\pi(A_i|X_i)} \right).
\]

It is well known that optimization over indicator functions is NP hard. Alternatively, similar to Zhou et al. (2017), we can replace the 0–1 loss function by the following smooth truncated loss,

\[
S(u) = \begin{cases} 
0 & \text{if } u \leq -\delta \\
(1 + u/\delta)^2/2 & \text{if } 0 > u \geq -\delta \\
1 - (1 - u/\delta)^2/2 & \text{if } \delta > u \geq 0 \\
1 & \text{if } u > \delta,
\end{cases}
\]

and then use a functional margin representation to express \( \mathbb{I}(A_i = \text{sign}(f(X_i))) \) as \( \mathbb{I}(A_i f(X_i) > 0) \) for each \( i \). The corresponding function plot of \( S(u) \) is shown in Figure 1 with \( \delta = 1 \). From the plot, we can see that the smooth approximation \( S(u) \) is very close to the 0–1 loss. The parameter \( \delta \) can control the closeness of this approximation. In practice, we can simply choose \( \delta = 1 \). By using the surrogate function, we can estimate the optimal IDR under \( M_{0, \gamma}(d) \) via computing

\[
\min_{\alpha \in \mathcal{H}, \gamma} \frac{1}{n} \sum_{i=1}^{n} S(A_i f(X_i)) \left( -\alpha + \frac{\alpha - R_i}_{+} \right) + \frac{\lambda n}{2} ||f||^2_{\mathcal{H}},
\]

where \( ||f||^2_{\mathcal{H}} \) is some convex penalty function on \( f \). For example, if we consider \( \mathcal{H} \) as a Reproducing Kernel Hilbert Space (RKHS), then \( ||f||^2_{\mathcal{H}} \) could be the square of the RKHS norm of \( f \). The estimated IDR is given by \( \tilde{d}_{0, \gamma}(X) = \text{sign}(\hat{f}(X)) \). Note that Problem (17) involves a nonconvex and potentially non-smooth optimization problem. Recent developments in difference-of-convex (DC) optimization (Pang, Razaviyayn, and Alvarado 2016) motivate us to use DC programming to efficiently solve this problem. Note that \( S(u) \) can be expressed as a difference of convex differentiable functions: \( S_1(u) - S_2(u) \), where

\[
S_1(u) = \begin{cases} 
0 & \text{if } u \leq -\delta \\
(1 + u/\delta)^2/2 & \text{if } -\delta < u \leq 0 \\
1/2 + u/\delta & \text{if } 0 < u \\
1 & \text{if } u > \delta
\end{cases}
\]

\[
S_2(u) = \begin{cases} 
0 & \text{if } u \leq 0 \\
(u/\delta)^2/2 & \text{if } 0 < u \leq \delta \\
u/\delta - 1/2 & \text{if } u > \delta
\end{cases}
\]

Define

\[
G^{(1)}(f, \alpha) := \frac{1}{n} \sum_{i=1}^{n} S(A_i f(X_i)) \left( -\alpha + \frac{\alpha - R_i}_{+} \right) + \frac{\lambda n}{2} ||f||^2_{\mathcal{H}},
\]

and

\[
G^{(1)}_j(f) := \frac{1}{n} \sum_{i=1}^{n} S(A_i f(X_i)) \left( -R_j + \frac{R_j - R_i}_{+} \right) + \frac{\lambda n}{2} ||f||^2_{\mathcal{H}}.
\]

The following proposition gives us a way to express (17) as a DC function.

**Proposition 3.1.** The following two optimization problems have the same optimal value, i.e.,

\[
\min_{\alpha \in \mathcal{H}, f \in \mathcal{H}} G^{(1)}(f, \alpha) = \min_{f \in \mathcal{H}} \tilde{G}^{(1)}(f),
\]

where \( \tilde{G}^{(1)}(f) := \min_{1 \leq j \leq n} (G^{(1)}_j(f)) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}} \). More importantly, the optimal solution sets of \( f \) to both problems are the same.

Based on Proposition 3.1, instead of solving (17), we can equivalently solve the optimization problem on the right hand side of (20). Let \( c_{ij} := \frac{-(R_j - R_i)_{+}}{\pi(A_i|X_i)} \) for \( i = 1, \ldots, n \) and \( j = 1, \ldots, n \), and note that \( c_{ij} \) is not necessarily nonnegative. Recall that \( S(u) = S_1(u) - S_2(u) \). We can further rewrite \( G^{(1)}_j(f) \) as

\[
G^{(1)}_j(f) = \frac{1}{n} \sum_{i=1}^{n} \left( \max(c_{ij}, 0) S_1(A_i f(X_i)) \right) + \frac{\lambda n}{2} ||f||^2_{\mathcal{H}}
\]

where both \( F_j(f) \) and \( H_j(f) \) are convex functions with respect to \( f \) for \( j = 1, \ldots, n \). Then we can further decompose

\[
\tilde{G}^{(1)}(f) = \min_{1 \leq j \leq n} \left( F_j(f) - H_j(f) \right) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}}
\]

\[
= \sum_{j=1}^{n} F_j(f) - \max_{1 \leq j \leq n} (H_j(f) + \sum_{k \neq j} F_k(f) ) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}}
\]

\[
:= F(f) - \max_{1 \leq j \leq n} (h_j(f) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}})
\]
as a DC function, where $h_l(f) := H_l(f) + \sum_{k \neq l} F_k(f)$. Note that $\tilde{G}^{(1)}(f)$ is a potentially non-smooth function if there exits multiple $k$’s such that $h_k(f) = \max_{1 \leq k \leq n} h_l(f)$. As discussed by Pang, Razaviyayn, and Alvarado (2016), traditional DC programming may converge to nonsense points. To address this issue, motivated by Pang, Razaviyayn, and Alvarado (2016), we define $\mathcal{M}_\epsilon(f) := \{j \mid h_j(f) \geq \max_{1 \leq k \leq n} h_k(f) - \epsilon\}$, that is, “e-argmax” index set, and use the following enhanced probabilistic DC algorithm summarized in Table 1 below to solve Problem (17).

Algorithm 1 Algorithm for solving (17)

1. Given a fixed $\epsilon > 0$, let $f^{(\nu)}$ be the solution at the $\nu$ iteration for $\nu = 1, 2, \ldots$
2. Randomly select $j \in \mathcal{M}_\epsilon(f^{(\nu)})$, and compute
   \[ f^{(\nu+1)} = \arg\min_{f \in \mathcal{H}} \{F(f) - \frac{\partial h_j(f^{(\nu)})}{\partial f}(f - f^{(\nu)}) + \frac{\epsilon}{2} \|f\|_2^2 \}. \] (21)
3. The algorithm stops when $|\tilde{G}^{(1)}(f^{(\nu)}) - \tilde{G}^{(1)}(f^{(\nu+1)})| < \kappa$, for some pre-specified positive constant $\kappa$.

The proof of the convergence to sharp stationary points by the above algorithm can be found in Pang, Razaviyayn, and Alvarado (2016). For the computation of the subproblem (21), efficient algorithms such as the quasi-Newton method can be used. Compared with the algorithm proposed in Qi et al. (2019a), by using the surrogate loss, we avoid directly solving a discontinuous optimization problem. While Qi et al. (2019a) transformed the problem into a constraint optimization by using the epigraph representation, it can be less efficient than our proposed algorithm as the number of constraints is proportional to the number of observations. In addition, the convergence to a shaper stationary point can be guaranteed by our algorithm. For a general treatment on non-convex and non-smooth optimization problems, we refer to Cui and Pang (2021).

### 4. Analysis of Statistical Convergence

In this section, we discuss the statistical theory related to our estimation under $M_{0,Y}(d)$. For the ease of presentation, we assume that Assumptions 2.1–2.3 always hold. With some abuse of notation, we first define

\[
M_{0,Y}(d, \alpha) = E \left[ \frac{I(A = d(X))}{\pi(A|X)} \left( \alpha - \frac{1}{\gamma} (\alpha - R)_+ \right) \right], \tag{22}
\]

denote the optimal solution of maximizing (22) as $d_{0,Y} = \text{sign}(f_0)$ and $\alpha_0$. Since we use the surrogate loss function $S(u)$, we further define

\[ M_S(f, \alpha) = E \left[ \frac{S(Af(X))}{\pi(A|X)} (\alpha - \frac{1}{\gamma} (\alpha - R)_+) \right] \]

as the surrogate value function. Our theoretical results are similar to the standard regret analysis such as Zhou et al. (2017). The additional challenge is to handle the “nuisance” parameter $\alpha$, which requires additional efforts. For example, Fisher consistency cannot be directly obtained by using the standard proof such as that in Zhou et al. (2017).

#### 4.1. Fisher Consistency

We first establish Fisher consistency of estimating optimal ITRs under $M_S(f, \alpha)$ to justify the use of the surrogate loss $S(u)$, compared with $M_{0,Y}(d, \alpha)$. The proof is different from the classical Fisher consistency in classification, which only involves one functional class of interest. Here we need to consider the effect of the surrogate function on estimating $\alpha$, which is thus more involved.

**Theorem 4.1.** For any measurable function $f$ and $\alpha$ and a given $0 < \gamma \leq 1$, if $(f_0^*, \alpha_0^*)$ maximizes $M_S(f, \alpha)$, then $\text{sign}(f_0^*)$, $\alpha_0^*$ maximizes $M_{0,Y}(d, \alpha)$.

Based on Theorem 4.1, instead of $M_{0,Y}(d, \alpha)$, we can target on $M_S(d, \alpha)$ equivalently.

#### 4.2. Excess Value Bound

Based on Theorem 4.1, we can further justify the use of the surrogate function $S(u)$ by establishing the following excess value bound for the 0–1 loss in $M_{0,Y}(d, \alpha)$.

**Theorem 4.2.** For any measurable function $f$, $\alpha$ and any probability distribution over $(X, A, R)$,

\[
M_{0,Y}(d_{0,Y}, \alpha_0) - M_{0,Y}(\text{sign}(f), \alpha) \leq 2(M_S(f_0^*, \alpha_0^*) - M_S(f, \alpha)).
\]

Theorem 4.2 gives us a way of bounding the difference between the optimal IDR and the estimated IDR under $M_{0,Y}(d, \alpha)$ by using $M_S(d, \alpha)$ instead.
4.3. Convergence Rate

In order to obtain the finite sample performance of our estimated optimal IDR under \( M_{0,y}(d, \alpha) \), it is enough to focus on the difference of \( M_2(d, \alpha) \) between the estimated optimal IDR and the optimal ITR based on Theorem 4.2. Define

\[
(\hat{f}, \hat{\alpha}) = \arg\min_{f \in \mathcal{H}, \alpha \in \mathbb{R}} O_n(f, \alpha) + \frac{\lambda_n}{2} ||f||^2_{\mathcal{H}},
\]

where \( O_n(f, \alpha) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > f(X_i)\}} (\gamma - f(X_i) - \alpha). \) For simplicity, in the following we consider \( \mathcal{H} \) be a RKHS. The results can be extended to other scenarios such as the class of linear functions with the \( l_1 \) penalty. Define \( O_T(f, \alpha) = -M_2(f, \alpha) \) and let \( (\hat{f}_n, \hat{\alpha}_n) = \arg\min_{f \in \mathcal{H}, \alpha \in \mathbb{R}} O_T(f, \alpha) + \frac{\lambda_n}{2} ||f||^2_{\mathcal{H}}. \) Then \( A(\lambda_n) := O_T(f_n, \alpha_{2n}) + \frac{\lambda_n}{2} ||f_n||^2_{\mathcal{H}} - O_T(f^*, \alpha^*_S) \) is considered to be the approximation error. The following theorem gives us a finite sample upper bound of our estimated optimal IDR and the optimal IDR based on \( M_{0,y}(d, \alpha) \).

Theorem 4.3. For any distribution \( P \) over \((X, A, R)\) such that the reward is uniformly bounded, that is, \(|R| \leq C_0\) for some positive constant \(C_0\), if \( A(\lambda_n) \leq C_1 \lambda_n^{w_1}\), where \(w_1 \in (0, 1)\), then with probability at least \(1 - \epsilon\),

\[
M_{0,y}(d_{0,y}, \alpha_0) - M_{0,y}(\text{sign}(\hat{f}), \hat{\alpha}) \leq C_2 \lambda_n^{-w_1/2\epsilon},
\]

for some positive constant \(C_2\).

The above theorem shows that the difference between our estimated IDR and the optimal IDR under \( M_{0,y}(d, \alpha) \) converges to 0 in probability under some conditions. The derivation requires additional control on the \( \alpha \), which can be shown to be uniformly bounded due to the optimization property. The upper bound assumption on the approximation error \( A(\lambda_n) \) is analogous to those in the statistical learning literature such as Steinwart and Scovel (2007) to derive the convergence rate. The convergence rate is the same as those in Zhao et al. (2012) and Zhou et al. (2017). This is not surprising because the proposed \( M_{0,y}(d) \) can be roughly regarded as the truncated mean.

5. Numerical Studies

We conduct extensive simulation studies and real data analysis to illustrate the performance of our proposed method. For all simulation settings, we consider binary randomized trials with equal probabilities of patients being assigned to each treatment group. The performance will be similar when the difference in terms of the treatment assigned probabilities is not too large. In the extremely unbalanced case, the variance of the empirical approximation (17) could be large, which thus requires a large sample to achieve desired performance. For the ease of presentation, we only present results of estimated optimal IDRs under \( M_{0,y}(d) \) with the \( l_2 \) penalty on \( f \) in Problem (17), where \( f \) is a linear function of \( X \). We denote this method as \( l_2\text{-CVaR}_\gamma \).

All tuning parameters are selected based on the 10-fold-cross-validation procedure. We select the tuning parameter that maximizes the empirical average of \( M_{0,y}(d) \) on the validation set. We compare our methods with the following four methods: (1) the \( l_1\text{-PLS} \) by Qian and Murphy (2011) with the basis function \((1, X, A, XA)\); (2) the RWL by Zhou et al. (2017) with the linear kernel and the \( l_2 \) penalty. Note that RWL used the same truncated hinge loss as ours; (3) the \( \gamma \)-quantile optimal treatment regime denoted by IPWEQ\(_\gamma\) (Wang et al. 2018); (4) the model-based \( \gamma \)-quantile regression for the optimal treatment regime denoted by QR\(_\gamma\) (Linn, Laber, and Stefanski 2017; Xiao, Zhang, and Lu 2019). All results are based on 100 replications.

Since we only consider the randomized design study where \( A \) is either \(1\) or \(-1\) with equal probabilities, we do not compare those methods designed for observational studies such as (Zhang et al. 2012).

5.1. A Motivating Example

We use one toy example to demonstrate the necessity of tail control. In particular, the categorical covariate gender \( X \) is generated by the uniform distribution over \([1, -1]\), where \(1\) and \(-1\) denote male and female, respectively. Based on the randomly assigned treatment, the corresponding outcome \(R\) is generated by the following model:

\[
R = I(XA = 1)\epsilon_1 + I(XA = -1)\epsilon_2,
\]

where \(\epsilon_1 \sim \mathcal{N}(-0.1, 1)\) and \(\epsilon_2 \sim \mathcal{N}(0, 0.5)\). The corresponding plot is given in Figure 2. We consider training data with the sample size \(n = 200\) and independently generated test data of size 10,000. Based on test data, in Figure 3, we plot boxplots of five different outcome distributions if treatments follow estimated IDRs by \(l_1\)-PLS, linear-RWL, IPWEQ\(_{0.5}\), QR\(_{0.5}\), and \(l_2\text{-CVaR}_{0.5}\), respectively. Based on these boxplots, we can observe that since there is not much difference between these two treatments based on the expected outcome, the empirical mean of value functions resulted from these five methods are indistinguishable. However, besides improving the expected or median outcome for each individual, our methods also control the tails of individuals, thus prefer treatments with less variability. The resulting outcome distribution by our method is more stable, and thus has less variability than the other four methods. If we choose different \(\gamma\) for IPWEQ\(_\gamma\) and QR\(_\gamma\), for example,
assign the treatment IDs given by our methods (or quantile-based methods) tend to report smaller coefficients of all estimated IDRs, it seems that robust under distributional shifts of both covariates.

Next we demonstrate the superior performance of our methods under distributional shifts of both covariates $X$ and outcome $R$ based on the dual representations of $M_{0,Y}(d)$ in (15). Since $0.25$, their resulting outcome distributions will be similar to ours since they focus more on the tail like our proposed method.

5.2. An Illustrative Example

To further illustrate how our proposed method differs from the mean and quantile optimal treatment regimes, we consider another simulating example, motivated by Wang et al. (2018). Consider the reward $R = 1 + X + (3 - 5X)A + (1 + A + 2AX)\epsilon$, where $\epsilon \sim \mathcal{N}(0,1)$ and $X \sim \text{uniform}(0,1)$. We implement the above mentioned five methods with $\gamma = 0.1, 0.25, 0.5$ using 50,000 observations. Then we report their performance on 106 testing observations under different criteria in Table 1. As expected, methods perform the best under their corresponding criteria. For the QR$_\gamma$ method, while the goal is to find the best treatment with the largest conditional quantile treatment effect, it is not necessary to maximize the population quantile, that is, $Q_{\gamma}(R(d))$. But their performances are also robust. In general, under the mean evaluation criterion, since our proposed methods focus on outcomes lower than certain quantiles, the performances are not as good as those mean-optimal treatment regimes, especially when $\gamma$ is small. However, when the evaluating criterion is on the low tails of outcomes, our methods are the most robust among all compared methods as $M_{0,Y}(d) \leq \min[V(d), Q_{\gamma}(R(d))]$. Furthermore, based on the reported coefficients of all estimated IDRs, it seems that robust IDRs given by our methods (or quantile based methods) tend to assign the treatment $A = -1$ as the majority treatment as it is better in improving lower tails of the outcome. Here we normalize the coefficients of IDRs so that all intercepts equal to 1.

5.3. Distributional Shift Examples

(15) is related to the expected outcome under some class of distributions, we only compare with mean-optimal IDRs such as $l_1$-PLS and RWL. For the purpose of illustration, we only consider the sample size $n = 200$ and the dimension $p = 20$. The outcome $R$ is generated by the model: $R = 1 + X_1 + X_2 + A(X_1 - X_2 + X_3) + \epsilon$. We consider the following two distributional shift scenarios:

1. Each training covariate follows a two-component Gaussian mixture distribution of $\mathcal{N}(0,1)$ and asymmetric log-normal distribution lognorm($0,2$) with probabilities of the two mixture components to be 0.7 and 0.3, respectively and $\epsilon$ follows the standard Gaussian distribution. Each test covariate and the test $\epsilon$ follow the standard Gaussian distribution.

2. Covariates $X$ are generated by the uniform distribution between $-1$ and 1 and $\epsilon$ follows a two-component mixture distribution of $\mathcal{N}(0,1)$ and asymmetric log-normal distribution lognorm($0,2$) with probabilities of the two mixture component to be 0.7 and 0.3, respectively. For the test data, covariates $X$ follow uniform distribution from $-1$ to 1 and the test $\epsilon$ follows the standard Gaussian distribution.

The first scenario considers the covariate distributional shift and the second scenario considers the outcome distributional shift. For simplicity, we only report misclassification error rates given by $l_1$-PLS, RWL and $l_2$-CVaR$_{0.5}$ in Table 2. For Scenario 1, since $l_1$-PLS assumes a linear model, its performance is not affected by the distributional shift of covariates. In con-

| Scenario (1)          | Scenario (2)          |
|-----------------------|-----------------------|
| $l_1$-PLS             | 0.07(0.01)            |
| RWL                   | 0.14(0.006)           |
| $l_2$-CVaR$_{0.5}$    | 0.08(0.01)            |
| 1-PLS                 | 0.006                 |
| 1-PLS                 | 0.007                 |
| 1-PLS                 | 0.01                  |

Bold entries represent smallest values in each column.

Figure 3. Boxplots of value functions computed by five methods and two reference distributions. Methods starting from left to right are $l_1$-PLS, linear RWL, IPWEQ$_{0.5}$ QR$_{0.5}$, $l_2$-CVaR$_{0.5}$, respectively. The last two boxplots correspond to two reference normal distributions, that is, $\mathcal{N}(-0.1, 0.5)$ and $\mathcal{N}(0, 1)$.
Table 3. Setting 1 with $n = 1000$ and $p = 10$. Mean, different quantiles, and CVaR of the outcomes under 11 different treatment regimes given by different methods are reported.

| Criterion | $l_1$-PLS | RWL | IPWEQ$_{0.5}$ | IPWEQ$_{0.25}$ | IPWEQ$_{0.10}$ |
|-----------|------------|-----|----------------|----------------|----------------|
| mean      | 2.162 (0.002) | 2.152 (0.003) | 2.081 (0.011) | 2.108 (0.009) | 2.110 (0.008) |
| $Q_{0.5}$ | 2.151 (0.001) | 2.141 (0.002) | 2.089 (0.009) | 2.107 (0.007) | 2.107 (0.007) |
| $Q_{1}$   | 0.192 (0.006) | 0.175 (0.005) | 0.016 (0.026) | 0.079 (0.021) | 0.088 (0.016) |
| CVaR$_{0.5}$ | 0.931 (0.004) | 0.916 (0.004) | 0.798 (0.019) | 0.843 (0.015) | 0.849 (0.012) |
| CVaR$_{0.1}$ | -0.482 (0.008) | -0.5 (0.007) | -0.718 (0.038) | -0.63 (0.028) | -0.611 (0.022) |
| overall-error-rate | 0.029 (0.005) | 0.05 (0.004) | 0.117 (0.008) | 0.097 (0.008) | 0.095 (0.006) |
| upper-error-rate | 0.006 (0.001) | 0.014 (0.002) | 0.032 (0.003) | 0.027 (0.003) | 0.029 (0.003) |
| lower-error-rate | 0.010 (0.002) | 0.013 (0.001) | 0.049 (0.007) | 0.034 (0.005) | 0.03 (0.003) |

| Criterion | $QR_{0.5}$ | $QR_{0.25}$ | $QR_{0.1}$ | $CVaR_{0.5}$ | $CVaR_{0.25}$ | $CVaR_{0.1}$ |
|-----------|------------|-------------|------------|--------------|--------------|--------------|
| mean      | 2.163 (0.001) | 2.163 (0.001) | 2.160 (0.001) | 2.143 (0.004) | 2.141 (0.004) | 2.131 (0.007) |
| $Q_{0.5}$ | 2.151 (0.001) | 2.151 (0.001) | 2.148 (0.001) | 2.135 (0.004) | 2.132 (0.004) | 2.124 (0.006) |
| $Q_{1}$   | 0.199 (0.000) | 0.197 (0.002) | 0.191 (0.002) | 0.156 (0.01) | 0.15 (0.009) | 0.132 (0.014) |
| CVaR$_{0.5}$ | 0.935 (0.001) | 0.933 (0.001) | 0.929 (0.002) | 0.902 (0.007) | 0.897 (0.007) | 0.883 (0.011) |
| CVaR$_{0.1}$ | -0.47 (0.002) | -0.473 (0.002) | -0.481 (0.004) | -0.527 (0.013) | -0.532 (0.011) | -0.551 (0.016) |
| overall-error-rate | 0.026 (0.002) | 0.028 (0.002) | 0.035 (0.002) | 0.061 (0.006) | 0.065 (0.005) | 0.075 (0.006) |
| upper-error-rate | 0.007 (0.001) | 0.008 (0.001) | 0.009 (0.001) | 0.016 (0.002) | 0.017 (0.002) | 0.022 (0.003) |
| lower-error-rate | 0.007 (0.001) | 0.008 (0.001) | 0.011 (0.001) | 0.018 (0.002) | 0.019 (0.002) | 0.02 (0.002) |

Note: The corresponding standard errors are in the parentheses. Bold entries represent smallest values in each column.

5.4. Additional Simulation Scenarios

We further study the performance of our proposed methods via three additional simulation examples, where the treatment-covariate interactions are linear. Nonlinear scenarios can be found in the supplementary material. We consider different combinations of $n$ and $p$, that is, $n = 500, 1000$ and $p = 5, 10, 20$, respectively. For the ease of presentation, we only present results of $n = 1000$ and $p = 10$ since it is close to our real data size. Other results can be found in the supplementary material. We consider covariates $X$ generated by the uniform distribution between $[-1,1]$. The outcome $R$ is generated by the model: $R = m(X) + \delta(X) + \epsilon$. We consider the following three different combinations of $m(X)$, $\delta(X)$ and $\epsilon$:

1. $m(X) = 1 + X_1 + X_2$, $\delta(X) = X_1 - X_2 + X_3$, and $\epsilon \sim \mathcal{N}(0, 2)$;
2. $m(X) = 1 + X_1 + X_2$, $\delta(X) = X_1 - X_2 + X_3$, and $\log(\epsilon) \sim \mathcal{N}(0, 1)$;
3. $m(X) = 1 + X_1 + X_2 + X_3 + X_4$, $\delta(X) = \exp(2X_1)(-1 + X_1 - X_2 + X_3)$, and $\log(\epsilon) \sim \mathcal{N}(0, 2)$.

In the first scenario, we consider the standard normal error distribution. This setting is oracle for $l_1$-PLS and $QR_{0.5}$. The second scenario considers the error generated by log-normal distribution, which has a heavy right tail. This setting is used to test the robustness of different IDRs and is oracle for $QR_{\gamma}$. The last setting is used to show the advantage of direct methods over model-based methods when the model is misspecified but the decision function is correctly specified. In addition, this scenario considers the same error distribution as the second one. Note that in all these three scenarios, the treatment only interacts with $\delta(X)$ in the mean model. Therefore, the global optimal decision rule is always sign($\delta(X)$).

For our proposed method, IPWEQ$_{\gamma}$ and $QR_{\gamma}$, three different $\gamma$’s, that is, $\gamma = 0.1, 0.25, 0.5$ are considered. In order to evaluate different methods, we generate test data and compute mean, different quantiles and CVaRs on the test data under all estimated IDRs. More results can be found in supplementary material. Since the global optimal IDR is known, that is, sign($\delta(X)$), we can compute misclassification error rates for different methods on the test data, denoted by overall-error-rate in all tables below. In addition, to further show the robustness of different IDRs, we also compute the upper-error-rate and lower-error-rate. The upper-error-rate refers to misclassification error of observations with outcomes larger than 90% quantiles and the lower-error-rate is for observations with outcomes smaller than 10% quantiles. The results are summarized in Tables 3-5. Overall, our methods show competitive performances among all methods. In particular, for Scenario 1, which is the standard simulation setting in the literature, our proposed methods perform well in finding optimal IDRs, despite the slightly worse performance than mean-optimal IDRs. Since we focus more on the lower-tail, this is not surprising. For Scenario 2, as the error distribution is heavily right-tailed, methods under the value function framework ignore individuals with potentially high risks, that is, outcome lower than the threshold, while only focusing on maximizing the value function. Therefore, the resulting performances could be worse than robust methods. Among three robust methods, $QR_{\gamma}$ performs the best as it is oracle in this scenario. It can also be seen that our method performs better than that of Wang et al. (2018) in improving outcomes of lower tails. One potential reason is due to the optimization instability in training IPWEQ$_{\gamma}$. In terms of misclassification errors, all robust methods have smaller lower-error-rates than mean-optimal methods, demonstrating the superior performance of these estimated IDRs on the individuals with poor outcomes.
Table 4. Setting 2 with $n = 1000$ and $p = 10$. Mean, different quantiles, and CVaR of the outcomes under 11 different treatment regimes given by different methods are reported.

| Criterion | $I_1$-PLS | RWL | IPWEQ$_{0.5}$ | IPWEQ$_{0.25}$ | IPWEQ$_{0.10}$ |
|-----------|-----------|-----|---------------|----------------|---------------|
| mean      | 8.875(0.122) | 8.933(0.106) | 9.316(0.029) | 9.42(0.017) | 9.459(0.011) |
| $Q_{0.5}$ | 3.208(0.107) | 3.268(0.092) | 3.574(0.011) | 3.62(0.006) | 3.635(0.004) |
| CVaR$_{0.5}$ | 1.454(0.136) | 1.514(0.119) | 1.961(0.034) | 2.094(0.02) | 2.148(0.012) |
| CVaR$_{0.1}$ | -0.337(0.187) | -0.301(0.157) | 0.281(0.084) | 0.559(0.049) | 0.686(0.027) |
| overall-error-rate | 0.336(0.039) | 0.318(0.034) | 0.18(0.012) | 0.127(0.009) | 0.099(0.007) |
| upper-error-rate | 0.334(0.039) | 0.317(0.034) | 0.177(0.011) | 0.125(0.009) | 0.098(0.007) |
| lower-error-rate | 0.265(0.06) | 0.247(0.052) | 0.096(0.019) | 0.047(0.011) | 0.027(0.005) |

Table 5. Setting 3 with $n = 1000$ and $p = 10$. Mean, different quantiles, and CVaR of the outcomes under 11 different treatment regimes given by different methods are reported.

| Criterion | $I_1$-PLS | RWL | IPWEQ$_{0.5}$ | IPWEQ$_{0.25}$ | IPWEQ$_{0.10}$ |
|-----------|-----------|-----|---------------|----------------|---------------|
| mean      | 9.504(0.006) | 9.512(0.003) | 9.507(0.002) | 9.485(0.008) | 9.501(0.006) |
| $Q_{0.5}$ | 3.656(0.001) | 3.662(0.0) | 3.663(0.0) | 3.647(0.003) | 3.652(0.002) |
| $Q_{0.1}$ | 1.405(0.001) | 1.408(0.0) | 1.335(0.011) | 1.363(0.007) | 1.364(0.007) |
| CVaR$_{0.5}$ | 2.211(0.003) | 2.231(0.00) | 2.233(0.0) | 2.183(0.007) | 2.201(0.005) |
| CVaR$_{0.1}$ | 0.818(0.008) | 0.861(0.001) | 0.865(0.0) | 0.76(0.016) | 0.797(0.011) |
| overall-error-rate | 0.051(0.004) | 0.014(0.001) | 0.006(0.0) | 0.077(0.006) | 0.062(0.005) |
| upper-error-rate | 0.051(0.004) | 0.014(0.001) | 0.006(0.0) | 0.076(0.006) | 0.061(0.005) |
| lower-error-rate | 0.031(0.002) | 0.003(0.0) | 0.001(0.0) | 0.015(0.003) | 0.011(0.002) |

In addition, we can observe that our method is more accurate than IPWEQ$_{y}$ in identifying optimal treatments for individuals with small outcomes. For Scenario 3, our method performs the best among all methods. This is not surprising as our method does not model the outcome directly and thus does not suffer from model misspecification like $Q_{y}$. In addition, our method is more robust than those under the value function framework, thus shows better empirical performance in this setting. Note that IPWEQ$_{y}$ performs similar to ours as it is also a direct method.

5.5. Real Data Applications

We perform a real data analysis to further evaluate our proposed robust criterion for estimating optimal IDRs. We use the clinical trial dataset from “AIDS Clinical Trials Group (ACTG) 175” in Hammer et al. (1996) to study whether there exists some subpopulations that are suitable for different combinations of treatments for AIDS. In this study, a total number of 2139 patients with HIV infection is randomly assigned into four treatment groups: zidovudine (ZDV) monotherapy, ZDV combined with didanosine (ddI), ZDV combined with zalcitabine (ZAL), and ddI monotherapy with equal probabilities. In this data application, we focus on finding optimal IDRs between two treatments: ZDV with ddI and ZDV with ZAL as our interest. The total number of patients receiving these two treatments is 1046.

Similar to the previous studies by Lu, Zhang, and Zeng (2013) and Fan et al. (2017), we select 12 baseline covariates for our model: age (year), weight (kg), CD4 T cells amount at baseline, Karnofsky score (scale at 0-100), CD8 amount at baseline, gender (1 = male, 0 = female), homosexual activity (1 = yes, 0 = no), race (1 = non white, 0 = white), history of intravenous drug use (1 = yes, 0 = no), symptomatic status (1=symptomatic, 0=asymptomatic), antiretroviral history (1=experienced, 0=naive) and hemophilia (1=yes, 0=no). The first five covariates are continuous and have been scaled before estimation. The remaining seven covariates are binary categorical variables. We consider the outcome as the difference between early stage (around 25 weeks) CD4+ T (cells/mm$^3$) cell amount and the baseline. Using this outcome, we can estimate the optimal IDR under our proposed robust criterion. To
evaluate the performance of our proposed methods under the robust criterion, we randomly divide the dataset into five folds and use four of them to estimate optimal IDR s by different methods. The remaining one fold of data is used to evaluate the performances of different methods. We repeat this procedure 200 times. For each method, we report same quantities as before except misclassification errors as we do not know the truth.

From Table 6, we can see that our proposed methods are competitive. Due to the possible right tail distribution of $R$ as shown in Figure 4, we observe the similar pattern as Scenario 2 where the outcome distribution has potential right tails. Robust methods seem to outperform mean-based methods in this application as mean-based methods tend to ignore individuals with small outcomes due to the individuals with particularly large outcomes. Among all robust methods we compare, it seems that $l_2$-CVaR$_{0.25}$ overall performs the best. It is somewhat surprising to see that while our methods in general have better performances in terms of quantiles and mean than other robust methods, $QR_{\gamma}$ methods show better performance than ours in terms of the CVaR criteria. This is possibly due to the difference between model-based and direct methods as observed in our simulation Scenario 2. After comparison, we then apply our method on the whole dataset using $\gamma = 0.25$ and find that the estimated optimal rule tends to assign all individuals to the combined treatment: ZDV with ZAL, which indicates ZDV with ZAL has uniformly better treatment effect than ZDV with ddI under this criterion. This implies that if one considers personalized treatments for individuals at risk (below 25% in this case), it may be more desirable to always assign the second treatment. In contrast, there seems to have a subgroup that can benefit from the treatment 1 according to the results from $l_1$-PLS and RWL. We also examine the estimated IDR using our method with $\gamma = 0.5$. The most important variables that contribute to our estimated IDR are age, Karnofsky score, gender, race, antiretroviral history, and CD4 T cells at the baseline. In particular, our estimated IDR recommends ZDV with ZAL to old patients, but ZDV with ddI to young patients. This is consistent with the finding in the existing literature such as Fan et al. (2017). To further investigate whether the subgroup is valid or not under the proposed criterion, it will be interesting to develop and conduct a similar hypothesis testing under the CVaR criteria to that in Shi, Lu, and Song (2019).

### 6. Discussion and Extensions

In this article, we discuss the robustness issue for ITR problems and propose new criteria to control tails of the individuals’ outcomes. Our framework is broad and can cover a wide range of criteria. In this section, we further discuss two extensions of our proposed criteria in terms of modelings. More discussions and results can be found in the supplementary material.

#### 6.1. Control of Two-sided Truncated Outcome

In this subsection, we extend the $\gamma$-CVaR criterion to the truncated mean over a certain range of quantiles. Consider two quantiles $y_1$ and $y_2$ with $y_1 < y_2$, we evaluate each IDR $d$ by considering its truncated mean between $y_2$ and $y_1$ quantiles, that is,

$$
E_{d}[\mathbb{I}(R \leq Q_{y_1}(R(d)))] - E_{d}[\mathbb{I}(R \leq Q_{y_2}(R(d)))]
$$

By the property of truncated mean, similar as $\gamma$-CVaR, we can use the following measure to evaluate each IDR $d$:

$$
\text{hence, }
$$

![Figure 4. Histogram of the difference between early stage CD4+T cell and the baseline. It exhibits a right tail behavior.](image)
\begin{equation}
M_{\gamma_1,\gamma_2}(d) = \sup_{\alpha_1 \in \mathbb{R}} \left\{ \gamma_1 \alpha_1 - \mathbb{E} \left[ \frac{1}{\pi(A|X)} (A - d(X)) (\alpha_1 - R)_+ \right] \right\} \\
- \sup_{\alpha_2 \in \mathbb{R}} \left\{ \gamma_2 \alpha_2 - \mathbb{E} \left[ \frac{1}{\pi(A|X)} (A - d(X)) (\alpha_2 - R)_+ \right] \right\},
\end{equation}

which is the same as (24) when the outcome has a continuous distribution. Then the optimal IDR under this criterion is defined as

\begin{equation}
d_3 \in \arg\max_{d \in D_0} M_{\gamma_1,\gamma_2}(d).
\end{equation}

In order to estimate \(d_3\), we can also replace the 0-1 loss by the smooth surrogate loss function \(S(u)\) and solve the following optimization problem.

\begin{equation}
\begin{aligned}
&\min_{f \in \mathcal{H}} \left\{ \left( \min_{\alpha_1 \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} S(A_i f(X_i)) (\gamma_1 \alpha_1 + (\alpha_1 - R_i)_+) \right) \\
&\quad \quad - \left( \min_{\alpha_2 \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} S(A_i f(X_i)) (\gamma_2 \alpha_2 + (\alpha_2 - R_i)_+) \right) \\
&\quad + \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2 \right\}.
\end{aligned}
\end{equation}

One can similarly use Proposition 3.1 to decompose the objective function in (27) into a DC function and develop a corresponding DC algorithm. In addition, we note that \(d_3\) indeed can approximate the optimal quantile IDR \(d_\gamma\) when \(\gamma_1\) and \(\gamma_2\) are chosen to be close to \(\gamma\).

### 6.2. Flexible Models

In previous applications of our proposed method, we only use the CVaR criterion to estimate a robust IDR. In practice, one can combine it with other criteria. For example, in order to have tail controls and more mean-outcome improvement, we can consider to use \(M_{1,\gamma}(d)\) defined below by combining our proposed \(M_{0,\gamma}(d)\) and \(V(d)\) together:

\begin{equation}
M_{1,\gamma}(d) := 0.5V(d) + 0.5M_{0,\gamma}(d).
\end{equation}

All our results in Sections 3 and 4 can be naturally extended. This is indeed roughly equivalent to maximizing \(V(d)\) among all the decision rules with \(M_{0,\gamma}(d)\) larger than some threshold. Such an interpretation motivates us to consider another approach to find the optimal IDR when observing multiple outcomes after receiving treatments. Without loss of generality, suppose we can observe a risk outcome \(H\) in addition to \(R\). In general, we prefer a smaller risk outcome. By some modification on \(M_{0,\gamma}(d)\), we can search the optimal IDRs by

\begin{equation}
\begin{aligned}
&\max_{d \in D_0} V(d) \\
&\quad \text{subject to } \min_{\alpha \in \mathbb{R}} \left\{ \alpha + \frac{1}{\gamma} \mathbb{E}[(H(d) - \alpha)_+] \right\} \leq t,
\end{aligned}
\end{equation}

for some prespecified constant \(t\), so that we can control the risk outcome with the CVaR criterion. By a similar analysis as in Section 2, the resulting IDRs using (29) are the best IDRs among all the decision rules with the risk outcome of \((1 - \gamma) \times 100\%\) of the population being less than some threshold \(t\). Given the data, we could use the same techniques in Section 3 to estimate the IDR. Specifically, by using a surrogate function \(S(u)\), we can solve the following optimization problem:

\begin{equation}
\begin{aligned}
&\max_{d \in D_0} \frac{1}{n} \sum_{i=1}^{n} R_i S(A_i f(X_i)) - \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2 \\
&\quad \text{subject to } \min_{\gamma \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( (H_i + \frac{1}{\gamma} (H_i - H_j)_+) S(A_i f(X_i)) \right) \right\} \leq t,
\end{aligned}
\end{equation}

which can be formulated as optimizing a DC function with a DC constraint as we can see that the minimum of a finite number of DC functions in the constraint is still a DC function. However, several issues need to be solved before proceeding. The existence of the optimal solution for the optimization problem (30) should be demonstrated. The use of a surrogate function to replace the indicator function in (29) needs to be further justified. Another challenge is to establish a regret bound for the optimal IDR obtained from Problem (30). Existing techniques in statistical learning theory may not be directly used since there is a stochastic term in the constraint of (29) and (30). Nevertheless, this can be an interesting direction to pursue in the future.

### Supplementary Materials

In the supplement, we provide proofs for the results along with additional simulation results and discussions.

### 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. Yufeng Liu’s research was supported in part by NSF grants DMS-1821231 and DMS-2100729, and NIH grant R01GM126550.

### ORCID

Zhengling Qi [http://orcid.org/0000-0003-0270-7969](http://orcid.org/0000-0003-0270-7969)

Yufeng Liu [http://orcid.org/0000-0002-1686-0545](http://orcid.org/0000-0002-1686-0545)

### References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999), “Coherent Measures of Risk,” Mathematical Finance, 9, 203–228. [2146]

Athey, S., and Wager, S. (2021), “Policy Learning with Observational Data,” Econometrica, 89, 133–161. [2144,2145]

Ben-Tal, A., and Teboulle, M. (1986), “Expected Utility, Penalty Functions, and Duality in Stochastic Nonlinear Programming,” Management Science, 32, 1445–1466. [2146]

Beygelzimer, A., and Langford, J. (2009), “The Offset Tree for Learning,” in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’09, 129–138, New York, NY. ACM. ISBN 978-1-60558-495-9. doi:10.1145/1557019.1557040. [2144]

Bhattacharya, D., and Dupas, P. (2012), “Inferring Welfare Maximizing Treatment Assignment Under Budget Constraints,” Journal of Econometrics, 167, 168–196. [2144]

Chamberlain, G. (2011). “Bayesian Aspects of Treatment Choice,” In The Oxford Handbook of Bayesian Econometrics. Oxford: Oxford University Press. [2144]
Xiao, W., Zhang, H. H., and Lu, W. (2019), “Robust Regression for Optimal Individualized Treatment Rules,” Statistics in Medicine, 38, 2059–2073. [2144,2150]
Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012), “A robust Method for Estimating Optimal Treatment Regimes,” Biometrics, 68, 1010–1018. [2150]
Zhang, C., Chen, J., Fu, H., He, X., Zhao, Y., and Liu, Y. (2020), “Multicategory Outcome Weighted Margin-Based Learning for Estimating Individualized Treatment Rules,” Statistica Sinica, 20, 1857–1879. [2144]
Zhang, Y., Laber, E. B., Tsiatis, A., and Davidian, M. (2015), “Using Decision Lists to Construct Interpretable and Parsimonious Treatment Regimes,” Biometrics, 71, 895–904. [2144]
Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), “Estimating Individualized Treatment Rules Using Outcome Weighted Learning,” Journal of the American Statistical Association, 107, 1106–1118. [2144,2150]
Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015), “Doubly Robust Learning for Estimating Individualized Treatment with Censored Data,” Biometrika, 102, 151. [2144]
Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017), “Residual Weighted Learning for Estimating Individualized Treatment Rules,” Journal of the American Statistical Association, 112, 169–187. [2144,2145,2148,2149,2150]