On Validation and Planning of An Optimal Decision Rule with Application in Healthcare Studies

Hengrui Cai [1] Wenbin Lu [1] Rui Song [1]

Abstract

In the current era of personalized recommendation, one major interest is to develop an optimal individualized decision rule that assigns individuals with the best treatment option according to their covariates. Estimation of optimal decision rules (ODR) has been extensively investigated recently, however, at present, no testing procedure is proposed to verify whether these ODRs are significantly better than the naive decision rule that always assigning individuals to a fixed treatment option. In this paper, we propose a testing procedure for detecting the existence of an ODR that is better than the naive decision rule under the randomized trials. We construct the proposed test based on the difference of estimated value functions using the augmented inverse probability weighted method. The asymptotic distributions of the proposed test statistic under the null and local alternative hypotheses are established. Based on the established asymptotic distributions, we further develop a sample size calculation formula for testing the existence of an ODR in designing A/B tests. Extensive simulations and a real data application to a schizophrenia clinical trial data are conducted to demonstrate the empirical validity of the proposed methods.

1. Introduction

In the current era of personalized recommendation, one major interest is to develop an optimal individualized decision rule that assigns individuals with the best treatment option according to their covariates. Due to individuals' heterogeneity in outcome to different treatment options, it is common that there may not exist a unified best decision for all individuals. A number of methods have been developed for estimating optimal decision rules (ODR), which include Q-learning (Watkins & Dayan, 1992; Zhao et al., 2009), A-learning (Murphy, 2003; Robins, 2004; Shi et al., 2018a), direct value search methods (Zhang et al., 2012; 2013), outcome-weighted learning (Zhao et al., 2012; Zhou et al., 2017), targeted minimum loss-based estimator (TMLE) (van der Laan & Luedtke, 2015), concordance-assisted learning (Fan et al., 2017; Liang et al., 2017), and maximin-projection learning (Shi et al., 2018b). Although estimation of ODRs has been extensively studied in recent years, to the best of our knowledge, testing the existence of an ODR that is better than the naive decision rule which always assigning individuals to a fixed treatment has been less studied. Moreover, it lacks a simple sample size calculation method for designing randomized trials to test such a hypothesis.

In this paper, we propose a test for the existence of an ODR that is better than always assigning individuals to a fixed treatment (the naive decision rule) in terms of values and derive its associated sample size calculation method. Here, we are interested in the randomized trial settings where the likelihood of assignment is assumed known as constant. Our test statistic is constructed based on the difference of estimated value functions under the estimated ODR and the naive decision rule. Here, the value functions under a given decision rule are estimated nonparametrically using the augmented inverse probability weighted (AIPW) estimator proposed by Zhang et al. (2012), and the decision rule is searched within a class of the basis function of the baseline covariates. However, a challenge is that the asymptotic distributions of the corresponding value difference based test statistic may become degenerate under the null hypothesis that there does not exist an ODR. To overcome this difficulty, we modify the way of estimating the value function under the naive decision rule so that asymptotic normal distributions of the resulting test statistic can be derived under both the null and local alternative hypotheses. Based on the established asymptotic distributions, we also derive the sample size calculation method.

Our contributions can be summarized in the following aspects:

[1] Department of Statistics, North Carolina State University, Raleigh, USA. Correspondence to: Hengrui Cai <hcai5@ncsu.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020. Copyright 2020 by the author(s).

• To the best of our knowledge, the proposed simple testing procedure for detecting the existence of an ODR is the first work that forms the hypothesis testing by proposing the non-degenerate value difference as the test statistic, which is a cutting edge work to the personalized recommendation; • The propose testing has novel yet effective sample size calculation method for designing studies in healthcare, which also contributes to the policy evaluation literature from a unique angle; • Our test method gives clear instruction on validation of a personalized optimal decision making between two competitive treatment options, which has great potential towards developing an automatic decision-making system that is capable of filtering ineffective rules and planing the ODR.

The rest of the paper is organized as follows. Section 3 introduces the statistical framework for testing the existence of an ODR. In section 4, we present the proposed test statistic and establish its asymptotic distributions under both the null and local alternative hypotheses. Section 5 gives the sample size calculation procedure based on the established asymptotic distributions. The finite sample performance of the proposed test and the associated sample size calculation method are evaluated by simulation studies in Section 6. An application of the proposed method to a dataset from a randomized schizophrenia study is illustrated in Section 7. In section 8, we conclude our paper with discussions. The proofs of all the theorems are given in the supplementary.

2. Related Work

In the literature, many tests and associated sample size methods have been developed for designing A/B tests to compare adaptive treatment strategies. For example, Murphy (2005) advocated the use of sequential multiple assignment randomized trials to develop adaptive treatment strategies and derived associated sample size method. Dawson & Lavori (2010) derived sample size methods for evaluating decision rules in multi-stage randomized trails. Kang et al. (2017) proposed a score-type test statistics by using a semi-parametric approach for detecting the existence of the subgroup and developed a novel procedure to calculate the sample size based on the proposed test. However, these sample size methods were designed for comparing adaptive treatment strategies or specified decision rules in multi-stage randomized trails, but not for testing the existence of an optimal decision rule.

On the other hand, there are some recent works closely related to the choice of test statistics. Laber et al. (2016) developed a method to size a two-arm randomized trial for finding a nearly optimal decision rule by using pilot data via Q-learning. Luedtke & Van Der Laan (2016) studied the statistical inference of the TMLE that mainly handles the non-regular case for the online estimation. Whereas, those methods either has very complicated asymptotic distribution or is difficult to implement in practice with additional tuning hyperparameters. To propose an efficient test statistic with a simple sample size calculation for healthcare studies, we mainly focus on the doubly robust estimator studied in Zhang et al. (2012).

3. Statistical Framework

In a randomized experiment/trial, suppose there are two treatment options, e.g., control (treatment 0) and experimental treatment (treatment 1). Let A, taking values 0 or 1 in accordance with the two options, denote the assigned treatment. In addition, let X be a $p \times 1$ vector of baseline covariates and Y be the observed outcome of interest. The observed data consist of $\{O_i = (Y_i, A_i, X_i), i = 1, \dots, n\}$, which are independent and identically distributed across i.

To introduce the optimal decision rule (ODR), we define the potential outcomes $Y^\star(0)$ and $Y^\star(1)$ as the outcomes that would be observed were a subject receiving treatment 0 or 1, respectively. As standard in causal inference (Rubin, 1978), we make the following assumptions: (i) stable unit treatment value assumption (SUTVA): $Y = AY^\star(1) + (1-A)Y^\star(0)$; (ii) no unmeasured confounder assumption: $\{Y^\star(1), Y^\star(0)\}$ are independent of A conditional on X. A decision rule is a deterministic function $d(\cdot)$ that maps X to $\{0,1\}$. Define the potential outcome of interest under $d(\cdot)$ as

$$Y^*(d) = Y^*(0)\{1 - d(X)\} + Y^*(1)d(X),$$

which would be observed if a randomly chosen individual had received a treatment according to $d(\cdot)$, where we suppress the dependence of $Y^*(d)$ on X. We then define the value function under $d(\cdot)$ as the expectation of the potential outcome as

$$V(d) = \mathrm{E}\{Y^*(d)\} = \mathrm{E}[Y^*(0)\{1 - d(X)\} + Y^*(1)d(X)].$$

Suppose the decision rule $d(\cdot)$ relies on a model parameter β, denoted as $d(X, \beta) = I\{g(X)^\top \beta > 0\}$, where $I(\cdot)$ is an indicator function and $g(\cdot)$ is an unknown function. We use $\phi_X(\cdot)$ to denote a set of basis functions of X with length v, which are "rich" enough to approximate the underlying function $g(\cdot)$. Here, for notational simplicity, we include 1 in $\phi_X(\cdot)$ so that the parameter $\beta \in \mathbb{R}^{v+1}$ includes intercept. Given a decision rule $d(X, \beta)$, we use a shorthand to write its value function $V(d)$ as $V(\beta) = E\{Y^\star(d(X, \beta))\}$. As a result, we have the optimal decision rule (ODR) of interest defined to maximize the value function among the class of $I\{\phi_X(X)^\top \beta > 0\}$ as $d(X, \beta_0)$, where $\beta_0 = \arg\max_{||\beta||=1} V(\beta)$. Then, the value function under the ODR $d(X, \beta_0)$ is $V(\beta_0)$. Here, we make the following assumption

$$P(g(X)^\top \beta = 0) = 0. \tag{1}$$

That is, we only consider the regular case in this paper. Such an assumption usually holds when covariates X contain some continuous variables. In addition, the class of the decision rules include the naive decision rules that assign all individuals to treatment 1 or 0 as special cases. To see this, setting $\beta = (1, 0, ..., 0)^\top$ gives the naive decision rule $d(X, \beta) \equiv 1$; while setting $\beta = (-1, 0, ..., 0)^\top$ gives the naive decision rule $d(X, \beta) \equiv 0$. Let V_1 and V_0 denote the values under the two naive decision rules, respectively.

Our goal here is to test whether there exists an ODR that is better than the two naive decision rules in terms of values. Without loss of generality, we assume that treatment 1 is no worse than treatment 0 on average, i.e. $V_1 \geq V_0$. This can be easily validated by conducting the two-sample t-test. Then, the considered null and alternative hypotheses can be described as follows

$$H_0: \quad V(\beta_0) = V_1 \quad \text{vs.} \quad H_a: \quad V(\beta_0) > V_1. \qquad (2)$$

Note that the true optimal decision rule may not fall in the class of the basis function. For example, consider the following regression model $E(Y|A, X) = U(X) + AC(X)$, where $U(\cdot)$ is the baseline mean function and $C(\cdot)$ is the contrast function that describes the treatment-covariates interaction. Under the SUTVA and no unmeasured confounder assumptions, it can be shown that the true optimal decision rule is given by $d^{opt}(X) = I\{C(X) > 0\}$, which may not be contained in the class of $I\{\phi_X(X)^\top\beta > 0\}$ if the basis function is not chosen appropriately. The null hypothesis may then correspond to the following two situations. First, the contrast function $C(X) \geq 0$. Under such a situation, the naive decision rule with $d(X) \equiv 1$ is the best and the ODR is obtained by choosing $\beta_0 = (1, 0, ..., 0)^\top$. Second, the contrast function $C(X)$ has a positive probability to be positive and negative across X, where the true optimal decision rule is given by $d^{opt}(X) = I\{C(X) > 0\}$. However, there dose not exist an ODR within a class of $I\{\phi_X(X)^\top\beta > 0\}$ that is strictly better than always assigning all individuals to treatment 1. To better approximate the true ODR, we use cross-validation to choose an appropriate set of basis function, under which we obtain an ODR that achieves the highest empirical value function (See details in Section 6.2).

4. Proposed Test

Define the propensity score $\pi = P(A = 1)$ as the likelihood of assignment, which is assumed known as constant in a randomized trial. To test the hypotheses given in (2), it is natural to construct test statistics based on the difference of estimated value functions under the estimated ODR and the naive decision rule. Here, the value functions under a given decision rule can be estimated nonparametrically using the augmented inverse probability weighted (AIPW) estimator proposed by Zhang et al. (2012). Given a decision rule $d(X, \beta)$, its value function $V(\beta)$ can be consistently estimated by

$$\widehat{V}(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{I\{A_i = d(X_i, \beta)\}}{\pi A_i + (1-\pi)(1-A_i)}\{Y_i - \widehat{\mu}(X_i, \beta)\} + \widehat{\mu}(X_i, \beta),$$

where the augmented term $\widehat{\mu}(X, \beta)$ is an estimator for $\mu(X, \beta) \equiv E\{Y|A = d(X, \beta), X\}$ by a regression model or nonparametric model such as the random forest. Define $\widehat{\beta} = \arg\max_{||\beta||=1} \widehat{V}(\beta)$, which can be calculated by the direct value search through a global optimization algorithm. The estimated value function under the estimated ODR $d(X, \widehat{\beta})$ is then given by $\widehat{V}(\widehat{\beta})$. Similarly, the value V_1 under the naive decision rule $d(X) \equiv 1$ can be estimated by

$$\begin{aligned}\widehat{V}^1 =& \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = 1)\{Y_i - \widehat{\mu}_1(X_i)\}}{\pi A_i + (1-\pi)(1-A_i)} + \widehat{\mu}_1(X_i)\\ =& \frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\pi}\{Y_i - \widehat{\mu}_1(X_i)\} + \widehat{\mu}_1(X_i),\end{aligned}$$

where $\widehat{\mu}_1(X)$ is an estimator for $\mu_1(X) \equiv E(Y|A = 1, X)$. A natural test statistic can be obtained as $\sqrt{n}\{\widehat{V}(\widehat{\beta}) - \widehat{V}^1\}$. However, under the assumption (1), it can be shown that the asymptotic distribution of $\sqrt{n}\{\widehat{V}(\widehat{\beta}) - \widehat{V}^1\}$ is degenerate under the null, i.e. $\sqrt{n}\{\widehat{V}(\widehat{\beta}) - \widehat{V}^1\}$ converges in distribution to 0 (see the proof in the supplementary article). Such test statistic becomes invalid as the type I error cannot maintain the nominal level. Here, we provide more intuition of the degeneration issue as follows. Under H_0, treatment 1 is the optimal choice ($d(X, \beta_0) \equiv 1$), which leads to the highest value ($V(\beta_0) = V_1$). When sample size n increases to infinity, the estimated ODR will become closer and closer to the true ODR, which will assign everyone to treatment 1 with probability one under the regular setting considered in this work. Thus, the statistics will become 0 with probability one when n is large (i.e. degenerate).

To overcome this difficulty and also keep the good properties of the AIPW estimator under the estimated ODR, we consider the following modified estimator for V_1, i.e.

$$\widehat{V}_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\pi},$$

which can be viewed as the inverse probability weighted estimator of the value function under the naive decision rule. Then, our test statistic can be constructed as

$$\begin{aligned}&\widehat{\Delta}_n = \sqrt{n}\{\widehat{V}(\widehat{\beta}) - \widehat{V}_1\}\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\frac{I\{A_i = d(X_i, \widehat{\beta})\}}{\pi A_i + (1-\pi)(1-A_i)}\{Y_i - \widehat{\mu}(X_i, \widehat{\beta})\} \qquad (3)\\ &+ \widehat{\mu}(X_i, \widehat{\beta}) - \frac{A_i Y_i}{\pi}\Bigg].\end{aligned}$$

To establish the asymptotic distribution of $\widehat{\Delta}_n$, we give Lemma 1 and Proposition 1 under the following regularity conditions (see the proofs in the supplementary article): (C1.) The support of X and Y are bounded; (C2.) The density function of the population covariates $f_X(\cdot)$ is bounded away from 0 and ∞ and is twice continuously differentiable with bounded derivatives; (C3.) Mean function $\mu(x, \beta)$ are smooth bounded functions with its first derivative exist and bounded; (C4.) The true value function $V(\beta)$ is twice continuously differentiable at a neighborhood of β_0.

Lemma 1 *Under the regularity conditions (C1-C4), we have*

$$n^{\frac{1}{3}}(\widehat{\beta} - \beta_0) = O_p(1), \tag{4}$$

where $O_p(1)$ means the random variable is stochastically bounded.

The above lemma shows that the convergence rate of $\widehat{\beta}$ is $n^{1/3}$ through a global optimization algorithm, which leads to the proposition below.

Proposition 1 *Under the regularity conditions (C1-C4) and Lemma 1, we have*

$$\sqrt{n}\{\widehat{V}(\widehat{\beta}) - \widehat{V}(\beta_0)\} = o_p(1), \tag{5}$$

where $o_p(1)$ means the random variable converges in probability to zero.

Using the above results, we next derive the analytical form of the standard deviation of the proposed test statistics under both null and alternative hypothesis, which, to the best of our knowledge, is new to the literature, and can be used for deriving the sample size formula and confidence interval further.

The following theorem shows the asymptotic distribution of $\widehat{\Delta}_n$ under the null hypothesis (see the proof in the supplementary article).

Theorem 1 *Under H_0, $\widehat{\Delta}_n$ converges in distribution to a normal random variable with mean 0 and variance*

$$\sigma_0^2 = \frac{1-\pi}{\pi} Var\{E(Y|A=1, X)\},$$

as $n \to \infty$.

Here, σ_0^2 can be consistently estimated by

$$\widehat{\sigma}_0^2 = \frac{1-\pi}{\pi} \widehat{Var}\{\widehat{\mu}_1(X)\}.$$

At level α, we reject the null hypothesis when $\widehat{\Delta}_n / \widehat{\sigma}_0 \geq z_\alpha$, where z_α is an upper α-quantile of the standard normal distribution. Therefore, a two-sided $1-\alpha$ confidence interval (CI) for the difference $V(\beta_0) - V_1$ under the null is given by $\widehat{V}(\widehat{\beta}) - \widehat{V}_1 \pm z_{\alpha/2}\widehat{\sigma}_0/\sqrt{n}$.

Next, we establish the asymptotic distribution of $\widehat{\Delta}_n$ under the local alternative $H_{a,n} : V(\beta_0) = V_1 + \Delta/\sqrt{n}$, where $\Delta > 0$. The proof is given in the supplementary article.

Theorem 2 *Under $H_{a,n}$, we have*

$$\widehat{\Delta}_n = \Delta + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_i + o_p(1),$$

where

$$\begin{aligned}\phi_i =& \frac{I\{A_i = d(X_i, \beta_0)\}}{\pi A_i + (1-\pi)(1-A_i)} \{Y_i - \mu(X_i, \beta_0)\} \\ &+ \mu(X_i, \beta_0) - V(\beta_0) - \left(\frac{A_i}{\pi} Y_i - V_1\right).\end{aligned}$$

It follows that $\widehat{\Delta}$ converges in distribution to a random variable with mean Δ and variance $\sigma_\phi^2 = E(\phi_i^2)$.

Moreover, σ_ϕ^2 can be consistently estimated by $\widehat{\sigma}_\phi^2 = n^{-1}\sum_{i=1}^n \widehat{\phi}_i^2$, where

$$\begin{aligned}\widehat{\phi}_i =& \frac{I\{A_i = d(X_i, \widehat{\beta})\}}{\pi A_i + (1-\pi)(1-A_i)} \{Y_i - \widehat{\mu}(X_i, \widehat{\beta})\} \\ &+ \widehat{\mu}(X_i, \widehat{\beta}) - \widehat{V}(\widehat{\beta}) - \left(\frac{A_i}{\pi} Y_i - \widehat{V}_1\right).\end{aligned}$$

Based on Theorems 1 and 2, the asymptotic power of the proposed α-level test under the local alternative hypothesis $H_{a,n}$ is given by $1 - \Phi\{(z_\alpha \widehat{\sigma}_0 - \Delta)/\widehat{\sigma}_\phi\}$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variable. Similarly, a one-sided $1-\alpha$ CI for the difference $V(\beta_0) - V_1$ under the local alternative is given by $[\widehat{V}(\widehat{\beta}) - \widehat{V}_1 - z_\alpha \widehat{\sigma}_\phi/\sqrt{n}, \infty]$.

5. Sample Size Calculation

Based on the established asymptotic power under the local alternative, we are able to derive a sample size formula to detect a pre-specified important difference $\delta_a = V(\beta_0) - V_1$ with a desired power at least $1-\beta$ for a one-sided level-α test. Specifically, setting $1 - \Phi\{(z_\alpha \widehat{\sigma}_0 - \Delta)/\widehat{\sigma}_\phi\} = 1 - \beta$, we have the required sample size as follows

$$n^\star = \frac{(Z_\alpha \sigma_0 + Z_\beta \sigma_\phi)^2}{\delta_a^2}. \tag{6}$$

In practice, based on a pilot study data, which is assumed to follow the same model as the follow-up study we are trying to size, we could obtain the estimated value difference $\widehat{\delta}_a$, and the variance estimates $\widehat{\sigma}_0^2$ and $\widehat{\sigma}_\phi^2$. Then, the estimated sample size $\widehat{n}^\star$ is given by

$$\widehat{n}^\star = \frac{(Z_\alpha \widehat{\sigma}_0 + Z_\beta \widehat{\sigma}_\phi)^2}{\widehat{\delta}_a^2}. \tag{7}$$

6. Simulation Study

We have conducted extensive simulation studies to investigate the finite sample performance of the proposed test for the existence of an ODR. We consider a randomized trial with underlying linear and nonlinear decision rules separately in the following two subsections. The computing infrastructure used is a Linux cluster with 32 processor cores and 300GB quota. Source code can be found in the supplementary material.

6.1. Testing and evaluation with linear decision rule

Suppose each element of the covariates $X = [X_1, X_2, \cdots, X_p]^\top$ is generated independently from a uniform distribution on $[-1, 1]$ and the treatment assignment indicator A is from a Bernoulli distribution with the success probability $\pi = 0.5$. Assume that the outcome Y follows a regression model,

$$Y = hU(X) + \gamma AC(X) + E, \qquad (8)$$

where $U(\cdot)$ is the baseline function, $C(\cdot)$ is the contrast function that describes the treatment-covariates interaction, $E{\sim}N(0, 0.5)$ is the random error, and h and γ are the tuning parameters that control the effect of the baseline and the treatment-covariates interaction, respectively. In the clinical trials, we usually have a dataset that has a few covariates to construct the decision rule, as described in our real data analysis section. Therefore, we set the dimension of the covariates as $p = 4$ in our simulation study for a better visual illustration (See details in Section 7). We consider $U(X) = 2 + 0.5X_1 - 0.2X_3 - 0.3X_4$, and $C(X) = 2 + X_3 - X_4 - c$, where c is chosen from the set $\{0, 1, 1.5\}$, denoted as Scenarios 1 to 3, respectively. And we choose $\gamma = 1$ or 2, and $h = 0.5$ or 1.

For Scenario 1 ($c = 0$), the ODR is the same as the naive decision rule, $d(X, \beta_0) \equiv 1$. However, for Scenarios 2 and 3 ($c \in \{1, 1.5\}$), it can be shown that they are from the alternative hypothesis, where the corresponding true ODRs can be obtained directly from their contrast functions: $I(0.577 + 0.577x_3 - 0.577x_4 > 0)$ for Scenarios 2, and $I(0.333 + 0.667x_3 - 0.667x_4 > 0)$ for Scenarios 3. The true values of the ODR ($V(\beta_0)$) and the naive decision rule (V_1) can be calculated using the Monte Carlo simulations, and reported in Table 1.

For each setting, we conduct 500 replications with sample size $n = 1000$. In our estimation, we use three different heuristic optimization algorithms, the Nelder-Mead Method (NM), the Simulated Annealing (SA), and the Genetic Algorithm (GA), to search for the ODR within a class of $d(X, \beta) = I(X^\top \beta > 0)$ subjecting to $||\beta||_2 = 1$. Both the 'NM' and 'SA' methods and are implemented in the R package `optimization`, while the 'GA' method is implemented in the R package `GA`. We set the search domain as [-10,10] and the starting values as a zero vector. Here, the augmented term is estimated through a linear regression. The simulation results under the 'GA' are summarized in Table 1, including the type I error under the null ('ERR.'), the power under the alternative ('POW.'), and the mean of the estimates $\widehat{\beta}$ in the ODR. For comparison, the results under the 'NM' and the 'SA' method are also reported in the supplementary article Section B, which are shown close to the results based on the 'GA' method.

Table 1. Simulation results of the proposed test under the Genetic Algorithm.

		$h = 0.5$		$h = 1$	
SCEN.	RESULTS	$\gamma = 1$	$\gamma = 2$	$\gamma = 1$	$\gamma = 2$
1	V_1	3.00	5.00	4.00	6.00
	$V(\beta_0)$	3.00	5.00	4.00	6.00
	ERR.	5.4%	5.8%	5.8%	5.4%
2	V_1	2.00	3.00	3.00	4.00
	$V(\beta_0)$	2.04	3.08	3.04	4.08
	POW.	24.0%	27.4%	15.8%	22.2%
	$\widehat{\beta}_1$	0.598	0.589	0.595	0.587
	$\widehat{\beta}_2$	0.007	-0.001	0.003	-0.001
	$\widehat{\beta}_3$	-0.001	0.001	-0.001	-0.001
	$\widehat{\beta}_4$	0.569	0.570	0.570	0.573
	$\widehat{\beta}_5$	-0.564	-0.573	-0.566	-0.572
3	V_1	1.50	2.00	2.50	3.00
	$V(\beta_0)$	1.64	2.28	2.64	3.28
	POW.	91.8%	99.2%	60.6%	90.8%
	$\widehat{\beta}_1$	0.364	0.353	0.364	0.352
	$\widehat{\beta}_2$	-0.004	-0.001	0.007	-0.002
	$\widehat{\beta}_3$	0.002	-0.001	0.003	0.002
	$\widehat{\beta}_4$	0.655	0.661	0.659	0.663
	$\widehat{\beta}_5$	-0.662	-0.662	-0.658	-0.661

From Table 1, we observe that type I errors are generally close to the nominal level. In addition, the proposed test has reasonable power, especially when the constant c increases. For example, when $c = 1.5$ in Scenario 3, the power increases to above 90% except for the setting with $h = 1$ and $\gamma = 1$. This may be partly due to the fact that the treatment-covariate effect is less informative compared with the baseline mean effect in this setting than other considered settings. Moreover, under the alternative hypothesis (Scenarios 2 and 3), the estimated optimal rule is close to the true optimal rule in all cases. Finally, we plot the density function of the standardized proposed test statistics in Figure 1 for the three scenarios under the Genetic Algorithm. From the plot, it can be seen that the density curves are close to the standard normal distribution under the null (Scenario 1) and shift toward the right with some scaling under the alternative (Scenarios 2 and 3). These findings are consistent with our Theorems 1 and 2. In addition, when the treatment-covariate effect is more informative compared

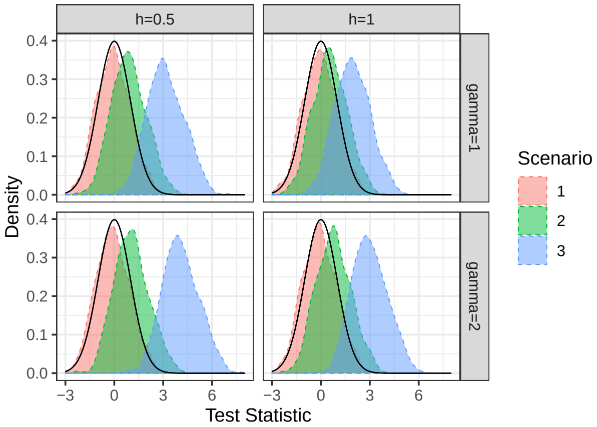

Figure 1. The density function of the standardized proposed test statistics under the Genetic Algorithm: the red, green, and blue filled curves present the cases when $C(X) = X_3 - X_4 + 2$, $X_3 - X_4 + 1$, $X_3 - X_4 + 0.5$, respectively; and the black curves presents the standard normal density. The top two panels stand for the cases when $\gamma = 1$ ($h = 0.5$, and $h = 1$, respectively), and the bottom two panels shows the results for $\gamma = 2$ ($h = 0.5$, and $h = 1$, respectively).

with the baseline (i.e. under $h = 0.5$ and $\gamma = 2$), we can observe that the density curves for the three scenarios are easier to be separated.

6.2. Testing and evaluation with nonlinear decision rule

Next, we consider a randomized trial with nonlinear decision rule and $\pi = 0.5$. We use the same model (8) but with $U(X) = \log\{(X_1 + 2)(X_2 + 2)\}$ and $C(X) = (X_3 - 1)^2 - 2X_4 + 2 - c$, where c is chosen from the set $\{0, 1, 2\}$, denoted as Scenarios 4 to 6, respectively. Since $(X_3 - 1)^2 - 2X_4 + 2 \geq 0$, Scenario 4 is from the null hypothesis. It can be shown that Scenarios 5 and 6 are all from the alternative hypothesis. And as c increases, the alternative hypothesis moves further away from the null. Under the alternatives, the corresponding ODRs can be similarly obtained from their contrast functions by $I\{C(X) > 0\}$. The values of the obtained ODR ($V(\beta_0)$) and the naive decision rule (V_1) are also calculated and reported in Table 2.

For each setting, we conduct 500 replications with sample size $n = 1000$, and apply the proposed testing procedure with the polynomial basis of the covariates as $\phi_X(\cdot)$. The degree ($\in \{1, 2, 3\}$) for the polynomial basis are selected based on five-fold cross validation to maximum the value function. Here, the augmented term is fitted with the selected basis function, and the decision rule are searched within a class of $I\{\phi_X(X)^\top \beta > 0\}$. We apply three different optimization algorithms (the 'NM', the 'SA', and the 'GA' method, respectively), and report the simulation results in Table 2 for comparison, including the bias between the estimated value under the estimated ODR ($\widehat{V}(\widehat{\beta})$) and the true value, the type I error under the null ('ERR.'), and the power under the alternative ('POW.'). In addition, we plot the bias of the estimated value, and the type I error / power of the proposed test statistics under three optimization methods with $h = 0.5$ and $\gamma = 1$ for illustration as Figure 2.

Table 2. Simulation results of the proposed test with nonlinear decision rule.

SCEN.	RESULTS	$h = 0.5$		$h = 1$	
		$\gamma = 1$	$\gamma = 2$	$\gamma = 1$	$\gamma = 2$
4	V_1	3.98	7.31	4.63	7.96
	$V(\beta_0)$	3.98	7.31	4.63	7.96
	BIAS(NM)	0.01	0.03	0.01	0.03
	BIAS(SA)	0.01	0.03	0.01	0.03
	BIAS(GA)	0.01	0.03	0.01	0.03
	ERR.(NM)	5.4%	5.0%	4.8%	5.0%
	ERR.(SA)	5.4%	5.2%	5.2%	5.2%
	ERR.(GA)	5.4%	5.0%	5.0%	5.2%
5	V_1	2.98	5.31	3.63	5.96
	$V(\beta_0)$	3.01	5.38	3.66	6.03
	BIAS(NM)	0.02	0.05	0.02	0.05
	BIAS(SA)	0.01	0.01	0.01	0.01
	BIAS(GA)	0.01	0.01	0.01	0.01
	POW.(NM)	5.0%	5.2%	4.8%	5.2%
	POW.(SA)	7.8%	7.4%	7.4%	7.2%
	POW.(GA)	10.2%	10.4%	8.6%	9.6%
6	V_1	1.98	3.31	2.63	3.96
	$V(\beta_0)$	2.17	3.67	2.82	4.34
	BIAS(NM)	0.18	0.34	0.18	0.36
	BIAS(SA)	0.04	0.05	0.04	0.07
	BIAS(GA)	0.01	0.02	0.01	0.01
	POW.(NM)	6.0%	6.0%	5.6%	5.8%
	POW.(SA)	64.8%	75.6%	46.6%	66.0%
	POW.(GA)	80.0%	86.6%	63.4%	79.6%

Based on the results, we observe that type I errors of the proposed test are all close to the nominal level. Among all three optimization algorithms, the Genetic Algorithm performs the best while the Nelder-Mead Method is the worst. Under the Genetic Algorithm, the estimated values for the estimated ODRs are all close to the true and the power increase as c increases. In particular, when $c = 1$, the alternative is very close to the null (the value difference $V(\beta_0) - V_1 \leq 0.07$), so the power is slightly larger than the nominal level. On the other hand, when $c = 2$, there is a relatively large value difference and the power achieves 86.6% with $h = 0.5$ and $\gamma = 2$. It is shown in both Table 2 and Figure 2 that the estimated values by the 'NM' and the 'SA' methods are much smaller compared to the true values, which indicates that both methods fail to find the ODR under the nonlinear setting. The poor performance of the Nelder-Mead Method may be due to the curse of dimensionality when using high order basis function to search the

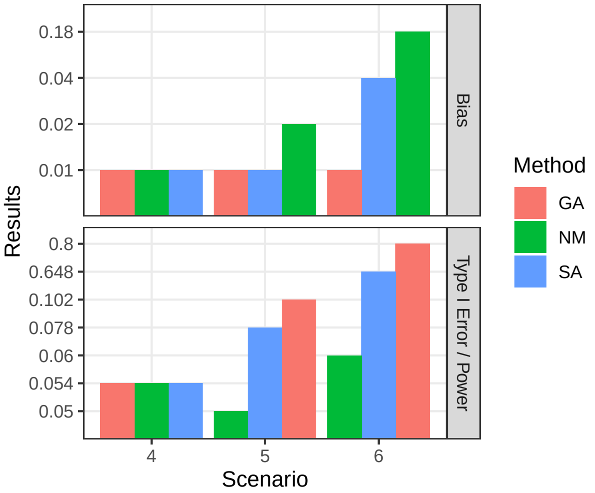

Figure 2. The bias between the estimated value under the estimated ODR and the true value, and the type I error / power of the proposed test statistics, under three optimization methods with $h = 0.5$ and $\gamma = 1$: the red, green, and blue bars present the results for the Genetic Algorithm, the Nelder-Mead Method, and the Simulated Annealing, respectively. Note the y-axis in the figure is in a relative scale.

decision rule; while for the Simulated Annealing, the reason can be the existence of multiple local optimums since our underlying true ODR is nonlinear.

6.3. Power and sample size calculation

In this section, we examine the performance of the proposed sample size calculation method for testing the existence of an ODR. Here, we consider the same model as in Section 6.2, i.e.

$$\begin{aligned} Y =& h\log\{(X_1+2)(X_2+2)\} \\ &+ \gamma A\{(X_3-1)^2 - 2X_4 + 2 - c\} + E, \end{aligned}$$

where c is chosen as 2 or 2.5, h is chosen as 0, 0.1, 0.2 or 0.4, and $\gamma = 1$ or 2.

Based on a pilot data, which is generated from the true model above, we first obtain the estimated value difference $\widehat{\delta}_a$, and the variance estimates $\widehat{\sigma}_0^2$ and $\widehat{\sigma}_\phi^2$. Then, we compute the required sample size $\widehat{n}^\star$ based on equation (6). We choose $\alpha = 0.05$ and $\beta = 0.1$, and apply the same testing procedure as described in Section 6.2 through the Genetic Algorithm with the degree of the polynomial basis set as 2. The sample size results are shown in Table 3.

It can be seen from the results that as h increases, the required sample size increases; while as γ increases, the required sample size decreases. This is expected since when h is larger of γ is smaller, the treatment-covariate interaction effect is weaker compared with the baseline mean effect. In addition, the case with $c = 2$ requires much larger sample size than $c = 2.5$ since the former is more closer to the null. Based on the estimated sample size, we conduct 500 replications for each setting to compute the empirical power of the proposed test. The powers are also reported in Table 3. We observe that the empirical powers are generally close to the nominal level of 90%, which shows the validity of our proposed sample size method.

Table 3. The result of the theoretical sample size and the related empirical power.

Scen.	$h =$	$\gamma = 1$		$\gamma = 2$	
		$n^\star$	Pow.	$n^\star$	Pow.
$c = 2.5$	0	248	92.1%	235	89.1%
	0.1	270	94.6%	249	92.3%
	0.2	289	90.8%	253	87.2%
	0.4	336	92.3%	275	92.6%
$c = 2$	0	956	95.5%	942	91.8%
	0.1	1073	93.2%	982	91.1%
	0.2	1162	92.8%	1019	89.3%
	0.4	1378	91.3%	1114	89.1%

7. A Data Application

Here, we illustrate our method with application in a healthcare study: a schizophrenia data. Due to privacy, we do not provide the source data, but a full data description can be found in Tarrier et al. (2004), where they specifically conducted a randomized trial with an 18 month follow-up period to examine the effectiveness of cognitive-behavioral therapy for schizophrenia. The Positive and Negative Syndrome Scale (PANSS, (Kay et al., 1987)) was employed to measure the symptoms of individuals. Patients were randomized to three treatment options, including the cognitive-behavioral therapy plus treatment as usual (CBT) as treatment group 1 with 44 subjects, supportive counseling plus treatment as usual (SC) as treatment group 2 with 41 subjects and treatment as usual (TAU) as treatment group 0 with 70 subjects. The reduction of PANSS score at 18th month's visit was set as patient's outcome Y. Two major information of subjects related to the schizophrenia are used as the covariates $X = (X_1, X_2)$, where X_1 is the log duration of untreated psychosis at baseline, and X_2 is the PANSS score at the baseline visit. The proposed test is conducted for comparing two treatments at a time: CBT vs. TAU, SC vs. TAU, and CBT vs. SC.

First, we compute treatment-specific means for the three treatment groups, and get $\widehat{\mu}_{TAU} = 21.96$, $\widehat{\mu}_{CBT} = 27.34$ and $\widehat{\mu}_{SC} = 28.76$. So, on average, treatment SC is the best while treatment TAU is the worst. Moreover, the mean outcomes of SC and CBT are comparable, which are much

Table 4. Data analysis on the schizophrenia study.

Test Pair	CBT vs. TAU	SC vs. TAU	CBT vs. SC
superior	CBT	SC	SC
$\widehat{V}_1$	27.34	28.76	28.76
$\widehat{V}(\widehat{\beta})$	30.35	33.06	34.70
P-value	0.190	0.125	0.039

larger than that of TAU. In our pairwise comparison, we denote the treatment with the larger mean outcome as treatment 1 (i.e. the superior treatment in Table 4) and the other as treatment 0. Then, the estimated value $\widehat{V}_1$ under the naive decision rule $d(X) \equiv 1$ is the same as the mean outcome of treatment 1. We also compute the estimated value $\widehat{V}(\widehat{\beta})$ under the estimated ODR. We apply the proposed test for checking whether there is an ODR that is better than the naive decision rule $d(X) \equiv 1$ in all three pairwise comparisons. Here, we use the random forest to fit the augmented term, and use the tensor-product B-splines as the basis function, where the degree and knots $(\in \{3, 4, 5, 6\})$ of B-splines are chosen via five-fold cross validation to maximum the value function as described in Section 6.2. The decision rule is searched within the class of $I\{\phi_X(X)^\top \beta > 0\}$ through the Genetic Algorithm. We report the corresponding p-values for each test in Table 4.

Based on the results, we observe the following findings. First, when comparing CBT vs. TAU and SC vs. TAU, the p-values are not significant. This implies that there doesn't exist an ODR that is better than assigning all individuals to CBT or SC when comparing with TAU. This finding is consistent with the literature because CBT and SC are known to be better treatments than TAU for all individuals. Second, when comparing CBT and SC, the p-value is significant (< 0.05). This implies that SC is not a uniformly better treatment than CBT and there exists an optimal treatment that is better than assigning all individuals to SC. Under such a situation, the estimated value difference is $\widehat{V}(\widehat{\beta}) - \widehat{V}_1 = 34.70 - 28.76 = 5.94$.

In addition, in Figure 3, we plot the treatment assignment given by the estimated ODR $\widehat{d}(X)$ when comparing SC and CBT. From the figure, we can see that individuals with median log durations of untreated psychosis at baseline and median PANSS score at baseline will benefit more from the CBT treatment (a total of 20 assignments under $\widehat{\beta}$). However, patients with extreme low or high log durations and PANSS score should receive the SC treatment (a total of 65 assignments under $\widehat{\beta}$). Finally, based on the data for comparing CBT and SC, we design an A/B test with the sample size calculation for testing the existence of an ODR. Here, we consider one-sided test with $\alpha = 0.05$ and a desired power at least $1 - \beta = 90\%$. Based on the pilot study, we can compute the standard error estimates $\widehat{\sigma}_0 = 31.17$ and $\widehat{\sigma}_\phi = 38.95$, and the value difference $\widehat{\delta}_a = \widehat{V}(\widehat{\beta}) - \widehat{V}_1 = 5.94$. Therefore, the required sample size for detecting such a value difference can be calculated by equation (7), which is $\widehat{n}^\star = 290$.

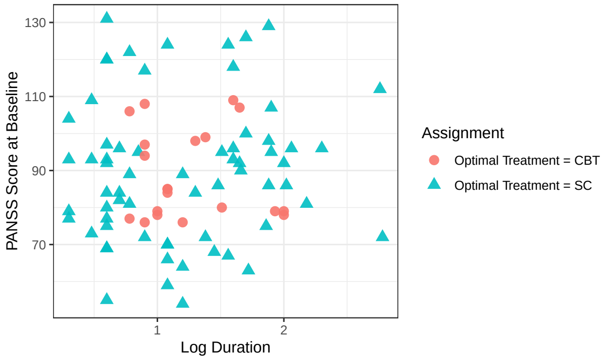

Figure 3. The plot for the treatment assignment given by the estimated optimal decision rule.

8. Discussion

We conclude our paper by following possible extensions. First, the proposed test only consider randomized trials. We may follow the studies of Kallus (2018) and Ozery-Flato et al. (2018) to use the generative adversarial networks to extend the proposed test for the observational studies. Second, the Assumption (1) usually holds when covariates X contain some continuous variables. However, we may extend the proposed test to incorporate the nonregular case by using some sample splitting method. Third, we may consider a more general test with multiple treatment options or continuous treatment. Specifically, for cases with multiple treatments where A could take multiple values, our framework can be easily extended by replacing the current decision class, i.e. the indicator function $d(X) = I\{C(X) > 0\}$, with a general classification function that maps features X to a treatment $a \in A$. Athey & Wager (2017) addressed the theoretical properties of the AIWP estimator under multiple treatments and restricted policy class in decision trees. Finally, we can extend the proposed test for detecting the existence of a dynamic ODR in multiple stage studies.

Acknowledgements

The authors HC and RS acknowledge funding from the National Science Foundation (NSF) under the Division of Mathematical Sciences (DMS) award No. 1555244; and the authors WL and RS acknowledge support from the National Institutes of Health (NIH) under the National Cancer Institute (NCI) grant agreement No. 142538. We also thank the anonymous reviewers for their valuable feedback.

References

Athey, S. and Wager, S. Efficient policy learning. *arXiv preprint arXiv:1702.02896*, 2017.

Dawson, R. and Lavori, P. W. Sample size calculations for evaluating treatment policies in multi-stage designs. *Clinical Trials*, 7:643–652, 2010.

Fan, C., Lu, W., Song, R., and Zhou, Y. Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1565–1582, 2017.

Kallus, N. Deepmatch: Balancing deep covariate representations for causal inference using adversarial training. *arXiv preprint arXiv:1802.05664*, 2018.

Kang, S., Lu, W., and Song, R. Subgroup detection and sample size calculation with proportional hazards regression for survival data. *Statistics in medicine*, 36:4646–4659, 2017.

Kay, S. R., Fiszbein, A., and Opler, L. A. The positive and negative syndrome scale (panss) for schizophrenia. *Schizophrenia bulletin*, 13(2):261–276, 1987.

Laber, E. B., Zhao, Y.-Q., Regh, T., Davidian, M., Tsiatis, A., Stanford, J. B., Zeng, D., Song, R., and Kosorok, M. R. Using pilot data to size a two-arm randomized trial to find a nearly optimal personalized treatment strategy. *Statistics in medicine*, 35:1245–1256, 2016.

Liang, S., Lu, W., Song, R., and Wang, L. Sparse concordance-assisted learning for optimal treatment decision. *The Journal of Machine Learning Research*, 18(1): 7375–7400, 2017.

Luedtke, A. R. and Van Der Laan, M. J. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713, 2016.

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

Murphy, S. A. An experimental design for the development of adaptive treatment strategies. *Statistics in medicine*, 24:1455–1481, 2005.

Ozery-Flato, M., Thodoroff, P., and El-Hay, T. Adversarial balancing for causal inference. *arXiv preprint arXiv:1810.07406*, 2018.

Robins, J. M. Optimal structural nested models for optimal sequential decisions. In In: Lin D.Y., H. P. (ed.), *Proceedings of the second seattle Symposium in Biostatistics*, pp. 189–326, New York, 2004. Springer.

Rubin, D. B. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 6:34–58, 1978.

Shi, C., Fan, A., Song, R., and Lu, W. High-dimensional a-learning for optimal dynamic treatment regimes. *Annals of statistics*, 46(3):925–957, 2018a.

Shi, C., Song, R., Lu, W., and Fu, B. Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):681–702, 2018b.

Tarrier, N., Lewis, S., Haddock, G., Bentall, R., Drake, R., Kinderman, P., Kingdon, D., Siddle, R., Everitt, J., and Leadley, K. Cognitive–behavioural therapy in first-episode and early schizophrenia: 18-month follow-up of a randomised controlled trial. *The British Journal of Psychiatry*, 184:231–239, 2004.

van der Laan, M. J. and Luedtke, A. R. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of causal inference*, 3(1):61–95, 2015.

Watkins, C. and Dayan, P. Q-learning. *Machine Learning*, 8:279–292, 1992.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. A robust method for estimating optimal treatment regimes. *Biometrics*, 68:1010–1018, 2012.

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

Zhao, Y., Kosorok, M. R., and Zeng, D. Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28:3294–3315, 2009.

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

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