Published in final edited form as:
Stat Methods Med Res. 2013 October ; 22(5): 493–504. doi:10.1177/0962280211428383.

Variable Selection for Optimal Treatment Decision

Wenbin Lu[1], **Hao Helen Zhang**[1,3], and **Donglin Zeng**[2]
[1]Department of Statistics, North Carolina State University
[2]Department of Biostatistics, University of North Carolina at Chapel Hill
[3]Department of Mathematics, University of Arizona

Abstract

In decision-making on optimal treatment strategies, it is of great importance to identify variables that are involved in the decision rule, i.e. those interacting with the treatment. Effective variable selection helps to improve the prediction accuracy and enhance the interpretability of the decision rule. We propose a new penalized regression framework which can simultaneously estimate the optimal treatment strategy and identify important variables. The advantages of the new approach include: (i) it does not require the estimation of the baseline mean function of the response, which greatly improves the robustness of the estimator; (ii) the convenient loss-based framework makes it easier to adopt shrinkage methods for variable selection, which greatly facilitates implementation and statistical inferences for the estimator. The new procedure can be easily implemented by existing state-of-art software packages like LARS. Theoretical properties of the new estimator are studied. Its empirical performance is evaluated using simulation studies and further illustrated with an application to an AIDS clinical trial.

Keywords

A-learning; Optimal treatment strategy; Personalized drugs; Shrinkage method; Variable selection

1 Introduction

An optimal treatment strategy is a set of treatment decision rules tailored for individuals, to maximize long-term clinical outcomes and reduce the risk of over- or under- treatment for individual patients. Personalized medicine takes into account individual heterogeneity in clinical, genetic, social, environmental, behavior characteristics, and so on, and has gained much attention in many disease studies like cancer. As in the lymphoma study [1], the patient subtypes identified by tumor gene expression profiles showed different responses to CHOP and RCHOP treatments and thus individuals' tumor subtype should be considered for treatment assignment in addition to other clinical information.

Let Y denote the real-valued response, $A \in \mathcal{A}$ denote the treatment received by the patient, where $\mathcal{A}$ is the set of available treatment methods, and $\mathbf{X} \in \mathcal{X} \subset R^p$ denote the baseline covariates such as clinical measurements and medical history, which can be used for treatment assignment. We focus on a simple two-treatment regime, $\mathcal{A} = \{0, 1\}$: 0 is for the control/standard treatment and 1 for the new treatment. A treatment regime is a mapping g: $\mathcal{X} \to \{0, 1\}$. The optimal treatment regime is a decision rule g^{opt} that assigns the best treatment to a patient based on the observed covariates $\mathbf{X}$. In practice, we collect data (Y_i,

Corresponding Author: Hao Helen Zhang, Department of Statistics, North Carolina State University, Raleigh, NC, 27695, USA. hzhang@stat.ncsu.edu.

A_i, $\mathbf{X}_i$), $i = 1, \cdots, n$ from a randomized clinical trial and the goal is to estimate g^{opt} from data. Estimation of an optimal treatment strategy can be challenging, as the underlying relationship between the response and relevant prognostic factors may be quite complicated. For randomized studies, the potential outcome model [2] provides an effective tool for analyzing the causal effect of time-independent treatment. [3] and [4] extend the potential outcome model for observational studies. Since then, there are a large number of works on optimal dynamic treatment regimes using Q- or A-learning algorithms, including [5], [6], [7], [8], [9], [10] and [11].

With rapid advances in technology and combinations of diverse data sources, a very large number of prognostic factors such as clinical measurements, tumor pathology, and genetic information are available for estimating the optimal treatment strategy. However, many of them might not be related to the disease or the treatment assignment. As such, there is a lot of redundant information, and variable selection becomes necessary and plays an important role for making an optimal decision rule that is interpretable and efficient. In this paper, we focus on variable selection for optimal treatment strategies. In the context of linear regression models, various methods have been developed for selecting variables that are important for prediction. These methods often lead to a better predictive model in practice. Recent developments in variable selection include shrinkage regression methods such as least absolute shrinkage and selection operator (LASSO) penalty [12], smoothly clipped absolute deviation (SCAD) penalty ([13], [14]), and adaptive LASSO penalty ([15], [16], [17]). The SCAD and adaptive LASSO are shown to be oracle when the tuning parameter is properly chosen.

However, there is scarce research on variable selection for optimal decision making on treatment strategies. Compared to standard regression problems, the main goal here is to identify important variables involved in treatment decision rules. Recently, in the framework of *Q*-learning, [18] developed a two-step procedure which estimates the conditional means first and then derived the treatment rule based on estimated conditional means, and l_1 penalty was employed for variable selection. The paper [19] proposed a new ranking method to variable selection in this context, in which they discussed the concepts of *predictive* variables and *prescriptive* variables: the former refers to variables which reduce the variability and increase the accuracy of the estimator, and the latter refers to variables which help prescribe the optimal action. In this article, we propose a new loss-based framework to estimate the optimal treatment strategy. The new method is equipped with a convenient quadratic loss, which greatly facilitates the variable selection process by incorporating shrinkage penalties in the estimation. Moreover, the new loss function corresponds to a form of A-learning, therefore the estimation does not require a correct specification of the baseline mean function and is robust. The remainder of the paper is organized as follows. In Section 2 we introduce the new loss function and propose the penalized regression framework. We also study large-sample properties of the estimator and present a computational algorithm. We demonstrate simulation results in Section 3 and apply the method to data from an AIDS study in Section 4. Section 5 contains some discussions. All the proofs are relegated to the Appendix Section.

2 Method

2.1 New Estimation Framework

We first give a brief review on the potential outcome. Based on [3], the potential outcome $Y^*(a)$ is the outcome value that would result if a patient were assigned to the treatment $a \in \mathcal{A}$. For a patient with covariates $\mathbf{X} = \mathbf{x}$, the goal is to find the optimal treatment regime that maximizes the expected outcome, i.e. $g^{opt}(\mathbf{X}) = \arg\max_{g \in \mathcal{G}} E[Y^*(g(\mathbf{X}))]$, where $\mathcal{G}$ denote

the set of all possible treatment regimes. Following [3], two assumptions are typically required for computing the expectation of the potential outcome:

(C1) The outcome of one patient is not influenced by the treatment allocation of other subjects. Or equivalently, $Y = I(A = 0)\,Y^*(0) + I(A = 1)\,Y^*(1)$. This is also known as consistency assumption;

(C2) The treatment assignment for an individual is independent of the potential outcomes conditional on **X**. In other words, $A \perp \{Y^*(a)\}_{a \in \mathcal{A}} | \mathbf{X}$. This essentially assumes no unmeasured confounders.

Under these two assumptions, it is easy to show that

$$E[Y^*(g(\mathbf{X}))] = E_{\mathbf{X}}[E(Y|\mathbf{X}, A=1)g(\mathbf{X}) + E(Y|A=0, \mathbf{X})\{1 - g(\mathbf{X})\}].$$

Therefore g^{opt} can be expressed as

$$g^{opt}(\mathbf{X}) = I\{E(Y|\mathbf{X}, A=1) - E(Y|\mathbf{X}, A=0) > 0\}.$$

Consider the following general model $E(Y|\mathbf{X}, A) = h_0(\mathbf{X}) + Af(\mathbf{X})$. Here $h_0(\mathbf{X})$ presents the baseline effects of **X** on Y and $f(\mathbf{X})$ describes the combination of marginal treatment effect and its interaction effects with covariates. It is easy to show that $E(Y|\mathbf{X} = \mathbf{x}, A = 1) - E(Y|\mathbf{X} = \mathbf{x}, A = 0) = f(\mathbf{x})$. Therefore, for a patient with covariates $\mathbf{X} = \mathbf{x}$, the optimal treatment is $g^{opt}(\mathbf{x}) = I\{f(\mathbf{x}) > 0\}$. Let $\pi(\mathbf{x})$ denote the propensity score, i.e. $\pi(\mathbf{x}) = P(A = 1|\mathbf{X} = \mathbf{x})$. For consistent estimation of the optimal treatment rule, it is usually assumed

(C3) $0 < \pi(\mathbf{x}) < 1$, $\forall \mathbf{x} \in \mathcal{X}$ and $E[\pi(\mathbf{X})(1 - \pi(\mathbf{X}))\mathbf{X}\mathbf{X}^T]$ is finite and nondegenerate. In randomized studies, $\pi(\mathbf{x})$ is actually known and it is the treatment assignment probability pre-determined by design. Throughout the paper, we assume that conditions (C1)–(C3) hold.

To simplify the optimal treatment strategy, we consider the linear form for the inter-action effect, also known as the contrast, i.e.

$$E(Y|\mathbf{X}, A) = h_0(\mathbf{X}) + A(\beta^T \tilde{\mathbf{X}}), \quad (2.1)$$

where $\tilde{\mathbf{X}} = (1, \mathbf{X}^T)^T$ and $\beta = (\beta_1, \cdots, \beta_{p+1})^T$. Let $\beta_0 = (\beta_{10}, \cdots, \beta_{p+1,0})^T$ denote the true parameters of β in (2.1). The primary interest is to estimate the contrast or inter-action function $\beta^T \tilde{\mathbf{X}}$, but not the baseline $h_0(\mathbf{X})$. Given the observations $\{Y_i, \mathbf{X}_i, A_i; i = 1, \cdots, n\}$, we propose to minimize the following loss function

$$L_{n,\varphi}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Big[Y_i - \varphi(\mathbf{X}_i) - \beta^T \tilde{\mathbf{X}}_i\{A_i - \pi(\mathbf{X}_i)\}\Big]^2,$$

where $\phi(\mathbf{x})$ is an arbitrary function. It is interesting to note that, when taking the derivative of $L_{n,\phi}(\beta)$ with respect to β, the resulting estimating equation has a form of A-learning [7]. Therefore $L_{n,\phi}(\beta)$ provides a loss function in the framework of A-learning. In practice, we suggest to use a parametric form for $\phi(\cdot)$ and minimizes

$$L_n(\beta,\gamma)=\frac{1}{n}\sum_{i=1}^{n}\Big[Y_i-\varphi(\mathbf{X}_i;\gamma)-\beta^T\tilde{\mathbf{X}}_i\{A_i-\pi(\mathbf{X}_i)\}\Big]^2. \quad (2.2)$$

Two feasible choices of ϕ are: the constant model $\phi(\mathbf{x};\gamma)=\gamma$ and the linear model $\phi(\mathbf{x};\gamma)=\gamma^T\tilde{\mathbf{x}}$. Denote the solution to (2.2) as $(\hat{\gamma}^T,\hat{\beta}^T)^T$. Asymptotic properties of $\tilde{\beta}$ are studied in the next session. Essentially, if $\pi(\mathbf{x})$ is known as in randomized studies, we can show that $\tilde{\beta}$ is a consistent estimator for β_0, regardless of the choice of $\phi(\mathbf{x};\gamma)$. This robustness is a desired property for both model estimation and variable selection.

The optimal decision rule only depends on the treatment and treatment-covariates interaction effects $\beta^T\tilde{\mathbf{X}}$, and so the important variables are those with nonzero coefficients. The convenient loss form in (2.2) makes it easy to adopt shrinkage penalties for variable selection. In order to select important prescriptive variables, we propose to solve

$$\min_{\beta} L_n(\beta,\tilde{\gamma})+\lambda_n\sum_{j=1}^{p+1}J(|\beta_j|), \quad (2.3)$$

where λ_n is a tuning parameter and J is a shrinkage penalty. There are plenty of choices for J, such as SCAD, adaptive LASSO, and minimax concave penalty [20]. In this article, we employ the adaptive lasso penalty for variable selection and solves

$$\min_{\beta} L_n(\beta,\tilde{\gamma})+\lambda_n\sum_{j=1}^{p+1}w_j|\beta_j|. \quad (2.4)$$

As pointed out in [15], the values of weights w_j's are crucial to effective selection in practice. In general, large penalties are desired for unimportant covariates and small penalties for important ones. In this work, we use $w_j^{-1}=|\tilde{\beta}_j|$, $j=1,\cdots,p+1$, where $\tilde{\beta}=(\tilde{\beta}_1,\cdots,\tilde{\beta}_{p+1})^T$. Denote the solution to (2.4) as $\hat{\beta}=(\hat{\beta}_1,\cdots,\hat{\beta}_{p+1})^T$. In the next session, we study the asymptotic properties of $\tilde{\beta}$ and $\hat{\beta}$.

2.2 Asymptotic properties

We study asymptotic properties of $\tilde{\beta}$ and $\hat{\beta}$. Proofs of theorems are given in the Appendix. Let $\mathcal{S}=\{j:\beta_{j0}\neq 0, j=1,\cdots,p+1\}$ denote the true set of important variables for the optimal decision. Without loss of generality, write $\beta_0=(\beta_{0,\mathcal{S}}^T,\beta_{0,\mathcal{S}^c}^T)^T$. Let $\hat{\mathcal{S}}=\{j:\hat{\beta}_j\neq 0, j=1,\cdots,p+1\}$ be the set of selected important variables. We have

Theorem 1—If regularity conditions (A1)–(A4) in the Appendix hold, the linear treatment-covariates interaction term in model (2.1) is correctly specified and $\pi(\mathbf{x})$ is known, then we have $\sqrt{n}(\tilde{\beta}-\beta_0)\to N(0,V)$ as $n\to\infty$, where V is given in the Appendix.

Theorem 2—Assume that $\sqrt{n}\lambda_n\to 0$ and $n\lambda_n\to\infty$. Then, under the conditions of Theorem 1, we have: (i) (selection consistency) $P(\hat{\mathcal{S}}=\mathcal{S})\to 1$ as $n\to\infty$; (ii) (asymptotic normality) $\sqrt{n}(\hat{\beta}_{\mathcal{S}}-\beta_{0,\mathcal{S}})\to N(0,\sum)$ as $n\to\infty$.

Remark—In observational studies, the propensity score $\pi(\mathbf{x})$ is usually not known in advance. A parametric model $\pi(\mathbf{x};\alpha)$, such as logistic regression can be used to estimate $\pi(\mathbf{x})$. As long as the parametric model is correctly specified, the parameter α can be

consistently estimated by the maximum likelihood estimator $\hat{\alpha}$. By replacing $\pi(\mathbf{X}_i)$ in (2.2) and (2.4) by $\pi(\mathbf{X}_i; \hat{\alpha})$, similar results as given in Theorems 1 and 2 can also be established for the resulting estimators.

2.3 Computation and Tuning

Let $\mathbb{Y}_i = Y_i - \phi(\mathbf{X}_i; \tilde{\gamma})$ and $\mathbb{X}_i = \tilde{\mathbf{X}}_i\{A_i - \pi(\mathbf{X}_i)\}$. The loss function $L_n(\tilde{\gamma}, \beta) = (1/n)\sum_{i=1}^{n}(\mathbb{Y}_i - \beta^T\mathbb{X}_i)^2$ has a standard quadratic form. The LARS algorithm [21] can be adapted to compute the entire solution path of (2.4). The following gives the algorithm:

- Step 1: Minimize (2.2). Denote the minimizers as $(\tilde{\beta}, \tilde{\gamma})$.
- Step 2: Construct the weights $w_j^{-1} = |\tilde{\beta}_j|$ for $j = 1, \cdots, p+1$.
- Step 3: Compute $\mathbb{Y}_i$ and $\mathbb{X}_i$, $i = 1, \cdots, n$. Solve the penalized least squared estimation in (2.4) using the LARS to obtain the whole solution path of β. For a fixed λ, denote the solution by $\hat{\beta}(\lambda)$.

We use a BIC-type criteria [15] to select the tuning parameter λ. Specifically, we minimize $L_n(\hat{\beta}(\lambda), \tilde{\gamma})/L_n(\tilde{\beta}, \tilde{\gamma}) + d(\lambda)\log(n)/n$ to obtain an estimator of λ, where $d(\lambda)$ is the number of non-zeros in $\hat{\beta}(\lambda)$.

3 Simulations

We evaluate the empirical performance of the new method in terms of estimation accuracy and variable selection under various settings. Assume the randomized trial with $\pi = 0.5$. We consider different function forms for the baseline h_0, including a simple linear form, a complex nonlinear form, and a function containing interactions between the covariates. Also, we allow important variables in the baseline to be different from those in the contrast function. Define $\tilde{\mathbf{X}} = (1, \mathbf{X}^T)^T$ and $\mathbf{0}_d$ for the zero vector of length d.

3.1 Low Dimension Examples

We consider the following three models with $p = 10$,

- Model I: $Y = 1 + \gamma_1^T\mathbf{X} + A\beta^T\tilde{\mathbf{X}} + \varepsilon$, $\mathbf{X} = (X_1, \cdots, X_{10})^T$ are multivariate normal with mean 0, variance 1, and the correlation $\mathrm{Corr}(X_j, X_k) = 0.5^{|j-k|}$. The error term $\varepsilon \sim N(0, 0.5^2)$. The coefficients $\gamma_1 = (1, -1, \mathbf{0}_8)^T$ and $\beta = (1, 1, \mathbf{0}_7, -0.9, 0.8)$.
- Model II: $Y = 1 + 0.5(\gamma_1^T\mathbf{X})(\gamma_2^T\mathbf{X}) + A\beta^T\tilde{\mathbf{X}} + \varepsilon$, $\gamma_1 = (1, -1, \mathbf{0}_8)^T$, $\gamma_2 = (1, \mathbf{0}_2, -1, \mathbf{0}_5, 1)^T$, and $\mathbf{X}$ and ε are same as Model I.
- Model III: $Y = 1 + 0.5\sin(\pi\gamma_1^T\mathbf{X}) + 0.25(1 + \gamma_2^T\mathbf{X})^2 + A\beta^T\tilde{\mathbf{X}} + \varepsilon$, γ_1 and γ_2 are same as in Model II, and other parameters are the same as Model I.

To evaluate the model estimation performance of the estimator, we report its mean squared error $\mathrm{MSE} = \|\hat{\beta} - \beta_0\|^2$. The average MSE over 500 realizations are reported and so are the corresponding standard errors (in parentheses). To evaluate variable selection performance, we summarize the number of correct zero coefficients identified (denoted as "Corr0"), the number of nonzero effects incorrectly identified as zero (denoted as "Incorr0"), and the proportion of selecting exactly the correct model (denoted as "Exact") among 500 replications. We also report the frequency of being selected for each variable. To evaluate the accuracy of a treatment assignment rule $I(\beta^T\tilde{X} > 0)$, we calculate the average percentage of making correct decisions (PCD) over 500 simulation runs, i.e. mean of

$1-\sum_{i=0}^{n}|I(\beta^T\tilde{X}_i>0)-I(\beta_0^T\tilde{X}_i>0)|/n$. For comparison, we report the PCDs of both the unpenalized estimator $\tilde{\beta}$ (denoted as "Unpen.") and the penalized estimator $\hat{\beta}$ (denoted as "Penalized").

We compare two cases which correspond to different working models for ϕ:

- Case 1: Set $\phi(\mathbf{X};\gamma)\equiv\gamma$, a constant model.
- Case 2: Set $\phi(\mathbf{X};\gamma)\equiv\gamma^T\tilde{\mathbf{X}}$, a linear model.

Table 1 summarizes the estimation, selection, and PCD results for Model I under two cases. We consider three different sample sizes: n = 100, 200, 400. For each case, both the MSE and classification error improve as the sample increases, which is expected. The proposed method gives an overall good performance in variable selection, especially when the sample size is large. For example, when n = 400, the frequencies of selecting the exact true model are respectively 70.6% and 91.0% in Case 1 and Case 2. The estimator in Case 2 consistently shows better performance than that in Case 1, in terms of both model estimation and variable selection. With regard to the PCD, the fit in Case 2 again yields higher accuracy than Case 1. Furthermore, the penalized estimator overall gives smaller PCD than the unpenalized estimator, except in Case 1 when the sample size is small. From Table 2, we observe that the new procedure is very effective in retaining important variables: intercept, X_1, X_9, and X_{10} in the model and removing noise variables from the model, especially when the sample size is moderately large.

Tables 3 to 6 summarizes the estimation, selection, and PCD results for Models II and III. Overall, the new procedure performs well for variable selection, and the penalized estimator produces smaller PCDs than the unpenalized estimator. In both models, the fit in Case 2 gives better performance than Case 1 with regard to model estimation, variable selection and PCD. These simulation results suggest that a posited model with a rich structure generally works better than a simple model.

3.2 Large Dimension Examples

We now increase the input dimension to p = 50 and check the performance of the new procedure under larger dimensional settings. We consider Model IV and Model V,

- Model IV: $Y=1+0.5(\gamma_1^T\mathbf{X})(\gamma_2^T\mathbf{X})+A\beta^T\tilde{\mathbf{X}}+\varepsilon$, $\mathbf{X}=(X_1,\cdots,X_{50})^T$ are multivariate normal with mean 0, variance 1, and the correlation $\mathrm{Corr}(X_j, X_k) = 0.5^{|j-k|}$, $\gamma_1=(1,-1,\mathbf{0}_{48})^T$, $\gamma_2=(1,\mathbf{0}_2,-1,\mathbf{0}_{45},1)^T$ and $\beta=(1,1,\mathbf{0}_{46},-0.9,0.8)^T$. Other settings are the same as in Model I.
- Model V: $Y=1+0.5\sin(\pi\gamma_1^T\mathbf{X})+0.25(1+\gamma_2^T\mathbf{X})^2+A\beta^T\tilde{\mathbf{X}}+\varepsilon$, all of the parameters and variable distributions are the same as in Model IV.

with n = 200, 400. Tables 7 and 8 summarize variable selection and estimation results respectively for each model. In these large dimensional settings, we observe the significant gain in PCD for the penalized estimator compared with the unpenalized estimator. Also the new procedure is effective in identifying important variables. The estimator in Case 2 generally works better than in Case 1 when the sample size is reasonably large.

4 Application to AIDS study (ACTG175)

We apply our method to data from AIDS Clinical Trials Group Protocol 175 (ACTG175), which contains 2139 HIV-infected subjects. In ACTG175, study subjects were randomized to four different treatment groups: zidovudine (ZDV) monotherapy, ZDV+didanosine (ddI), ZDV+zalcitabine, and ddI monotherapy [22]. As in [23] and [24], we chose the CD4 count

(cells/mm^3) at 20 ± 5 weeks post-baseline as the primary continuous response Y. Besides the treatment indicator, we included the same 12 baseline covariates as considered by [23] and [24] in our model, which consist of 5 continuous covariates: age (years), weight (kg), Karnofsky score (scale of 0–100), CD4 count (cells/mm^3) at baseline and CD8 count (cells/mm^3) at baseline, and 7 binary covariates: hemophilia (0=no, 1=yes), homosexual activity (0=no, 1=yes), history of intravenous drug use (0=no, 1=yes), race (0=white, 1=non-white), gender (0=female, 1=male), antiretroviral history (0=naive, 1=experienced) and symptomatic status (0=asymptomatic, 1=symptomatic). The goal of our study is to find the optimal treatment to maximize the expected CD4 count (cells/mm^3) at 20±5 weeks post-baseline. We fit model (2.1) with $\mathbf{X}$ being the 12 baseline covariates. For the treatment indicator A, we considered the following four analyses:

- Analysis 1: $A = 0$ for zidovudine (ZDV) monotherapy (532 subjects) vs. $A = 1$ for the other three treatments combined together (1607 subjects). Here $\pi(\mathbf{X}_i) \equiv 0.75$.
- Analysis 2: Consider the subset for patients receiving the treatment ZDV +didanosine (ddI) or ZDV+zalcitabine. $A = 0$ for ZDV+zalcitabine (524 subjects) vs. $A = 1$ for ZDV+didanosine (ddI) (522 subjects). Here $\pi(\mathbf{X}_i) \equiv 0.5$.
- Analysis 3: Consider the subset for patients receiving the treatment ZDV +didanosine (ddI) or ddI monotherapy. $A = 0$ for ddI monotherapy (561 subjects) vs. $A = 1$ for ZDV+didanosine (ddI) (522 subjects). Here $\pi(\mathbf{X}_i) \equiv 0.5$.
- Analysis 4: Consider the subset for patients receiving the treatment ZDV +zalcitabine or ddI monotherapy. $A = 0$ for ddI monotherapy (561 subjects) vs. $A = 1$ for ZDV+zalcitabine (524 subjects). Here $\pi(\mathbf{X}_i) \equiv 0.5$.

In our analysis, we assumed the linear model for $h(\mathbf{x}; \gamma)$. For analysis 1, the adaptive LASSO estimator $\hat{\beta} = (48.46, 0, \cdots, 0)^T$, implying that the other three treatments are uniformly better than zidovudine (ZDV) monotherapy and all the patients should be assigned to the other three treatments. For analysis 2, except for the intercept, the new estimator also selects two important covariates: age and homosexual activity (homo), and their corresponding estimates are −44.92, 2.52, and −21.31, respectively. Therefore, the optimal treatment rule is $I(-44.92 + 2.52 * \text{age} - 21.31 * \text{homo} > 0)$ for comparing treatments ZDV+didanosine (ddI) ($A = 1$) and treatment ZDV+zalcitabine ($A = 0$), or equivalently, for a patient with homo = 0, assign to treatment 1 if age > 17.8 and treatment 0 otherwise; while for a patient with homo = 1, assign to treatment 1 if age > 26.3 and treatment 0 otherwise. Note that the age of study subjects ranges from 12 to 70. According to the obtained optimal treatment rule, 978 out of 1046 patients (93.5%) in this subset should be assigned to treatment ZDV+didanosine (ddI). It is also noted that treatment ZDV +zalcitabine is more favorable to young patients with AIDS. For analysis 3, except for the intercept, the new estimator also selects three important covariates: age, CD4 count at baseline (CD40) and homo, and their corresponding estimates are 71.59, 1.07, −0.18, and −33.57, respectively. This leads to the optimal treatment rule $I(71.59 + 1.07 * \text{age} - 0.18 * \text{CD40} - 33.57 * \text{homo} > 0)$ when comparing treatments ZDV+didanosine (ddI) ($A = 1$) and ddI monotherapy ($A = 0$). According to the obtained optimal treatment rule, 878 out of 1083 patients (81.1%) in this subset should be assigned to treatment ZDV+didanosine (ddI). For analysis 4, the adaptive LASSO selects no covariates including intercept, i.e. $\hat{\beta} = \mathbf{0}$, suggesting that treatments ZDV+zalcitabine and ddI monotherapy are equally good for all patients in this subset.

5 Discussion

In this article, we propose a new loss-based estimation framework for estimating the optimal treatment strategy, which naturally leads to a penalized framework for variable selection and

sparse estimation. One desired property of the new approach is that it does not require the correct specification of the baseline function in order to produce consistent estimation and variable selection for the optimal treatment rule, as long as the interaction function form is correct. Numerical results suggest that the posited baseline model with a richer structure tends to improve the estimation efficiency.

This work focuses on a two-treatment setup. In practice, the number of treatment options can be more than two, and the optimal treatment amounts to selecting the treatment which produces the largest gain in the outcome. It would be interesting to extend this methodology to multiple-treatment settings. Variable selection is more complicated for multiple treatments, since the decision rule involves multiple comparisons among candidate treatments. Moreover, for dynamic treatment regimes, a sequence of decision rules are needed, one per time interval, throughout an individual's disease course. We study the variable selection problem at one decision point in this work, and plan to extend the new framework to dynamic treatment regimes.

Acknowledgments

We thank Professors Marie Davidian and Anastasios Tsiatis for inspiring discussions on the paper. We acknowledge support from National Institutes of Health RO1 CA140632, P01 CA142538, R01 CA085848, and National Science Foundation grant DMS-0645293.

References

1. Lenz G, et al. Stromal gene signatures in large-B-cell lymphomas. New England Journal of Medicine. 2008; 359(22):2313–2323. [PubMed: 19038878]
2. Neyman J. On the application of probability theory to agricultural experiments. Statistical Science. 1923; 5:465–480.
3. Rubin DB. Estimating causal effects of treatments in randomized and non-randomized studies. Journal of Educational Psychology. 1974; 66:688–701.
4. Rubin DB. Bayesian inference for causal effects: the role of randomization. Statistical Science. 1978; 6:34–58.
5. Watkins, CJCH. PhD thesis. Cambridge University; 1989. Learning from Delayed Rewards.
6. Watkins CJCH, Dayan P. Q-learning. Machine Learning. 1992; 8:279–292.
7. Murphy SA. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society, Series B. 2003; 65(2):331–366.
8. Murphy SA. An experimental design for the development of adaptive treatment strategies. Statistics in Medicine. 2005; 24:1455–1481. [PubMed: 15586395]
9. Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Statistics in Medicine. 2009; 28:3294–3315. [PubMed: 19750510]
10. Brinkley J, Tsiatis AA, Anstrom KJ. A generalized estimator of the attributable benefit of an optimal treatment regime. Biometrics. 2010; 66(2):512–522. [PubMed: 19508237]
11. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. Statistical Methods in Medical Research. 2010; 19:317–343. [PubMed: 19608604]
12. Tibshirani RJ. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, B. 1996; 58:267–288.
13. Fan J, Li R. Variable selection via non-concave penalized likelihood and its oracle properties. Journal of American Statistical Association. 2001; 96:1348–1360.
14. Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of American Statistical Association. 2004; 99:710–723.
15. Zou H. The adaptive lasso and its oracle properties. Journal of American Statistical Association. 2006; 101:1418–1429.

16. Zhang HH, Lu W. Adaptive-lasso for Cox's proportional hazards model. Biometrika. 2007; 94:691–703.
17. Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection via the LAD-LASSO. Journal of Business & Economics Statistics. 2007; 20:347–355.
18. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Annals of Statistics. 2010 to appear.
19. Gunter L, Zhu J, Murphy SA. Variable selection for qualitative interactions. Statistical Methodology. 2011; 8:42–55. [PubMed: 21179592]
20. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics. 2010; 38:894–942.
21. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Annals of Statistics. 2004; 32(2):407–499.
22. Hammer SM, Katzenstein DA, Hughes MD, Gundaker H, Schooley RT, Haubrich RH, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. for the AIDS Clinical Trials Group Study 175 Study Team. A trial comparingn ucleoside monotherapy with combination therapy in HIVinfected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New EnglandJournal of Medicine. 1996; 335:1081–1089. [PubMed: 8813038]
23. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. Statistics in Medicine. 2007; 27:4658–4677. [PubMed: 17960577]
24. Zhang M, Tsiatis AA, Davidian M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Biometrics. 2008; 64:707–715. [PubMed: 18190618]

Appendix

Define $\mathbf{Z} = (Y, \mathbf{X}, A)$ and denote the observations by $\mathbf{Z}_i = (Y_i, \mathbf{X}_i, A_i)$, $i = 1, \cdots, n$. Let $\theta = (\beta^T, \gamma^T)^T$. Define $l(\theta(\mathbf{z}) = y - \phi(\mathbf{x}; \gamma) - \beta^T \tilde{\mathbf{x}}(A - \pi(\mathbf{x}))$. Before we present the proofs, we first state regularity conditions.

Regularity Conditions

(A1) The observations $\mathbf{Z}_i$, $i = 1, \cdots, n$ are independent and identically distributed.

(A2) $\phi(\mathbf{X}, \gamma)$ is continuously differentiable with respect to γ, and γ belongs to a compact set in R^q with q being finite.

(A3) U exists and is non-singular, where U is defined in the proof of Theorem 1.

(A4) $|\partial^3 l(\mathbf{x})/\partial_i\theta_j\theta_k| < g(\mathbf{x})$ for all i, j, k and θ in a neighborhood of $\theta^* = (\beta^*, \gamma^*)$ for some integrable g, where (β^*, γ^*) is defined in the proof of Theorem 1.

Proof of Theorem 1

Define

$$G(\beta,\gamma) = E\left[Y - \varphi(\mathbf{X};\gamma) - \beta^T\tilde{\mathbf{X}}\{A - \pi(\mathbf{X})\}\right]^2,$$
$$\equiv E\left[Y - \varphi(\mathbf{X};\gamma) - \beta^T\tilde{\mathbf{X}}W(\pi, A\}\right]^2,$$

where $W(\pi, A) = A - \pi(\mathbf{X})$. Recall that the underlying true model is $Y = h_0(\mathbf{X}) + A(\beta_0^T\tilde{\mathbf{X}}) + \varepsilon$. Under regularity conditions which allow differentiation under the integral sign, we have

$$\begin{aligned}\frac{\partial G}{\partial \beta} &= -2E\left\{\left[Y-\varphi(\mathbf{X};\gamma)-\beta^T\tilde{\mathbf{X}}W(\pi,A)\right]\tilde{\mathbf{X}}W(\pi,A)\right\}\\ &= 2E\left\{\pi(\mathbf{X})\{1-\pi(\mathbf{X})\}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\right\}(\beta-\beta_0),\\ \frac{\partial G}{\partial \gamma} &= -2E\left\{\left[Y-\varphi(\mathbf{X};\gamma)-\beta^T\tilde{\mathbf{X}}W(\pi,A)\right]\frac{\partial\varphi}{\partial\gamma}\right\}\\ &= -2E\left\{\left[h_0(\mathbf{X})+\beta_0^T\tilde{\mathbf{X}}\pi(\mathbf{X})-\varphi(\mathbf{X};\gamma)\right]\frac{\partial\varphi}{\partial\gamma}\right\},\end{aligned}$$

where we use the double expectation rule and the fact $E(\varepsilon|A, \mathbf{X}) = 0$. Assume that the function $G(\beta, \gamma)$ achieves its maximum at (β^*, γ^*). Under assumptions (C1)–(C3), we can show that $\beta^* = \beta_0$ and γ^* satisfies the following equation:

$$E\left\{\left[h_0(\mathbf{X})+\beta_0^T\tilde{\mathbf{X}}\pi(\mathbf{X})-\varphi(\mathbf{X};\gamma)\right]\frac{\partial\varphi}{\partial\gamma}\right\}=\mathbf{0}. \quad (5.1)$$

Let $(\tilde{\beta}, \tilde{\gamma})$ be the minimizer of $L_n(\beta, \gamma)$ defined in (2.2). By the law of large numbers, $L_n(\beta, \gamma) \to G(\beta, \gamma)$ in probability as $n \to \infty$ for any (β, γ). By the argmax continuous mapping theorem, the minimizer of L_n converges to the minimizer of G in probability. In other words, $(\tilde{\beta}, \tilde{\gamma})$ is a consistent estimator of (β^*, γ^*), or equivalently, (β_0, γ^*). Define

$$q_n(\beta,\gamma)=\begin{bmatrix}\frac{\partial L_n}{\partial\beta}\\ \frac{\partial L_n}{\partial\gamma}\end{bmatrix}=\begin{bmatrix}\frac{1}{n}\sum_{i=1}^n\{-2[Y_i-\varphi(\mathbf{X}_i;\gamma)-\beta^T\tilde{\mathbf{X}}_iW_i]\tilde{\mathbf{X}}_iW_i\}\\ \frac{1}{n}\sum_{i=1}^n\{-2[Y_i-\varphi(\mathbf{X}_i;\gamma)-\beta^T\tilde{\mathbf{X}}_iW_i]\frac{\partial\varphi(\mathbf{X}_i;\gamma)}{\partial\gamma}\}\end{bmatrix}\equiv\frac{1}{n}\sum_{i=1}^n q_{n,i}(\beta,\gamma)$$

and

$$Q_n(\beta,\gamma)=\begin{bmatrix}\frac{\partial^2 L_n}{\partial\beta\partial\beta^T} & \frac{\partial^2 L_n}{\partial\beta\partial\gamma^T}\\ \frac{\partial^2 L_n}{\partial\gamma\partial\beta^T} & \frac{\partial^2 L_n}{\partial\gamma\partial\gamma^T}\end{bmatrix},$$

where $W_i = W(\pi(\mathbf{X}_i), A_i)$ for $i = 1, \cdots, n$. Using the Taylor expansion of L_n around (β_0, γ^*), we have

$$\mathbf{0}=q_n(\tilde{\beta},\tilde{\gamma})=q_n(\beta_0,\gamma^*)+Q_n(\beta_0,\gamma^*)\begin{bmatrix}\tilde{\beta}-\beta_0\\ \tilde{\gamma}-\gamma^*\end{bmatrix}+R_n,$$

where R_n is the remainder term. Therefore we have

$$\sqrt{n}\begin{bmatrix}\tilde{\beta}-\beta_0\\ \tilde{\gamma}-\gamma^*\end{bmatrix}=-Q_n^{-1}(\beta_0,\gamma^*)\,[\,\sqrt{n}q_n(\beta_0,\gamma^*)]+\sqrt{n}R_n.$$

Under regularity conditions (A1) to (A4), we can show that $\sqrt{n}R_n \to 0$. By the law of large numbers, we have $Q_n \to U$ in probability as $n \to \infty$, where

$$U=\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}=2\begin{bmatrix} E\{\pi(\mathbf{X})(1-\pi(\mathbf{X}))\mathbf{X}\mathbf{X}^T\} & 0 \\ 0 & E\left\{\left[\frac{\partial\varphi(\mathbf{X},\gamma)}{\partial\gamma}\right]^2_{\gamma=\gamma^*}\right\} \end{bmatrix}.$$

According to (5.1), we have $E\{q_n(\beta_0, \gamma^*)\} = \mathbf{0}$. By the central limit theorem, we have $\sqrt{n}q_n(\beta_0,\gamma^*) \to N(\mathbf{0},\Omega)$, where $\Omega=E\{q_{n,1}(\beta_0,\gamma^*)q_{n,1}^T(\beta_0,\gamma^*)\}$. It follows by Slutsky's theorem that

$$\sqrt{n}(\tilde{\beta}-\beta_0) \to N(0,V),$$

where $V=U_{11}^{-1}\Omega_{11}U_{11}^{-1}$ and Ω_{11} is the first $p \times p$ submatrix of Ω.

Proof of Theorem 2

Note that the objective function (2.4) can be written as a penalized least squared estimation, and thus the proofs follow [15] and [16].

Table 1

Estimation, selection and classification results for Model I.

Case	*n*	MSE	Selection			Percent. Correct Decision	
			Incor0 (0)	Corr0 (7)	Exact	Unpen.	Penalized
1	100	1.48 (0.21)	1.10	6.29	0.180	83.9	81.3
	200	0.69 (0.13)	0.40	6.42	0.384	87.8	88.3
	400	0.23 (0.05)	0.03	6.64	0.706	91.4	93.3
2	100	0.10 (0.02)	0	6.32	0.590	93.6	95.4
	200	0.04 (0.01)	0	6.71	0.778	95.8	97.1
	400	0.02 (0.00)	0	6.90	0.910	97.1	98.1

Table 2

Selection frequency results for Model I.

Case	*n*	int	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	100	0.886	0.826	0.138	0.096	0.086	0.086	0.094	0.102	0.106	0.622	0.562
	200	0.994	0.942	0.136	0.076	0.080	0.072	0.072	0.078	0.070	0.872	0.788
	400	1	1	0.076	0.054	0.042	0.044	0.038	0.05	0.058	0.992	0.978
2	100	1	1	0.090	0.086	0.102	0.070	0.120	0.100	0.112	1	1
	200	1	1	0.038	0.054	0.042	0.032	0.028	0.052	0.044	1	1
	400	1	1	0.020	0.016	0.016	0.014	0.010	0.012	0.014	1	1

Stat Methods Med Res. Author manuscript; available in PMC 2013 May 23.

Table 3

Estimation, selection and classification results for Model II.

Case	*n* Size	MSE	Selection			Percent. Correct Decision	
			Incor0 (0)	Corr0 (7)	Exact	Unpen.	Penalized
1	100	0.98 (0.15)	0.51	6.07	0.244	86.0	86.3
	200	0.42 (0.09)	0.14	6.44	0.526	89.7	91.4
	400	0.17 (0.03)	0.01	6.63	0.692	92.6	94.5
2	100	0.72 (0.12)	0.23	5.68	0.228	87.2	88.3
	200	0.29 (0.06)	0.04	6.33	0.508	90.8	92.5
	400	0.12 (0.03)	0.00	6.61	0.692	93.6	95.2

Table 4

Selection frequency results for Model II.

Case	*n*	int	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	100	0.980	0.916	0.242	0.104	0.152	0.078	0.116	0.122	0.120	0.832	0.762
	200	1	0.976	0.148	0.066	0.094	0.064	0.062	0.066	0.056	0.968	0.920
	400	1	0.998	0.130	0.038	0.058	0.036	0.038	0.028	0.040	1	0.996
2	100	1	0.958	0.316	0.168	0.198	0.152	0.158	0.160	0.172	0.936	0.874
	200	1	0.992	0.166	0.086	0.132	0.076	0.068	0.076	0.068	0.996	0.968
	400	1	0.998	0.134	0.048	0.074	0.044	0.038	0.024	0.032	1	1

Table 5

Estimation, selection and classification results for Model III.

Case	*n*	MSE	Selection			Percent. Correct Decision	
			Incor0 (0)	Corr0 (7)	Exact	Unpen.	Penalized
1	100	1.90 (0.23)	1.56	6.33	0.120	81.7	74.9
	200	1.04 (0.18)	0.79	6.43	0.264	85.9	84.5
	400	0.44 (0.09)	0.19	6.54	0.544	89.9	91.5
2	100	0.93 (0.16)	0.41	5.79	0.198	86.0	86.3
	200	0.38 (0.08)	0.12	6.29	0.454	89.8	91.3
	400	0.14 (0.03)	0.01	6.68	0.732	92.8	94.6

Table 6

Selection frequency results for Model III.

Case	*n*	int	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	100	0.754	0.714	0.088	0.088	0.152	0.070	0.090	0.088	0.092	0.536	0.440
	200	0.946	0.886	0.088	0.074	0.120	0.062	0.074	0.076	0.074	0.750	0.628
	400	1	0.998	0.058	0.056	0.120	0.054	0.058	0.050	0.066	0.938	0.872
2	100	0.982	0.952	0.144	0.152	0.274	0.150	0.140	0.176	0.170	0.882	0.778
	200	0.998	0.990	0.080	0.082	0.182	0.078	0.084	0.090	0.116	0.980	0.912
	400	1	1	0.032	0.030	0.118	0.040	0.028	0.040	0.034	1	0.994

Table 7

Estimation, selection and classification results for Model IV.

Case	*n*	MSE	Selection			Percent. Correct Decision	
			Incor0 (0)	Corr0 (47)	Exact	Unpen.	Penalized
1	200	0.63 (0.10)	0.22	44.5	0.194	80.7	88.9
	400	0.23 (0.05)	0.02	46.1	0.460	85.4	93.7
2	200	1.08 (0.16)	0.03	35.3	0.014	79.8	86.1
	400	0.17 (0.03)	0.00	45.3	0.376	86.5	94.1

Table 8

Estimation, selection and classification results for Model V.

Case	*n*	MSE	Selection			Percent. Correct Decision	
			Incor0 (0)	Corr0 (47)	Exact	Unpen.	Penalized
1	200	1.41 (0.18)	0.94	44.6	0.046	74.7	81.3
	400	0.65 (0.11)	0.31	45.8	0.284	80.5	89.1
2	200	1.38 (0.21)	0.10	35.3	0.002	78.3	84.4
	400	0.23 (0.04)	0.02	45.1	0.334	85.1	93.1