Multiple-stage Dynamic Treatment Regimes under Constraints

Shuping Ruan, Eric Laber
Department of Statistics, North Carolina State University

March 23, 2018

1 Introduction

Dynamic treatment regimes (DTRs), also known as adaptive treatment strategies or policies, are sequences of decision rules of which input is time-varying patient information and output is a recommended treatment at each intervention point [3, 15, 16]. These decision rules can be used to inform treatment decision for chronic conditions, e.g., depression, alcohol and drug abuse, HIV infection, cancer, diabetes etc., where clinicians have to make decisions at each stage based on evolving patient histories. A handful of methods have been developed to estimate the optimal treatment regimes. For example, indirect methods include Q-learning [17], penalized Q-learning [25], interactive Q-learning [13], A-learning [24], regret-regression [5], g-estimation [21] and so on. Policy search methods include marginal structural mean models [19, 18], outcome weighted learning [30, 27, 29], doubly robust estimators [28], and so forth. However, these methods only take a single clinical outcome into consideration, and neglect the clinical need to balance several competing outcomes. For example, a clinician may have to balance treatment effectiveness, side-effect burden, and cost while developing a treatment strategy for a patient with a chronic disease; or maximize the expected time to an adverse event while controlling the variance of the time to the adverse event.

Although handling the trade-off among multiple competing outcomes is important in practice, there has been little work done on this issue. Lizotte et al. proposed to compute the optimal treatment regimes of all the possible linear combinations of two competing outcomes [14]. However, only considering linear trade-off between two competing outcomes may not be sufficient to describe all possible patient preferences [10]. Wang et al. considered a compound score or "expert score" by numerically combining information on treatment efficacy, toxicity, and the risk of disease progression [26]. Unfortunately, it can be difficult to elicit a good composite outcome, and the quality of the estimated treatment regime maybe severely affect by the misspecification of a composite

outcome [11]. Some methods do not require the formation of composite outcomes. For example, set-valued dynamic treatment regimes proposed by Laber et al. inputs current patient information and outputs a set of recommended treatments. Multiple treatments may included in the set recommended, unless there exists a treatment that is best across all outcomes. Domain expertise is needed for tie breaking when a set of several treatments are recommended. Also, it needs to specify "clinically significant differences" for competing outcomes [10].

In this chapter, we continue the work previously done by Linn at el. [9], and propose a new statistical framework to tackle the problem of balancing multiple competing outcomes using constrained estimation. By restricting the values of secondary ones, we search for the feasible regimes with the maximized value of the primary outcome, ie., constrained optimal regimes. This method is useful, for example, when the clinicians need to find an adaptive intervention strategy that maximize the effectiveness and controls the side-effect burden simultaneously. This chapter focuses on constrained optimal regimes under the multiple stage setting. Data are assumed to be from Sequential, Multiple Assignment, Randomized Trials [12]. Observational data can also fit in our framework if the additional assumptions about the treatment assignment mechanism are tenable. However, precaution is needed when using data from observational studies, as one key assumption, the no unmeasured confounder assumption, can not be verified [2].

2 Methodology

2.1 Define multi-stage constrained optimal treatment regimes

2.1.1 Dataset

The dataset is denoted by

$$\{(\boldsymbol{X}_1^i, A_1^i, \boldsymbol{X}_2^i, A_2^i, \cdots, \boldsymbol{X}_T^i, A_T^i, \boldsymbol{Y}^i)\}_{i=1}^n,$$

which is composed of n identically, independently distributed patient trajectories
$\{(\boldsymbol{X}_1, A_1, \boldsymbol{X}_2, A_2, \cdots, \boldsymbol{X}_T, A_T, \boldsymbol{Y})\}$. Capital letters denote random variables; lower case letters denote realized values of these random variables. Let $\boldsymbol{X}_1$ be a patient baseline covariate, A_1 be the first-stage treatment variable, $\boldsymbol{X}_2$ be the patient covariate collected between first decision point and second decision point, A_2 be the second-stage treatment variable. So on, and so forth. Finally, $\boldsymbol{X}_T$ is the patient intermediate outcomes collected at the final decision point T, A_T is the treatment assignment at that time point, and $\boldsymbol{Y}$ is the final outcome vector. For $t = 1, \cdots, T$, $\boldsymbol{X}_t \in \boldsymbol{\mathcal{X}}_t \subseteq \mathbb{R}^{p_t}$, $A_t \in \mathcal{A}_t = \{1, 2, \cdots, m_t\}$, and $\boldsymbol{Y} \in \mathbb{R}^J$. The first component Y_1 denote the primary outcome of interest, which is coded so that larger values are more desirable. Meanwhile, $Y_2, \cdots, Y_J$

are the secondary outcomes of interest, which are coded so that lower is better. Let $\boldsymbol{H}_t$ denote the patient history information up to the decision point t, i.e., $\boldsymbol{H}_1^\intercal = (1, \boldsymbol{X}_1^\intercal)$, $\boldsymbol{H}_2^\intercal = (\boldsymbol{H}_1^\intercal, A_1, \boldsymbol{X}_2^\intercal)$, $\cdots$, $\boldsymbol{H}_t^\intercal = (\boldsymbol{H}_{t-1}^\intercal, A_{t-1}, \boldsymbol{X}_t^\intercal)$, $\cdots$, $\boldsymbol{H}_T^\intercal = (\boldsymbol{H}_{T-1}^\intercal, A_{T-1}, \boldsymbol{X}_T^\intercal)$. Besides, let $\bar{A}_t = (A_1, A_2, \cdots, A_t)$ denotes a sequence of treatment history up to time point t, and $\bar{A}_t \in \bar{\mathcal{A}}_t$, where $\bar{\mathcal{A}}_t = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_t$, $t = 2, \cdots, T$.

2.1.2 Potential outcomes

The potential outcome or counter-factual framework by Neyman, Rubin and Robins are adopted to identify the causal effect of a regime. The set of potential outcomes is $\boldsymbol{W}^* = \{\boldsymbol{X}_2^*(a_1), \boldsymbol{X}_3^*(\bar{a}_2), \cdots, \boldsymbol{X}_T^*(\bar{a}_{T-1}), \boldsymbol{Y}_T^*(\bar{a}_T), \text{for all } \bar{a}_t \in \bar{\mathcal{A}}_t, t = 1, 2, \cdots, T\}$, where $\boldsymbol{X}_t^*(\bar{a}_{t-1})$ is the potential outcome that would have been observed if the patient followed the treatment history sequence $\bar{a}_{t-1}$. The following three necessary assumptions are necessary to connect observed data with potential outcomes [8, 23, 22, 20, 6].

- *B1. Consistency:* $\boldsymbol{Y} = \boldsymbol{Y}^*(\bar{A}_T)$, and $\boldsymbol{X}_t = \boldsymbol{X}_t^*(\bar{A}_{t-1})$, $t = 2, \cdots, T$.
- *B2. Sequential randomization assumption:* $A_t \perp\!\!\!\perp \boldsymbol{W}^* \mid \boldsymbol{H}_t$ for $t = 1, 2, \cdots, T$.
- *B3. Positivity assumption:* $\exists\, \epsilon_t > 0$, such that $\Pr(A_t = a_t \mid \boldsymbol{H}_t = \boldsymbol{h}_t) > \epsilon_t$, for all $a_t \in \mathcal{A}_t$, $t = 1, 2, \cdots, T$.

B1) states that the intermediate and final outcomes observed equal to the patient's intermediate and final potential outcomes under the sequence of treatment actually assigned. It also implies no interference among individuals. B2) mean that conditional on the observed patient history $\boldsymbol{H}_t$, the treatment at time point t is assigned independently of the his or her potential outcomes. B3) guarantees a positive possibility for any $a_t \in \mathcal{A}_t$ having been assigned to patients with $\boldsymbol{H}_t = \boldsymbol{h}_t$. These assumptions imply that $\Pr(\boldsymbol{Y}^*(\bar{a}_T) \le \boldsymbol{y} \,|\, \boldsymbol{H}_T^*(\bar{a}_{T-1}) = \boldsymbol{h}_T) = \Pr(\boldsymbol{Y} \le \boldsymbol{y} \,|\, \boldsymbol{H}_T = \boldsymbol{h}_T, A_T = a_T)$ and $\Pr(\boldsymbol{X}_{t+1}^*(\bar{a}_t) \le \boldsymbol{x}_{t+1} \,|\, \boldsymbol{H}_t^*(\bar{a}_{t-1}) = \boldsymbol{h}_t^*) = \Pr(\boldsymbol{X}_{t+1} \le \boldsymbol{x}_{t+1} \mid \boldsymbol{H}_t = \boldsymbol{h}_t, A_t = a_t)$ for $t = 1, 2, \cdots, T-1$. Hence, we can estimate the values of a regime using the observed dataset.

2.1.3 Define constrained optimal dynamic treatment regimes

A dynamic treatment regime, $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_T)$, is a sequence of decision rules. Each decision rule, $\pi_t : \text{supp}(\boldsymbol{H}_t) \to \mathcal{A}_t$, is a function that maps the support of patient history information $\boldsymbol{H}_t$ to the set of all possible treatments at time point t. The final potential outcome under the regime $\boldsymbol{\pi}$ is $\boldsymbol{Y}^*(\boldsymbol{\pi}) = \sum_{\bar{a}_T \in \bar{\mathcal{A}}_T} \boldsymbol{Y}^*(\bar{a}_T)\mathbb{I}(\boldsymbol{\pi} = \bar{a}_T)$, and the intermediate potential outcome under that regime is $\boldsymbol{X}_{t+1}^*(\boldsymbol{\pi}_t) = \sum_{\bar{a}_t \in \bar{\mathcal{A}}_t} \boldsymbol{X}_{t+1}^*(\bar{a}_t)\mathbb{I}(\boldsymbol{\pi}_t = \bar{a}_t)$, where $\boldsymbol{\pi}_t = (\pi_1, \pi_2, \cdots, \pi_t)$. The value of a dynamic treatment regime, $\boldsymbol{V}(\boldsymbol{\pi}) = \mathbb{E}\boldsymbol{Y}^*(\boldsymbol{\pi})$, is defined as the expected final outcome if each patient in the population of interest is treated according to $\boldsymbol{\pi}$. Each component of $\boldsymbol{V}(\boldsymbol{\pi})$ is denoted by $V_j(\boldsymbol{\pi}) = \mathbb{E}Y_j^*(\boldsymbol{\pi})$, for $j = 1, \cdots, J$. Our goal is to find a constrained optimal regime, $\boldsymbol{\pi}_{\boldsymbol{\nu}}^*$, that

maximizes the expectation of the primary final potential outcome $V_1(\boldsymbol{\pi})$, subject to an upper bound constraints on the expectation of the secondary final potential outcomes $V_j(\boldsymbol{\pi})$, for $j = 2, \cdots, J$. The $J-1$ dimensional vector of upper bounds is denoted as $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_{J-1})$, which can be determined by the preference of clinicians or patients. Therefore, a multi-stage constrained optimal regime problem is defined as

$$\begin{aligned} &\max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} V_1(\boldsymbol{\pi}) \\ &\text{subject to } V_j(\boldsymbol{\pi}) \leq \nu_{j-1}, \end{aligned} \tag{1}$$

where $j = 2, 3, \cdots, J$ and $\boldsymbol{\Pi}$ is the class of dynamic treatment regimes under consideration. The feasible space of the class of regimes, $\mathcal{F}(\boldsymbol{\Pi})$, is the set of all regimes satisfying the constraints. For each $\boldsymbol{\pi} \in \mathcal{F}(\boldsymbol{\Pi})$, $V_j(\boldsymbol{\pi}) \leq \nu_{j-1}$, for $j = 2, \cdots, J$. Then, a multi-stage constrained optimal regime can also be written as $\boldsymbol{\pi}^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\pi} \in \mathcal{F}(\boldsymbol{\Pi})} V_1(\boldsymbol{\pi})$.

We choose the class of regime to be the class of linear decision rules, where each mapping function π_t at time point t is indexed by $\boldsymbol{\theta}_t$. More specifically, $\pi_t(\boldsymbol{h}_t) = \text{sgn}(\boldsymbol{h}_t^\intercal \boldsymbol{\theta}_t)$. Hence, all $V_j(\boldsymbol{\pi})$'s can be considered as functions of $\boldsymbol{\theta}$ and can be exchangeably denoted $V_j(\boldsymbol{\theta})$'s. As only the directions of $\boldsymbol{h}_t^\intercal \boldsymbol{\theta}_t$ matters, we restrict $\boldsymbol{\theta}_t$ to be unit vectors, i.e., $\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t = 1$, for $t = 1, \cdots, T$. Then, problem (2.1) above can be written as

$$\begin{aligned} &\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} V_1(\boldsymbol{\theta}) \\ &\text{subject to } V_j(\boldsymbol{\theta}) - \nu_j \leq 0, \\ &\qquad \boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t^\intercal - 1 = 0. \end{aligned} \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$, $j = 2, \cdots, J$ and $t = 1, \cdots, T$. Let the feasible set of $\boldsymbol{\Theta}$ be $\mathcal{F}(\boldsymbol{\Theta})$, such that $V_j(\boldsymbol{\theta}) \leq \nu_j$, for any $\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})$ and $j = 2, \cdots, J$. Then, the corresponding index parameter of a constrained optimal dynamic treatment regime is $\boldsymbol{\theta}^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})} V_1(\boldsymbol{\theta})$, and $\boldsymbol{\theta}^*_{\boldsymbol{\nu}} = (\boldsymbol{\theta}^*_{\boldsymbol{\nu},1}, \boldsymbol{\theta}^*_{\boldsymbol{\nu},2}, \cdots, \boldsymbol{\theta}^*_{\boldsymbol{\nu},T})$.

2.2 Re-define constrained optimal regimes via penalization

Problem (2.2) is a nonlinear constrained continuous optimization task. It is solved via interior point method, where we re-formalize the problem via quadratic-barrier penalization. To re-formalize the problem, we let $v_1(\boldsymbol{\theta}) = -V_1(\boldsymbol{\theta})$ and $v_j(\boldsymbol{\theta}) = V_j(\boldsymbol{\theta}) - \nu_j$ for $j = 2, \cdots, J$. Moreover, $h_t(\boldsymbol{\theta}_t) = \boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1$. Hence, we have problem (2.2) equivalent to the following

$$\begin{aligned} &\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} v_1(\boldsymbol{\theta}) \\ &\text{subject to } v_j(\boldsymbol{\theta}) \leq 0, \\ &\qquad h_t(\boldsymbol{\theta}_t) = 0. \end{aligned} \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$, $j = 1, \cdots, J$ and $t = 1, \cdots, T$. Interior point method approximate the solution of problem (2.3) by solving a sequence of the following problem (2.4), where μ is positive and approaches to zero in the limit. For each $\mu > 0$, the approximate problem is

$$\min_{\boldsymbol{\theta}, \boldsymbol{z}} \phi_\mu(\boldsymbol{\theta}, \boldsymbol{z}) = \min v_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln z_j, \text{subject to } v_j(\boldsymbol{\theta}) + z_j = 0, h_t(\boldsymbol{\theta}_t) = 0 \tag{4}$$

where $j = 2, \cdots, J$ and $t = 1, \cdots, T$. z_j's are the slack variables, which are restricted to be positive due the ln operator. The logarithmic terms, $\ln z_j$'s, are the barrier functions, which enforce the solution path to be within the feasible region of the problem (2.3). More details on interior points method can be found at section 2.1.2.

The sequence of solutions to problem (2.4) forms a trajectory path $\left\{\boldsymbol{\theta}^*_\nu(\mu)\right\}_{\mu \to 0+}$ that convergence to the solution to problem (2.3) as $\mu \to 0$, i.e., $\lim_{\mu \to 0} \boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu) = \boldsymbol{\theta}^*_{\boldsymbol{\nu}}$. The conditions for its convergence can be found at section 2.1.3. Let $\widehat{\boldsymbol{V}}(\theta)$ be a consistent estimator of the values of a regime π. Then, correspondingly, $\widehat{v}_1(\boldsymbol{\theta}) = -\widehat{V}_1(\boldsymbol{\theta})$ and $\widehat{v}_j(\boldsymbol{\theta}) = \widehat{\boldsymbol{V}}_j(\boldsymbol{\theta}) - \nu_j$ for $j = 2, \cdots, J$. Then, problem (2.4) with the plugin estimator is

$$\min_{\boldsymbol{\theta}, \boldsymbol{z}} \widehat{\phi}_\mu(\boldsymbol{\theta}, \boldsymbol{z}) = \min \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln z_j, \text{ subject to } \widehat{v}_j(\boldsymbol{\theta}) + z_j = 0, h_t(\boldsymbol{\theta}_t) = 0, \tag{5}$$

for $j = 2, \cdots, J$ and $t = 1, \cdots, T$. The solution to problem (2.5) is equivalent to the solution to penalty-barrier function below.

$$\min_{\boldsymbol{\theta}} \widehat{\phi}^{BP}_\mu(\boldsymbol{\theta}) = \min \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln(-\widehat{v}_j(\boldsymbol{\theta})) + \frac{1}{2\mu} \sum_{t=1}^{T} (\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1)^2 \tag{6}$$

Denote the solution to (2.5/2.6) as $\widehat{\boldsymbol{\theta}}_\nu(\mu)$.We have proven that $\widehat{\boldsymbol{\theta}}_\nu(\mu) \xrightarrow{p} \boldsymbol{\theta}^*_\nu(\mu)$. For the consistency of this estimator, the details and proof are provided in section 2.1.4 and appendix A.2.

2.3 Estimation of the values of a regime

To estimate the values of a regime, we use the G-computation formula by Robins, etc [4]. For any arbitrary regime $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_T)$, assume the three causal assumptions B1)-B3) are satisfied, then for each component of $\boldsymbol{Y}$

$$\begin{aligned}
&\Pr\left(Y_j^*(\boldsymbol{\pi}) \leq y_j\right) = F_{Y_j^*(\boldsymbol{\pi})}\left(y_j\right) \\
&= \int \cdots \int F_{Y_j|\boldsymbol{H}_T, A_T}\left(y_j|\boldsymbol{h}_T, \pi_T(\boldsymbol{h}_T)\right)\, dF_{\boldsymbol{H}_T|\boldsymbol{H}_{T-1}, A_{T-1}}\left(\boldsymbol{h}_T|\boldsymbol{h}_{T-1}, \pi_{T-1}(\boldsymbol{h}_{T-1})\right) \\
&\quad dF_{\boldsymbol{H}_{T-1}|\boldsymbol{H}_{T-2}, A_{T-2}}\left(\boldsymbol{h}_{T-1}|\boldsymbol{h}_{T-2}, \pi_{T-2}(\boldsymbol{h}_{T-2})\right) \cdots dF_{\boldsymbol{H}_2|\boldsymbol{H}_1, A_1}\left(\boldsymbol{h}_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right)\, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}$$

where $F_{Y_j|\boldsymbol{H}_T,A_T}\left(\cdot \mid \cdot,\cdot\right)$ is the conditional cumulative density function of Y_j conditioning on $\boldsymbol{H}_T$ and A_T, $F_{\boldsymbol{H}_t|\boldsymbol{H}_{t-1},A_{t-1}}\left(\cdot \mid \cdot,\cdot\right)$ the conditional cumulative density function of $\boldsymbol{H}_t$ conditioning on $\boldsymbol{H}_{t-1}, A_{t-1}$, and $F_{\boldsymbol{H_1}}(\cdot)$ the cumulative density function of $\boldsymbol{H}_1$. Thus, the marginal distribution of the potential outcomes under any regime $\boldsymbol{\pi}$ can be estimated from observed data, if we can estimate the conditional distributions involved. However, the estimation of the sequence of conditional distribution could being a daunting task. Linn et al. used a two-step estimator via mean and variance modeling to construct two-stage constrained optimal dynamic treatment regimes [9], and it is demonstrated in the following simulation studies. However, the modeling becomes complex rapidly as the number of stages increases. Note the regime is index by $\boldsymbol{\theta}$, we use $F_{Y_j^*(\boldsymbol{\pi})}\left(y_j\right)$ and $F_{Y_j^*(\boldsymbol{\theta})}\left(y_j\right)$ interchangeably.

2.4 Asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ here is similar to the corresponding part for one-stage problem in Chapter 1 (Section 1.1.6).

2.4.1 Limiting distribution of $\nabla\widehat{V}_j(\boldsymbol{\theta})$

Before we derive the limiting distribution of the estimator $\widehat{\boldsymbol{\theta}}_{\kappa}(\mu)$, we need to examine, for any fixed value of $\boldsymbol{\theta}$: $\boldsymbol{\theta}_1^{\intercal}\boldsymbol{\theta}_1 = 1$ and $\boldsymbol{\theta}_2^{\intercal}\boldsymbol{\theta}_2 = 1$, the limiting distribution of $\nabla\widehat{V}_j(\boldsymbol{\theta})$, where

$$\begin{aligned}
\nabla\widehat{V}_j(\boldsymbol{\theta}) =& \frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y) \\
=& \frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{F}_{Y^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right) \\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i}),
\end{aligned}$$

where $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(\cdot)$ denote the estimator of $F_{Y_j^*(\boldsymbol{\theta})}(\cdot)$

Lemma 2.1. *Suppose the following conditions hold.*

1. $\forall\boldsymbol{a}\in\mathbb{R}^p$,$\exists\delta>0$,*such that*

   (a) $\mathbb{E}\left|\boldsymbol{a}^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right|^{2+\delta}<\infty$

   (b) $\left\{\boldsymbol{a}^{\intercal}\,Var\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}<\infty.$

Then, we have, for any fixed $\boldsymbol{\theta}$,

The proof of this is similar to the proof of Lemma 1.1.3 and is shown in Appendix B.1.

Assume $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ is consistent, and the following corollary shows that the estimations do not effect the limiting distribution obtained above.

Corollary 2.2. *Suppose all the assumptions in Lemma 2.2.1 hold, and* $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ *is a consistent estimator of* $F_{Y_j^*(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$*. Then, we have*

$$\sqrt{n}\left(\nabla\widehat{V}_j(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)) - \nabla V_j(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu))\right) \xrightarrow{d} \mathcal{N}\left(0, Avar\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)}\right)\right)$$

See Appendix B.2 for proof.

2.4.2 Limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

Now, we investigate the limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$.

Theorem 2.3. *Suppose all the assumptions above hold. Then we have, as* $n \to \infty$

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) - \boldsymbol{\theta}_{\boldsymbol{\nu}}(\mu)^*) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,
$\boldsymbol{C}^* = \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right) - \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right)\mathbb{E}\left(\nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right)$,
and $\boldsymbol{D}^* = \nabla^2\phi_\mu^{BP}(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu))$.

The proof is similar to the proof of Theorem 1.1.5, and is presented in Appendix B.3.

3 Simulation

3.1 Simulation design

We demonstrate our proposed method using the toy example presented by Linn et al [9], where there are two competing outcomes Y and Z. The goal is to maximize the mean of Y, subject to an upper bound on the mean of Z. Y is coded so that the higher the value the better, such as the effectiveness of the treatment regimes. Meanwhile, Z is coded the lower the better, such as the side-effect burden. The model for generating the patient trajectories $(X_1, A_1, X_2, A_2, Y, Z)$ are

as follow:

$$
\begin{aligned}
&X_1 \sim \text{Normal}(1,1),\\
&\boldsymbol{H}_1 = (1, X_1)^\intercal,\\
&A_1 \sim \text{Uniform}\left\{-1,1\right\},\\
&X_2 = \boldsymbol{H}_1^\intercal \boldsymbol{\beta}_{1,0} + A_1 \boldsymbol{H}_1^\intercal \boldsymbol{\beta}_{1,1} + \epsilon,\\
&\epsilon \sim \text{Normal}(0,1),\\
&\boldsymbol{H}_2 = (1, X_2)^\intercal,\\
&A_2 \sim \text{Uniform}\left\{-1,1\right\},\\
&Y = \boldsymbol{H}_2^\intercal \boldsymbol{\beta}_{2,0,Y} + A_2 \boldsymbol{H}_2^\intercal \boldsymbol{\beta}_{2,1,Y} + \epsilon_Y\\
&Z = \boldsymbol{H}_2^\intercal \boldsymbol{\beta}_{2,0,Z} + A_2 \boldsymbol{H}_2^\intercal \boldsymbol{\beta}_{2,1,Z} + \epsilon_Z\\
&(\epsilon_Y, \epsilon_Z)^\intercal \sim \text{Normal}(\mathbf{0}_2, \Sigma_{Y,Z})
\end{aligned}
$$

This model is a simple representation of the data from a two-stage randomized SMART. Variable X_1 represents the summary of patient status before the first treatment assignment A_1. Variable X_2 represents the summary of patient status before the second treatment assignment A_2. The parameters involved are set to the following,

$$
\begin{aligned}
&\boldsymbol{\beta}_{1,0} = (0.5, 0.75)^\intercal\\
&\boldsymbol{\beta}_{1,1} = (0.25, 0.5)^\intercal\\
&\boldsymbol{\gamma}_0 = (0.25, -0.05)^\intercal\\
&\boldsymbol{\gamma}_1 = (0.1, -0.05)^\intercal\\
&\boldsymbol{\beta}_{2,0,Y} = (30, 2)^\intercal\\
&\boldsymbol{\beta}_{2,1,Y} = (5, -1.5)^\intercal\\
&\boldsymbol{\beta}_{2,0,Z} = (15, 1)^\intercal\\
&\boldsymbol{\beta}_{2,1,Z} = (3, -0.5)^\intercal\\
&\Sigma_{Y,Z} = \begin{bmatrix} 1.0, & 0.7 \\ 0.7, & 1.0 \end{bmatrix}
\end{aligned}
$$

The class of regimes under consideration is restricted to linear decision rules at each stage. That is $\pi_1 = \text{sgn}(\boldsymbol{h}_1^\intercal \boldsymbol{\theta}_1)$ and $\pi_2 = \text{sgn}(\boldsymbol{h}_2^\intercal \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the index parameters for the regimes. The true optimal regimes are denoted by $\pi_1^* = \text{sgn}(\boldsymbol{h}_1^\intercal \boldsymbol{\theta}_1^*)$ and $\pi_2^* = \text{sgn}(\boldsymbol{h}_2^\intercal \boldsymbol{\theta}_2^*)$. The estimated optimal regimes are denoted by $\widehat{\pi}_1 = \text{sgn}(\boldsymbol{h}_1^\intercal \widehat{\boldsymbol{\theta}}_1)$ and $\widehat{\pi}_2 = \text{sgn}(\boldsymbol{h}_2^\intercal \widehat{\boldsymbol{\theta}}_2)$. Here, the sgn function is defined as

$$
\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}
$$

3.2 Modeling and estimation

Modeling for and estimation of the distributions of potential outcomes

The distribution of potential outcomes are unknown. The two major quantities under an arbitrary regime $\boldsymbol{\pi}$ involved, $\mathbb{E}Y^*(\boldsymbol{\pi})$ and $\mathbb{E}Z^*(\boldsymbol{\pi})$, need to be estimated from the observed data. Our strategy for estimating these two quantities is to model the marginal distribution of each potential outcome, and then draw random samples from the estimated marginal distributions to calculate their expectations numerically. To connect observed data with potential outcomes, three necessary causal inference assumptions *B1)-B3).* are assumed to hold.

Following the G-computation formula [4], we have, for any arbitrary regime $\boldsymbol{\pi} = (\pi_1, \pi_2)$, that

$$\Pr\{Y^*(\boldsymbol{\pi})) \le y\} = \mathbb{E}_{\boldsymbol{H}_1}\left\{\mathbb{E}_{\boldsymbol{H}_2}\left[\Pr\left\{Y \le y \mid \boldsymbol{H}_2, A_2 = \pi_2(\boldsymbol{H}_2), \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right\} \mid \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right]\right\}$$

and similarly,

$$\Pr\{Z^*(\boldsymbol{\pi}) \le z\} = \mathbb{E}_{\boldsymbol{H}_1}\left\{\mathbb{E}_{\boldsymbol{H}_2}\left[\Pr\left\{Z \le z \mid \boldsymbol{H}_2, A_2 = \pi_2(\boldsymbol{H}_2), \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right\} \mid \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right]\right\}$$

Hence, we can estimate the probabilty function of the potential outcomes under a regime $\boldsymbol{\pi}$, $\Pr\{Y^*(\boldsymbol{\pi}) \le y\}$ and $\Pr\{Z^*(\boldsymbol{\pi}) \le z\}$, using observed data by modeling and estimating the conditional distributions involved, and hence, $\mathbb{E}Z^*(\boldsymbol{\pi})$ and $\mathbb{E}Z^*(\boldsymbol{\pi})$.

Following the modeling tactic in "Constrained estimation for competing outcomes" by Linn et al [9]. We assume the following model,

$$\begin{aligned}
&Y = \mathbb{E}(Y|\boldsymbol{H}_2, A_2) + \varepsilon_Y,\\
&\mathbb{E}(Y|\boldsymbol{H}_2, A_2) = m_Y(\boldsymbol{H}_2) + A_2 c_Y(\boldsymbol{H}_2),\\
&\text{where } \mathbb{E}(\varepsilon_Y) = 0, \mathrm{Var}(\varepsilon_Y) = \sigma^2, \text{and } \varepsilon_Y \perp\!\!\!\perp (\boldsymbol{H}_2, A_2).
\end{aligned}$$

Define $F_{\varepsilon_Y}(\cdot)$ to be the distribution of ε_Y; $F_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}(\cdot \mid \boldsymbol{h}_1, a_1)$ to be the conditional distribution of $\boldsymbol{H}_2$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$; $F_{\boldsymbol{H}_1}(\cdot)$ to be the distribution of $\boldsymbol{H}_1$. Again, we have $\boldsymbol{H}_1^\intercal = (1, \boldsymbol{X}_1^\intercal)$, $\pi_1(\boldsymbol{H}_1)$, $\boldsymbol{H}_2 = \{\boldsymbol{H}_1^\intercal, \pi_1(\boldsymbol{H}_1), \boldsymbol{X}_2^\intercal\}^\intercal$.

$$\begin{aligned}
&\Pr\left\{Y \le y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\}\\
=&\Pr\left\{m(\boldsymbol{H}_2) + \pi_2(\boldsymbol{H}_2)c_Y(\boldsymbol{H}_2) + \varepsilon_Y \le y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\}\\
=&\Pr\left\{\varepsilon_Y \le y - m(\boldsymbol{H}_2) - \pi_2(\boldsymbol{H}_2)c_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\}\\
=&F_{\varepsilon_Y}\left\{y - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2)c_Y(\boldsymbol{h}_2)\right\}\\
=&F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \mathrm{sgn}\left\{r_2(\boldsymbol{h}_2; \boldsymbol{\theta}_2)\right\} c_Y(\boldsymbol{h}_2)\right]
\end{aligned}$$

Hence, we have

$$
\begin{aligned}
&\Pr\left\{Y^*(\boldsymbol{\pi}) \leq y\right\} \\
&= \iint \Pr\left\{Y \leq y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, A_2 = \pi_2(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2|\pi_1(\boldsymbol{h_1}), \boldsymbol{h}_1\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left\{y - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2) c_Y(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2 \mid \boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \operatorname{sgn}\left\{r_2(\boldsymbol{h}_2;\boldsymbol{\theta}_2)\right\} c_Y(\boldsymbol{h}_2)\right] \, dG_Y\left\{m_Y, c_Y, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \operatorname{sgn}(r_2) c_Y(\boldsymbol{h}_2)\right] \, dG_Y\left\{m_Y, c_Y, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}
$$

where $G_Y\left\{m_Y, c_Y, r_2 \mid \boldsymbol{h}_1, a_1\right\}$ is the joint conditional distribution of $m_Y\left(\boldsymbol{H}_2\right)$, $c_Y\left(\boldsymbol{H}_2\right)$ and $r_2(\boldsymbol{H}_2;\boldsymbol{\theta}_2)$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$. The second equality is due to
$\int z(x,y)\, dF_{X|Y}(x|y) = \mathbb{E}(z|y) = \int z\, dF_{Z|Y}(z|y)$.

Same applies to Z:

$$
\begin{aligned}
&Z = \mathbb{E}(Z|\boldsymbol{H}_2, A_2) + \epsilon, \\
&\text{where } \mathbb{E}(\epsilon) = 0, \operatorname{Var}(\epsilon) = \sigma^2, \text{and } \epsilon \perp\!\!\!\perp (\boldsymbol{H}_2, A_2) \\
&\mathbb{E}(Z|\boldsymbol{H}_2, A_2) = m_Z(\boldsymbol{H}_2) + A_2 c_Z(\boldsymbol{H}_2)
\end{aligned}
$$

$$
\begin{aligned}
&\Pr\left\{Z^*(\boldsymbol{\pi}) \leq z\right\} \\
&= \iint \Pr\left\{Z \leq z \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2|d_1(\boldsymbol{h_1}), \boldsymbol{h}_1\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left\{z - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2) c_Z(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2 \mid \boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left[z - m(\boldsymbol{h}_2) - \operatorname{sgn}\left\{r_2(\boldsymbol{h}_2;\boldsymbol{\theta}_2)\right\} c_Z(\boldsymbol{h}_2)\right] \, dG_Z\left\{m_Z, c_Z, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left[z - m(\boldsymbol{h}_2) - \operatorname{sgn}(r_2) c_Z(\boldsymbol{h}_2)\right] \, dG_Z\left\{m_Z, c_Z, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}
$$

where $G_Z\left\{m_Z, c_Z, r_2 \mid \boldsymbol{h}_1, a_1\right\}$ is the joint conditional distribution of $m_Z\left(\boldsymbol{H}_2\right)$, $c_Z\left(\boldsymbol{H}_2\right)$ and $r_2(\boldsymbol{H}_2;\boldsymbol{\theta}_2)$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$.

We model the joint distribution of $\left\{m_Y(\boldsymbol{H}_2), c_Y(\boldsymbol{H}_2), m_Z(\boldsymbol{H}_2), c_Z(\boldsymbol{H}_2)\right\}$ by modeling the joint distribution of the standardized residuals obtained from

the mean and variance modeling of each component for given $\boldsymbol{H}_1$ and A_1

$$
\begin{aligned}
e_Y^m &= \frac{m_Y(\boldsymbol{H}_2) - \mu_Y^m(\boldsymbol{H}_1, A_1)}{\sigma_Y^m(\boldsymbol{H}_1, A_1)} \\
e_Y^c &= \frac{c_Y(\boldsymbol{H}_2) - \mu_Y^c(\boldsymbol{H}_1, A_1)}{\sigma_Y^c(\boldsymbol{H}_1, A_1)} \\
e_Z^m &= \frac{m_Z(\boldsymbol{H}_2) - \mu_Z^m(\boldsymbol{H}_1, A_1)}{\sigma_Z^m(\boldsymbol{H}_1, A_1)} \\
e_Z^c &= \frac{c_Z(\boldsymbol{H}_2) - \mu_Z^c(\boldsymbol{H}_1, A_1)}{\sigma_Z^c(\boldsymbol{H}_1, A_1)} \\
e_{f_2} &= \frac{f_2(\boldsymbol{H}_2) - \mu_{f_2}(\boldsymbol{H}_1, A_1)}{\sigma_{f_2}(\boldsymbol{H}_1, A_1)}
\end{aligned}
$$

The mean functions are defined as

$$
\begin{aligned}
\mu_Y^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{m_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Y^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{c_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Z^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{m_Z(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Z^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{c_Z(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_{f_2}(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{f_2(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\}
\end{aligned}
$$

and the standard deviation functions are defined as

$$
\begin{aligned}
\sigma_Y^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{m_Y(\boldsymbol{H}_2) - \mu_Y^m(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Y^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{c_Y(\boldsymbol{H}_2) - \mu_Y^c(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Z^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{m_Z(\boldsymbol{H}_2) - \mu_Z^m(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Z^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{c_Z(\boldsymbol{H}_2) - \mu_Z^c(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_{f_2}(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{f_2(\boldsymbol{H}_2) - \mu_{f_2}(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2}
\end{aligned}
$$

Therefore, we model the joint distribution of the standardized residuals $(e_Y^m, e_Y^c, e_Z^m, e_Z^c, e_{f_2})$ to obtain an estimator of $G_{Y,Z}^{\pi}(\cdot, \cdot, \cdot, \cdot, \cdot \mid x_1, a_1)$.

Due to the cost of clinical data, sample sizes are usually small. We consider parametric models for $m_Y(\boldsymbol{H}_2), c_Y(\boldsymbol{H}_2), m_Z(\boldsymbol{H}_2), c_Z(\boldsymbol{H}_2)$ and $f_2(\boldsymbol{H}_2)$. Here, we model $m_Y(\boldsymbol{H}_2) = \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_1 + A_1 \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_2 + \varepsilon$, where ε is a mean-zero error term. Then, $\mu_Y^m(\boldsymbol{H}_1, A_1) = \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_1 + A_1 \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_2$.

To estimate, we fit the corresponding least squares regressions, and estimate the residuals empirically. For more details, see reference [9].

3.3 Summary of simulation results

We summarize the simulation results here. Figure 2. below shows the estimated optimal regime values and their standard deviation. Figure 2.1 is the efficient

frontier plot. The red dashed line represents $\widehat{V}_1$ under estimated constrained optimal regime, and the blue dash-dotted line represents $\widehat{V}_2$ under that regime. The plot represents the best possible value of the primary potential outcome for its level of risk, which is the value of the secondary potential outcome. In the plot, the value of the primary outcome increases as the constraint bound gets looser. Meanwhile the value of the secondary outcome keep up with the constraint, until the constraint is not active. Once the constraint gets larger than the maximum value of the secondary potential outcome, the constrained problem becomes an unconstrained problem.

Table 1: Simulation results

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$
12.86	27.48	0.46	12.86	0.23
13.45	28.90	0.42	13.37	0.15
14.03	30.38	0.64	13.89	0.17
14.62	31.85	0.83	14.46	0.16
15.21	33.53	0.64	15.05	0.13
15.79	34.47	0.70	15.64	0.14
16.38	35.47	0.94	16.15	0.29
16.97	36.33	0.93	16.68	0.40
17.55	37.08	0.87	17.31	0.41
18.14	37.62	0.51	17.89	0.31
18.72	37.79	0.72	18.32	0.44
19.31	38.03	0.17	18.91	0.10
19.90	38.03	0.17	18.91	0.10
20.48	38.03	0.17	18.91	0.10
21.07	38.03	0.17	18.91	0.10
21.66	38.03	0.17	18.91	0.10
22.24	38.03	0.17	18.91	0.10
22.83	38.03	0.17	18.91	0.10
23.41	38.03	0.17	18.91	0.10
24.00	38.03	0.17	18.91	0.10

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest.

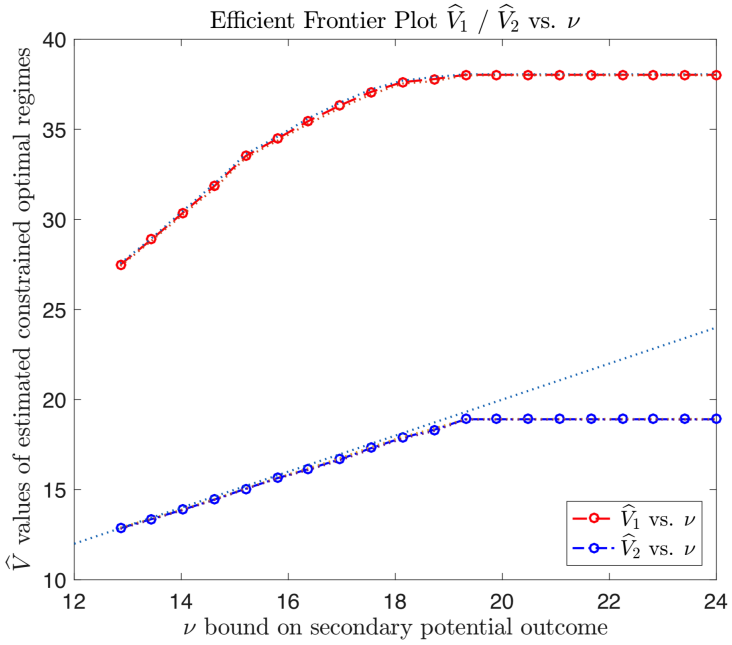

Figure 1: Efficient frontier for estimated constrained optimal regimes (multi-stage)

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

4 Conclusion

Focusing on optimizing a single scalar outcome may be an oversimplification of the goals of practical clinical decision making. In this chapter, a new method is proposed to handle multiple competing outcomes in the multi-stage setting. Estimating an optimal treatment regime with competing outcomes is cast as a constrained optimization problem. We maximize the primary outcome of interest, subject to the constraints on the secondary outcomes of interest. Our estimator of a constrained optimal treatment regime has the properties of consistency and asymptotic normality under mild regularity conditions. The efficient frontier plots provide an intuitive visualization for clinicians to examine the trade-off between two competing outcomes.

References

[1] Patrick Billingsley. Probability & Measure. page 362, 1995.

[2] Bibhas Chakraborty and Erica E.M. Moodie. *Statistical Methods for Dynamic Treatment Regimes*. 2013.

[3] Bibhas Chakraborty and Susan A. Murphy. Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application*, 1(1):447–464, 2014.

[4] Richard D. Gill and James M. Robins. Causal inference for complex longitudinal data: The continuous case. *Annals of Statistics*, 29(6):1785–1811, 2001.

[5] Robin Henderson, Phil Ansell, and Deyadeen Alshibani. Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–201, December 2010.

[6] Miguel A Hernan and James M Robins. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health*, (7):578–86, July.

[7] David R Hunter. Notes for a graduate-level course in asymptotics for statisticians. page 97, 2014.

[8] Jerzy Splawa-Neyman, D. M. Dabrowska and T. P. Speed. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1990.

[9] Linn KA, Laber EB, and Stefanski LA. Constrained estimation for competing outcomes. *Chapter in Adaptive Treatment Strategies In Practice, ASA-SIAM Statistics and Applied Probability Series, 2015*, 29, 2001.

[10] Eric B. Laber, Daniel J. Lizotte, and Bradley Ferguson. Set-valued dynamic treatment regimes for competing outcomes. *Biometrics*, 70(1):53–61, 2014.

[11] Eric B Laber, Daniel J Lizotte, Min Qian, William E Pelham, and Susan A Murphy. Dynamic treatment regimes : technical challenges and applications. 2014.

[12] H. Lei, I. Nahum-Shani, K. Lynch, D. Oslin, and S.A. Murphy. A "SMART" Design for Building Individualized Treatment Sequences. *Annual Review of Clinical Psychology*, 8(1):21–48, 2012.

[13] Kristin A Linn, Eric B Laber, and Leonard A Stefanski. Interactive Q-learning for Probabilities and Quantiles. 2014.

[14] Daniel J Lizotte, Michael H. Bowling, and Susan A. Murphy. Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. *in Proc. of Int. Conf. on Machine Learning*, pages 695–702, 2010.

[15] Erica Moodie. Dynamic treatment regimes. *Clinical trials (London, England)*, 1(5):471, 2004.

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

[17] Inbal Nahum-Shani, Min Qian, Daniel Almirall, William E. Pelham, Beth Gnagy, Gregory A. Fabiano, James G. Waxmonsky, Jihnhee Yu, and Susan A. Murphy. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods*, 17(4):478–494, 2012.

[18] Liliana Orellana, Andrea Rotnitzky, and James M Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content. *The international journal of biostatistics*, 6(2):Article 8, January 2010.

[19] J M Robins, M A Hernán, and B Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[20] James M. Robins. Causal Inference from Complex Longitudinal Data, 1997.

[21] James M. Robins, Donald Blevins, Grant Ritter, and Michael Wulfsohn. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of aids patients. *Epidemiology*, 3(4):319–336, 1992.

[22] D. B. Rubin. Discussion of Randomized analysis of experimental data: The Fisher randomization test by D. Basu. *Journal of the American Statistical Association*, (75):591–593, 1980.

[23] Donald B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[24] Phillip J. Schulte, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. Q- and A-Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science*, 29(4):640–661, 2014.

[25] Rui Song, Weiwei Wang, Donglin Zeng, and Michael R. Kosorok. Penalized Q-Learning for Dynamic Treatment Regimes. August 2011.

[26] Lu Wang, Andrea Rotnitzky, Xihong Lin, Randall E Millikan, and Peter F Thall. Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer. *Journal of the American Statistical Association*, 107(498):493–508, June 2012.

[27] Baqun Zhang, Anastasios A. Tsiatis, Marie Davidian, Min Zhang, and Eric Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012.

[28] Baqun Zhang, Anastasios a. Tsiatis, Eric B. Laber, and Marie Davidian. A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics*, 68(4):1010–1018, 2012.

[29] Ying-Qi Zhao, Donglin Zeng, Eric B. Laber, and Michael R. Kosorok. New Statistical Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

[30] Yingqi Zhao, Donglin Zeng, a John Rush, and Michael R Kosorok. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, 107(449):1106–1118, 2012.

Appendices

A Proof of Lemma 2.1.1

Lemma A.1. *Suppose the following conditions hold.*

1. $\forall \boldsymbol{a} \in \mathbb{R}^p$,$\exists \delta > 0$ *,such that*

   (a) $\mathbb{E}\left|\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right|^{2+\delta} < \infty$

   (b) $\left\{\boldsymbol{a}^\intercal V \left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right] \boldsymbol{a}\right\}^{1+\frac{\delta}{2}} < \infty.$

Then, we have, for any fixed $\boldsymbol{\theta}$*,*

$$\sqrt{n}\left(\nabla \widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left(\nabla \widehat{V}_j(\boldsymbol{\theta})\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right)\right)$$

The proof of this is similar to the proof of Lemma 1.1.3 and is shown in APPENDIX.

Proof. For any $\boldsymbol{a} \in \mathbb{R}^p$, we let $W_{ni} = \boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$. For each value of n, $w_{n1}, w_{n2}, \cdots, w_{nn}$ are i.i.d, and functions of the sample size n. This is because that $\boldsymbol{X}_i$ are assumed to be i.i.d., and h is a function of sample size n. Then, we have

$$\mu_n := \mathbb{E}W_{ni} = \mathbb{E}\left(\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right),$$

and

$$\sigma_n^2 := V(W_{ni}) = \boldsymbol{a}^\intercal V\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right) \boldsymbol{a}$$

We let $G_{ni} = W_{ni} - \mu_n$, and $T_n = \sum_{i=1}^n G_{ni}$. Also, we let $s_n^2 = V(T_n) = \sum_{i=1}^n V(G_{ni}) = \sum_{i=1}^n \sigma_n^2 = n\sigma_n^2$, where the second equality is because of independence, and the last equality is due to identicalness. Therefore, ${}^{T_n}/_{s_n}$ has mean 0, and variance 1. If we can show G_{ni} satisfying the Lyapunov condition, then we have

$$\frac{T_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1), \text{ as } n \to \infty$$

,

Now, we check the Lyapunov condition, that is, [1, 7]

$$\exists \delta > 0, \text{ such that } \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} \mid G_{n,i} \mid^{2+\delta} \to 0, \text{ as } n \to 0.$$

We define, for any $\boldsymbol{a}$,

$$C_1 \triangleq \mathbb{E}\,|G_{ni}|^{2+\delta} = \mathbb{E}\,|W_{ni} - \mu_n|^{2+\delta} = \mathbb{E}\left|\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) - \mu_n\right|^{2+\delta},$$

and

$$C_2 \triangleq s_n^{2+\delta} = n^{1+\frac{\delta}{2}}\sigma_n^{2+\delta} = n^{1+\frac{\delta}{2}}\left\{\boldsymbol{a}^\intercal V\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}.$$

Then, we have

$$\begin{aligned}
&\frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} \mid G_{n,i} \mid^{2+\delta} \\
=&\frac{\mathbb{E}\left|\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) - \mu_n\right|^{2+\delta}}{n^{\frac{\delta}{2}}\left\{\boldsymbol{a}^\intercal V\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}\left(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}\right)\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}} \\
=&\frac{C_1}{n^{\frac{\delta}{2}} C_2}.
\end{aligned}$$

As long as $\delta > 0$, for finite C_1 and finite C_2, we have ${}^{C_1}/_{n^{\frac{\delta}{2}}C_2} \to 0$, as $n \to \infty$. This means that the Lyapunov condition is satisfied, if $\mathbb{E}\,|G_{ni}|^{2+\delta}$ and $s_n^{2+\delta}$ are finite. Then, by Lyapunov Central Limit Theorem, we have

$$\frac{T_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

As this hold for any arbitary non-random vector $\boldsymbol{a} \in \mathbb{R}^p$, we have, by Cramer-Wold Theorem, that

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) - \mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right\}\right] \xrightarrow{d} \mathcal{N}\left(0, V\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]\right),$$

as $n \to \infty$. We denote $\boldsymbol{L}_{ni} = \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$, then this is written as

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right] \xrightarrow{d} \mathcal{N}\left(0, V\left[\boldsymbol{L}_{n1}\right]\right).$$

Then, we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{[V(\boldsymbol{L}_{n1})/n]^{1/2}} \frac{[V(\boldsymbol{L}_{n1})/n]^{1/2}}{\left[AV(\boldsymbol{L}_{n1})/n\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1).$$

As $n \to \infty$,

$$\frac{V(\boldsymbol{L}_{n1})^{1/2}}{AV(\boldsymbol{L}_{n1})^{1/2}} \to 1,$$

then we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{[AV(\boldsymbol{L}_{n1})/n]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),$$

i.e.,

$$\sqrt{n}\left[1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right] \xrightarrow{d} N\left(0, AV(\boldsymbol{L}_{n1})\right).$$

As $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) = \nabla\widehat{V}_j(\boldsymbol{\theta})$, we have

$$\sqrt{n}\left[\nabla\widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left\{\nabla\widehat{V}_j(\boldsymbol{\theta})\right\}\right] \xrightarrow{d} \mathcal{N}\left(0, AV\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]\right)$$

■

B Proof of Corollary 2.1.2

Corollary B.1. *Suppose all the assumptions in Lemma 3 hold, and* $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ *is a consistent estimator of* $F_{Y_j^*(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$*. Then, we have*

$$\sqrt{n}\left(\nabla\widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) - \nabla V_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)}\right)\right)$$

Proof. We write

$$\begin{aligned}&\nabla\widehat{V}_j(\boldsymbol{\theta}) - \nabla V_j^*(\boldsymbol{\theta})\\ =&\nabla\widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left(\nabla\widehat{V}_j(\boldsymbol{\theta})\right) + \mathbb{E}\left(\nabla\widehat{V}_j(\boldsymbol{\theta})\right) - \nabla V_j^*(\boldsymbol{\theta}),\end{aligned}$$

where $\mathbb{E}\left(\nabla \widehat{V}_j(\boldsymbol{\theta})\right) - \nabla V_j^*(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right) -$
$\mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, dF_{Y_j^* x(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right) = o_p(1)$, due to the consistency of $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ and dominated convergence theorem.

In lemma 2.1.1, let $\boldsymbol{\theta} = \boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)$ and then

$$\sqrt{n}\left(\nabla \widehat{V}_j(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)) - \nabla \mathbb{E}\left(\widehat{V}_j(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu))\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)}\right)\right).$$

As $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ is consistent, we have

$$\frac{AV\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]}{AV\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, dF_{Y^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]} \xrightarrow{p} 1.$$

Then, we have

$$\sqrt{n}\left(\nabla \widehat{V}_j(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)) - \nabla V_j\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)}\right)\right).$$

■

C Proof of Theorem 2.1.3

Theorem C.1. *Suppose all the assumptions above hold. Then we have, as* $n \to \infty$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) - \boldsymbol{\theta}_{\boldsymbol{\nu}}(\mu)^*\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,
$\boldsymbol{C}^* = \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right) \nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right) - \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right)\mathbb{E}\left(\nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right)$,
and $\boldsymbol{D}^* = \nabla^2 \phi_\mu^{BP}(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu))$.

Proof. For notation simplicity in this proof, let $\phi(\boldsymbol{\theta}) = \phi_\mu^{PB}(\boldsymbol{\theta})$ and $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{\phi}_\mu^{PB}(\boldsymbol{\theta})$ for this proof. Also, let $\boldsymbol{\theta}^* = \boldsymbol{\theta}_\nu^*(\mu)$ and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_\nu(\mu)$ here. Recall $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^J \ln \widehat{v}_j(\boldsymbol{\theta}) + \frac{1}{2\mu} \sum_{t=1}^T (\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1)^2$. As $\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1 = 0$ is always satisfied as a constraint, the gradient is $\nabla \widehat{\phi}(\boldsymbol{\theta}) = \nabla \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^J {}^{\nabla \widehat{v}_j(\boldsymbol{\theta})}/_{\widehat{v}_j(\boldsymbol{\theta})}$. Taylor expansion of $\nabla \widehat{\phi}\left(\boldsymbol{\theta}^*\right)$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ shows that

$$\nabla \widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla \widehat{\phi}(\widehat{\boldsymbol{\theta}}) - \nabla^2 \widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + o_p(1),$$

where $\tilde{\boldsymbol{\theta}}$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. As $\widehat{\boldsymbol{\theta}}$ is the maximizer of $\widehat{\phi}(\boldsymbol{\theta})$, it satisfies the first order condition that $\nabla \widehat{\phi}(\widehat{\boldsymbol{\theta}}) = 0$. Therefore,

$$\sqrt{n} \nabla \widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n} \nabla^2 \widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{7}$$

where $\nabla\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J}{}^{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}/_{\widehat{v}_j\left(\boldsymbol{\theta}\right)}$. Recall $v_1\left(\boldsymbol{\theta}\right) = -V_1\left(\boldsymbol{\theta}\right)$ and $v_j\left(\boldsymbol{\theta}\right) = V_j\left(\boldsymbol{\theta}\right) - \nu_j$, for $j = 2, \cdots, J$. Due to Corollary 2.1.2, together with (A.4) and (A.5),

$$\sqrt{n}\Big(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \nabla v_1(\boldsymbol{\theta}^*)\Big) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right), \tag{8}$$

where $\boldsymbol{C}^* = AV\Big(\nabla v_1(\boldsymbol{\theta}^*)\Big) = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\nabla v_1(\boldsymbol{\theta}^*)\mathbb{E}\nabla^\intercal v_1(\boldsymbol{\theta}^*)$
$= AV\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i})\right)$. That is, Then, due to (B.1) and (B.2), we have

$$\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}{\widehat{v}_j\left(\boldsymbol{\theta}\right)} - \sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta})}{v_j(\boldsymbol{\theta})} = o_p(1). \tag{9}$$

Note $v_j(\boldsymbol{\theta}) > 0$, for $j = 2, \cdots, J$, is implied by the log barrier operator. Put (B.2) and (B.3) together by Slutsky's theorem, we have

$$\sqrt{n}\left\{\left(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}^*\right)}{\widehat{v}_j(\boldsymbol{\theta}^*)}\right) - \left(\nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta}^*)}{v_j(\boldsymbol{\theta}^*)}\right)\right\} \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

Due to the stationarity of $\boldsymbol{\theta}^*$, $\nabla\phi(\boldsymbol{\theta}^*) = \nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}{}^{\nabla v_j(\boldsymbol{\theta}^*)}/_{v_j(\boldsymbol{\theta}^*)} = 0$. Together with Sluskty's theorem, we have

$$\sqrt{n}\nabla\widehat{\phi}(\boldsymbol{\theta}^*) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

where $\boldsymbol{C}^* = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\right\}\mathbb{E}\left\{\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\}$.

As $\sqrt{n}\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ stated in (A.7),we have

$$\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, \boldsymbol{C}^*) \tag{10}$$

The Hessian is $\nabla^2\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla^2\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J}{}^{\left(\nabla^2\widehat{v}_j\left(\boldsymbol{\theta}\right)\widehat{v}_j\left(\boldsymbol{\theta}\right) - \left(\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)\right)^2\right)}/_{\widehat{v}_j^2\left(\boldsymbol{\theta}\right)}$. Based on (A.4) and (A.5), we have

$$\boldsymbol{D}^* \triangleq_p \lim_{n\to\infty}\nabla^2\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla^2\phi(\boldsymbol{\theta}^*) = \nabla^2 v_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla^2 v_j\left(\boldsymbol{\theta}^*\right)v_j\left(\boldsymbol{\theta}^*\right) - \left\{\nabla v_j\left(\boldsymbol{\theta}^*\right)\right\}^2}{v_j^2\left(\boldsymbol{\theta}^*\right)}. \tag{11}$$

As $\tilde{\boldsymbol{\theta}}$ is a vector in-between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$, we have $\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}}) = \nabla^2\widehat{\phi}(\boldsymbol{\theta}^*) + o_p(1)$. Therefore, based on (A.10) and (A.11), we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$, $\boldsymbol{C}^* = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\nabla v_1(\boldsymbol{\theta}^*)\mathbb{E}\nabla^\intercal v_1(\boldsymbol{\theta}^*)$ and $\boldsymbol{D}^* = \nabla^2\phi(\boldsymbol{\theta}^*)$. ■