Supplemental Material for "Estimation of Optimal Treatment Regimes Using Lists"

1 Introduction

In Section 2, we present the proofs of Theorems 1 and 2 in the main text. In Section 3, we describe our estimation algorithm in detail and prove Proposition 1. We also briefly discuss the extension of our algorithm to handle an arbitrary number of covariates per clause and summarize the algorithm proposed by Zhang et al. (2015). In Section 4, we present additional simulation results, including the computation time and the optimal treatment selection rate. In Section 5, we describe the covariates used in the real data example.

2 Proofs of the Theorems

2.1 Overview

An overview of our proofs is as follows. We first derive risk bounds on $\widehat{Q}_T$ and $\widehat{\pi}_T$ and then recursively derive bounds for $t = T-1, T-2, \ldots, 1$. In deriving each of these bounds, we first establish a bound on the difference $\widehat{Q}_t - Q_t$, and subsequently infer bounds on the distance between $\widehat{\pi}_t$ and π_t^*. Due to their discrete nature, there is no natural distance between two decision lists. Instead, we make a clause-by-clause comparison between $\widehat{\pi}_t$ and π_t^*, starting with the first clause. At each clause, we examine the distance between $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$ and $(R_{t\ell}^*, a_{t\ell}^*)$, as measured by the ρ_t-distance defined in Section 3 of the main text. We are thereby able to obtain the bound

$$\Pr_{\boldsymbol{X}}\{\widehat{\pi}_t(\boldsymbol{X}_t) \neq \pi_t^*(\boldsymbol{X}_t)\} \leq \sum_{\ell=1}^{L_{\max}} \left\{\Pr(\widehat{a}_{t\ell} \neq a_{t\ell}^*) + \rho_t(\widehat{R}_{t\ell}, R_{t\ell}^*)\right\},$$

where $\Pr_{\boldsymbol{X}}$ is the probability measure with respect to $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$ only. An upper bound on the reduction in value $V_t(\pi_t^*) - V_t(\widehat{\pi}_t)$ is obtained similarly. In the following paragraphs, we provide an overview of the derivation of the risk bounds for the estimated Q-functions and the distance between $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$ and $(R_{t\ell}^*, a_{t\ell}^*)$.

In order to establish the risk bound for $\widehat{Q}_t$, which is estimated via kernel ridge regression, we first present some auxiliary results on the properties of the RKHS induced by Gaussian kernel with multiple scaling factors (in Section 2.4). Though most of the results are straightforward extensions of known results for the Gaussian kernel with a single scaling factor, to the best of our knowledge they are not found elsewhere in the literature and so we include them for completeness. After reviewing these auxiliary results, we decompose the difference between $\widehat{Q}_t$ and Q_t into two terms, known as the approximation error (in Section 2.5) and the estimation error (in Section 2.6), which we bound separately. Our bounds employ a technique similar to that used in Steinwart and Christmann (2008), except to for two major changes to accommodate the following issues:
(1) the estimation of $\widehat{Q}_t(\cdot, a)$ only utilizes a random subset of samples with $A_{it} = a$, and hence the randomness in A_{it} must be handled properly;
(2) the estimation of $\widehat{Q}_t(\cdot, a)$ with $t < T$ is based on the extrapolated responses using the estimated Q-functions in later stages, rather than based on the (unobserved) true responses.

In order to derive the convergence properties of the estimated list $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$, we first transform the criterion function so that the estimation of $(R_{t\ell}, a_{t\ell})$ falls into the M-estimation framework (in Section 2.7). We emphasize that optimization with respect to $R \in \mathcal{R}_t$ involves the form of R, the indexes j_1, j_2 and the threshold values τ_1, τ_2. Thus, it will not only determine which variables should be used to define R, but also the corresponding threshold values and the direction of the inequalities. Therefore, the variables used in each clause are chosen in a data-driven way, and different variables may be chosen for different clauses.

Though we put the computation of $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$ into an M-estimation formulation, we observe the following facts that prevents us from using the well established theory of M-estimator:
(1) at each stage t, the product space $\mathcal{R}_t \times \mathcal{A}_t$ is neither a vector space not equipped with a natural definition of norm or distance;
(2) except for the first clause, the M-function involves estimated quantities such as previously estimated clauses. Because the estimated clauses do not have an influence function expansion, they must be handled using new techniques.
In order to handle these two issues, we establish a few auxiliary inequalities for decision lists (in Section 2.8).

After laying the groundwork as described above, the proofs of Theorem 1 (in Section 2.9) and Theorem 2 (in Section 2.10) follow directly.

2.2 Notation

For vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^q$, define component-wise operations $\boldsymbol{u}^p = (u_1^p, \ldots, u_q^p)^T$, $p \in \mathbb{R}$, and $\boldsymbol{u} \circ \boldsymbol{v} = (u_1 v_1, \ldots, u_q v_q)^T$. For $V \subset \mathbb{R}^q$, define $u \circ V = \{u \circ v : v \in V\}$. In addition, $\boldsymbol{u}$ is said to be positive if each of its components is positive.

Let $\boldsymbol{O}_i$ be the collection of random variables associated with the ith subject. For any function f, define $\mathbb{P}_n(f) = n^{-1} \sum_{i=1}^n f(\boldsymbol{O}_i)$. For any measurable function f defined on $D \subset \mathbb{R}^q$, we write $\|f\|_2 = \big(\int_D f^2 d\mu \big)^{1/2}$ and $\|f\|_\infty = \inf\{t \in \mathbb{R} : \mu(|f| > t) = 0\}$, where μ is the Lebesgue measure on D. Let (T, d) be a metric space and S be a subset of T. For $\varepsilon > 0$, the ε-covering number of S is defined by $\mathcal{N}(S,d,\varepsilon)=\inf\{n\geq 1:$ there exists $t_1,\cdots,t_n\in T$ such that $S\subset\bigcup_{i=1}^{n}B(t_i,\varepsilon)\}$, where $\inf\emptyset=\infty$ and $B(t,\varepsilon)=\{u\in T: d(u,t)\leq\varepsilon\}$ is the a ball with center t and radius ε. If $(T,\|\cdot\|)$ is a normed vector space, the ε-covering number is defined by viewing T as a metric space with induced metric $d(s,t)=\|s-t\|$. Let $(T,\|\cdot\|)$ be a normed vector space. The unit ball of T is defined by $\mathcal{B}_T=\{t:\|t\|\leq 1\}$. Given a scalar $w\in\mathbb{R}$ and a set $S\subset T$, define $wS=\{ws: s\in S\}$.

In the following proofs, c and c_i denote generic constants.

2.3 Concentration inequalities

We first state Talagrand's inequality (Bousquet, 2002, Theorem 2.3; see also Massart, 2000, Theorem 3 and Boucheron et al., 2013, Theorem 12.5).

Proposition 1. *Let $\mathcal{F}$ be a countable set of functions. Suppose* $\mathrm{E}(f)=0$, $\mathrm{E}(f^2)\leq V$, $\|f\|_\infty\leq B$ *for all* $f\in\mathcal{F}$. *Denote* $Z=\sup_{f\in\mathcal{F}}|\mathbb{P}_n(f)|$. *Then for all* $\tau>0$,

$$\Pr\left[Z\geq \mathrm{E}(Z)+\left\{\frac{2V\tau+4B\tau\cdot\mathrm{E}(Z)}{n}\right\}^{1/2}+\frac{B\tau}{3n}\right]\leq e^{-\tau}.$$

Corollary 2. *Under the conditions in Proposition 1,*

$$\Pr\left\{Z\geq 2\,\mathrm{E}(Z)+\left(\frac{2V\tau}{n}\right)^{1/2}+\frac{2B\tau}{n}\right\}\leq e^{-\tau}.$$

Proof. It is clear that

$$\left\{\frac{2V\tau+4B\tau\cdot\mathrm{E}(Z)}{n}\right\}^{1/2}\leq\left(\frac{2V\tau}{n}\right)^{1/2}+\left\{\frac{4B\tau\cdot\mathrm{E}(Z)}{n}\right\}^{1/2}\leq\left(\frac{2V\tau}{n}\right)^{1/2}+\frac{B\tau}{n}+\mathrm{E}(Z).$$

Note that we use a larger constant for simplicity. □

When the variance of f is not available, we have the following proposition (Boucheron et al., 2013, Theorem 12.1).

Proposition 3. *Let $\mathcal{F}$ be a countable set of functions. Suppose* $\mathrm{E}(f) = 0$, $\|f\|_\infty \le B$ *for all* $f \in \mathcal{F}$. *Denote* $Z = \sup_{f \in \mathcal{F}} |\mathbb{P}_n(f)|$. *Then for all* $\tau > 0$, *we have*

$$\Pr\left\{ Z \ge \mathrm{E}(Z) + \left(\frac{2B^2\tau}{n}\right)^{1/2} \right\} \le e^{-\tau}.$$

Next, we establish bounds on $\mathrm{E}(Z)$.

Proposition 4. *Let $\mathcal{F}$ be a countable set of functions which contains the zero function. Assume*

$$\sup_Q \log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(Q)}, \varepsilon) \le \psi(\varepsilon)$$

for some function $\psi(\cdot)$, *where the supremum is taken over all discrete probability measures* Q. *Suppose* $\mathrm{E}(f) = 0$, $\mathrm{E}(f^2) \le V$, $\|f\|_\infty \le B$ *for all* $f \in \mathcal{F}$. *Denote* $Z = \sup_{f \in \mathcal{F}} |\mathbb{P}_n(f)|$. *Then we have*

$$\mathrm{E}(Z) \le 1024 \left(\frac{BJ_V}{n}\right) + 64 \left(\frac{VJ_V}{n}\right)^{1/2},$$

where $J_V = \int_0^1 \psi(V^{1/2}\varepsilon)\, d\varepsilon$.

Proof. Without loss of generality, we assume $B = 1$. The general case can be obtained by scaling f. The proof extends the idea in Boucheron et al. (2013, Lemma 13.5).

Let $\sigma_1, \dots, \sigma_n$ be i.i.d. Rademacher random variables, i.e., $\Pr(\sigma = 1) = \Pr(\sigma = -1) = 1/2$. By the symmetrization inequality (van der Vaart and Wellner, 1996, Lemma 2.3.1), we have $\mathrm{E}(n^{1/2} Z) \le 2\, \mathrm{E}\left\{ \sup_f |n^{1/2}\, \mathbb{P}_n(\sigma f)| \right\}$.

Conditional on all random variables except σ_is, by Hoeffding's inequality, the process $n^{1/2}\, \mathbb{P}_n(\sigma f)$ is subgaussian with respect to the metric $\|f - g\|_{L^2(\mathbb{P}_n)} = \left[\mathbb{P}_n\{(f-g)^2\}\right]^{1/2}$. Hence the chaining technique (van der Vaart and Wellner, 1996, Corollary 2.2.8) implies

$$\mathrm{E}_\sigma \left\{ \sup_f |n^{1/2}\, \mathbb{P}_n(\sigma f)| \right\} \le 4 \int_0^{\eta_n} \left\{ \log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mathbb{P}_n)}, \varepsilon) \right\}^{1/2} d\varepsilon,$$

where E_σ denote the expectation with respect to $\sigma_1, \ldots, \sigma_n$ only and $\eta_n^2 = \max\{\sup_f \mathbb{P}_n(f^2), V\}$. Hence, we obtain

$$\mathrm{E}_\sigma \left\{ \sup_f |n^{1/2}\, \mathbb{P}_n(\sigma f)| \right\} \leq 4 \int_0^{\eta_n} \psi^{1/2}(\varepsilon)\, d\varepsilon = 4\eta_n \int_0^1 \psi^{1/2}(\eta_n \varepsilon)\, d\varepsilon \leq 4\eta_n \int_0^1 \psi^{1/2}(V^{1/2}\varepsilon)\, d\varepsilon.$$

Because $\log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mathbb{P}_n)}, \varepsilon) \leq \psi(\varepsilon)$ and $\psi(\varepsilon)$ is a decreasing function in ε.

Taking the other layer of expectation, we get

$$\mathrm{E}(n^{1/2} Z) \leq 8\{\mathrm{E}(\eta_n)\} \int_0^1 \psi^{1/2}(V^{1/2}\varepsilon)\, d\varepsilon \leq 8 J_V^{1/2} \{\mathrm{E}(\eta_n^2)\}^{1/2}$$

by Jensen's inequality. Also, we have $\mathrm{E}(\eta_n^2) \leq \mathrm{E}\left\{ \sup_f |\mathbb{P}_n(f^2) - \mathrm{E}(f^2)| \right\} + V$, since $\mathrm{E}(f^2) \leq V$ for all f. By the symmetrization inequality (van der Vaart and Wellner, 1996, Lemma 2.3.1), we have $\mathrm{E}\left\{\sup_f |\mathbb{P}_n(f^2) - \mathrm{E}(f^2)|\right\} \leq 2\,\mathrm{E}\left\{\sup_f |\mathbb{P}_n(\sigma f^2)|\right\}$. By the contraction inequality (van der Vaart and Wellner, 1996, Proposition A.3.2) and $\|f\|_\infty \leq 1$, we have $\mathrm{E}\left\{ \sup_f |\mathbb{P}_n(\sigma f^2)| \right\} \leq 4\,\mathrm{E}\left\{ \sup_f |\mathbb{P}_n(\sigma f)| \right\}$. By the desymmetrization inequality (van der Vaart and Wellner, 1996, Lemma 2.3.6), we have $\mathrm{E}\{\sup_f |\mathbb{P}_n(\sigma f)|\} \leq 2\,\mathrm{E}\{\sup_f |\mathbb{P}_n(f)|\}$. Combining these inequalities yields $\mathrm{E}(\eta_n^2) \leq 16\,\mathrm{E}(Z) + V$.

Therefore,

$$n^{1/2}\,\mathrm{E}(Z) \leq 8\{16\,\mathrm{E}(Z) + V\}^{1/2} J_V^{1/2}.$$

Solving for $\mathrm{E}(Z)$, shows $\mathrm{E}(Z) \leq (2n)^{-1}\{a + (a^2 + 4nb)^{1/2}\} \leq n^{-1}a + n^{-1/2}b^{1/2}$ with $a = 1024 J_V$ and $b = 64 V J_V$. Hence, $\mathrm{E}(Z) \leq 1024 n^{-1} J_V + 64 n^{-1/2} V^{1/2} J_V^{1/2}$. □

Proposition 5. *Let* $\mathcal{F}$ *be a countable set of functions which contains the zero function. Assume*

$$\sup_Q \log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(Q)}, \varepsilon) \leq \psi(\varepsilon)$$

for some function $\psi(\cdot)$*, where the supremum is taken over all discrete probability measures* Q*. Suppose* $\mathrm{E}(f) = 0$*,* $\|f\|_\infty \leq B$ *for all* $f \in \mathcal{F}$*. Denote* $Z = \sup_{f\in\mathcal{F}} |\mathbb{P}_n(f)|$*. Then we* *have*

$$\mathrm{E}\, Z \le 8 \left(\frac{B^2 J_B}{n} \right)^{1/2},$$

where $J_B = \int_0^1 \psi(B\varepsilon)\, d\varepsilon$.

Proof. Just apply the trivial bound $|\eta_n| \le B$ in the proof of Proposition 4. $\square$

Though all the propositions in this subsection assume that $\mathcal{F}$ is countable, they all apply if $\mathcal{F}$ is uncountable and separable as $\Pr\left\{\sup_{f\in\mathcal{F}} |\mathbb{P}_n(f)| = \sup_{f\in\mathcal{F}'} |\mathbb{P}_n(f)|\right\} = 1$ for some countable subset $\mathcal{F}' \subset \mathcal{F}$.

2.4 Properties of the RKHS

We establish several useful properties of the RKHS $\mathbb{H}$ induced by the Gaussian kernel with individual scaling factors for each dimension

$$K_{\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{z}) = \exp\left\{ -\sum_{j=1}^{q} \gamma_j (x_j - z_j)^2 \right\},$$

where $\boldsymbol{x}, \boldsymbol{z} \in D \subset \mathbb{R}^q$. When all γ_js are identical, the properties of $\mathbb{H}$ are well studied (see, e.g., Steinwart and Christmann, 2008). The lemmas below extend those properties to RKHS induced by Gaussian kernel with multiple scaling factors.

From here to Section 2.6, we use q to denote dimension, whose value will depends on the context. For example, when we analyze the estimated Q-function at stage t, then $q = d_t$.

We may omit $\boldsymbol{\gamma}$ and write $K(\cdot, \cdot)$ when the value of $\boldsymbol{\gamma}$ is clear from the context. Similarly, to emphasize the dependence of $\mathbb{H}$ on the parameter $\boldsymbol{\gamma}$ and the domain D, we may write $\mathbb{H}_{\boldsymbol{\gamma}}$, $\mathbb{H}(D)$, or $\mathbb{H}_{\boldsymbol{\gamma}}(D)$.

The following lemma provides a feature map of the Gaussian kernel.

Lemma 6. *Define the function* $\phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}} : \mathbb{R}^q \to L^2(\mathbb{R}^q)$ *by*

$$\phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}(\boldsymbol{u}) = \left(\frac{4}{\pi}\right)^{q/4} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/4} \exp\left\{-\sum_{j=1}^{q} 2\gamma_j (x_j - u_j)^2\right\},\ \boldsymbol{x} \in D,\ \boldsymbol{u} \in \mathbb{R}^q.$$

Then $\phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}$ *is a feature map of* $K_{\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{z})$*.*

Proof. Straightforward calculation similar to Steinwart and Christmann (2008, Lemma 4.45) gives $\langle \phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}, \phi_{\boldsymbol{\gamma}}^{\boldsymbol{z}} \rangle_{L^2(\mathbb{R}^q)} = K_{\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{z})$. By definition, $\phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}$ is a feature map. □

The following lemma shows that $\mathbb{H}_{\boldsymbol{\gamma}}(D)$ can be embedded into $\mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}(D)$ if $\gamma_j < \widetilde{\gamma}_j$ for all $j = 1, \ldots, q$.

Lemma 7. *Let* $\boldsymbol{\gamma}$*,* $\widetilde{\boldsymbol{\gamma}}$ *be two positive vectors satisfying* $\gamma_j < \widetilde{\gamma}_j$ *for all* j*. If* $f \in \mathbb{H}_{\boldsymbol{\gamma}}$*, then* $f \in \mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}$ *and* $\|f\|_{\mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}} \le \left(\prod_{j=1}^{q} \widetilde{\gamma}_j\right)^{1/4} \left(\prod_j \gamma_{j=1}^{q}\right)^{-1/4} \|f\|_{\mathbb{H}_{\boldsymbol{\gamma}}}$*.*

Proof. We follow the same strategy as in Steinwart and Christmann (2008, Theorem 4.46). Because $f \in \mathbb{H}_{\boldsymbol{\gamma}}$, by Steinwart and Christmann (2008, Theorem 4.21), there exists $g \in L^2(\mathbb{R}^q)$ such that $f(\boldsymbol{x}) = \langle \phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}, g \rangle_{L^2(\mathbb{R}^q)}$ for all $\boldsymbol{x} \in D$.

Given $\boldsymbol{s} \in \mathbb{R}^q$ with $s_j > 0$ for all j, define the operator $W_s : L^2(\mathbb{R}^q) \to L^2(\mathbb{R}^q)$ by

$$(W_{\boldsymbol{s}} g)(\boldsymbol{v}) = \int_{\mathbb{R}^q} \pi^{-q/2} \left(\prod_{j=1}^{q} s_j\right)^{-1/2} \exp\left\{-\sum_{j=1}^{q} s_j^{-1} (v_j - u_j)^2\right\} g(\boldsymbol{u}) d\boldsymbol{u},\ \text{for } \boldsymbol{v} \in \mathbb{R}^q.$$

For any $g \in L^2(\mathbb{R}^q)$ and any $\boldsymbol{v} \in \mathbb{R}^q$, straightforward calculation using properties of normal densities shows $(W_{\boldsymbol{s}_1} W_{\boldsymbol{s}_2} g)(\boldsymbol{v}) = (W_{\boldsymbol{s}_1 + \boldsymbol{s}_2} g)(\boldsymbol{v})$, hence, $W_{\boldsymbol{s}_1} W_{\boldsymbol{s}_2} = W_{\boldsymbol{s}_1 + \boldsymbol{s}_2}$.

Define $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^{\mathrm{T}}$ and $\widetilde{\boldsymbol{\tau}} = (\widetilde{\tau}_1, \ldots, \widetilde{\tau}_q)^{\mathrm{T}}$, where $\tau_j = 1/\gamma_j$ and $\widetilde{\tau}_j = 1/\widetilde{\gamma}_j$. The assumption $\gamma_j < \widetilde{\gamma}_j$ implies $\tau_j > \widetilde{\tau}_j$. We observe that

$$f = \langle \phi_{\boldsymbol{\gamma}}^{\boldsymbol{x}}, g \rangle_{L^2(\mathbb{R}^q)} = (W_{\boldsymbol{\tau}/2} g) \cdot \pi^{q/4} \left(\prod_{j=1}^{q} \gamma_j\right)^{-1/4}.$$

Because

$$W_{\boldsymbol{\tau}/2}g = W_{\widetilde{\boldsymbol{\tau}}/2}W_{(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}})/2}g = \langle \phi^{\boldsymbol{x}}_{\widetilde{\boldsymbol{\gamma}}}, W_{(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}})/2}g\rangle_{L^2(\mathbb{R}^q)} \cdot \pi^{-q/4}\left(\prod_{j=1}^{q}\widetilde{\gamma}_j\right)^{1/4},$$

it follows that

$$f = \langle \phi^{\boldsymbol{x}}_{\widetilde{\boldsymbol{\gamma}}}, W_{(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}})/2}g\rangle_{L^2(\mathbb{R}^q)} \cdot \left(\prod_{j=1}^{q}\gamma_j\right)^{-1/4}\left(\prod_{j=1}^{q}\widetilde{\gamma}_j\right)^{1/4}.$$

By Steinwart and Christmann (2008, Theorem 4.21), $f \in \mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}$.

Moreover, $\|f\|_{\mathbb{H}_{\boldsymbol{\gamma}}} = \|g\|_{L^2(\mathbb{R}^q)}$ and $\|f\|_{\mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}} = \|W_{(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}})/2}g\|_{L^2(\mathbb{R}^q)} \cdot \left(\prod_{j=1}^{q}\gamma_j\right)^{-1/4}\left(\prod_{j=1}^{q}\widetilde{\gamma}_j\right)^{1/4}$. By Young's inequality, $\|W_{(\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}})/2}g\|_{L^2(\mathbb{R}^q)} \leq \|g\|_{L^2(\mathbb{R}^q)}$. Hence,

$$\|f\|_{\mathbb{H}_{\widetilde{\boldsymbol{\gamma}}}} \leq \|g\|_{L^2(\mathbb{R}^q)}\left(\prod_{j=1}^{q}\gamma_j\right)^{-1/4}\left(\prod_{j=1}^{q}\widetilde{\gamma}_j\right)^{1/4} \leq \|f\|_{\mathbb{H}_{\boldsymbol{\gamma}}}\left(\prod_{j=1}^{q}\gamma_j\right)^{-1/4}\left(\prod_{j=1}^{q}\widetilde{\gamma}_j\right)^{1/4}.$$

□

The following lemma establishes an isometric isomorphism between $\mathbb{H}_{\boldsymbol{\alpha}^{-2}\circ\boldsymbol{\gamma}}(\boldsymbol{\alpha}\circ D)$ and $\mathbb{H}_{\boldsymbol{\gamma}}(D)$ for any fixed $\boldsymbol{\alpha}$.

Lemma 8. *Let $\boldsymbol{\alpha}$ be an arbitrary positive vector. We define a mapping $\tau_{\boldsymbol{\alpha}} : L^\infty(D) \to L^\infty(\boldsymbol{\alpha}\circ D)$ as follows: given a function $f \in L^\infty(D)$, let $\tau_{\boldsymbol{\alpha}}(f)(\boldsymbol{x}) = f(\boldsymbol{\alpha}^{-1}\circ\boldsymbol{x})$ for $\boldsymbol{x} \in \boldsymbol{\alpha}\circ D$. Then, for all $f \in \mathbb{H}_{\boldsymbol{\gamma}}(D)$, we have $\tau_{\boldsymbol{\alpha}}(f) \in \mathbb{H}_{\boldsymbol{\alpha}^{-2}\circ\boldsymbol{\gamma}}(\boldsymbol{\alpha}\circ D)$ and $\|\tau_{\boldsymbol{\alpha}}(f)\|_{\mathbb{H}_{\boldsymbol{\alpha}^{-2}\circ\boldsymbol{\gamma}}(\boldsymbol{\alpha}\circ D)} = \|f\|_{\mathbb{H}_{\boldsymbol{\gamma}(D)}}$.*

Proof. It is easy to verify that the arguments in Steinwart and Christmann (2008, Proposition 4.37) remain valid when scalar multiplication is replaced by component-wise multiplication between vectors. □

The following lemma computes the covering number of the unit ball in $\mathbb{H}_{\boldsymbol{\gamma}}(D)$.

Lemma 9. *Suppose $D \subset s\mathcal{B}_{\mathbb{R}^q}$. For any integer $m \geq 1$,*

$$\log \mathcal{N}\{\mathcal{B}_{\mathbb{H}_{\boldsymbol{\gamma}}(D)}, \|\cdot\|_\infty, \varepsilon\} \leq c_{m,q,s} \prod_{j=1}^{q} (1+\gamma_j)^{1/2} \varepsilon^{-q/m},$$

where $c_{m,q,s}$ is a constant that depends on m, q and s only.

Proof. Let $\mathbf{1}$ be the vector of ones. By Lemma 3, $\mathbb{H}_{\boldsymbol{\gamma}}(D)$ is isometric isomorphic to $\mathbb{H}_{\mathbf{1}}(\boldsymbol{\gamma}^{1/2} \circ D)$. Thus, it suffices to compute the covering number for $\mathbb{H}_{\mathbf{1}}(\boldsymbol{\gamma}^{1/2} \circ D)$.

Define $\widetilde{D} = \boldsymbol{\gamma}^{1/2} \circ D$. It is shown that $\mathbb{H}_{\mathbf{1}}(\widetilde{D})$ can be embedded into $\mathbb{C}^m(\widetilde{D})$ (Steinwart and Christmann, 2008, Theorem 6.26). By Steinwart and Christmann (2008, Corollary 4.36), the embedding map from $\mathbb{H}_{\mathbf{1}}(\widetilde{D})$ to $\mathbb{C}^m(\widetilde{D})$ is continuous, and hence bounded. Thus, there exists a constant c_1 which depends only on m such that $\|f\|_{\mathbb{C}^m(\widetilde{D})} \leq c_1 \|f\|_{\mathbb{H}_{\mathbf{1}}(\widetilde{D})}$ for all $f \in \mathbb{H}_{\mathbf{1}}(\widetilde{D})$. Hence, we have

$$\mathcal{N}\{\mathcal{B}_{\boldsymbol{H}_{\mathbf{1}}(\widetilde{D})}, \|\cdot\|_\infty, \varepsilon\} \leq \mathcal{N}(c_1 \mathcal{B}_{\mathbb{C}^m(\widetilde{D})}, \|\cdot\|_\infty, \varepsilon\} = \mathcal{N}(\mathcal{B}_{\mathbb{C}^m(\widetilde{D})}, \|\cdot\|_\infty, \varepsilon/c_1).$$

By Theorem 2.7.1 in van der Vaart and Wellner (1996), there exists a constant c_2 that depends only on m and q such that

$$\log \mathcal{N}\{\mathcal{B}_{\mathbb{C}^m(\widetilde{D})}, \|\cdot\|_\infty, \varepsilon) \leq c_2 \mu(\{\boldsymbol{x} : \|\boldsymbol{x} - \widetilde{D}\| \leq 1\}) \varepsilon^{-q/m},$$

where μ is the Lebesgue measure on $\mathbb{R}^q$. Because $D \subset s\mathcal{B}_{\mathbb{R}^q}$ and $(1+su^{1/2}) \leq (1+s)(1+u)^{1/2}$ for all $u \geq 0$,

$$\mu(\{\boldsymbol{x} : \|\boldsymbol{x} - \widetilde{D}\| \leq 1\}) \leq \prod_{j=1}^{q} (1 + s\lambda_j^{1/2}) \leq (1+s)^q \prod_{j=1}^{q} (1+\lambda_j)^{1/2}.$$

$\square$

2.5 Approximation error in kernel ridge regression

Define $\widetilde{Y}_T = Y_T$ and $\widetilde{Y}_t = Y_t + Q_{t+1}\{\boldsymbol{X}_{t+1}, \pi^*_{t+1}(\boldsymbol{X}_{t+1})\}$ for $t < T$. Then, $Q_t(\boldsymbol{x}, a) = E(\widetilde{Y}_t | \boldsymbol{X}_t = \boldsymbol{x}, A_t = a)$ for all t. Fix a stage t and a treatment $a \in \mathcal{A}_t$. For notational simplicity, we shall omit the subscripts t and a hereafter. Let q be the dimension of $\boldsymbol{X}$. Given a function $f \in L^\infty(D)$, we define

$$\mathcal{L}(f) = \mathrm{E}\left[I(A = a)\big\{\widetilde{Y} - f(\boldsymbol{X})\big\}^2\right]$$

and

$$f_0 = \mathop{\arg\min}_{f: D \to \mathbb{R}, \text{ measurable}} \mathcal{L}(f).$$

Simple calculations show that $f_0(\boldsymbol{x}) = E(\widetilde{Y} | \boldsymbol{X} = \boldsymbol{x}, A = a)$ almost surely with respect to the distribution of $\boldsymbol{X}$, say $P_{\boldsymbol{X}}$. Hence, f_0 is exactly $Q_t(\cdot, a)$. In addition,

$$\mathcal{L}(f) - \mathcal{L}(f_0) = E\left[I(A = a)\left\{f(\boldsymbol{X}) - f_0(\boldsymbol{X})\right\}^2\right].$$

The function f_0 need not belong to the RKHS $\mathbb{H}_{\boldsymbol{\gamma}}$. Nevertheless, the estimator must belong to $\mathbb{H}_{\boldsymbol{\gamma}}$. The following proposition shows that it is always possible to find an $f \in \mathbb{H}_{\boldsymbol{\gamma}}$ such that f and f_0 are close. The following proposition is a stronger version of Eberts and Steinwart (2013, Theorems 2.2 and 2.3) which allows multiple scaling factors and separates signal and noise variables.

Proposition 10. *Suppose f_0 satisfies the modulus of smoothness condition $\omega_r(f_0, s) \le c_1 s^r$ for some positive integer r and $\|f_0\|_\infty \le B$ for some constant B. Let $\mathcal{S}$ denote the indices of signal variables in f_0, i.e., the value of $f(\boldsymbol{x})$ only depends on $\boldsymbol{x}_{\mathcal{S}}$. Then, there exists some $f \in \mathbb{H}_{\boldsymbol{\gamma}}$ such that*

$$\lambda \|f\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \|f - f_0\|^2_\infty \le c\left\{\lambda\big(\max_{j\in\mathcal{S}} \gamma_j\big)^{|\mathcal{S}|/2}\big(\max_{j\in\mathcal{S}^c} \gamma_j\big)^{|\mathcal{S}^c|/2} + \big(\min_{j\in\mathcal{S}} \gamma_j\big)^{-r}\right\}$$

and $\|f\|_\infty \le 2^r B$, where c is some constant that depends on c_1, r, B and $|\mathcal{S}|$ only.

Proof. Define

$$W(\boldsymbol{x}, \boldsymbol{u}) = \sum_{i=1}^{r} \binom{r}{i} (-1)^{i-1} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} i^{-q} \exp\left\{-\sum_{j=1}^{q} 2\gamma_j (x_j - u_j)^2 / i^2\right\},$$

where $\boldsymbol{x}, \boldsymbol{u} \in \mathbb{R}^q$. Let $f(\boldsymbol{x}) = \int_{\mathbb{R}^q} W(\boldsymbol{x}, \boldsymbol{u}) f_0(\boldsymbol{u})\, d\boldsymbol{u}$, $\boldsymbol{x} \in D$.

Then, for every $\boldsymbol{x} \in D$,

$$f(\boldsymbol{x}) = \sum_{i=1}^{r} \binom{r}{i} (-1)^{i-1} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} \int_{\mathbb{R}^q} i^{-q} \exp\left\{-\sum_{j=1}^{q} 2\gamma_j (x_j - u_j)^2 / i^2\right\} f_0(\boldsymbol{u}) d\boldsymbol{u}.$$

Apply the change of variables $h_j = (u_j - x_j)/i$ so that

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{r} \binom{r}{i} (-1)^{i-1} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} \int_{\mathbb{R}^q} \exp\left\{-\sum_{j=1}^{q} 2\gamma_j h_j^2\right\} f_0(\boldsymbol{x} + i\boldsymbol{h}) d\boldsymbol{h} \\ &= \int_{\mathbb{R}^q} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} \exp\left(-\sum_{j=1}^{q} 2\gamma_j h_j^2\right) \sum_{i=1}^{r} \binom{r}{i} (-1)^{i-1} f_0(\boldsymbol{x} + i\boldsymbol{h}) d\boldsymbol{h}. \end{aligned}$$

Note that

$$f_0(\boldsymbol{x}) = \int_{\mathbb{R}^q} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} \exp\left(-\sum_{j=1}^{q} 2\gamma_j h_j^2\right) f_0(\boldsymbol{x}) d\boldsymbol{h},$$

therefore

$$|f(\boldsymbol{x}) - f_0(\boldsymbol{x})| \leq \int_{\mathbb{R}^q} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} \exp\left(-\sum_{j=1}^{q} 2\gamma_j h_j^2\right) |\Delta_{\boldsymbol{h}}^r (f_0, \boldsymbol{x})|\, d\boldsymbol{h}.$$

Because $f_0(\boldsymbol{x}) = f_0^*(\boldsymbol{x}_{\mathcal{S}})$ for some function $f_0^* : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}$,

$$|\Delta_{\boldsymbol{h}}^r (f_0, \boldsymbol{x})| = \left|\Delta_{\boldsymbol{h}_{\mathcal{S}}}^r (f_0^*, \boldsymbol{x}_{\mathcal{S}})\right| \leq \omega_r(f_0^*, \|\boldsymbol{h}_{\mathcal{S}}\|_2) = \omega_r(f_0, \|\boldsymbol{h}_{\mathcal{S}}\|_2).$$

Thus,

$$|f(\boldsymbol{x}) - f_0(\boldsymbol{x})| \leq \int_{\mathbb{R}^{|\mathcal{S}|}} \left(\frac{2}{\pi}\right)^{|\mathcal{S}|/2} \left(\prod_{j} \gamma_{\mathcal{S},j}\right)^{1/2} \exp\left(-\sum_{j} 2_{\mathcal{S},j} h_{\mathcal{S},j}^2\right) \omega_r(f_0, \|\boldsymbol{h}_{\mathcal{S}}\|_2) d\boldsymbol{h}_{\mathcal{S}}.$$

Because $\omega_r(f_0,t) \le (1+t/s)^r\omega_r(f_0,s)$ for all $s,t>0$, it follows that

$$\begin{aligned}\omega_r(f_0,\|\boldsymbol{h}_{\mathcal{S}}\|_2) &\le \left\{1+\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{1/2}\|\boldsymbol{h}_{\mathcal{S}}\|_2\right\}^r \omega_r\left\{f_0,\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1/2}\right\}\\ &\le (1+\|\boldsymbol{\gamma}_{\mathcal{S}}\circ\boldsymbol{h}_{\mathcal{S}}\|_2)^r\,\omega_r\left\{f_0,\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1/2}\right\}.\end{aligned}$$

Combining these inequalities,

$$\begin{aligned}|f(\boldsymbol{x})-f_0(\boldsymbol{x})| \le \omega_r&\left\{f_0,\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1/2}\right\}\cdot\\ &\int_{\mathbb{R}^{|\mathcal{S}|}}\left(\frac{2}{\pi}\right)^{|\mathcal{S}|/2}\left(\prod_j\gamma_{\mathcal{S},j}\right)^{1/2}\exp\left\{-\sum_j 2\gamma_{\mathcal{S},j}h_{\mathcal{S},j}^2\right\}(1+\|\boldsymbol{\gamma}_{\mathcal{S}}\circ\boldsymbol{h}_{\mathcal{S}}\|_2)^r\,d\boldsymbol{h}_{\mathcal{S}}\end{aligned}$$

Using the change of variables $t_j=\gamma_{\mathcal{S},j}h_{\mathcal{S},j}$, we can see that the integral above is a constant that depends only on $|\mathcal{S}|$. Denote this integral by c_2, then

$$|f(\boldsymbol{x})-f_0(\boldsymbol{x})| \le c_2\omega_r\left\{f_0,\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1/2}\right\} \le c_1c_2\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1/2}.$$

Note that $W(\boldsymbol{x},\boldsymbol{u})=\sum_{i=1}^r\binom{r}{i}(-1)^{i-1}\pi^{-q/4}i^{-q/2}\left(\prod_{j=1}^q\gamma_j\right)^{1/4}\phi^{\boldsymbol{x}}_{\boldsymbol{\gamma}/i^2}(\boldsymbol{u})$, where ϕ is the feature map defined in Lemma 1. Let $g_i(\boldsymbol{x})=\int_{\mathbb{R}^q}\phi^{\boldsymbol{x}}_{\boldsymbol{\gamma}/i^2}(\boldsymbol{u})f_0(\boldsymbol{u})d\boldsymbol{u}$, then $g_i\in\mathbb{H}_{\boldsymbol{\gamma}/i^2}$. By Lemma 2, we have $g_i\in\mathbb{H}_{\boldsymbol{\gamma}}$ and the $\mathbb{H}_{\boldsymbol{\gamma}}$ norm of g_i is at most $i^{q/2}$ times its $\mathbb{H}_{\boldsymbol{\gamma}/i^2}$ norm. Thus,

$$\|f\|_{\mathbb{H}} \le \sum_{i=1}^r\binom{r}{i}\pi^{-q/4}\left(\prod_{j=1}^q\gamma_j\right)^{1/4}\|f_0\|_2 \le 2^r\pi^{-q/4}\big(\max_{j\in\mathcal{S}}\gamma_j\big)^{|\mathcal{S}|/4}\big(\max_{j\in\mathcal{S}^c}\gamma_j\big)^{|\mathcal{S}^c|/4}\|f_0\|_2.$$

Therefore,

$$\begin{aligned}\lambda\|f\|_{\mathbb{H}}^2+\mathcal{L}(f)-\mathcal{L}(f_0) &= \lambda\|f\|_{\mathbb{H}}^2+\mathrm{E}\left\{f(\boldsymbol{X})-f_0(\boldsymbol{X})\right\}^2\\ &\le 2^{2r}\pi^{-q/2}B^2\lambda\big(\max_{j\in\mathcal{S}}\gamma_j\big)^{|\mathcal{S}|/2}\big(\max_{j\in\mathcal{S}^c}\gamma_j\big)^{|\mathcal{S}^c|/2}+c_1^2c_2^2\big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-1}.\end{aligned}$$

In addition, for any $\boldsymbol{x} \in D$, it follows that

$$\begin{aligned}|f(\boldsymbol{x})| &\leq \sum_{i=1}^{r} \binom{r}{i} \int_{\mathbb{R}^q} \left(\frac{2}{\pi}\right)^{q/2} \left(\prod_{j=1}^{q} \gamma_j\right)^{1/2} i^{-q} \exp\left\{-\sum_{j=1}^{q} 2\gamma_j (x_j - u_j)^2 / i^2\right\} d\boldsymbol{u} \cdot \|f_0\|_\infty \\ &= \sum_{i=1}^{r} \binom{r}{i} \|f_0\|_\infty \leq 2^r B.\end{aligned}$$

$\square$

2.6 Risk bounds for kernel ridge regression

Define $\widehat{Y}_T = Y_T$ and $\widehat{Y}_t = Y_t + \widehat{Q}_{t+1}\{\boldsymbol{X}_{t+1}, \widehat{\pi}_{t+1}(\boldsymbol{X}_{t+1})\}$ for $t < T$. By Assumption 1, we have $Y_t \in [-b, b]$. Hence it is safe to restrict the estimated Q-functions within the interval $[-B, B]$ for some sufficiently large but fixed B. To this end, define the truncation operator $\mathcal{T}_B : L^\infty(D) \to L^\infty(D)$ as

$$\mathcal{T}_B(f)(\boldsymbol{x}) = f(\boldsymbol{x}) I\{-B \leq f(\boldsymbol{x}) \leq B\} + B I\{f(\boldsymbol{x}) > B\} + (-B) I\{f(\boldsymbol{x}) < -B\}, \quad \boldsymbol{x} \in D.$$

For any function f, g, we have $|\mathcal{T}_B(f)(\boldsymbol{x}) - \mathcal{T}_B(g)(\boldsymbol{x})| \leq |f(\boldsymbol{x}) - g(\boldsymbol{x})|$. Hence, we have $\|\mathcal{T}_B(f) - \mathcal{T}_B(g)\|_\infty \leq \|f - g\|_\infty$. As a consequence, for any $B \geq \|f_0\|_\infty$, we have

$$\mathcal{L}\{\mathcal{T}_B(f)\} - \mathcal{L}(f_0) = \mathrm{E}\{\mathcal{T}_B(f)(\boldsymbol{X}) - f_0(\boldsymbol{X})\}^2 \leq \mathrm{E}\{f(\boldsymbol{X}) - f_0(\boldsymbol{X})\}^2 = \mathcal{L}(f) - \mathcal{L}(f_0).$$

Given sequences $\boldsymbol{\gamma}_n$ and λ_n, the estimator of the Q-function is $\widehat{Q}_t(\cdot, a) = \mathcal{T}_B(\widehat{f}_n)$, where

$$\widehat{f}_n = \operatorname*{arg\,min}_{f \in \mathbb{H}_{\boldsymbol{\gamma}}} \mathbb{P}_n \left[I(A = a) \big\{\widehat{Y} - f(\boldsymbol{X})\big\}^2 \right] + \lambda \|f\|_{\mathbb{H}_{\boldsymbol{\gamma}}}^2.$$

To facilitate our analysis, we define

$$d_n = \operatorname*{arg\,min}_{f \in \mathbb{H}_{\boldsymbol{\gamma}}} \mathbb{P}_n \left[I(A = a) \big\{\widetilde{Y} - f(\boldsymbol{X})\big\}^2 \right] + \lambda \|f\|_{\mathbb{H}_{\boldsymbol{\gamma}}}^2.$$

Note that we omit the subscript n in $\boldsymbol{\gamma}_n$ and λ_n for simplicity. The difference between $\widehat{f}_n$ and d_n is that we use $\widetilde{Y}_t = Y_t + Q_{t+1}\{\boldsymbol{X}_{t+1}, \pi^*_{t+1}(\boldsymbol{X}_{t+1})\}$ for $t < T$ when defining $\widehat{f}_n$, which is an unobserved quantity as it relies on π^*_{t+1} and Q_{t+1}. In contrast, we replace π^*_{t+1} and Q_{t+1} by their estimates $\widehat{\pi}_{t+1}$ and $\widehat{Q}_{t+1}$ to obtain $\widehat{Y}_t$. Hence $\widehat{Q}_t(\cdot, a)$ is based on observed quantities only.

In this Section, we will show that the difference between $\mathcal{T}_B(\widehat{f}_n)$ and $f_0 = Q_t(\cdot, a)$ is small. To be precise, define $\mathcal{E}(f) = \lambda\|f\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathcal{L}\{\mathcal{T}_B(f)\} - \mathcal{L}(f_0)$. Our goal is to show that $\mathcal{E}(\widehat{f}_n)$ is small with large probability. The proof below follows the idea in Steinwart and Christmann (2008, Theorem 7.20) while accounting for the error in the responses. For notational convenience, define $\overline{\gamma}_{\mathcal{S}} = 1+\max_{j\in\mathcal{S}}\gamma_j$, $\underline{\gamma}_{\mathcal{S}} = \min_{j\in\mathcal{S}}\gamma_j$ and $\overline{\gamma}_{\mathcal{S}^c} = 1+\max_{j\in\mathcal{S}^c}\gamma_j$. For any f, define $\ell_f = I(A=a)\big\{\widetilde{Y} - f(\boldsymbol{X})\big\}^2$ and $h_f = \ell_f - \ell_{f_0}$. Then, $\mathcal{L}(f) - \mathcal{L}(f_0) = \mathrm{E}(h_f)$. Thus, $\mathrm{E}(h_f) \geq 0$ for all f.

Lemma 11. *For any $f \in \mathbb{H}_{\boldsymbol{\gamma}}$, we have*

$$\mathcal{E}(\widehat{f}_n) \leq \lambda\|f\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n(h_f) - \mathbb{P}_n(h_{\widehat{f}_n}) + \mathrm{E}\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\} + 2\,\mathbb{P}_n\left(\widehat{Y} - \widetilde{Y}\right)^2.$$

Proof. By the definition of $\widehat{f}_n$ and d_n, we have

$$\begin{aligned}
\lambda\|\widehat{f}_n\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n\left[I(A=a)\big\{\widehat{Y} - \widehat{f}_n(\boldsymbol{X})\big\}^2\right] &\leq \lambda\|d_n\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n\left[I(A=a)\big\{\widehat{Y} - d_n(\boldsymbol{X})\big\}^2\right],\\
\lambda\|d_n\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n\left[I(A=a)\big\{\widetilde{Y} - d_n(\boldsymbol{X})\big\}^2\right] &\leq \lambda\|f\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n\left[I(A=a)\big\{\widetilde{Y} - f(\boldsymbol{X})\big\}^2\right].
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\lambda\|\widehat{f}_n\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} \leq{}& \lambda\|f\|^2_{\mathbb{H}_{\boldsymbol{\gamma}}} + \mathbb{P}_n(h_f) - \mathbb{P}_n(h_{d_n})\\
&+ \mathbb{P}_n\left[I(A=a)\big\{\widehat{Y} - d_n(\boldsymbol{X})\big\}^2\right] - \mathbb{P}_n\left[I(A=a)\big\{\widehat{Y} - \widehat{f}_n(\boldsymbol{X})\big\}^2\right].
\end{aligned}$$

For any real number a_1, a_2, b_1, b_2, it follows that

$$\begin{aligned}(a_1-b_1)^2-(a_1-b_2)^2 &= (2a_1-b_1-b_2)(b_2-b_1)\\ &= (2a_2-b_1-b_2)(b_2-b_1)+2(a_1-a_2)(b_2-b_1)\\ &\le (a_2-b_1)^2-(a_2-b_2)^2+(a_1-a_2)^2+(b_1-b_2)^2.\end{aligned}$$

Hence,

$$\begin{aligned}&\mathbb{P}_n\left[I(A=a)\big\{\widehat{Y}-d_n(\boldsymbol{X})\big\}^2\right]-\mathbb{P}_n\left[I(A=a)\big\{\widehat{Y}-\widehat{f}_n(\boldsymbol{X})\big\}^2\right]\\ &\le \mathbb{P}_n\left(h_{d_n}\right)-\mathbb{P}_n\left(h_{\widehat{f}_n}\right)+\mathbb{P}_n\left[I(A=a)\big(\widehat{Y}-\widetilde{Y}\big)^2\right]+\mathbb{P}_n\left[I(A=a)\big\{\widehat{f}_n(\boldsymbol{X})-d_n(\boldsymbol{X})\big\}^2\right].\end{aligned}$$

Let $\widehat{\boldsymbol{Y}}$ be the vector of $\widehat{Y}_i$, $i\in\mathcal{I}_a$, $\widetilde{\boldsymbol{Y}}$ the vector of $\widetilde{Y}_i$, $i\in\mathcal{I}_a$ and $\boldsymbol{K}$ the matrix of $K(\boldsymbol{X}_i,\boldsymbol{X}_j)$, $i,j\in\mathcal{I}_a$, where $\mathcal{I}_a=\{i:A_i=a\}$. By the representer theorem and the fact that all the eigenvalues of $\boldsymbol{K}(\boldsymbol{K}+\lambda\boldsymbol{I})^{-1}$ are less than one, it follows that

$$\big\|\{\widehat{f}_n(\boldsymbol{X}_i)\}_{i\in\mathcal{I}_a}-\{d_n(\boldsymbol{X}_i)\}_{i\in\mathcal{I}_a}\big\|_2=\|\boldsymbol{K}(\boldsymbol{K}+\lambda\boldsymbol{I})^{-1}(\widehat{\boldsymbol{Y}}-\widetilde{\boldsymbol{Y}})\|_2\le\|\widehat{\boldsymbol{Y}}-\widetilde{\boldsymbol{Y}}\|_2.$$

Thus, the inequality in the lemma follows from

$$\mathbb{P}_n\left[I(A=a)\big\{\widehat{f}_n(\boldsymbol{X})-d_n(\boldsymbol{X})\big\}^2\right]\le\mathbb{P}_n\left[I(A=a)\big(\widehat{Y}-\widetilde{Y}\big)^2\right].$$

□

Proposition 12. *Suppose* $\Pr\left\{\mathbb{P}_n(\widehat{Y}-\widetilde{Y})^2\ge c_1n^{-\alpha}+c_2n^{-\beta}\tau\right\}\le e^{-\tau}$ *for some* $\alpha,\beta>0$, *and* f_0 *satisfies the conditions in Proposition 10. Then for any* $\delta>0$ *and* $\tau>0$,

$$\begin{aligned}\Pr\Bigg[\mathrm{E}_{\boldsymbol{X}}&\left\{\mathcal{T}_B(\widehat{f}_n)(\boldsymbol{X})-f_0(\boldsymbol{X})\right\}^2\ge\\ &c\left\{\lambda\bar{\gamma}_{\mathcal{S}}^{|\mathcal{S}|/2}\bar{\gamma}_{\mathcal{S}^c}^{|\mathcal{S}^c|/2}+\underline{\gamma}_{\mathcal{S}}^{-r}+\bar{\gamma}_{\mathcal{S}}^{|\mathcal{S}|/2}\bar{\gamma}_{\mathcal{S}^c}^{|\mathcal{S}^c|/2}\lambda^{-\delta}n^{-1}+n^{-\alpha}+n^{-\min(\beta,1)}\tau\right\}\Bigg]\le e^{-\tau},\end{aligned}$$

where c *is a constant that depends on* δ, q, r, B *and* ϖ *only, and* $\mathrm{E}_{\boldsymbol{X}}$ *denotes the expectation with respect to* $\boldsymbol{X}$ *only.*

Proof. By Proposition 10 and the inequality $E\left[I(A=a)\left\{f(\boldsymbol{X})-f_0(\boldsymbol{X})\right\}^2\right] \le \|f-f_0\|_\infty^2$, there exists some function $f_n \in \mathbb{H}_{\gamma}$ such that

$$\lambda\|f_n\|_{\mathbb{H}_\gamma}^2 + \mathrm{E}\left(h_{f_n}\right) \le c\left\{\lambda\big(\max_{j\in\mathcal{S}}\gamma_j\big)^{|\mathcal{S}|/2}\big(\max_{j\in\mathcal{S}^c}\gamma_j\big)^{|\mathcal{S}^c|/2} + \big(\min_{j\in\mathcal{S}}\gamma_j\big)^{-r}\right\} \tag{1}$$

for some constant c independent of n, and $\|f\|_\infty \le 2^r B$.

By the property of the truncation operator and the fact that $\|\widehat{Y}\|_\infty \le B$ with probability 1, we have $\mathbb{P}_n\, h_{\mathcal{T}_B(\widehat{f}_n)} \le \mathbb{P}_n\, h_{\widehat{f}_n}$. We apply Lemma 11 with $f = f_n$ to obtain

$$\begin{aligned}
\mathcal{E}(\widehat{f}_n) &\le \lambda\|f_n\|_{\mathbb{H}_\gamma}^2 + \mathbb{P}_n\left(h_{f_n}\right) - \mathbb{P}_n\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\} + \mathrm{E}\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\} + \mathbb{P}_n\left\{(\widehat{Y}-\widetilde{Y})^2\right\}\\
&\le \left\{\lambda\|f_n\|_{\mathbb{H}_\gamma}^2 + \mathrm{E}(h_{f_n})\right\} + \left|\mathbb{P}_n\left(h_{f_n}\right) - \mathrm{E}\left(h_{f_n}\right)\right|\\
&\quad + \left|\mathrm{E}\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\} - \mathbb{P}_n\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\}\right| + \mathbb{P}_n\left\{(\widehat{Y}-\widetilde{Y})^2\right\}.
\end{aligned}$$

Note that $\mathrm{E}\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\}$ is defined as first computing $h_{\mathcal{T}_B(f)}$ and then plugging in $f = \widehat{f}_n$. Thus, $\mathrm{E}\left\{h_{\mathcal{T}_B(\widehat{f}_n)}\right\}$ is a random variable.

We will consider the four terms in the right hand side of the above display formula separately. The first term can be bounded above using equation (1). The fourth term is controlled by the condition. So we will focus on the second and the third terms.

For the second term, we first observe that

$$|h_{f_n}| \le |\{Y-f_n(\boldsymbol{X})\}^2 - \{Y-f_0(\boldsymbol{X})\}^2| = |\{f_n(\boldsymbol{X})+f_0(\boldsymbol{X})-2Y\}\{f_n(\boldsymbol{X})-f_0(\boldsymbol{X})\}|.$$

Because $\|f_0\|_\infty \le \widetilde{B}$ and $\|f_n\|_\infty \le \widetilde{B}$ for $\widetilde{B} = 2^r B$, we have $E\big(h_{f_n}^2\big) \le 16\widetilde{B}^2\,\mathrm{E}\left[\{f_n(\boldsymbol{X}) - f_0(\boldsymbol{X})\}^2\right] = 16\widetilde{B}^2\,\mathrm{E}\left(h_{f_n}\right)$ and $|h_{f_n}| \le 8\widetilde{B}^2$. By Bernstein's inequality (Steinwart and Christmann, 2008, Theorem 6.12), we obtain

$$\Pr\left[\left|\mathbb{P}_n(h_{f_n}) - \mathrm{E}(h_{f_n})\right| \ge \frac{16\widetilde{B}^2\tau}{3n} + \left\{\frac{32\widetilde{B}^2\tau\,\mathrm{E}(h_{f_n})}{n}\right\}^{1/2}\right] \le 2e^{-\tau}.$$

Using $2(uv)^{1/2} \le u + v$, it follows that

$$\left\{\frac{32\widetilde{B}^2\tau\,\mathrm{E}(h_{f_n})}{n}\right\}^{1/2} \le \frac{8\widetilde{B}^2\tau}{n} + \mathrm{E}(h_{f_n}) \le \frac{8\widetilde{B}^2\tau}{n} + \mathrm{E}(h_{f_n}) + \lambda\|f_n\|^2_{\mathbb{H}_\gamma}.$$

Therefore,

$$\Pr\left\{\left|\,\mathbb{P}_n(h_{f_n}) - \mathrm{E}(h_{f_n})\right| \ge \frac{14\widetilde{B}^2\tau}{n} + \mathrm{E}(h_{f_n}) + \lambda\|f_n\|^2_{\mathbb{H}_\gamma}\right\} \le 2e^{-\tau}. \tag{2}$$

Bounding the third term is a little bit more involved. Let $s > 0$ be fixed; for any $f \in \mathbb{H}_\gamma$, define

$$m_f = \frac{h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}}{\mathcal{E}(f) + s} = \frac{h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}}{\lambda\|f\|^2_{\mathbb{H}_\gamma} + \mathrm{E}\{h_{\mathcal{T}_B(f)}\} + s}.$$

Because $\|\mathcal{T}_B(f)\|_\infty \le B$, we have $\|m_f\|_\infty \le 16B^2/s$. Furthermore, because $\mathrm{E}\{h^2_{\mathcal{T}_B(f)}\} \le 16B^2\,\mathrm{E}\{h_{\mathcal{T}_B(f)}\}$, we have

$$\mathrm{E}(m_f^2) \le \frac{\mathrm{E}\{h^2_{\mathcal{T}_B(f)}\}}{4s\,\mathrm{E}\{h_{\mathcal{T}_B(f)}\}} \le \frac{4B^2}{s},$$

when $\mathrm{E}\{h_{\mathcal{T}_B(f)}\} > 0$, and $\mathrm{E}\,h^2_{\mathcal{T}_B(f)} = 0 \le 4B^2/s$ when $\mathrm{E}\{h_{\mathcal{T}_B(f)}\} = 0$.

Define $\mathcal{F}_s = \{f \in \mathbb{H}_\gamma : \mathcal{E}(f) \le s\} \cup \{0\}$, where 0 denotes the zero function. By Corollary 2, it follows that

$$\Pr\left\{\sup_{f\in\mathcal{F}_s}|m_f| \ge 2\,\mathrm{E}\left(\sup_{f\in\mathcal{F}_s}|m_f|\right) + \left(\frac{8B^2\tau}{ns}\right)^{1/2} + \frac{32B^2\tau}{ns}\right\} \le e^{-\tau}.$$

We shall derive an upper bound for $\mathrm{E}\left(\sup_{f\in\mathbb{H}_\gamma}|m_f|\right)$ based on an upper bound for $\mathrm{E}\left\{\sup_{f\in\mathcal{F}_s}|h_{\mathcal{T}_B(f)} - \mathrm{E}\,h_{\mathcal{T}_B(f)}|\right\}$. To this end, we compute the covering number for $\mathcal{G}_s = \{h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\} : f \in \mathcal{F}_s\}$.

For any $f \in \mathcal{F}_s$, we have $\|f\|_{\mathbb{H}_\gamma} \le s^{1/2}\lambda^{-1/2}$. Hence,

$$\mathcal{N}(\mathcal{F}_s, \|\cdot\|_\infty, \varepsilon) \le \mathcal{N}\{(s^{1/2}\lambda^{-1/2})\mathcal{B}_{\mathbb{H}_\gamma}, \|\cdot\|_\infty, \varepsilon\} = \mathcal{N}(\mathcal{B}_{\mathbb{H}_\gamma}, \|\cdot\|_\infty, s^{-1/2}\lambda^{1/2}\varepsilon).$$

By the fact that $\|\mathcal{T}_B(f) - \mathcal{T}_B(g)\|_\infty \le \|f - g\|_\infty$, we have

$$\|h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\} - h_{\mathcal{T}_B(g)} + \mathrm{E}\{h_{\mathcal{T}_B(g)}\}\|_\infty \le 8B\|f - g\|_\infty.$$

Hence, $\mathcal{N}(\mathcal{G}_s, \|\cdot\|_\infty, \varepsilon) \le \mathcal{N}\{\mathcal{F}_s, \|\cdot\|_\infty, \varepsilon/(8B)\}$. Combining these inequalities and applying Lemma 9, shows

$$\log \mathcal{N}(\mathcal{G}_s, \|\cdot\|_\infty, \varepsilon) \le \log \mathcal{N}\{\mathcal{B}_{\mathbb{H}_\gamma}, \|\cdot\|_\infty, (8B)^{-1} s^{-1/2} \lambda^{1/2} \varepsilon\} \le c_1 a_{\boldsymbol{\gamma}} s^{q/(2m)} \lambda^{-q/(2m)} \varepsilon^{-q/m},$$

where $m \ge 1$ is an arbitrary integer, c_1 is a constant that depends on m, q, B, r only, and $a_{\boldsymbol{\gamma}} = \prod_{j=1}^q (1 + \gamma_j)^{1/2} \le \overline{\gamma}_{\mathcal{S}}^{|\mathcal{S}|/2} \overline{\gamma}_{\mathcal{S}^c}^{|\mathcal{S}^c|/2}$.

For any $f \in \mathcal{F}_s$, we have $\|h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}\|_\infty \le 16B^2$ and $\operatorname{Var} h_{\mathcal{T}_B(f)} \le \mathrm{E}\, h^2_{\mathcal{T}_B(f)} \le 16B^2 s$. Apply Proposition 4 to obtain

$$\mathrm{E}\left[\sup_{f \in \mathcal{F}_s} \left|h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}\right|\right] \le 1024(16B^2 J n^{-1}) + 64(16B^2 J s n^{-1})^{1/2},$$

where $J = \int_0^1 c_1 a_{\boldsymbol{\gamma}} (16B^2)^{q/(2m)} \lambda^{-q/(2m)} \varepsilon^{-q/m}\, d\varepsilon \le c_2 a_{\boldsymbol{\gamma}} \lambda^{-q/(2m)}$. Thus,

$$\mathrm{E}\left[\sup_{f \in \mathcal{F}_s} |h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}|\right] \le c_3 \left\{a_{\boldsymbol{\gamma}} \lambda^{-q/(2m)} n^{-1} + a_{\boldsymbol{\gamma}}^{1/2} \lambda^{-q/(4m)} s^{1/2} n^{-1/2}\right\}.$$

Hence, by the peeling technique (Steinwart and Christmann, 2008, Theorem 7.7), we obtain

$$\mathrm{E}\left(\sup_{f \in \mathbb{H}_\gamma} |m_f|\right) \le 4c_3 \left\{a_{\boldsymbol{\gamma}} \lambda^{-q/(2m)} s^{-1} n^{-1} + a_{\boldsymbol{\gamma}}^{1/2} \lambda^{-q/(4m)} s^{-1/2} n^{-1/2}\right\}.$$

Combine the bound of $\mathrm{E}(\sup_{f \in \mathbb{H}_\gamma} |m_f|)$ and the tail bound of $\sup_{f \in \mathbb{H}_\gamma} |m_f|$ to obtain

$$\Pr\left[\sup_{f \in \mathbb{H}_\gamma} \frac{|h_{\mathcal{T}_B(f)} - \mathrm{E}\{h_{\mathcal{T}_B(f)}|\}}{\mathcal{E}(f) + s} \ge c_4 \left\{\frac{a_{\boldsymbol{\gamma}}}{\lambda^{q/(2m)} s n} + \frac{a_{\boldsymbol{\gamma}}^{1/2}}{\lambda^{q/(4m)} s^{1/2} n^{1/2}} + \frac{\tau^{1/2}}{s^{1/2} n^{1/2}} + \frac{\tau}{sn}\right\}\right] \le e^{-\tau},$$

where $c_4 > 0$ is some constant that depends on m, q, B, r only. Without loss of generality, we assume $c_4 \ge 1$.

Let

$$s = 64 c_4^2 \max\left\{ \frac{a_{\gamma}}{\lambda^{q/(2m)} n}, \frac{\tau}{n} \right\},$$

then

$$\frac{c_4^2 a_{\gamma}}{\lambda^{q/(2m)} s n} \leq \left(\frac{c_4^2 a_{\gamma}}{\lambda^{q/(2m)} s n} \right)^{1/2} \leq \frac{1}{8}, \quad \frac{c_4^2 \tau}{s n} \leq \left(\frac{c_4^2 \tau}{s n} \right)^{1/2} \leq \frac{1}{8}.$$

Therefore, we have

$$\Pr\left[|\mathbb{P}_n\{h_{\mathcal{T}_B(f)}\} - \mathrm{E}\{h_{\mathcal{T}_B(f)}\}| \geq \mathcal{E}(f)/2 + s/2 \text{ for some } f \in \mathbb{H}_{\gamma} \right] \leq e^{-\tau}. \tag{3}$$

Plug-in $f = \widehat{f}_n$ in equation (3) and combine equations (1), (2), (3) and the condition on $\mathbb{P}_n(\widehat{Y} - \widetilde{Y})^2$ to obtain

$$\Pr\left\{ \mathcal{E}(\widehat{f}_n) \leq c_6 \left(\lambda \overline{\gamma}_{\mathcal{S}}^{|\mathcal{S}|/2} \overline{\gamma}_{\mathcal{S}^c}^{|\mathcal{S}^c|/2} + \underline{\gamma}_{\mathcal{S}}^{-r} + \overline{\gamma}_{\mathcal{S}}^{|\mathcal{S}|/2} \overline{\gamma}_{\mathcal{S}^c}^{|\mathcal{S}^c|/2} \lambda^{-q/(2m)} n^{-1} + n^{-1}\tau + n^{-\alpha} + n^{-\beta}\tau \right) \right\} \leq e^{-\tau}.$$

Because m can be arbitrarily large, $\delta = q/(2m)$ can be arbitrarily small.

The final result follow from

$$\begin{aligned} \mathcal{E}(\widehat{f}_n) &\geq \mathrm{E}_{\boldsymbol{X}} \left[I(A = a) \left\{ \mathcal{T}_B(\widehat{f}_n)(\boldsymbol{X}) - f_0(\boldsymbol{X}) \right\}^2 \right] \\ &= \mathrm{E}_{\boldsymbol{X}} \left[\Pr(A = a | \boldsymbol{X}) \left\{ \mathcal{T}_B(\widehat{f}_n)(\boldsymbol{X}) - f_0(\boldsymbol{X}) \right\}^2 \right] \\ &\geq \varpi \, \mathrm{E}_{\boldsymbol{X}} \left\{ \mathcal{T}_B(\widehat{f}_n)(\boldsymbol{X}) - f_0(\boldsymbol{X}) \right\}^2, \end{aligned}$$

where in the last inequality we use $\Pr(A = a | \boldsymbol{X}) \geq \varpi$ in view of Assumption 2. □

Recall that $\widehat{Q}_t(\cdot, a) = \mathcal{T}_B(\widehat{f}_n)(\cdot)$ and $Q_t(\cdot, a) = f_0(\cdot)$. We immediately obtain the following corollaries.

Corollary 13. *Assume the conditions in Proposition 12 hold. Furthermore, suppose* $\overline{\gamma}_{\mathcal{S}} = \overline{\theta}_{\mathcal{S}} n^{2/(2r+q)}$, $\underline{\gamma}_{\mathcal{S}} = \underline{\theta}_{\mathcal{S}} n^{2/(2r+q)}$, $\overline{\gamma}_{\mathcal{S}^c} = \overline{\theta}_{\mathcal{S}^c} n^{2/(2r+q)}$, *and* $\lambda = \theta_{\lambda} n^{-1}$, *where* q *is the dimension of* $\boldsymbol{X}$ *and* r *is the degree of the modulus of smoothness of* $Q(\cdot, a)$. *Then, for any* $\xi > 0$,

$$\Pr\left(\mathrm{E}_{\boldsymbol{X}} \left\{ \widehat{Q}_t(\boldsymbol{X}, a) - Q_t(\boldsymbol{X}, a) \right\}^2 \geq c \left[n^{-\min\{2r/(2r+q)+\xi, \alpha\}} + n^{-\min(\beta, 1)} \tau \right] \right) \leq e^{-\tau}.$$

Corollary 14. *Assume the conditions in Proposition 12 hold. Furthermore, suppose* $\overline{\gamma}_{\mathcal{S}} = \overline{\theta}_{\mathcal{S}} n^{2/(2r+|\mathcal{S}|)}$, $\underline{\gamma}_{\mathcal{S}} = \underline{\theta}_{\mathcal{S}} n^{2/(2r+|\mathcal{S}|)}$, $\overline{\gamma}_{\mathcal{S}^c} = \overline{\theta}_{\mathcal{S}^c}$, *and* $\lambda = \theta_\lambda n^{-1}$, *where* $|\mathcal{S}|$ *is the number of signal variables of* $Q(\cdot, a)$ *and* r *is the degree of the modulus of smoothness of* $Q(\cdot, a)$. *Then, for any* $\xi > 0$,

$$\Pr\left(\mathrm{E}_{\boldsymbol{X}}\left\{\widehat{Q}_t(\boldsymbol{X}, a) - Q_t(\boldsymbol{X}, a)\right\}^2 \geq c\left[n^{-\min\{2r/(2r+|\mathcal{S}|)+\xi, \alpha\}} + n^{-\min(\beta,1)}\tau\right]\right) \leq e^{-\tau}.$$

The convergence rate of $\widehat{Q}_t$ depends on two factors. First, the term $2r/(2r+q)$ or $2r/(2r+|\mathcal{S}|)$ reflects the curse of dimensionality and the smoothness of Q-function. If the dimension is smaller and the Q-function is smoother, the convergence is faster. Moreover, if the tuning parameters are chosen in a clever way in the sense of the conditions in *Corollary* 14, then the number of noise variables won't affect the convergence rate. This consequence is a major difference between Gaussian kernel with a single scaling factor and that with multiple scaling factors. Second, the terms α and β reflects the non-random errors in the response. The larger the values of α and β, the smaller the magnitude of non-random errors and hence the faster the convergence rate.

Hereafter, we shall only presents results assuming the tuning parameters are chosen following Corollary 13. Parallel results can be obtained by replacing q by $|\mathcal{S}|$.

2.7 Clause estimator as an M-estimator

In this Section, we will show that the estimator $\widehat{R}_{t\ell}, \widehat{a}_{t\ell}$ can be expressed as an M-estimator.

Define $U_t(\boldsymbol{x}, a) = \max_{a' \in \mathcal{A}_t} Q_t(\boldsymbol{x}, a') - Q_t(\boldsymbol{x}, a)$ and $\widehat{U}_t(\boldsymbol{x}, a) = \max_{a' \in \mathcal{A}_t} \widehat{Q}_t(\boldsymbol{x}, a') - Q_t(\boldsymbol{x}, a)$. Because

$$\left|\max_{a' \in \mathcal{A}_t} \widehat{Q}_t(\boldsymbol{x}, a') - \max_{a' \in \mathcal{A}_t} Q_t(\boldsymbol{x}, a)\right| \leq \max_{a' \in \mathcal{A}_t} \left|\widehat{Q}_t(\boldsymbol{x}, a') - Q_t(\boldsymbol{x}, a)\right|,$$

it follows that

$$\left|\widehat{U}_t(\boldsymbol{x}, a) - U_t(\boldsymbol{x}, a)\right| \leq 2 \max_{a' \in \mathcal{A}_t} \left|\widehat{Q}_t(\boldsymbol{x}, a') - Q_t(\boldsymbol{x}, a')\right|.$$

Thus, for any $p \geq 1$

$$\mathbb{P}_n \left|\widehat{U}_t(\boldsymbol{X_t}, a) - U_t(\boldsymbol{X_t}, a)\right|^p \leq 2 \sum_{a' \in \mathcal{A}_t} \left|\widehat{Q}_t(\boldsymbol{X_t}, a') - Q_t(\boldsymbol{X_t}, a')\right|^p. \tag{4}$$

By equation (4), we know that Corollaries 13 and 14 hold when $\widehat{Q}_t, Q_t$ are replaced by $\widehat{U}_t, U_t$, with a possibly larger constant c'.

Following the notation used in the main text, define

$$\widehat{\Omega}_{t\ell}(R, a) = I(\boldsymbol{X}_t \in \widehat{G}_{t\ell}, \boldsymbol{X}_t \in R) \left\{\widehat{U}_t(\boldsymbol{X}_t, a) - \zeta\right\} - \eta \left\{2 - V(R)\right\}$$

and

$$\Omega_{t\ell}(R, a) = I(\boldsymbol{X}_t \in G^*_{t\ell}, \boldsymbol{X}_t \in R) \left\{U_t(\boldsymbol{X}_t, a) - \zeta\right\} - \eta \left\{2 - V(R)\right\}.$$

By the definition of $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$ in the main article, we have

$$\begin{aligned}(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}) = \operatorname*{arg\,max}_{R \in \mathcal{R}_t, a \in \mathcal{A}_t} \mathbb{P}_n\, I(\boldsymbol{X}_t \in \widehat{G}_{t\ell}) \widehat{Q}_t\{\boldsymbol{X}_t, \widehat{\pi}^Q_t(\boldsymbol{X}_t)\} \\ - \mathbb{P}_n\, I(\boldsymbol{X}_t \in \widehat{G}_{t\ell}, \boldsymbol{X}_t \in R) \widehat{U}_t(\boldsymbol{X}_t, a) \\ + \mathbb{P}_n\, \zeta I\{\boldsymbol{X}_t \in \widehat{G}_{t\ell}, \boldsymbol{X}_t \in R\} + \eta\{2 - V(R)\}.\end{aligned}$$

Thus, we observe that $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}) = \arg\min_{R \in \mathcal{R}_t, a \in \mathcal{A}_t} \mathbb{P}_n\, \widehat{\Omega}_{t\ell}(R, a)$. Similarly,

$$(R^*_{t\ell}, a^*_{t\ell}) = \operatorname*{arg\,max}_{R \in \mathcal{R}_t, a \in \mathcal{A}_t} \Psi_{t\ell}(R, a) = \operatorname*{arg\,min}_{R \in \mathcal{R}_t, a \in \mathcal{A}_t} \mathrm{E}\, \Omega_{t\ell}(R, a).$$

In addition, because $0 \leq \widehat{U}_t \leq B$, for any $\zeta \geq B$ we will always end up with $\widehat{R}_t = \mathbb{R}^{d_t}$. Thus, hereafter we assume $\zeta \in [0, B]$.

2.8 Auxiliary results for the analysis of decision lists

Lemma 15. *$\mathcal{R}_t$ is a Vapnik-Cervonenkis class (VC class, hereafter).*

Proof. Recall that $\mathcal{R}_t$ consists of rectangles in $\mathbb{R}^{d_t}$ defined using at most two variables. Hence $\mathcal{R}_t$ is a subset of the set of all the intervals $\{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{d_t}\}$, where $(\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R}^{d_t} : a_j \leq x_j \leq b_j \text{ for all } j\}$. Hence $\mathcal{R}_t$ is a Vapnik-Cervonenkis (van der Vaart and Wellner, 1996, Example 2.6.1). $\square$

The following lemma gives an upper bound for

$$\sup_{R \in \mathcal{R}_t} \big| \mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R, a)\} - \mathrm{E}\{\Omega_{t\ell}(R, a)\} \big|$$

for any given $a \in \mathcal{A}_t$.

Lemma 16. *Let $\epsilon = \big[\mathrm{E} \big\{ \widehat{U}_t(\boldsymbol{X}_t, a) - U_t(\boldsymbol{X}_t, a) \big\}^2 \big]^{1/2} + B \sum_{k<\ell} \rho_t\big(\widehat{R}_{tk}, R^*_{tk}\big)$. We have, for any $\tau > 0$,*

$$\Pr\left\{ \sup_{R \in \mathcal{R}_t} |\mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R, a)\} - \mathrm{E}\{\Omega_{t\ell}(R, a)\}| \geq c(\epsilon + n^{-1/2} + n^{-1/2}\tau^{1/2}) \right\} \leq e^{-\tau},$$

where c is some constant independent of ϵ, n and τ.

Proof. We have

$$\begin{aligned} &\sup_{R \in \mathcal{R}_t} \big| \mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R, a)\} - \mathrm{E}\{\Omega_{t\ell}(R, a)\} \big| \\ &\qquad \leq \sup_{R \in \mathcal{R}_t} \big| \mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R, a)\} - \mathbb{P}_n\{\Omega_{t\ell}(R, a)\} \big| + \sup_{R \in \mathcal{R}_t} \big| \mathbb{P}_n\{\Omega_{t\ell}(R, a)\} - \mathrm{E}\{\Omega_{t\ell}(R, a)\} \big|. \end{aligned}$$

For the first term, we observe that

$$\begin{aligned}
&\sup_{R\in\mathcal{R}_t}\big|\,\mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R,a)\}-\mathbb{P}_n\{\Omega_{t\ell}(R,a)\}\big|\\
&\quad\le \sup_{R\in\mathcal{R}_t}\Big|\mathbb{P}_n\left[I(\boldsymbol{X}_t\in\widehat{G}_{t\ell}\cap G^*_{t\ell},\boldsymbol{X}_t\in R)\big\{\widehat{U}_t(\boldsymbol{X}_t,a)-U_t(\boldsymbol{X}_t,a)\big\}\right]\Big|\\
&\qquad+\sup_{R\in\mathcal{R}_t}\mathbb{P}_n\left[I(\boldsymbol{X}_t\in\widehat{G}_{t\ell}\triangle G^*_{t\ell},\boldsymbol{X}_t\in R)\big|\widehat{U}_t(\boldsymbol{X}_t,a)-\zeta\big|\right]\\
&\quad\le \mathbb{P}_n\left|\widehat{U}_t(\boldsymbol{X}_t,a)-U_t(\boldsymbol{X}_t,a)\right|+B\,\mathbb{P}_n\left\{I(\boldsymbol{X}_t\in\widehat{G}_{t\ell}\triangle G^*_{t\ell})\right\},
\end{aligned}$$

By the definition of $G_{t\ell}$, we have

$$\mathbb{P}_n\left\{I(\boldsymbol{X}_t\in\widehat{G}_{t\ell}\triangle G^*_{t\ell})\right\}\le\sum_{k<\ell}\mathbb{P}_n\left\{I(\boldsymbol{X}_t\in\widehat{R}_{tk}\triangle R^*_{tk})\right\}.$$

This concludes the proof of the first inequality. Since functions $\widehat{U}_t$, U_t and indicator functions are bounded, using concentration inequalities, we have

$$\Pr\left[\sup_{R\in\mathcal{R}_t}\big|\,\mathbb{P}_n\{\widehat{\Omega}_{t\ell}(R,a)\}-\mathbb{P}_n\{\Omega_{t\ell}(R,a)\}\big|\ge\epsilon+cn^{-1/2}\tau^{1/2}\right]\le e^{-\tau}.$$

For the second term, by the VC preservation properties (van der Vaart and Wellner, 1996, Lemma 2.6.18), the set

$$\mathcal{F}=\{I(\boldsymbol{X}_t\in R)I(\boldsymbol{X}_t\in G^*_{t\ell})\{U_t(\boldsymbol{X}_t,a)-\zeta\}:R\in\mathcal{R}_t\}$$

is also a VC class. Let ν be its VC index. Then, by Theorem 2.6.7 in van der Vaart and Wellner (1996),

$$\sup_Q\mathcal{N}(\mathcal{G},\|\cdot\|_{L^2(Q)},\varepsilon)\le c_1\varepsilon^{-2\nu},$$

where Q is any probability measure and c_1 is a constant that depends on ν only. For any $f\in\mathcal{F}$, it can be seen that $\|f\|_\infty\le B$. Thus, by Propositions 3 and 5, since $\int_0^1\log(\varepsilon^{-2\nu})<\infty$, we have

$$\Pr\left\{\sup_{f\in\mathcal{F}}|\,\mathbb{P}_n(f)-\mathrm{E}(f)|\ge c\left(\frac{B^2}{n}\right)^{1/2}+c\left(\frac{B^2\tau}{n}\right)^{1/2}\right\}\le e^{-\tau}$$

for any $\tau > 0$, where c is a constant that depends on ν.

Combining both terms and adjusting τ leads to the final inequality. $\square$

Recall that $\rho_t(R_1, R_2) = \Pr\{\boldsymbol{X}_t \in (R_1 \triangle R_2)\}$. The following lemma gives an upper bound on

$$\sup_{R\in\mathcal{R}_t,\rho_t(R,R^*_{t\ell})\le\delta} \left|\left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R,a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R,a^*_{t\ell})\right\} - \left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R^*_{t\ell},a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R^*_{t\ell},a^*_{t\ell})\right\}\right|.$$

This result will be useful for deriving the convergence rate of $\widehat{R}_{t\ell}$.

Lemma 17. *Let* $\epsilon = \left[\mathrm{E}\left\{\widehat{U}_t(\boldsymbol{X}_t,a) - U_t(\boldsymbol{X}_t,a)\right\}^2\right]^{1/2} + B\sum_{k<\ell}\left\{\rho_t\big(\widehat{R}_{tk},R^*_{tk}\big)\right\}^{1/2}$, *and*

$$J_\delta = \sup_{R\in\mathcal{R}_t,\rho_t(R,R^*_{t\ell})\le\delta} \left|\left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R,a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R,a^*_{t\ell})\right\} - \left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R^*_{t\ell},a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R^*_{t\ell},a^*_{t\ell})\right\}\right|.$$

We have, for any $\beta > 0$ *and* $\tau > 0$,

$$\Pr\left\{J_\delta \ge c\delta^{1/2-\beta}(\epsilon + n^{-1/2} + n^{-1/2}\tau^{1/2}) + c\delta^{-\beta}(n^{-1} + n^{-1}\tau)\right\} \le e^{-\tau},$$

where c *is a constant that is independent of* ϵ, n *and* τ.

Proof. We have

$$\begin{aligned}
&\sup_{R\in\mathcal{R}_t,\rho_t(R,R^*_{t\ell})\le\delta} \left|\left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R,a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R,a^*_{t\ell})\right\} - \left\{\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R^*_{t\ell},a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R^*_{t\ell},a^*_{t\ell})\right\}\right| \\
\le &\sup_{R\in\mathcal{R}_t,\rho_t(R,R^*_{t\ell})\le\delta} \left|\mathbb{P}_n\,\Omega_{t\ell}(R,a^*_{t\ell}) - \mathrm{E}\,\Omega_{t\ell}(R,a^*_{t\ell}) - \mathbb{P}_n\,\Omega_{t\ell}(R^*_{t\ell},a^*_{t\ell}) + \mathrm{E}\,\Omega_{t\ell}(R^*_{t\ell},a^*_{t\ell})\right| \\
&+ \sup_{R\in\mathcal{R}_t,\rho_t(R,R^*_{t\ell})\le\delta} \left|\mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R,a) - \mathbb{P}_n\,\widehat{\Omega}_{t\ell}(R^*_{t\ell},a) - \mathbb{P}_n\,\Omega_{t\ell}(R,a) + \mathbb{P}_n\,\Omega_{t\ell}(R^*_{t\ell},a)\right|.
\end{aligned}$$

The first term can be bounded above using properties of VC classes. For any $\delta > 0$, define

$$\begin{aligned}
\mathcal{F}_\delta = \big\{&I(\boldsymbol{X}_t \in R)I(\boldsymbol{X}_t \in G^*_{t\ell})\left\{U_t(\boldsymbol{X}_t,a) - \zeta\right\} \\
&- I(\boldsymbol{X}_t \in R^*_{t\ell})I(\boldsymbol{X}_t \in G^*_{t\ell})\left\{U_t(\boldsymbol{X}_t,a) - \zeta\right\} : R \in \mathcal{R}_t, \rho_t(R,R^*_{t\ell}) \le \delta\big\}.
\end{aligned}$$

Because $\mathcal{R}_t$ is a VC class, $\mathcal{F}_\delta$ is a VC class for any δ. In addition, $\sup_Q \mathcal{N}(\mathcal{F}_\delta, \|\cdot\|_{L^2(Q)}, \varepsilon) \leq c_1\varepsilon^{-2\nu}$ for some constants c_1 and ν independent of δ.

For any $f \in \mathcal{F}_\delta$, we have $\|f\|_\infty \leq B$ and $\mathrm{E}\, f^2 \leq B^2\delta$. Thus, by Propositions 1 and 3,

$$\Pr\left[\sup_{f\in\mathcal{F}_\delta} |\mathbb{P}_n\, f - \mathrm{E}\, f| \geq c_2\left\{\frac{\delta^{1/2}\log^{1/2}(1/\delta)}{n^{1/2}} + \frac{\log(1/\delta)}{n} + \frac{\delta^{1/2}\tau^{1/2}}{n^{1/2}} + \frac{\tau}{n}\right\}\right] \leq e^{-\tau},$$

where c_2 is some constant that depends on B. As $\delta \in (0, 1]$, it follows that $\log(1/\delta) \leq c_3\delta^{-\beta}$ for any $\beta > 0$, where c_3 is same constant that depends on β only. Thus,

$$\Pr\left\{\sup_{f\in\mathcal{F}_\delta} |\mathbb{P}_n\, f - \mathrm{E}\, f| \geq c_4\delta^{1/2-\beta}\left(n^{-1/2} + n^{-1/2}\tau^{1/2}\right) + c_4\delta^{-\beta}\left(n^{-1} + n^{-1}\tau\right)\right\} \leq e^{-\tau}.$$

For the second term, define

$$L = \left|\mathbb{P}_n\, \widehat{\Omega}_{t\ell}(R, a) - \mathbb{P}_n\, \widehat{\Omega}_{t\ell}(R^*_{t\ell}, a) - \mathbb{P}_n\, \Omega_{t\ell}(R, a) + \mathbb{P}_n\, \Omega_{t\ell}(R^*_{t\ell}, a)\right|.$$

we observe that

$$\begin{aligned} L = \Big|&\mathbb{P}_n\left\{I(\boldsymbol{X}_t \in R) - I(\boldsymbol{X}_t \in R^*_{t\ell})\right\} I(\boldsymbol{X}_t \in \widehat{G}_{t\ell})\left\{\widehat{U}_t(\boldsymbol{X}_t, a) - \zeta\right\} \\ &- \mathbb{P}_n\left\{I(\boldsymbol{X}_t \in R) - I(\boldsymbol{X}_t \in R^*_{t\ell})\right\} I(\boldsymbol{X}_t \in G^*_{t\ell})\left\{\widehat{U}_t(\boldsymbol{X}_t, a) - \zeta\right\}\Big| \\ \leq\ & \mathbb{P}_n\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})I(\boldsymbol{X}_t \in \widehat{G}_{t\ell} \cap G^*_{t\ell})\left|\widehat{U}_t(\boldsymbol{X}_t, a) - U_t(\boldsymbol{X}_t, a)\right|\} \\ &+ \mathbb{P}_n\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})I(\boldsymbol{X}_t \in \widehat{G}_{t\ell} \triangle G^*_{t\ell})B\}. \end{aligned}$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned} &\mathrm{E}\left\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})I(\boldsymbol{X}_t \in \widehat{G}_{t\ell} \cap G^*_{t\ell})\left|\widehat{U}_t(\boldsymbol{X}_t, a) - U_t(\boldsymbol{X}_t, a)\right|\right\} \\ \leq\ & \mathrm{E}\left\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})\left|\widehat{U}_t(\boldsymbol{X}_t, a) - U_t(\boldsymbol{X}_t, a)\right|\right\} \\ \leq\ & \left[\mathrm{E}\left\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})\right\}\right]^{1/2}\left[\mathrm{E}\left\{\widehat{U}_t(\boldsymbol{X}_t, a) - U_t(\boldsymbol{X}_t, a)\right\}^2\right]^{1/2}, \end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}\left\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})I(\boldsymbol{X}_t \in \widehat{G}_{t\ell} \triangle G^*_{t\ell})B\right\}\\ \leq& B\left[\mathrm{E}\left\{I(\boldsymbol{X}_t \in R \triangle R^*_{t\ell})\right\}\right]^{1/2}\left[\mathrm{E}\left\{I(\boldsymbol{X}_t \in \widehat{G}_{t\ell} \triangle G^*_{t\ell})\right\}\right]^{1/2}.\end{aligned}$$

Hence $\mathrm{E}(L) \leq \delta^{1/2}\epsilon$. In addition, it is easy to see $\mathrm{Var}(L) \leq E(L^2) \leq c_5\delta$.

Thus, by concentration inequalities,

$$\Pr\left\{L \geq c_6\delta^{1/2-\beta}\left(n^{-1/2} + n^{-1/2}\tau^{1/2}\right) + c_6\delta^{-\beta}\left(n^{-1} + n^{-1}\tau\right)\right\} \leq e^{-\tau}.$$

Combining two terms together gives the final inequality. □

The following lemma is useful for establishing the rate of convergence. It is a finite-sample version of Theorem 3.2.5 in van der Vaart and Wellner (1996). Though we state the lemma in terms of maximizing M_n, an analogous conclusion applies for minimizing M_n.

Lemma 18. *Let $\{M_n(\theta) : \theta \in \Theta\}$ be a stochastic process and $M(\theta)$ a deterministic function. Suppose $M(\theta) - M(\theta_0) \leq -\kappa d^2(\theta, \theta_0)$ for some non-negative function $d : \Theta \times \Theta \to \mathbb{R}$ and positive number κ. Let c_0 be some value that may depend on n. Suppose when $\eta \geq c_0$, we have*

$$\Pr\left\{\sup_{\theta:d(\theta,\theta_0)\leq\delta} |(M_\theta - M)(\theta) - (M_n - M)(\theta_0)| \geq a\delta^\xi\tau^{1/2}\right\} \leq e^{-\tau},$$

where $\xi \in (0, 1]$, a is an expression which is independent of δ and τ but may depend on n.

Let $\widehat{\theta}_n = \arg\max_{\theta\in\Theta} M_n(\theta)$. Define

$$\eta = \max\left\{4\kappa^{-1/(2-\xi)}a^{1/(2-\xi)}\tau^{1/(4-2\xi)}, c_0\right\}.$$

Then,

$$\Pr\left\{d(\widehat{\theta}_n, \theta_0) \geq \eta\right\} \leq 3e^{-\tau}.$$

Proof. Fix $\eta > 0$, define $\eta_j = \eta 2^{-j}$, $j \geq 0$, then

$$\Pr\left\{d(\widehat{\theta}_n, \theta_0) \geq \eta\right\} \leq \sum_{j=1}^{\infty} \Pr\left[\sup_{\theta:\eta_{j-1}\leq d(\theta,\theta_0)<\eta_j} \{M_n(\theta) - M_n(\theta_0)\} \geq 0\right].$$

We observe that

$$\begin{aligned} M_n(\theta) - M_n(\theta_0) &= \{(M_n - M)(\theta) - (M_n - M)(\theta_0)\} + \{M(\theta) - M(\theta_0)\} \\ &\leq |(M_n - M)(\theta) - (M_n - M)(\theta_0)| - \kappa d^2(\theta, \theta_0). \end{aligned}$$

Hence, we have

$$\begin{aligned} &\Pr\left[\sup_{\theta:\eta_{j-1}\leq d(\theta,\theta_0)<\eta_j} \{M_n(\theta) - M_n(\theta_0)\} \geq 0\right] \\ &\qquad\qquad \leq \Pr\left\{\sup_{\theta:d(\theta,\theta_0)\leq\eta_j} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \geq \kappa\eta_{j-1}^2\right\}. \end{aligned}$$

Let $\beta = 1/(2-\xi)$. Then $\eta = 4\kappa^{-\beta}a^{\beta}\tau^{\beta/2}$. Hence, $\eta^{2-\xi} \geq 4\kappa^{-1}a\tau^{1/2}$. Because $j2^{-j} \leq 1$, $j \geq j^{1/2} \geq 1$ for all $j \geq 1$ and $\xi - 2 \leq -1$,

$$\eta^{2-\xi} \geq \kappa^{-1}2^{-j+2}ja\tau^{1/2} \leq \kappa^{-1}2^{j(\xi-2)+2}aj^{1/2}\tau^{1/2}.$$

That is, $\kappa\eta^2 2^{2j-2} \geq \eta^{\xi}2^{j\xi}aj^{1/2}\tau^{1/2}$. By the definition of η_j and η_{j-1}, we have $\kappa\eta_{j-1}^2 \geq \eta_j^{\xi}aj^{1/2}\tau^{1/2}$. By the condition on $M_n - M$, we have

$$\Pr\left\{\sup_{\theta:d(\theta,\theta_0)\leq\eta_j} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \geq \kappa\eta_{j-1}^2\right\} \leq e^{-j\tau}.$$

Therefore, we have $\Pr\left\{d(\widehat{\theta}_n, \theta_0) \geq \eta\right\} \leq \sum_{j=1}^{\infty} e^{-j\tau} = e^{-\tau}/\left(1 - e^{-\tau}\right)$. Note that $e^{-\tau}/\left(1 - e^{-\tau}\right) \leq 3e^{-\tau}$ when $\tau \geq 1$ and $\Pr\left\{d(\widehat{\theta}_n, \theta_0) \geq \eta\right\} \leq 1 \leq 3e^{-\tau}$ when $\tau < 1$.

□

2.9 Proof of Theorem 1

Since we can always increase the constant terms c_1 and c_2 in Theorem 1 so that the bounds become trivial when $\tau \leq 1$, we can safely assume that $\tau > 1$. In this subsection, ξ and β denote arbitrary positive numbers. The value of ξ or β may be different at each occurrence.

Stage $t = T$. We start at the last stage $t = T$. Define $\varphi_T = r_T/(2r_T + d_T)$. Because $\widehat{Y}_T = \widetilde{Y}_T$ for any $a \in \mathcal{A}_T$, under the conditions on $\boldsymbol{\gamma}_T$ and λ_T, by Proposition 12 and its corollary, we have

$$\Pr\left[\mathrm{E}_{\boldsymbol{X}}\left\{\widehat{Q}_T(\boldsymbol{X}, a) - Q_T(\boldsymbol{X}, a)\right\}^2 \geq c_1\left(n^{-2\varphi_T+\xi} + n^{-1}\tau\right)\right] \leq e^{-\tau}.$$

This establishes the consistency and convergence rate for $\widehat{Q}_T$.

Next, we consider $(\widehat{R}_{T\ell}, \widehat{a}_{T\ell})$ for $\ell = 1, 2, \ldots$. In view of Assumption 4 (i) and (ii), by reducing κ, we can have Assumption 4 (i) hold for all R instead of only those R close to the true value.

When $\ell = 1$, we have $\widehat{G}_{T1} = G^*_{T1} = \mathcal{X}_T$. Thus, for any $a \in \mathcal{A}_T$, by equation (4) and Lemma 16, it follows that

$$\Pr\left\{\sup_{R\in\mathcal{R}_T} |\mathbb{P}_n\,\widehat{\Omega}_{T1}(R, a) - \mathrm{E}\,\Omega_{T1}(R, a)| \geq c_1 n^{-\varphi_T+\xi}\tau\right\} \leq e^{-\tau}.$$

By Assumption 4 (iii), we have $\inf_{R\in\mathcal{R}_T, a\neq a^*_{T1}} \mathrm{E}\,\Omega_{T1}(R, a) \geq \mathrm{E}\,\Omega_{T1}(R^*_{T1}, a^*_{T1}) + \varsigma$. Thus,

$$\begin{aligned}\Pr(\widehat{a}_{T1} \neq a^*_{T1}) &\leq \sum_{a\neq a^*_{T1}} \Pr\left\{\sup_{R\in\mathcal{R}_T} \mathbb{P}_n\,\widehat{\Omega}_{T1}(R, a) \geq \mathbb{P}_n\,\widehat{\Omega}_{T1}(R^*_{T1}, a^*_{T1})\right\}\\ &\leq \sum_{a} \Pr\left\{\sup_{R\in\mathcal{R}_T} \left|\mathbb{P}_n\,\widehat{\Omega}_{T1}(R, a) - \mathrm{E}\,\Omega_{T1}(R, a)\right| \geq \varsigma/2\right\}.\end{aligned}$$

Hence,

$$\Pr(\widehat{a}_{T1} \neq a^*_{T1}) \leq c_1 \exp(-c_2 n^{\varphi_T-\xi}),$$

where c_1 depends on $|\mathcal{A}_T|$ and c_2 depends on ς. Actually, as seen from the proof of Theorem 2, we are able to obtain a faster convergence rate for $\widehat{a}_{T1}$. However, this does not affect the final result because $\widehat{R}_{T1}$ converges at a much slower rate, as shown below.

We proceed to establish the convergence rate for $\widehat{R}_{T1}$. For any $\delta > 0$, by Lemma 17 and absorbing terms with faster convergence into those with slower convergence,

$$\Pr\left\{\sup_{R\in\mathcal{R}_T,\rho_T(R,R^*_{T1})\le\delta}\left|\mathbb{P}_n\,\widehat{\Omega}_{T1}(R,a^*_{T1}) - \mathbb{P}_n\,\widehat{\Omega}_{T1}(R^*_{T1},a^*_{T1}) - \mathrm{E}\,\Omega_{T1}(R,a^*_{T1}) + \mathrm{E}\,\Omega_{T1}(R^*_{T1},a^*_{T1})\right|\right.$$
$$\left.\ge c_1\delta^{1/2-\beta}n^{-\varphi_T+\xi}\tau\right\} \le e^{-\tau}.$$

Hence, by Lemma 18,

$$\Pr\left\{\rho_T(\widehat{R}_{T1},R^*_{T1}) \ge c_1 n^{-(2/3)\varphi_T+\xi}\tau\right\} \le c_2e^{-\tau}.$$

Note that we take β sufficiently small so that it can be absorbed into ξ.

We next proceed to $\ell = 2$. By equation (4) and Lemma 16, for any $a \in \mathcal{A}_T$,

$$\Pr\left\{\sup_{R\in\mathcal{R}_T}|\,\mathbb{P}_n\,\widehat{\Omega}_{T2}(R,a) - \mathrm{E}\,\Omega_{T2}(R,a)| \ge c_1n^{-(2/3)\varphi_T+\xi}\tau\right\} \le e^{-\tau}.$$

Similar to $\widehat{a}_{T1}$, we obtain

$$\Pr(\widehat{a}_{T2} \ne a^*_{T2}) \le c_1\exp\left\{-c_2n^{(2/3)\varphi_T-\xi}\right\}.$$

By equation (4) and Lemma 17, for any $\delta > 0$, we have

$$\Pr\left\{\sup_{R\in\mathcal{R}_T,\rho_T(R,R^*_{T2})\le\delta}\left|\mathbb{P}_n\,\widehat{\Omega}_{T2}(R,a^*_{T2}) - \mathbb{P}_n\,\widehat{\Omega}_{T2}(R^*_{T2},a^*_{T2}) - \mathrm{E}\,\Omega_{T2}(R,a^*_{T2}) + \mathrm{E}\,\Omega_{T2}(R^*_{T2},a^*_{T2})\right|\right.$$
$$\left.\ge c_1\delta^{1/2-\beta}n^{-(2/3)\varphi_T+\xi}\tau\right\} \le e^{-\tau}.$$

Hence, by Lemma 18,

$$\Pr\left\{\rho_T(\widehat{R}_{T2},R^*_{T2}) \ge c_1n^{-(2/3)^2\varphi_T+\xi}\tau\right\} \le c_2e^{-\tau}.$$

Again, β is chosen to be sufficiently small so as to be absorbed into ξ.

Using induction, for any ℓ, we obtain

$$\Pr(\widehat{a}_{T\ell} \neq a^*_{T\ell}) \leq c_1 \exp\left\{-c_2 n^{(2/3)^{\ell-1}\varphi_T}\right\}$$

and

$$\Pr\left\{\rho_T(\widehat{R}_{T\ell}, R^*_{T\ell}) \geq c_1 n^{-(2/3)^{\ell}\varphi_T}\tau\right\} \leq c_2 e^{-\tau}.$$

Make the change of variables $\tau \to c_1 n^{-(2/3)^{\ell}\varphi_T}\tau$ to obtain, for any ℓ,

$$\Pr\{\rho_T(\widehat{R}_{T\ell}, R^*_{T\ell}) \geq \tau\} \leq c_1 \exp\left\{-c_2 n^{(2/3)^{\ell}\varphi_T}\right\}.$$

Since

$$F_T(\widehat{\pi}_T) \leq \sum_{\ell=1}^{L^*_T}\left\{\Pr(\widehat{a}_{T\ell} \neq a^*_{T\ell}) + \rho_t(\widehat{R}_{T\ell}, R^*_{T\ell})\right\},$$

we obtain

$$\begin{aligned}\Pr\left\{F_T(\widehat{\pi}_T) \geq \tau\right\} &\leq \sum_{\ell=1}^{L^*_T}\Pr\left(\widehat{a}_{T\ell} \neq a^*_{T\ell}\right) + \sum_{\ell=1}^{L^*_T}\Pr\left\{\rho_T(\widehat{R}_{T\ell}, R^*_{T\ell}) \geq \tau/L^*_T\right\}\\ &\leq c_1\exp(-c_2 n^{\phi_T-\xi}\tau),\end{aligned}$$

where $\phi_T = (2/3)^{L^*_T}\varphi_T$. In the last inequality, the terms with faster convergence are absorbed into the term with the slowest convergence. Consequently,

$$\Pr\left\{V_T(\pi^*_T) - V_T(\widehat{\pi}_T) \geq \tau\right\} \leq \Pr\left\{F_T(\widehat{\pi}_T) \geq \tau/B\right\} \leq c_3\exp(-c_4 n^{\phi_T-\xi}\tau).$$

By the change of variable $\tau \to c_5 n^{-\phi_T+\xi}\tau$ and putting c_1 and c_3 into the exponential terms, we conclude that

$$\Pr\left\{F_T(\widehat{\pi}_T) \geq c_1 n^{-\phi_T+\xi}\tau \text{ or } V_T(\pi^*_T) - V_T(\widehat{\pi}_T) \geq c_2 n^{-\phi_T+\xi}\tau\right\} \leq e^{-\tau}.$$

Stage $t = T-1, \ldots, 1$. We now proceed to the earlier stages. Consider the $(T-1)$th stage. By the risk bounds of $\widehat{Q}_T$ and $\widehat{\pi}_T$,

$$\Pr\left\{\mathbb{P}_n\left(\widehat{Y}_T - \widetilde{Y}_T\right)^2 \geq c_1 n^{-\phi_T+\xi}\tau\right\} \leq c_2 e^{-\tau}.$$

Hence, by Proposition 12, for any $a \in \mathcal{A}_{T-1}$, we have

$$\Pr\left[E_{\boldsymbol{X}}\left\{\widehat{Q}_{T-1}(\boldsymbol{X}, a) - Q_{T-1}(\boldsymbol{X}, a)\right\}^2 \geq c_1 n^{-2\varphi_{T-1}+\xi}\tau\right] \leq c_2 e^{-\tau},$$

where $\varphi_{T-1} = \min\{\phi_T/2, r_{T-1}/(2r_{T-1}+d_{T-1})\}$, i.e., the convergence rate of $\widehat{Q}_{T-1}$ depends on two factors: the kernel regression convergence rate assuming the true response $\widetilde{Y}$ is observed, and the convergence rate of the surrogate response $\widehat{Y}$ towards $\widetilde{Y}$.

The analysis of clauses, $(\widehat{R}_{T-1,\ell}, \widehat{a}_{T-1,\ell})$, $\ell = 1, \ldots, L^*_{T-1}$ are in the same manner as in the last stage. Thus, with similar calculations we obtain

$$\Pr\left\{F_{T-1}(\widehat{\pi}_{T-1}) \geq c_1 n^{-\phi_{T-1}+\xi}\tau \text{ or } V_{T-1}(\pi^*_{T-1}) - V_{T-1}(\widehat{\pi}_{T-1}) \geq c_2 n^{-\phi_{T-1}+\xi}\tau\right\} \leq e^{-\tau}.$$

where $\phi_{T-1} = (2/3)^{L^*_{T-1}}\varphi_{T-1}$.

Using induction, the same inequality hold when $T-1$ is replaced by $t = T-2, \ldots, 1$.

2.10 Proof of Theorem 2

Stage $t = T$. At the last stage, by Proposition 12,

$$\Pr\left[\mathrm{E}_{\boldsymbol{X}}\left\{\widehat{Q}_T(\boldsymbol{X}, a) - Q_T(\boldsymbol{X}, a)\right\}^2 \geq c_1\left(n^{-2\varphi_T+\xi} + n^{-1}\tau\right)\right] \leq e^{-\tau},$$

where $\varphi_T = r_T/(2r_T + d_T)$ and $\xi > 0$ is arbitrary. Using a similar argument to the proof of Theorem 1,

$$\Pr\left\{\sup_{R\in\mathcal{R}_T} |\,\mathbb{P}_n\,\widehat{\Omega}_{T1}(R, a) - \mathrm{E}\,\Omega_{T1}(R, a)| \geq c_1\left(n^{-\varphi_T+\xi} + n^{-1/2}\tau^{1/2}\right)\right\} \leq e^{-\tau},$$

and

$$\Pr(\widehat{a}_{T1} \neq a^*_{T1}) \leq \sum_a \Pr\left\{ \sup_{R \in \mathcal{R}_T} \left| \mathbb{P}_n \widehat{\Omega}_{T1}(R,a) - \mathrm{E}\, \Omega_{T1}(R,a) \right| \geq \varsigma/2 \right\}.$$

Note that ς is a fixed number independent of n. Let $\tau^{1/2} = n^{1/2} \max(c_2\varsigma - n^{-\varphi_T+\xi}, 0)$ and choose c_2 such that $2c_1c_2 < 1$. Note that $\varphi_T \in (0,1)$. Then we have

$$\Pr(\widehat{a}_{T1} \neq a^*_{T1}) \leq c_3 \exp(-c_4 n).$$

Define $\vartheta = \inf_{R:\rho_T(R,R^*_{T1})>0} \rho_T(R, R^*_{T1})$. Because the covariates are discrete, ϑ is strictly positive. This is a major difference between the continuous covariates and the discrete covariates. By Assumption 4 (i), using an argument similar to that for $\Pr(\widehat{a}_{T1} \neq a^*_{T1})$, we have

$$\begin{aligned}\Pr\left\{\rho_T(\widehat{R}_{T1}, R^*_{T1}) > 0\right\} &\leq \Pr\left\{ \sup_{R\in\mathcal{R}_T:\rho_T(R,R^*_{T1})\geq\vartheta} \mathbb{P}_n \widehat{\Omega}_{T1}(R, a^*_{T1}) \geq \mathbb{P}_n \widehat{\Omega}_{T1}(R^*_{T1}, a^*_{T1}) \right\} \\ &\leq \Pr\left\{ \sup_{R\in\mathcal{R}_T} \left| \mathbb{P}_n \widehat{\Omega}_{T1}(R, a^*_{T1}) - \mathrm{E}\, \Omega_{T1}(R, a^*_{T1}) \right| \leq \kappa\vartheta^2/2 \right\} \\ &\leq c_5 \exp(-c_6 n).\end{aligned}$$

We next analyze $(\widehat{R}_{T2}, \widehat{a}_{T2})$. For any $a \in \mathcal{A}_T$,

$$\Pr\left\{ \sup_{R\in\mathcal{R}_T} |\, \mathbb{P}_n \widehat{\Omega}_{T2}(R,a) - \mathrm{E}\, \Omega_{T2}(R,a)| \geq c_1 \left(n^{-\varphi_T+\xi} + n^{-1/2}\tau^{1/2} \right) \right\} \leq e^{-\tau}.$$

Similar to $(\widehat{R}_{T1}, \widehat{a}_{T1})$,

$$\Pr(\widehat{a}_{T2} \neq a^*_{T2}) \leq c_1 \exp(-c_2 n)$$

and

$$\Pr\left\{\rho_T(\widehat{R}_{T2}, R^*_{T2}) > 0\right\} \leq c_3 \exp(-c_4 n).$$

As seen from this inequality, a notable difference is that the estimation error does not propagate along the list, compared to the general case where covariates can be continuous.

The tail probability decays at the same exponential rate for every ℓ. Therefore, we have

$$\Pr\left\{F_T(\widehat{\pi}_T) > 0\right\} \le \sum_{\ell=1}^{L_T^*} \Pr\left(\widehat{a}_{T\ell} \ne a_{T\ell}^*\right) + \sum_{\ell=1}^{L_T^*} \Pr\left\{\rho_T(\widehat{R}_{T\ell}, R_{T\ell}^*) > 0\right\} \le c_1 \exp(-c_2 n),$$

and consequently,

$$\Pr\left\{V_T(\pi_T^*) - V_T(\widehat{\pi}_T) > 0\right\} \le \Pr\left\{F(\widehat{\pi}_T) > 0\right\} \le c_1 \exp(-c_2 n).$$

Stage $t = T-1, \ldots, 1$. We then move to the $(T-1)$th stage. Conditional on the event $\{F_T(\widehat{\pi}_T) = 0\}$, which occurs with probability $1 - c_1 \exp(-c_2 n)$,

$$\Pr\left[\widehat{Q}\left\{\boldsymbol{X}_T, \widehat{\pi}_T(\boldsymbol{X}_T)\right\} = \widehat{Q}\left\{\boldsymbol{X}_T, \pi_T^*(\boldsymbol{X}_T)\right\}\right] = 1.$$

Hence,

$$\Pr\left\{\mathbb{P}_n\left(\widehat{Y}_{T-1} - \widetilde{Y}_{T-1}\right)^2 \ge c_1\left(n^{-2\varphi_T+\xi} + n^{-1}\tau\right)\right\} \le e^{-\tau}.$$

Namely, the error in the pseudo response $\widehat{Y}_{T-1}$ only comes from the non-parametric kernel regression. The error due to the estimation of regime is of higher order and can be absorbed.

Define $\varphi_{T-1} = \min\left\{r_{T-1}/(2r_{T-1} + d_{T-1}), \varphi_T\right\}$. By Proposition 12 and its corollaries,

$$\Pr\left[\mathrm{E}_{\boldsymbol{X}}\left\{\widehat{Q}_{T-1}(\boldsymbol{X}, a) - Q_{T-1}(\boldsymbol{X}, a)\right\}^2 \ge c_1\left(n^{-2\varphi_{T-1}+\xi} + n^{-1}\tau\right)\right] \le e^{-\tau}.$$

Compared to the counterpart inequality in the last stage, nothing is changed except that T is replaced by $T-1$. Using the same approach as in the Tth stage, conditional on the event $\{F_T(\widehat{\pi}_T) = 0\}$, we obtain

$$\Pr\left\{F_{T-1}(\widehat{\pi}_{T-1}) > 0\right\} \le c_1 \exp(-c_2 n),$$

and

$$\Pr\left\{V_{T-1}(\pi_{T-1}^*) - V_{T-1}(\widehat{\pi}_{T-1}) > 0\right\} \le c_1 \exp(-c_2 n).$$

Because the event $\{F_T(\widehat{\pi}_T) = 0\}$ occurs with probability $1 - c_1 \exp(-c_2 n)$, both inequalities hold unconditionally with larger constants c_1 and c_2.

Using induction, we can establish analogous inequalities for $t = T-2, \ldots, 1$.

3 Algorithm Discussions

3.1 Details of the Proposed Algorithm and Proof of Proposition 1

Fix an t and ℓ. Define

$$U_{iat\ell} = \left[\widehat{Q}_t \left\{\boldsymbol{X}_{it}, \widehat{\pi}_t^Q(\boldsymbol{X}_{it})\right\} - \widehat{Q}_t(\boldsymbol{X}_{it}, a) - \zeta\right] I\left(\boldsymbol{X}_{it} \in \widehat{G}_{t\ell}\right).$$

For notational simplicity, we shall omit the subscript t and ℓ and write U_{ia} and $\boldsymbol{X}_i$. By definition of $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$,

$$(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}) = \underset{R \in \mathcal{R}_t, a \in \mathcal{A}_t}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} U_{ia} I(\boldsymbol{X}_i \in R) - \eta\{2 - V(R)\}.$$

We will first fix the treatment a and the covariates involved in R, and focus on the computation of the optimal thresholds. Then we will loop over all covariate pairs and all treatment options.

Finding the threshold when R involves one covariate Without loss of generality, we assume $R = \{\boldsymbol{x} : x_j \le \tau\}$. The other situation $R = \{\boldsymbol{x} : x_j > \tau\}$ can be handled similarly. We want to compute

$$\widehat{\tau} = \underset{\tau}{\arg\min} \sum_{i=1}^{n} U_{ita} I(X_{ij} \le \tau),$$

where X_{ij} is the jth component of $\boldsymbol{X}_i$.

Let $i_1, \ldots, i_n$ be a permutation of $1, \ldots, n$ such that $X_{i_1 j} \leq \cdots \leq X_{i_n j}$. Because the objective function is piecewise constant, we only need to compute

$$F(\tau) = \sum_{i=1}^{n} U_{ia} I(X_{ij} \leq \tau)$$

when τ equals to some $X_{i_s j}$. We observe that

$$F(X_{i_s j}) = \sum_{h \leq s} U_{i_h a}.$$

Thus, it follows that when $s \geq 2$

$$F(X_{i_s j}) = F(X_{i_{s-1} j}) + U_{i_s a}.$$

Hence, by starting at $s = 1$ and using the recursive relationship, we can compute $F(X_{i_s j})$ for all s and pick the smallest one in $O(n)$ time.

Dealing with ties If $X_{i_s j} = X_{i_{s+1} j}$ for some $s \geq 1$, then $F(X_{i_s j})$ should not be counted when picking the minimum. This is because $F(X_{i_s j})$ has not included all subjects with $X_{ij} = X_{i_s j}$ yet.

To avoid this problem, when there are ties, we first aggregate the U_{ia} values for subjects having the same value of X_{ij}. Similar action can be taken when R involves two covariates, in which case the U_{ia} values for subjects having the same value for both covariates are aggregated.

Finding the threshold when R involves two covariates This situation is more complicated. Without loss of generality, we assume $R = \{\boldsymbol{x} : x_j \leq \tau \text{ and } x_k \leq \sigma\}$. We want to compute

$$(\widehat{\tau}, \widehat{\sigma}) = \arg\min_{\tau, \sigma} \frac{1}{n} \sum_{i=1}^{n} U_{ia} I(X_{ij} \leq \tau, X_{ik} \leq \sigma).$$

We cannot utilize the idea for one covariate as there is no natural ordering in two-dimensional space. Our solution is to sort in one dimension and to use binary tree for fast lookup and insertion in the other dimension.

We start with constructing a complete binary tree of at least n leaves. The height of such a tree is of order $O(\log_2 n)$.

Let $i_1, \ldots, i_n$ be a permutation of $1, \ldots, n$ such that $X_{i_1 j} \leq \ldots X_{i_n j}$. At each time s, we will insert $U_{i_s a}$ into the binary tree and search for the optimal threshold σ among $X_{ik}, i = 1, \ldots, n$. Note that at time s, values $U_{i_h a}, h \leq s$ are contained in the binary tree. So we are looking at the threshold $\tau = X_{i_s j}$. Specifically, if the rank of $X_{i_s k}$ among X_{ik}s is h, which means $X_{i_s k}$ is the hth smallest among X_{ik}s, then we put $U_{i_s a}$ in the hth leaf from the left in the tree.

In the tree, each node is associated with a subtree in which that node serves as the root. Each node contains two pieces of information. First, it computes the sum of all $U_{i_s a}$s in the associated subtree. Second, it computes the best thresholding sum in the associated subtree, which is the smallest value among the sum of all $U_{i_s a}$s that satisfies $X_{i_s k} \leq \sigma$ for some σ, where σ can take the value of any $X_{i_s k}$ in the associated subtree.

The binary tree structure enables us to update these two pieces of information effectively when a new value, $U_{i_s a}$, is inserted into the tree. We move from the leaf node to its parent, and then its ancestors, and finally the root. At each node, the sum of all $U_{i_s a}$s in the associated subtree is increased by $U_{i_s a}$. As for updating the best thresholding sum, because the thresholding condition is $X_{i_s k} \leq \sigma$, the best thresholding sum of a node can only be either the best thresholding sum in its left child, or, the sum of all $U_{i_s a}$ values in the left child plus the best thresholding sum in the right child, whichever is smaller.

Because the height of the tree is $O(\log_2 n)$, the updating process involves at most $O(\log_2 n)$ nodes and the time complexity at each node is constant. Therefore, when $U_{i_s a}$ is inserted into the tree, we are able to find the optimal σ that minimizes $\sum_{h\leq s} U_{i_h a} I(X_{i_h k} \leq \sigma)$ in $O(\log_2 n)$ time.

Then we let s run from 1 to n, and find the s that gives the minimum. In this way, we find the minimum of $\sum_{h=1}^{n} U_{i_h a} I(X_{i_h k} \leq \sigma, X_{i_h j} \leq X_{i_s j})$ with respect to σ and s, which is exactly the minimum of $\sum_{i=1}^{n} U_{i_s a} I(X_{ik} \leq \sigma, X_{ij} \leq \tau)$ with respect to σ and τ. And the time complexity for finding both τ and σ is $O(n \log_2 n)$.

Finding the covariate(s) and treatment Heretofore, we have discussed how to find the optimal thresholds when the covariates to use X_{ij}, X_{ik} and the treatment a are given. Certainly we need to explore all Rs defined using only one covariate, and all Rs defined using some pair of X_{ij} and X_{ik}. We also need to loop over all treatment options $a \in \mathcal{A}_t$.

Therefore, the overall time complexity is $O(n \log n d_t^2 m_t)$, where d_t is the dimension of $\boldsymbol{X}_i$ and $m_t = |\mathcal{A}_t|$ is the number of available treatment options.

3.2 Extension of the Proposed Algorithm

We present an extension of the proposed algorithm which is able to handle an arbitrary number of covariates per if-then clause. Let q be the number of covariates in each clause. Recall that

$$(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}) = \underset{R\in\mathcal{R}_t, a\in\mathcal{A}_t}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} U_{ia} I(\boldsymbol{X}_i \in R) - \eta\{q - V(R)\},$$

and $\mathcal{A}_t$ is the set of treatment options. In general, any element in $\mathcal{R}_t$ can be written as $R = \{\boldsymbol{x} : s_k(x_{j_k} - \tau_k) \leq 0,\ k = 1, \ldots, q\}$, where $s_k \in \{-1, 1\}$, $j_k \in \{1, \ldots, d_t\}$ and $\tau_k \in \mathbb{R}$. The important observation is that, since the objective function is piecewise constant, the maximum with respect to $\tau_k \in \mathbb{R}$ is the same as the maximum with respect to $\tau_k \in \mathcal{X}_k$, where the set $\mathcal{X}_k$ consists of all unique values in $X_{1k}, \ldots, X_{nk}$.

Hence, we can compute $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell})$ as follows.

1. Loop a over $\mathcal{A}_t$.

2. For each a, loop $j_1 < \cdots < j_q$ over all possible d_t-choose-q combinations in $\{1, \ldots, d_t\}$.

3. For each a and each $j_1 < \cdots < j_q$, loop $(s_1, \ldots, s_q)^{\mathrm{T}}$ over $\{-1, 1\}^q$. The usage of s_k is to control the direction of the inequality.

4. For each a, each $j_1 < \cdots < j_q$, and each $(s_1, \ldots, s_q)^{\mathrm{T}}$, loop $(\tau_1, \ldots, \tau_k)^{\mathrm{T}}$ over the Cartesian product $\mathcal{X}_{j_1} \times \cdots \times \mathcal{X}_{j_q}$.

5. Construct R as $R = \{\boldsymbol{x} : s_k(x_{j_k} - \tau_k) \leq 0,\ k = 1, \ldots, q\}$ and evaluate the objective function at R, a; keep track of the minimizer.

This algorithm has a time complexity of $O(2^q n^q d_t^q m_t)$.

3.3 A Description of the Algorithm in Zhang et. al. (2015)

For the sake of completeness, we present the algorithm in Zhang et al. (2015) using the notations in our paper.

Step 1. Choose a maximum list length $L_{\max}$ and a critical level $\alpha \in (0, 1)$. Set $\ell = 1$, and $\Pi_{t1} = \{\overline{\pi}\}$ with $\overline{\pi} = \{(\mathcal{X}_t, \overline{a}_{t1})\}$, where $\overline{a}_{t1} = \arg\max_{a \in \mathcal{A}_t} n^{-1} \sum_{i=1}^n \widehat{Q}_t(\boldsymbol{X}_{it}, a)$. Note that π is just the best single treatment. Define

$$\widehat{V}_t(\pi) = \frac{1}{n} \sum_{i=1}^n \sum_{a \in \mathcal{A}_t} I\{\pi(\boldsymbol{X}_{it}) = a\} \widehat{Q}_t(\boldsymbol{X}_{it}, a).$$

Step 2. Fix $\overline{\pi} \in \Pi_{t\ell}$. Then $\overline{\pi}$ can be represented by

$$\{(\overline{R}_{t1}, \overline{a}_{t1}), \ldots, (\overline{R}_{t,\ell-1}, \overline{a}_{t,\ell-1}), (\mathcal{X}_t, \overline{a}_{t\ell})\}.$$

Let $\overline{G}_{t1} = \mathcal{X}_t$, $\overline{G}_{t\ell} = \mathcal{X}_t \setminus \left(\bigcup_{k<\ell} \overline{R}_{tk}\right)$ for $\ell \geq 2$. Compute

$$(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}, \widetilde{a}_{t\ell}) \in \mathop{\arg\max}_{R \in \mathcal{R}_t, a \in \mathcal{A}_t, a' \in \mathcal{A}_t} \frac{1}{n} \sum_{i=1}^{n} \Big[I(\boldsymbol{X}_{it} \in \overline{G}_{t\ell}, \boldsymbol{X}_{it} \in R) \widehat{Q}_t(\boldsymbol{X}_{it}, a) + I(\boldsymbol{X}_{it} \in \overline{G}_{t\ell}, \boldsymbol{X}_{it} \notin R) \widehat{Q}_t(\boldsymbol{X}_{it}, a') \Big]. \quad (5)$$

Note that the maximizer won't be unique, since if $(\widehat{R}_{t\ell}, \widehat{a}_{t\ell}, \widetilde{a}_{t\ell})$ is a maximizer, then $(\widehat{R}_{t\ell}, \widetilde{a}_{t\ell}, \widehat{a}_{t\ell})$ is also a maximizer. Let $\Omega_{t\ell}$ be the sets that consists of all the maximizers. For each $(\widehat{R}_{t\ell}, \widetilde{a}_{t\ell}, \widehat{a}_{t\ell})$ in $\Omega_{t\ell}$, define a regime $\widehat{\pi}$ represented by

$$\{(\overline{R}_{t1}, \overline{a}_{t1}), \ldots, (\overline{R}_{t,\ell-1}, \overline{a}_{t,\ell-1}), (\widehat{R}_{t\ell}, \widehat{a}_{t\ell}), (\mathcal{X}_t, \widetilde{a}_{t\ell})\}.$$

Roughly speaking, $\widehat{\pi}$ adds one more layer of personalization to $\overline{\pi}$, because $\widehat{\pi}$ is a regime that has the same first $(\ell - l)$ if-then clauses as $\overline{\pi}$ and has one more if-then clause. Let $\Pi_{t,\ell+1}$ be the set

$$\{\widehat{\pi} : \overline{\pi} \in \Pi_{t\ell} \text{ and } \widehat{V}(\widehat{\pi}) - \widehat{V}(\overline{\pi}) > z_\alpha \widehat{\sigma}_{\widehat{\pi},\overline{\pi}}\},$$

where z_α is the upper α-quantile of standard normal, and $\widehat{\sigma}_{\widehat{\pi},\overline{\pi}}\}^2$ is some quantity that approximates the variance of $\widehat{V}(\widehat{\pi}) - \widehat{V}(\overline{\pi})$. The purpose of the inequality is to avoid scenarios that the additional if-then clause only captures random fluctuations.

Step 3. Increase ℓ by 1. If $\ell < L_{\max}$ and $\Pi_{t\ell}$ is not an empty set, repeat Step 2; otherwise, go to Step 4.

Step 4. Compute $\widehat{\pi}_t = \arg\max_{\pi \in \Pi_{t\ell}} \widehat{V}_t(\pi)$ and output $\widehat{\pi}_t$ as the estimated optimal regime.

4 Additional Simulation Results

In Section 4 in the main text, we present the mean outcome under the estimated treatment regime. In this section, in order to provide more insights, we shall present the probability that the estimated regime selects the best treatment option and the computation time of the proposed algorithm. The scenarios considered are the same as those in the main text.

The probabilities of correct treatment selection under regimes estimated by different methods are given in Table 1. We say that the treatment selection is correct if at every stage, the treatment dictated by the estimated regime coincides with the best option in that stage. The pattern in treatment selection accuracy is qualitatively very similar to that in the mean outcome. Special caution should be given to Scenario V, because there are ten stages in total, and seven treatment options per each stage. Hence there are altogether 7^{10} possible choices of treatment. Consequently, it is extremely difficult for the estimated regime to get treatment recommendations at all stages correct.

The computation time of the proposed algorithm are given in Table 2. In each cell, we report the time in seconds for estimating a dynamic treatment regime, which consists on T regimes, one per each stage. All tunings are included in the timing. Admittedly, due to the complexity of estimating a discrete structure, the proposed algorithm run the slowest among all approaches. Though slow, the proposed algorithm is able to estimate a dynamic treatment regime within a few minutes, which won't cause much burden in practice. Moreover, we believe that the interpretability brought by the if-then clauses outweigh such a computation burden.

5 Covariates in Real Data Example

In the first stage, we have the following variables:

Table 1: Simulation results. The number in each cell is the probability that the estimated regime selects the best treatment option, averaged over 1000 replications, with standard deviation in parentheses.

Scenario	n	DL	Q-lasso	Q-RF	BOWL	SOWL
I	100	0.36 (0.05)	0.45 (0.08)	0.36 (0.04)	0.35 (0.06)	0.36 (0.06)
I	200	0.41 (0.08)	0.52 (0.05)	0.36 (0.03)	0.35 (0.06)	0.36 (0.05)
I	400	0.50 (0.11)	0.54 (0.04)	0.36 (0.02)	0.35 (0.06)	0.36 (0.05)
II	100	0.88 (0.04)	0.89 (0.04)	0.81 (0.05)	0.74 (0.01)	0.69 (0.11)
II	200	0.91 (0.03)	0.92 (0.03)	0.87 (0.04)	0.75 (0.02)	0.70 (0.08)
II	400	0.93 (0.02)	0.94 (0.02)	0.91 (0.02)	0.79 (0.03)	0.71 (0.06)
III	100	0.76 (0.11)	0.42 (0.21)	0.64 (0.10)	0.58 (0.12)	0.56 (0.15)
III	200	0.86 (0.08)	0.55 (0.10)	0.77 (0.07)	0.66 (0.07)	0.70 (0.08)
III	400	0.92 (0.05)	0.59 (0.06)	0.88 (0.04)	0.74 (0.05)	0.83 (0.06)
IV	100	0.73 (0.15)	0.35 (0.22)	0.52 (0.10)	0.45 (0.10)	0.38 (0.12)
IV	200	0.88 (0.07)	0.50 (0.15)	0.64 (0.09)	0.50 (0.07)	0.47 (0.10)
IV	400	0.94 (0.05)	0.58 (0.08)	0.75 (0.08)	0.55 (0.04)	0.54 (0.10)
V	100	0.01 (0.01)	0.00 (0.00)	0.00 (0.00)	–	–
V	200	0.04 (0.02)	0.00 (0.00)	0.00 (0.00)	–	–
V	400	0.08 (0.03)	0.00 (0.00)	0.01 (0.00)	–	–

Table 2: Simulation results. The number in each cell is the computation time in seconds, averaged over 1000 replications, with standard deviation in parentheses.

Scenario	n	DL	Q-lasso	Q-RF	BOWL	SOWL
I	100	35.97 (4.53)	0.49 (0.08)	0.96 (0.05)	0.97 (0.09)	7.19 (0.35)
I	200	110.09 (9.89)	0.36 (0.05)	2.67 (0.17)	1.24 (0.14)	10.37 (0.52)
I	400	275.93 (37.54)	0.24 (0.02)	8.31 (0.65)	2.32 (0.11)	13.00 (0.61)
II	100	21.91 (1.42)	0.39 (0.06)	0.79 (0.03)	1.10 (0.07)	7.73 (0.30)
II	200	53.08 (3.97)	0.39 (0.11)	2.20 (0.06)	1.18 (0.06)	11.13 (1.23)
II	400	136.10 (12.38)	0.30 (0.02)	4.98 (0.23)	2.49 (0.07)	12.04 (1.19)
III	100	0.80 (0.22)	0.26 (0.06)	0.33 (0.07)	0.60 (0.04)	6.32 (0.64)
III	200	3.51 (1.45)	0.23 (0.03)	0.73 (0.08)	0.91 (0.09)	6.47 (1.35)
III	400	18.88 (9.50)	0.23 (0.01)	2.11 (0.26)	1.37 (0.09)	7.34 (0.63)
IV	100	37.20 (7.89)	1.11 (0.30)	1.18 (0.10)	1.53 (0.07)	12.89 (0.67)
IV	200	82.39 (11.16)	1.25 (0.16)	3.46 (0.13)	3.36 (0.54)	24.30 (1.84)
IV	400	193.65 (18.48)	1.06 (0.13)	7.86 (0.73)	12.04 (1.66)	32.14 (4.26)
V	100	47.03 (7.31)	3.38 (0.45)	1.00 (0.15)	–	–
V	200	92.40 (12.74)	9.07 (1.14)	2.78 (0.30)	–	–
V	400	195.55 (24.05)	4.45 (1.15)	6.50 (1.16)	–	–

1. age: integer;
2. gender: 1 for male, 0 for female;
3. race: 1 for white, 0 for others;
4. education level: 1 for high school or below, 2 for some college, 3 for bachelor or up;
5. work status: 1 for full time, 0.5 for part time, 0 for no work;
6. bipolar type: 1 or 2;
7. status prior to the onset of the current episode: 1 for remission longer than 8 weeks;
8. status prior to the onset of the current episode: 1 for manic/hypomanic;
9. status prior to the onset of the current episode: 1 for mixed/cycling;
10. SUM-D at week 0;
11. SUM-ME at week 0.

In the second stage, we have the following variables:

1. binary indicator for adverse effect tremor;
2. binary indicator for adverse effect dry mouth;
3. binary indicator for adverse effect sedation;
4. binary indicator for adverse effect constipation;
5. binary indicator for adverse effect diarrhea;

6. binary indicator for adverse effect headache;
7. binary indicator for adverse effect poor memory;
8. binary indicator for adverse effect sexual dysfunction;
9. binary indicator for adverse effect increase appetite;
10. SUM-D at week 6;
11. SUM-ME at week 6.

References

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford Univeristy Press.

Bousquet, O. (2002). A bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences, Series I 334*(6), 495–500.

Eberts, M. and I. Steinwart (2013). Optimal regression rates for SVMs using Gaussian kernels. *Electronic Journal of Statistics 7*, 1–42.

Massart, P. (2000). About the constants in talagrand's concentration inequalities for empirical processes. *The Annals of Probability 28*(2), 863–884.

Steinwart, I. and A. Christmann (2008). *Support vector machines*. Springer-Verlag.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes, With Applications to Statistics*. Springer-Verlag.

Zhang, Y., E. B. Laber, A. Tsiatis, and M. Davidian (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics 71*(4), 895–904.