Published in final edited form as:
Electron J Stat. 2017 ; 11(2): 3927–3953. doi:10.1214/17-EJS1305.

Tree based weighted learning for estimating individualized treatment rules with censored data

Yifan Cui,
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Ruoqing Zhu, and
Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Michael Kosorok
Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

Estimating individualized treatment rules is a central task for personalized medicine. [23] and [22] proposed outcome weighted learning to estimate individualized treatment rules directly through maximizing the expected outcome without modeling the response directly. In this paper, we extend the outcome weighted learning to right censored survival data without requiring either inverse probability of censoring weighting or semiparametric modeling of the censoring and failure times as done in [26]. To accomplish this, we take advantage of the tree based approach proposed in [28] to nonparametrically impute the survival time in two different ways. The first approach replaces the reward of each individual by the expected survival time, while in the second approach only the censored observations are imputed by their conditional expected failure times. We establish consistency and convergence rates for both estimators. In simulation studies, our estimators demonstrate improved performance compared to existing methods. We also illustrate the proposed method on a phase III clinical trial of non-small cell lung cancer.

Keywords and phrases

Individualized treatment rule; Nonparametric estimation; Right censored data; Excess value bound; Recursively imputed survival trees; Outcome weighted learning

1. Introduction

An individualized treatment regime provides a personalized treatment strategy for each patient in the population based on their individual characteristics. A significant amount of work has been devoted to estimating optimal treatment rules [17, 18, 22, 24, 23]. While each of these approaches has strengths and weaknesses, we highlight the approach in [23] because of its robustness to model misspecification (this is similarly true of the approach in [22]) combined with its ability to incorporate support vector machines through the recognition that optimizing the treatment rule can be recast as a weighted classification problem. This approach is commonly referred to as outcome weighted learning. In clinical trials, right censored survival data are frequently observed as primary outcomes. Adapting outcome weighted learning to the censored setting, [26] proposed two new approaches, inverse censoring weighted outcome weighted learning and doubly robust outcome weighted learning, both of which require semiparametric estimation of the conditional censoring probability given the patient characteristics and treatment choice. The doubly robust estimator additionally involves semiparametric estimation of the conditional failure time expectation but only requires that one of the two models, for either the failure time or censoring time, be correct. Potential drawbacks of these methods are that either or both models may be misspecified and inverse censoring weighting estimation can be unstable numerically [18, 28].

In this paper, we propose a nonparametric tree based approach for right censored outcome weighted learning which avoids both the inverse probability of censoring weighting and restrictive modeling assumptions for imputation through recursively imputed survival trees [28]. Since the true failure times T are only partially known, they cannot be used directly as weights in the outcome weighted learning [23] framework. However, recursively imputed survival trees [28] provide an alternative approach to weighting by using the conditional expectations of censored observations without requiring inverse weighting. Tree-based methods [4, 3] are a broad class of nonparametric estimators which have become some of the most popular machine learning tools. Its adaptation to the survival setting has also drawn a lot of interests in the literature [14, 9, 11], and it has also been used for interpretable prediction modeling in personalized medicine [12]. The recursively imputed survival tree approach [28] combines extremely randomized trees with a recursive imputation method, which has been shown to improve performance and reduce prediction error while avoiding estimation of inverse censoring weights without making parametric or semiparametric assumptions on the conditional probability distribution of the failure time. Numerical studies demonstrate that the proposed method outperforms existing alternatives in a variety of settings.

The proposed method uses these recursively imputed survival trees to impute the survival times nonparametrically in a manner suitable for implementation within outcome weighted learning. We verify this novel approach both theoretically and in numerical examples. As part of this, we also present for the first time consistency and rate results for tree-based survival models in a more general setting than the categorical predictors considered in [10].

The remainder of the article is organized as follows. In section 2, we present the mathematical framework for individualized treatment rules for right censored survival outcomes. In section 3 we establish consistency and an excess value bound for the estimated treatment rules. Extensive simulation studies are presented in Section 4. We also illustrate our method using a phase III clinical trial on non-small cell lung cancer in Section 5. The article concludes with a discussion of future work in Section 6. Some needed technical results are provided in the Appendix.

2. Methodology

2.1. Individualized treatment regime framework

Before characterizing the individualized treatment regime, we first introduce some general notation and introduce the value function, and then extend the notation and ideas to the censored data setting. Let $X \in \mathscr{X}$ be the observed patient-level covariate vector, where $\mathscr{X}$ is a d dimensional vector space, and let $A \in \{-1, +1\}$ be the binary treatment indicator. $\tilde{T}$ is the true survival time, however, we consider a truncated version at τ, i.e., $T = \min(\tilde{T}, \tau)$, where the maximum follow-up time $\tau < \infty$ is a common practical restriction in clinical studies. The goal in this framework is to maximize a reward R, which could represent any clinical outcome. Specifically, we wish to identify a treatment rule $\mathscr{D}$, which is a map from the patient-level covariate space $\mathscr{X}$ to the treatment space $\{+1, -1\}$ which maximizes the expected reward. In the survival outcome setting, we use $R = T$ or $\log(T)$ as done in [26].

To achieve this maximization, we define the value function as

$$V(\mathscr{D}) = E^{\mathscr{D}}(R) = E\left[RI\{A = \mathscr{D}(X)\}/\pi(A; X)\right],$$

where $I\{\cdot\}$ is an indicator function, $\pi(a; X) = \text{pr}(A = a|X) > M'$ *a.s.* for some $M' > 0$ and each $a \in \{+1, -1\}$. The function π is the propensity score and is known in a randomized trial setting, which we assume is the case for this paper, but needs to be estimated in a non-randomized, observational study setting. The individualized treatment regime we are most interested in is the optimal treatment rule $\mathscr{D}*$ which maximizes the value function, i.e.

$$\mathscr{D}* = arg\max_{\mathscr{D}} E\left[RI\{A = \mathscr{D}(X)\}/\pi(A; X)\right]. \quad (1)$$

After rewriting the value function as

$$V(\mathscr{D}) = E\left[E(R|A = 1, X)I\{\mathscr{D}(X) = 1\} + E(R|A = -1, X)I\{\mathscr{D}(X) = -1\}\right],$$

it is easy to see that

$$\mathscr{D}* = \text{sign}\{E(R|A = 1, X) - E(R|A = -1, X)\}.$$

Hence, the definition of $\mathscr{D}*$ is equivalent to $\mathscr{D}*(x) = arg\max_a E(R|A = a, X = x)$. Instead of maximization the objective function in (1), the outcome weighted learning approach searches for the optimal decision rule $\mathscr{D}*$ by minimizing the weighted misclassification error, i.e.,

$$\mathscr{D}* = arg\max_{\mathscr{D}} E\left[RI\{A \neq \mathscr{D}(X)\}/\pi(A; X)\right]. \quad (2)$$

In an ideal situation, we would replace R with T or $\log(T)$. However, this is not possible under right censoring.

2.2. Value function under right censoring

Consider a censoring time C that is independent of T given (X, A). We then have the observed time $Y = \min(T, C)$, and the censoring indicator $\delta = I(T \le C)$. Assume that n independent and identically distributed copies, $\{Y_i, \delta_i, X_i, A_i\}_{i=1}^{n}$, are collected. Since T is not fully observed we seek for a sensible replacement which maintains as close as possible the same value function. We propose two approaches in the following, denoted as R_1 and R_2 respectively. The first approach is to obtain a nonparametric estimated conditional expectation $\hat{E}(T|X, A)$. Letting $R_1 = E(T|\ X, A)$ and bringing the expectation of T inside, we have

$$E\left[TI\{A=\mathscr{D}(X)\}/\pi(A;X)\right]=E\left[R_1 I\{A=\mathscr{D}(X)\}/\pi(A;X)\right]. \quad (3)$$

Another approach is to replace only the censored observations conditioning on the observed data. It is interesting to observe that the conditional expectation of T, given Y and δ, can be written as

$$R_2 := E(T|X, A, Y, \delta) = I(\delta=1)Y + I(\delta=0)E(T|X, A, Y, \delta=0) = I(\delta=1)Y + I(\delta=0)E(T|X, A, C=Y, T>Y, Y) \quad (4)$$

An important property that we used in the last equality is the conditional independence between T and C. With the information of $Y = y$ given, and knowing that $\delta = 0$, the conditional distribution of T is defined on $(c, \tau]$ with density function proportional to the original density of T. In other words, the conditional survival function of T is $S(t|X, A)/S(c|X, A)$ for $t > c$, where $S(\cdot|X, A)$ is the conditional survival function of T. Hence, we can calculate the expectation of T accordingly. With the definition of R_2, it is easy to see that the corresponding value function is equivalent to the left side of equation (3) by further taking expectations with respect to Y and δ. Note that the above arguments remain unchanged if we replace T, C and Y with $\log(T)$, $\log(C)$, and $\log(Y)$, respectively: this equivalence will be tacitly utilized throughout the paper, except when the distinction is needed.

With our proposed two reward measures, the remaining challenge is to nonparametrically estimate the conditional expectations. To this end, we utilize the nonparametric tree based method proposed by [28]. It is worth noting that the conditional expectation of T defined in R_2 shares the same logical underpinnings as the imputation step in [28]. However, the goal of the imputation step is to replace the censored observations with a randomly generated conditional failure time which utilizes the same condition survival distribution of T given $T > C$. We will provide details of the estimation procedure in the next section. To conclude this section, we provide the empirical versions of the value function using the two rewards R_1 and R_2, respectively, which we solve for the optimal decision $\mathcal{D}*$ by minimization:

$$n^{-1}\sum_{i=1}^{n}\frac{\hat{E}(T_i|A_i,X_i)I\{A_i=\mathcal{D}(X_i)\}}{\pi(A_i;X_i)}, \qquad (5)$$

$$\text{and } n^{-1}\sum_{i=1}^{n}\frac{\{\delta_iY_i+(1-\delta_i)\hat{E}(T_i|X_i,A_i,T_i>Y_i,Y_i)\}I\{A_i=\mathcal{D}(X_i)\}}{\pi(A_i;X_i)}. \qquad (6)$$

2.3. Outcome weighted learning with survival trees

The recursively imputed survival trees method proposed by [28] is a powerful tool to estimate conditional survival functions for censored data. A brief outline of the algorithm is provided in the following. We refer interested readers to the original paper for details. To fit the model, we first generate extremely randomized survival trees for the training dataset. Secondly, we calculate conditional survival functions for each censored observation, which can be used for imputing the censored value to a random conditional failure time. Thirdly, we generate multiple copies of the imputed dataset, and one survival tree is fitted for each dataset. We repeat the last two steps recursively and the final nonparametric estimate of $\hat{E}(T|X,A)$ is obtained by averaging the trees from the last step.

Following [23], we next use support vector machines to solve for the optimal treatment rule. A decision function $f(x)$ is learned by replacing $I\{A_i=\mathcal{D}(X_i)\}$ in Equations (5) or (6) with $\phi\{A_if(X_i)\}$, where $\phi(x) = (1 - x)^+$ is the hinge loss and $x^+ = \max(x, 0)$. Furthermore, to avoid overfitting, a regularization term $\lambda_n\|f\|^2$ is added to penalize the complexity of the estimated decision function f. Here, $\|f\|$ is some norm of f, and λ_n is a tuning parameter. A high-level description of the proposed method is given in Algorithm 1 below. We consider both linear and nonlinear decision functions f when solving (7). For a linear decision function, $f(x) = \theta_0 + \theta^T x$ and we let $\|f\|$ be the Euclidean norm of θ. For nonlinear decision functions, we employ a universal kernel function $k:\mathscr{X} \times \mathscr{X} \to \mathbb{R}$, such as the Gaussian kernel, which is continuous, symmetric and positive semidefinite. The optimization problem is then equivalent to a dual problem that maximizes

$$\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jA_iA_jk(X_i,X_j),$$

subject to $0 \le \alpha_{\mathrm{i}} \le \gamma W_i/\pi_i$ and $\sum_{i=1}^{n}\alpha_iA_i=0$, where W_i is the numerator in either (5) or (6) and π_i is the respective denominator. Both settings can be efficiently solved by quadratic programming. For further details regarding solving weighted classification problems using support vector machines, we refer to [23, 26, 5].

Algorithm 1: Pseudo algorithm for the proposed method—Step 1. Use $\{(X_i^{\mathrm{T}}, A_i, A_i X_i^{\mathrm{T}})^{\mathrm{T}}, Y_i, \delta_i\}_{i=1}^{n}$ to fit recursively imputed survival trees. Obtain the estimation $\hat{E}(T_i|A_i, X_i)$ for reward R_1 or the estimation $\hat{E}(T_i|X_i, A_i, T_i>Y_i, Y_i)$ for reward R_2.

Step 2. Let the weights W_i be either $\hat{E}(T_i|A_i, X_i)$ or $\delta_i Y_i+(1-\delta_i)\hat{E}(T_i|A_i, X_i, T_i>Y_i, Y_i)$, depending on which of the two proposed approaches is used. Minimize the following weighted misclassification error:

$$\hat{f}(x)=arg\min_{f}\sum_{i=1}^{n}W_i\frac{\phi\{A_i f(X_i)\}}{\pi(A_i;X_i)}+\lambda_n\|f\|^2. \qquad (7)$$

Step 3. Output the estimated optimal treatment rule $\hat{\mathscr{D}}(x)=sign\{\hat{f}(x)\}$.

3. Theoretical results

3.1. Preliminaries

The risk function is defined as

$$R(f)=E\left[\frac{R}{\pi(A;X)}I\{A\neq\mathrm{sign}(f(X))\}\right],$$

where the reward $R = R_1 = E(T|X, A)$ for the first approach, or $R = R_2 = \delta Y + (1-\delta)E(T|X, A, T > Y, Y)$ for the second one. We define ϕ-risk for both the true and the working model as, respectively, $R_\phi(f) = E[R\phi\{Af(X)\}/\pi(A; X)]$ and $R_\phi^{'}(f)=E[\hat{R}\phi\{Af(X)\}/\pi(A;X)]$, where $\hat{R}$ is the estimated value of R based on one of the two proposed methods. We also define the hinge loss function for the true and working models as $L_\phi(f) = R\phi\{Af(X)\}/\pi(A; X)$ and $L_\phi^{'}(f)=\hat{R}\phi\{Af(X)\}/\pi(A;X)$, respectively.

The proposed estimator $\hat{\mathscr{D}}=\mathrm{sign}(\hat{f}_n(X))$, where $\hat{f}_n$ is solved by one of the following optimization problems within some reproducible kernel Hilbert space $\mathscr{H}_k$:

$$\hat{f}_n=arg\min_{f\in\mathscr{H}_k}n^{-1}\sum_{i=1}^{n}\frac{\hat{E}(T_i|X_i, A_i)}{\pi(A_i;X_i)}\phi\{f(X_i)A_i\}+\lambda_n\|f\|_n^2,$$

or

$$\hat{f}_n=arg\min_{f\in\mathscr{H}_k}n^{-1}\sum_{i=1}^{n}\frac{\delta_i Y_i+(1-\delta_i)\hat{E}(T_i|X_i, A_i, T_i>Y_i, Y_i)}{\pi(A_i;X_i)}\phi\{f(X_i)A_i\}+\lambda_n\|f\|_n^2.$$

3.2. Consistency of tree-based survival models

In this section, we provide the convergence bound of a simplified tree-based survival model, which is very close to the original algorithm in [28]. The purpose of this section and its main result, Theorem 1, is to demonstrate the existence of an accurate estimator of the underlying hazard function when tree-based methods are used. An earlier result developed in [10] considers only categorical feature variables. To the best of our knowledge, what we present below is the first consistency result for a tree-based survival model under general settings with restrictions only on the splitting rules, which is interesting in its own right.

For simplicity, we assume in this section that $\mathscr{Q}_n = \{(Y_i, \delta_i, X_i, A_i), i = 1, \ldots, n\}$ is the training sample, where X_i is independent uniformly distributed on $[0, 1]^d$. The result can be easily generated to distributions with bounded support and density function bounded above and below. For any fixed x, our goal is to estimate the cumulative hazard function of failure time $r(\cdot, X, A) = \Lambda_T(\cdot|X, A)$; hereinafter, we write it as $\Lambda(\cdot|X, A)$.

A random forest is a collection of randomized regression trees $\{\hat{r}_n(\cdot, X, A, \Theta_j, \mathscr{Q}_n), 1 \leq j \leq m\}$, where m is the number of trees. The randomizing variable Θ is used to indicate how the successive cuts are performed when an individual tree is built. Hence the forest version of the survival tree model can be expressed as

$$\hat{r}_n(\cdot, X, A, \mathscr{Q}_n) = \frac{1}{m} \sum_{j=1}^{m} \hat{r}_n(\cdot, X, A, \Theta_j, \mathscr{Q}_n).$$

Here, we consider a simplified scenario in which the selection of the coordinate is completely random and independent from the training data [1]. We only consider the consistency of a single tree and denote our tree estimator as $\hat{r}_n(\cdot, X, A)$. The result can be easily extended to the situation where m is finite.

A brief description of how each individual tree is constructed is provided in the appendix. Here we highlight some key assumptions and the main result. Our first assumption puts a lower bound on the probability of observing a failure at τ, and the second one assumes the smoothness of the hazard and cumulative hazard functions.

Assumption 1—For some $M > 0$, $S_Y(\tau|X, A) > M$ almost surely.

Assumption 2—For any fixed time point t and treatment decision A, the cumulative hazard function $\Lambda(t \mid X, A)$ is L-Lipschitz continuous in terms of X, and the hazard function $\lambda(t \mid X, A)$ is L'-Lipschitz continuous in terms of X, *i.e.*, $|\Lambda(t|X_1, A) - \Lambda(t|X_2, A)| \leq L\|X_1 - X_2\|$ and $|\lambda(t|X_1, A) - \lambda(t|X_2, A) \leq L'\|X_1 - X_2\|$, respectively, where $\|\cdot\|$ is the Euclidean norm.

The following theorem provides the bound of the proposed tree based survival model for each X. Details of the proof are collected in the Appendix.

Theorem 1—Assume that Assumptions 1–2 and the construction of a tree-based survival model described in the Appendix. Further assume that $k_n \to \infty$ and $n/k_n \to \infty$ as $n \to \infty$, where k_n is a tuning parameter denoting the number of terminal nodes. For any $b = n^{\zeta}$, where $\zeta > 0$, we have for each X,

$$pr\{\sup_{t<\tau}|\hat{r}_n(t,X,A)-r(t,X,A)| \le C[d^{1/2}2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d}+b^{1/2}\{(1-u)n2^{-\lceil log_2 k_n\rceil-1}\}^{-1/2}]\} \ge 1-w_n,$$

where r, $u \in (0, 1)$, $n \ge 288b/M^4$, C is some universal constant, and

$$w_n=16[(1-u)n2^{-\lceil log_2 k_n\rceil-1}+2]e^{-b}+e^{-u^2 n2^{-\lceil log_2 k_n\rceil-1}}+de^{-\lceil log_2 k_n\rceil r^2/2d}.$$

The ideal balance happens when $k_n=n^{(1+2/d)^{-1}}$. In this case, the optimal rate of the bound is close to $n^{-(d+2)^{-1}}$. The following theorem proves consistency of the proposed tree based survival model. Details of the proof are collected in the Appendix.

Theorem 2—Assume that Assumptions 1–2 and the construction of a tree-based survival model described in the Appendix. Further assume that $k_n = n^{\eta}$, where $0 < \eta < 1$. Then the estimator of the survival tree model is consistent. Moreover, for any $b = n^{\zeta}$, where $\zeta > 0$,

$$\sup_{t<\tau} E_X|\hat{r}_n(t,X,A)-r(t,X,A)| \le C[d^{1/2}2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d}]+b^{1/2}\{(1-u)n2^{-\lceil log_2 k_n\rceil-1}\}^{-1/2}+w_n O(\,ln\,(n)),$$

where $r,u \in (0,1)$, $n \ge 288b/M^4$, C is some universal constant, and

$$w_n=16[(1-u)n2^{-\lceil log_2 k_n\rceil-1}+2]e^{-b}+e^{-u^2 n2^{-\lceil log_2 k_n\rceil-1}}+de^{-\lceil log_2 k_n\rceil r^2/(2d)}.$$

3.3. Consistency and Excess Value Bound

Fisher consistency follows directly from Proposition 3.1 in [23], hence the proof is omitted. Here we restate the result as the following lemma. For the proposed method, we simply replace the reward R in $R_{\phi}(f)$ with R_1 or R_2. Note that both versions are equivalent to the reward function $R_{\phi}(f) = E[T\phi\{Af(X)\}/\pi(A; X)]$:

Lemma 1 (Proposition 3.1 in [23])—For any measurable function $\tilde{f}$, if $\tilde{f}$ minimizes $R_{\phi}(f)$, then $\mathscr{D}*(x)=\text{sign}(\tilde{f}(x))$.

Provided the Assumptions in Section 3.2 hold, the following lemma ensures the convergence of the estimated conditional expectations. The proof is given in Appendix.

Lemma 2—Based on Theorem 1, for each X the estimated conditional expectations converge in probability, *i.e.*,

$$pr\{|\hat{E}(T|X,A)-E(T|X,A)| \le C_1[2^{\{(1-r)\lceil log_2 k_n\rceil\}/d}+(b/\{(1-u)n2^{-\lceil log_2 k_n\rceil-1}\})^{1/2}]\} \ge 1-w_n,$$

$$pr\{|\hat{E}(T|X,A,T{>}Y,Y){-}E(T|X,A,T{>}Y,Y)| \leq C_2[2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d}{+}(b/\{(1-u)n2^{-\lceil log_2 k_n\rceil -1}\})^{1/2}]\} \geq 1{-}2w_n,$$

for some constant C_1, C_2 (depending on L, L', τ, M, d).

We will use the above lemmas to prove our main theorem based on the Gaussian kernel. Before we derive the convergence rate and excess value bound, we define the value function corresponding to the true and working model as $V(f) = E(RI[A = \text{sign}\{f(X)\}]/\pi(A;X))$ and $V'(f){=}E(\hat{R}I[A{=}\text{sign}\{f(X)\}]/\pi(A;X))$, respectively. We further define the empirical L_2–norm, $\|f-g\|_{L_2(P_n)}{=}(n^{-1}\sum_{i=1}^{n}|f(X_i-g(X_i))|^2)^{1/2}$, which also defines an ε-ball based on this norm. By Theorem 2.1 in [20], we restate the bound for covering numbers:

Lemma 3 (Theorem 2.1 in [20])—For any $\beta > 0$, $0 < v < 2$, $\varepsilon > 0$ we have $sup_{P_n} logN(B_{\mathscr{H}_k}, \varepsilon, L_2(P_n)) \leq c_{v,\beta,d}\sigma_n^{(1-v/2)(1+\beta)d}\varepsilon^{-v}$, where $B_{\mathscr{H}_k}$ is the closed unit ball of $\mathscr{H}_k$, and d is the dimension of $\mathscr{X}$.

Lastly, for $\tilde{f}{=}arg\min_{f\in\mathscr{F}}E\{L_\phi(f)\}$, we define the approximation error function

$$a(\lambda){=}\underset{f\in\mathscr{H}_k}{inf}[E\{L_\phi(f)\}{+}\lambda\|f\|_k^2 - E\{L_\phi(\tilde{f})\}].$$

Then we have following theorem, the proof of which is given in Appendix.

Theorem 3—Based on Theorem 2 and assuming that the sequence $\lambda_n > 0$ satisfies $\lambda_n \to 0$ and $\lambda_n \ln n \to \infty$, we have that

$$pr(V(f*) \leq V(\hat{f}_n){+}\varepsilon) \geq 1-2e^{-\rho},$$

where f^* maximize the true value function V,

$$\begin{aligned}\varepsilon{=}&a(\lambda_n){+}M_v(n\lambda_n/c_n)^{-2/(v+2)}\\ &{+}M_v\lambda_n^{-1/2}(c_n/n)^{2/(d+2)}\\ &{+}K\rho(n\lambda_n)^{-1}\\ &{+}2K\rho n^{-1}\lambda_n^{-1/2}\\ &{+}C\lambda_n^{-1/2}\{2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d}{+}(b/\{(1-u)n2^{-\lceil log_2 k_n\rceil -1}\})^{1/2}{+}16\,ln\,n[(1-u)]n2^{-\lceil log_2 k_n\rceil -1}{+}2]e^{-b}{+}e^{-u^2}\end{aligned}$$

, $c_n{=}c_{v,\beta,d}\sigma_n^{(1-v/2)(+\beta)d}$ and $\rho > 0$ for both methods; also, M_v is a constant depending on v, K is a sufficiently large positive constant, and C is a some large constant depending on d.

The rate consists of two parts. The first part is from the approximation error using $\mathscr{H}_k$. The second part controls the approximation error due to using the proposed tree-based method to estimate the conditional expectation.

4. Simulation studies

We perform simulation studies to compare the proposed method with existing alternatives, including the Cox proportional hazards model with covariate-treatment interactions, inverse censoring weighted outcome weighted learning, and doubly robust learning, both proposed in [26]. We use survival time on the log scale $\log(T)$ as outcome. We also present for comparison an "oracle" approach which uses the true failure time on the log scale $\log(T)$ as the weight in outcome weighted learning, although this would not be implementable in practice. However, this approach is a representation of the best possible performance under the outcome weighted learning framework.

We generate X_i's independently from a uniform distribution. Treatments are generated from $\{+1, -1\}$ with equal probabilities. We present four scenarios in this simulation study. The failure time T and censoring time C are generated differently in each scenario, including both linear and nonlinear decision rules. For each case, we learn the optimal treatment rule from a training dataset with sample size $n = 200$. A testing dataset with size 10000 is used to calculate the value function under the estimated rule. Each simulation is repeated 500 times.

Tuning parameters in the tree based methods need to be selected. We mostly use the default values. The number of variables considered at each split is the integer part of the square root of d as suggested by [11] and [7]. We set the total number of trees to be 50 as suggested by [28] and use one fold imputation. For the alternative approaches such as inverse censoring weighted outcome weighted learning and doubly robust learning, a Cox proportional hazards model with covariates (X, A, XA) is used to model T and C respectively. Note that when at least one of the two working models is correctly specified, the doubly robust method enjoys consistency. We implemented outcome weighted learning using a Matlab library for support vector machine [5]. Both linear and Gaussian kernels are considered for all methods except for the Cox model approach which could be directly inverted to obtain the decision rules. The parameter λ_n is chosen by ten-fold cross-validation.

4.1. Simulation settings

For all scenarios, we generate $\tilde{T}$ and C independently. The failure time $T = \min(\tau, \tilde{T})$. For all accelerated failure time models, ε is generated from a standard normal distribution. For all Cox proportional hazards models, the baseline hazard function $\lambda_0(t) = 2t$. For all simulation results presented in this section, we consider setting the censoring rates to approximately 45% for all scenarios. We also perform a sensitivity analysis for different censoring rates (30% and 60%) for each scenario. These additional results are presented in the Appendix.

Scenario 1—Both $\tilde{T}$ and C are generated from the accelerated failure time model. $\tau = 2.5$ and $d = 10$. The optimal decision function is linear. The value of the optimal treatment rule is approximately 0.031:

$$log(\tilde{T}) = -0.2 - 0.5X_1 + 0.5X_2 + 0.3X_3 + (0.5 - 0.1X_1 - 0.6X_2 + 0.1X_3)A + \varepsilon,$$

$$log(C)=0.1-0.8X_1+0.4X_2+0.4X_3+(0.5-0.1X_1-0.6X_2+0.3X_3)A+\varepsilon.$$

Scenario 2— $\tilde{T}$ is generated from a Cox model and C is generated from the accelerated failure time model. The optimal decision function is nonlinear. τ = 8 and d = 10. The value of the optimal treatment rule is approximately 0.181:

$$\lambda_{\tilde{T}}(t|A,X)=\lambda_0(t)exp\{-0.2-1.5X_1^{1.5}+0.5X_2+(0.8-0.7X_1^{0.5}-1.2X_2^2)A\},$$

$$log(C)=-0.5+0.7X_1+X_2^2+0.6X_3+0.1X_4+(0.2+X_1^{2.5}-2X_2+0.5X_3)A+\varepsilon.$$

Scenario 3— $\tilde{T}$ is generated from an accelerated failure time model with tree structured effects. C is generated from a Cox model with nonlinear effects. τ = 8 and d = 5. The value of the optimal treatment rule is approximately 1.079:

$$log(\tilde{T})=X_1+I(X_2>0.5)I(X_3>0.5)+(0.3-X_1)A+2\{I(X_4<0.3)I(X_5<0.3)+I(X_4>0.7)I(X_5>0.7)\}A+\varepsilon,$$

$$\lambda_C(t|A,X)=\lambda_0(t)exp\{-1.5+X_1+(1+0.6X_2^{1.5})A\}.$$

Scenario 4— $\tilde{T}$ is is generated from an accelerated failure time model. C is generated from a Cox model. τ = 2 and d = 10. The value of the optimal treatment rule is approximately −0.389:

$$log(\tilde{T})=-0.5-0.8X_1+0.7X_2+0.2X_3+(0.6-0.4X_1-0.2X_2-0.4X_3)A+\varepsilon,$$

$$\lambda_C(t|A,X)=\lambda_0(t)exp\{-0.5X_1-0.5X_2+0.2X_3-(1-0.5X_1+0.3X_2-0.5X_3)A\}.$$

4.2. Simulation results

Figure 1 shows the boxplot of values based on the logarithm of T calculated from the test data. The mean and standard deviation of values are shown in Table 1. In scenario 1, since the model is not correctly specified for inverse probability of censoring outcome weighted learning, the doubly robust estimator, or Cox regression, our method performs better than all other competitors.

In scenario 2, we added some nonlinear terms into both the Cox and accelerated failure time models. The model assumptions for inverse censoring outcome weighted learning and the doubly robust estimator are not satisfied. Our estimated treatment rule performs much better than these two. Compared with inverse censoring outcome weighted learning and doubly robust learning, both our approaches improve more than 0.1 for the mean. Since the true model for the failure time is the Cox model, Cox regression performs better here. In this case, the Gaussian kernel performs less well than the linear kernel for most methods since the true model structure is linear and the Gaussian kernel is too flexible.

For scenario 3, which has a more complicated tree structure, the Gaussian kernel performs better than the linear kernel for all outcome weighted learning approaches. The performance of the Gaussian kernel is enhanced since it can better address the true nonlinear model structure. We can see that with either a linear or Gaussian kernel, our estimators perform better than Cox regression. Compared with doubly robust learning, our two approaches improve 0.2 for the mean.

In scenario 4, we see that when the model is correctly specified for inverse probability of censoring outcome weighted learning and doubly robust learning, the performances of both approaches are satisfactory while our methods seem to be only a little better. The performances of our first approach, inverse probability of censoring outcome weighted learning and Cox regression are all similar. Our second approach has the best treatment effect among all estimators. Note that our second approach appears to perform as well as the first, oracle approach. Also, our two proposed methods have smaller standard errors in scenarios 1 and 3. The standard error is similar for all outcome weighted learning approaches in scenario 2 and 4. Overall, our proposed methods have generally lower variances.

Compared with results of censoring rates (30% and 60%) in the Appendix, we can observed a consistently pattern that lower censoring rate leads to higher performances in terms of both mean value and variance. The relative performances between the proposed and the competing methods remain similar across different censoring rates.

5. Data Analysis

We apply the proposed method to a non-small-cell lung cancer randomized trial dataset described in [19]. 228 subjects with complete information are used in this analysis. Each treatment arm contains 114 subjects. The censoring rate is 29%. Here we use five covariates: performance status (119 subjects ranging from 90% to 100% and 109 subjects ranging from 70% to 80%), cancer stage (31 subjects in stage 3 and 197 subjects in stage 4), race (167 white, 54 black and 7 others), gender (143 male and 85 female), age (ranging from 31 to 82 with median 63). The length of study is $\tau = 104$ weeks. We adopt the same tuning parameters used in the simulation study for this analysis. The value function is again calculated by using the logarithm of survival time $\log(T)$ (in weeks) as the reward.

We randomly divide the 228 patients into four equal proportions and use three parts as training data to estimate the optimal rule and calculate the empirical value based on the remaining part. We then permute the training and testing portions and average the four results. This procedure is then repeated 100 times and averaged to obtain the mean and standard deviation. To calculate the testing data performance, we consider two different measurements, both are calculated based on the formula

$\sum_{i=1}^{n} R_i I\{A_i = \mathscr{D}(X_i)\} / \sum_{i=1}^{n} I\{A_i = \mathscr{D}(X_i)\}$ for the testing samples, where two versions of R_i's are used. We first consider the procedure proposed in [26], where R is defined as

$$\frac{\Delta Y}{\hat{S}_C(Y|A,Y)} - \int \hat{E}_{\tilde{T}}\{T|T>t, A, X\} \left\{ \frac{dN_C(t)}{\hat{S}_C(t|A,X)} + I(Y_i \geq t) \frac{d\hat{S}_C(t|A,X)}{\hat{S}_C(t|A,X)^2} \right\}.$$

Here, $\hat{S}_C(t|A,X)$ and $\hat{E}_{\tilde{T}}(T|T>t, A, X)$ are estimated from the Cox model for simplicity. We also consider a more direct clinical measurement without the double robustness correction, which can be interpreted in a similar way as the expected survival time or the restricted mean survival time [6, 16, 21]. To be specific, we consider a restricted mean (log) survival time truncated at τ defined as $\delta T + (1 - \delta)E(T)$, and use this as a plug-in quantity of R in the testing performance calculation. To estimate this quantity, we use a recursively imputed survival trees (RIST) method to produce the expected survival time $E(T)$. The results are presented in Tables 2 and 3 and Figures 2 and 3.

The value function results are presented in Table 2 and Figure 2. Both proposed methods have higher values than the compared methods. Note that for the Gaussian kernel, our two new approaches are still better than Cox regression, however, inverse probability of censoring outcome weighted learning and doubly robust learning are not much different from Cox regression. The standard error is comparable among all four methods using the linear kernel. For the Gaussian kernel, the standard errors of the proposed methods and inverse probability of censoring weighted learning are similar. The standard error for the doubly robust method is slightly worse in this instance. Overall, the proposed methods seem to perform best.

The restricted log mean results are presented in Table 3 and Figure 3. Note for the linear kernel, the median of the proposed methods are higher than 3.6 and median of both inverse probability of censoring outcome weighted learning and doubly robust learning are lower. For the Gaussian kernel, the proposed methods are much better than inverse probability of censoring outcome weighted learning and doubly robust learning. Interestingly, under this measure, the performance of Cox regression is the best. A possible reason is that the true underlying model may not deviate much from the proportional hazard model, making the Cox model a better choice. This is also reflected by the fact that the results look similar to the simulation Scenario 2 plot, where the Cox model performs the best. Another possible reason is that the pseudo-outcome estimated from RIST may not be completely accurate and favors the Cox model in this particular dataset.

6. Discussion

We proposed a new method that redefines the reward function in a censored survival setting. The method works by replacing the censored observations (or all observations) by an estimated conditional expectation of the failure time. In practice, the failure time (or logarithm of the failure time) is commonly used in defining the reward function R, however, this choice could more flexible. For example, we may be interested in searching for a treatment rule that maximizes the median survival time or a certain quantile. Under our framework, this is achievable by replacing the censored observations with a suitable estimate of the quantile. This part of the work is currently under investigation.

The proposed methods may be improved or extended in multiple ways. The estimated treatment rule may be affected by the shift of the outcome. A potential extension is to combine our methods with residual weighted learning [27], which has been shown to reduce the total variation of the weights and improve stability. Trials with multiple treatment arms occur frequently. Thus a potential extension of our method is in the direction of multicategory classification [2, 15]. It is also interesting to extend our method to dynamic treatment regimes where a sequence of decision rules [17, 24, 13, 25] need to be learned in a censored survival outcome setting [8].

Acknowledgments

This research is supported in part by U.S. National Science Foundation grant DMS-1407732 and by U.S. National Institutes of Health grant P01 CA142538. We thank Yinqi Zhao for helpful conversations and suggestions. We thank the editor, associated editor, and reviewers for helpful comments which led to an improved manuscript.

Appendix

A simplified tree-based survival model used in Theorem 1

We consider a simplified version of a tree-based survival model. Starting from the root node $[0, 1]^d$, at each internal node, we randomly chose the j-th feature of X to split the node, while the splitting point is always at the midpoint of the range of the chosen feature. We repeat splitting $\lceil \log_2 k_n \rceil$ times, where k_n is a deterministic parameter which we can control. Hence, each individual tree has exactly $2^{\lceil log_2 k_n \rceil}$ terminal nodes, which is approximately k_n. In practice, we always chose k_n to go to infinity as n goes to infinity.

After we build an individual tree, let $B_i (i=1, 2, \ldots, 2^{\lceil log_2 k_n \rceil})$ be the rectangular cell of the random partition. We treat observations inside each leaf node as a group of homogeneous subjects and compute the Nelson-Aalen estimator $\hat{\Lambda}(\cdot|B_i)$ for each leaf node B_i. Hence, our estimator is essentially

$$\hat{r}_n(\cdot, X, A) = \sum_{i=1}^{2^{\lceil log_2 k_n \rceil}} I\{(X, A) \in B_i\} \hat{\Lambda}(\cdot|B_i).$$

Proof of Theorem 1

Proof

Since we always assume that the treatment variable A is important, and A has only two categories, we force a split on A at the root node. This is equivalent to fitting trees for $A = 1$ and $A = -1$ separately. In a balanced design, the problem reduces to estimating $\hat{r}(\cdot, X, 1)$ or $\hat{r}(\cdot, X, 1)$ with sample size $n/2$. Without the risk of ambiguities, the following results are developed for $\hat{r}(\cdot, X)$ with sample size n, where the results can be applied to either $A = 1$ or -1. Our proof utilizes two facts from [1]:

Fact 1 Let $K_{nj}\{B_i\}$ be the number of times the j-th coordinate ($j = 1, \ldots, d$) is split on to reach the terminal node B_i, $(i{=}1, 2, \ldots, 2^{\lceil log_2 k_n \rceil})$. Conditionally on X, $K_{nj}\{B_i\}$ is $Binomial(\lceil \log_2 k_n \rceil, 1/d)$. Moreover, $\sum_{j=1}^{d} K_{nj}\{B_i\}{=}\lceil log_2 k_n \rceil$.

Fact 2 Let $N_n(B_i)$ be the number of data points falling in the cell B_i, $(i{=}1, 2, \ldots, 2^{\lceil log_2 k_n \rceil})$. Conditionally on Θ, $N_n(B_i)$ follows $Binominal(n, 2^{-\lceil log_2 k_n \rceil})$.

The following lemma, for later reference, provides the deterministic limit of the Nelson-Aalen estimator in the independent non-identically distributed case. The proof can be found in an unpublished technical report by Mai Zhou at the University of Kentucky.

Lemma 4

Suppose we have two sets of non-negative random variables: $T_1, T_2, \ldots, T_n$ which are survival times, independent but non-identically distributed with continuous distribution $F_1(t), F_2(t), \ldots, F_n(t); C_1, C_2, \ldots, C_n$ which are censoring times, independent but non-identically distributed with continuous distribution $G_1(t), G_2(t), \ldots, G_n(t)$. We also assume the $T_i^{'}s$ and $C_i^{'}s$ are independent. The Nelson-Aalen estimator of data $Y_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$ is $\hat{\Lambda}(t)$. Provided Assumption 1, for $b = n^{\zeta}$, where $\zeta > 0$,

$$pr(\underset{t<\tau}{sup}|\hat{\Lambda}(t) - \int_0^t \frac{\sum_i \{1-G_i(s)\}dF_i(s)}{\sum_i \{1-G_i(s)\}\{1-F_i(s)\}}| > \frac{(1152 b^{1/2})}{n^{1/2} M^2}) < 16(n{+}2)e^{-b}. \qquad (8)$$

Now we start the proof of Theorem 1. Let the limit of the Nelson-Aalen estimator inside the cell B_i, $(i{=}1, 2, \ldots, 2^{\lceil log_2 k_n \rceil})$ be

$$\Lambda * (t|B_i){=}\int_0^t \frac{[\sum_{X_j \in B_i} \{1-G_j(s)\}dF_j(s)]}{[\sum_{X_j \in B_i} \{1-G_j(s)\}\{1-F_j(s)\}]}.$$

For any $t < \tau$, in order to bound the $|\hat{r}_n(t, X) - r(t, X)|$, we define

$$r_n^*(t, X){=}\sum_{i=1}^{2^{\lceil log_2 k_n \rceil}} I\{X \in B_i\}\Lambda * (t|B_i).$$

Then $|\hat{r}_n(t, X) - r(t, X)|$ can be decomposed as

$$|\hat{r}_n(t, X) - r(t, X)|{=}|\hat{r}_n(t, X) - r_n^*(t, X)|{+}|r_n^*(t, X) - r(t, X)|. \qquad (9)$$

We start with the first term in Equation (9). From Fact 2, we know the number of observations in each terminal node is $Binominal(n, 2^{-\lceil log_2 k_n \rceil})$. By the Chernoff bound, with probability larger than $1-e^{-u^2n2^{-\lceil log_2k_n\rceil-1}}$, in one terminal node we have at least $(1-u)n2^{-\lceil log_2k_n\rceil-1}$ observations for some $0 < u < 1$.

Combining Equation (8), with probability larger than $1-16[(1-u)n2^{-\lceil log_2k_n\rceil-1}+2]e^{-b}-e^{-u^2n2^{-\lceil log_2k_n\rceil-1}}$, the following equation holds:

$$|\hat{r}_n(t,X)-r_n^*(t,X)| \le \sum_{i=1}^{2^{\lceil log_2k_n\rceil}} I\{X\in B_i\}(1152b)^{1/2}\{(1-u)n2^{-\lceil log_2k_n\rceil-1}\}^{-1/2}M^{-2} \quad (10)$$

$$=(1152b)^{1/2}\{(1-u)n2^{-\lceil log_2k_n\rceil-1}\}^{-1/2}M^{-2}. \quad (11)$$

Before we bound the second term in Equation (9). We first show the bound for the difference between the true cumulative hazard function and aggregated estimator inside the cell B_i, $(i=1,2,\ldots,2^{\lceil log_2k_n\rceil})$, i.e. $|I\{X\in B_i\}\{\Lambda^*(t|B_i)-\Lambda(t|X)\}$.

From Fact 1, we know the number of times the terminal node B_i is split on the j-th coordinate ($j = 1, \cdots, d$) $K_{nj}\{B_i\}$ is $Binomial(\lceil\log_2 k_n\rceil, 1/d)$. By the Chernoff bound, $P(K_{nj}\{B_i\}\le(1-r)\lceil log_2k_n\rceil/d)\le e^{-\lceil log_2k_n\rceil r^2/(2d)}$ for some $0 < r < 1$. So with probability $(1-e^{-\lceil log_2k_n\rceil r^2/(2d)})^d \ge 1-de^{-\lceil log_2k_n\rceil r^2/(2d)}$, every dimension of B_i is less than $2^{-\{(1-r)\lceil log_2k_n\rceil\}/d}$. So with probability larger than $1-2^{\lceil log_2k_n\rceil}de^{-\lceil log_2k_n\rceil r^2/(2d)}$, for arbitrary i, $(i=1,2,\ldots,2^{\lceil log_2k_n\rceil})$, we have

$$\max_{X_1,X_2\in B_i}\|X_1-X_2\|\le d^{1/2}2^{-\{(1-r)\lceil log_2k_n\rceil\}/d}.$$

So for all the observations X_j inside the same cell as X, by Assumption 2, we have

$$|F_X(\cdot)-F_j(\cdot)|\le Ld^{1/2}2^{-\{(1-r)\lceil log_2k_n\rceil\}/d},$$

$$|f_X(\cdot)-f_j(\cdot)|\le(L^{'}+L^2)d^{1/2}2^{-\{(1-r)\lceil log_2k_n\rceil\}/d},$$

where $f_X(\cdot)$ and $F_X(\cdot)$ denote the true density function and distribution function at X, respectively. Then $\Lambda^*(t|B_i)$ has the upper bound and lower bound

$$\int_0^t[f_X(s)+b_1]/[1-F_X(s)-b_2]ds \text{ and } \int_0^t[f_X(s)+b_1]/[1-F_X(s)-b_2]ds,$$

respectively, where

$$b_1 = (L^{'} + L^2) d^{1/2} 2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d} \text{ and } b_2 = L d^{1/2} 2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d}.$$

Hence, $|I\{X \in B_i\}\{\Lambda^*(t \mid B_i) - \Lambda(t \mid X)\}|$ has the bound

$$\int_0^t \frac{b_1(1 - F(s)) + b_2 f(s)}{(1 - F(s) - b_2)(1 - F(s))} ds \le C_\tau d^{1/2} 2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d},$$

where C is some constant depending on L and L'. We then bound the second term of Equation (9) as follows:

$$|r_n^*(t, X) - r(t, X)| \le \sum_{i=1}^{2^{\lceil log_2 k_n \rceil}} I\{X \in B_i\} |\Lambda(t|X) \le C\tau d^{1/2} 2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d}. \quad (12)$$

Combining Equation (10) and (12), For each X, we have

$$\mathrm{pr}[\sup_{t<\tau} |\hat{r}_n(t, X) - r(t, X)| \le C[\tau d^{1/2} 2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d}] + (1152b)^{1/2} \{(1-u) n 2^{-\lceil log_2 k_n \rceil - 1}\}^{-1/2} M^{-2}] \ge 1 - w_n,$$

where

$$w_n = 16[(1-u) n 2^{-\lceil log_2 k_n \rceil - 1} + 2] e^{-b} + e^{-u^2 n 2^{-\lceil log_2 k_n \rceil - 1}} + d e^{-\lceil log_2 k_n \rceil r^2/(2d)}.$$

This completes the proof. □

Proof of Theorem 2

Proof

Based on Theorem 1, we now only need to establish the bound of $|\hat{r}_n(t, X, A) - r(t, X, A)|$ under the event with small probability w_n. Noticing that $\hat{r}_n(t, X, A)$ is simply the Nelson-Aalen estimator of the cumulative hazard function with at most n terms, for any $t < \tau$ we have

$$\hat{r}_n(t, X, A) \le \frac{1}{n} + \ldots + \frac{1}{1} = O(\ln(n)),$$

which implies that

$$|\hat{r}_n(t, X, A) - r(t, X, A)| \le O(\ln(n)).$$

Combining this with Theorem 1 completes the proof. □

Proof of Lemma 2

Proof

Our survival function estimator is $\hat{S}(t)=e^{-\hat{\Lambda}(t)}$. From Theorem 1, we know that for any $t < \tau$,

$$\begin{aligned} & pr(|\hat{S}(t|X,A) - S(t|X,A| \le C[2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d} \\ & +(b/\{(1-u)n2^{-\lceil log_2 k_n \rceil -1}\})^{1/2}]) \ge 1 - 16[(1 \\ & -u)n2^{-\lceil log_2 k_n \rceil -1} \\ & +2]e^{-b} \\ & - e^{-e^2 n2^{-\lceil log_2 k_n \rceil -1}} \\ & - de^{-\lceil log_2 k_n \rceil r^2/(2d)}. \end{aligned}$$

It is then easy to see that for R_1,

$$\begin{aligned} & \left|\hat{E}(T|X,A) - E(T|X,A)\right| \\ & = \left|\int_0^\tau \hat{S}(t|X,A)dt - \int_0^\tau S(t|X,A)dt\right| \le \int_0^\tau |\hat{S}(t|X,A) - S(t|X,A)|dt \le \tau C[2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d} + (b/\{(1-u)n2^{-\lceil log_2 k_n \rceil -1} \end{aligned}$$

with probability larger than $1 - w_n$. And for reward R_2, we have

$$\begin{aligned} & \left|\hat{E}(T|X,A,T>Y,Y) - E(T|X,A,T>Y,Y)\right| \\ & = \left|\int_Y^\tau \{\hat{S}(t|X,A)/\hat{S}(Y|X,A)\}dt\right. \\ & -\int_Y^\tau \{S(t|X,A)/S(Y|X,A)\}dt \le |\int_Y^\tau \{\hat{S}(t|X,A)/\hat{S}(Y|X,A)\}dt - \int_Y^\tau \{\hat{S}(t|X,A)/S(Y|X,A)\}dt| \\ & +|\int_Y^\tau \{\hat{S}(t|X,A)/S(Y|X,A)\}dt - \int_Y^\tau \{S(t|X,A)/S(X,A)\}dt|. \end{aligned}$$

Note that we can bound the distance between $\hat{S}(Y|X,A)$ and $S(Y|X, A)$ with probability no less than $1 - w_n$, which is further bounded above by

$$(1/M^2 + 1/M)\int_Y^\tau |\hat{S}(Y|X,A) - S(Y|X,A)|dt \le C_2[2^{-\{(1-r)\lceil log_2 k_n \rceil\}/d} + (b/\{(1-u)n2^{-\lceil log_2 k_n \rceil -1}\})^{1/2}],$$

for some constant C_2 with probability larger than $1 - 2w_n$. □

Proof of Theorem 3

Proof

We restate the value function corresponding to the true and working model as

$$V(f) = E(RI[A = \text{sign}\{f(x)\}]/\pi(A;X))$$

$$\text{and } V^{'}(f){=}E(\hat{R}I[A{=}\text{sign}\{f(x)\}]/\pi(A;X)),$$

respectively. Then we have

$$\begin{aligned} V(f*) - V(\hat{f}_n) \le V(f*) \\ - \underset{f\in\mathscr{F}}{sup} V^{'}(f) \\ + \underset{f\in\mathscr{F}}{sup} V^{'}(f) \\ - V^{'}(\hat{f}_n) \\ + V^{'}(\hat{f}_n) \\ - V(\hat{f}_n) \le V(f*) - V^{'}(f*) \\ + \underset{f\in\mathscr{F}}{sup} V^{'}(f) \\ - V^{'}(\hat{f}_n) \\ + V^{'}(\hat{f}_n) \\ - V(\hat{f}_n) \le \underset{f\in\mathscr{F}}{sup} V^{'}(f) \\ - V^{'}(\hat{f}_n) \\ + 2\underset{f\in\mathscr{F}}{sup} |V(f) - V^{'}(f)|. \end{aligned} \tag{13}$$

We start with the first term in Equation (13). From Lemma 1, we know that $sup_{f\in\mathscr{F}} V^{'}(f) - V^{'}(\hat{f}_n){=}V^{'}(\tilde{f}) - V^{'}(\hat{f}_n)$, where $\tilde{f}{=}argmin_{f\in\mathscr{H}_k} E\{L_\phi(f)\}$.

Let $\tilde{f}_{\lambda_n}{=}argmin_{f\in\mathscr{H}_k}[E\{R_\phi\{Af(X)\}/\pi(A;X)\}{+}\lambda_n\|f\|_k^2]$, then

$$n^{-1}\sum_{i=1}^{n}\frac{\hat{R}\phi\{A_i\hat{f}(X_i)\}}{\pi(A_i;X_i)}{+}\lambda_n\|\hat{f}\|_k^2 \le n^{-1}\sum_{i=1}^{n}\frac{\hat{R}\phi\{A_i\tilde{f}_{\lambda_n}(X_i)\}}{\pi(A_i;X_i)}{+}\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2. \tag{14}$$

By the definition of $a(\lambda)$, we have

$$a(\lambda_n){=}[E\{L_\phi(\tilde{f}_{\lambda_n})\}{+}\lambda\|\tilde{f}_{\lambda_n}\|_k^2 - E\{L_\phi(\tilde{f})\}],$$

and by Theorem 3.2 in [23], we further have

$$
\begin{aligned}
V(\tilde{f})-V(\hat{f}) \leq E\{L_\phi(\hat{f})\}\\
-E\{L_\phi(\tilde{f})\} \leq E\{L_\phi(\hat{f})\}\\
-E\{L_\phi(\tilde{f}_{\lambda_n})\}-\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2+E\{L_\phi(\tilde{f}_{\lambda_n})\}\\
-E\{L_\phi(\tilde{f})\}\\
+\lambda\|\tilde{f}_{\lambda_n}\|_k^2 \leq E\{L_\phi(\hat{f})\}\\
-E\{L_\phi(\tilde{f}_{\lambda_n})\}\\
-\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2\\
+\lambda_n\|\hat{f}\|_k^2+a(\lambda_n).
\end{aligned}
$$

Combined with (14),

$$
\begin{aligned}
&V(\tilde{f})-V(\hat{f}) \leq a(\lambda_n)\\
&+E\left[\frac{R\phi\{A\hat{f}(X)\}}{\pi(A;X)}-\frac{\hat{R}\phi\{A\hat{f}(X)\}}{\pi(A;X)}\right]\\
&+E\left[\frac{\hat{R}\phi\{A\tilde{f}_{\lambda_n}(X)\}}{\pi(A;X)}-\frac{R\phi\{A\tilde{f}_{\lambda_n}(X)\}}{\pi(A;X)}\right]\\
&-\left(n^{-1}\sum_{i=1}^{n}[\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2-\frac{\hat{R}\phi\{A_i\tilde{f}_{\lambda_n}(X_i)\}}{\pi(A_i;X_i)}]+E[\lambda_n\|\hat{f}\|_k^2+\frac{\hat{R}\phi\{A\hat{f}(X)\}}{\pi(A;X)}-\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2-\frac{\hat{R}\phi\{A_i\tilde{f}_{\lambda_n}(X_i)\}}{\pi(A;X)}]\right)\\
&=a(\lambda_n)\\
&+(\mathrm{I})+(\mathrm{II})+(\mathrm{III}).
\end{aligned}
$$

Since

$$
n^{-1}\sum_{i=1}^{n}\frac{\hat{R}\phi\{A_i\hat{f}(X_i)\}}{\pi(A_i;X_i)}+\lambda_n\|\hat{f}\|_k^2 \leq n^{-1}\sum_{i=1}^{n}\frac{\hat{R}\phi(0)}{\pi(A_i;X_i)}=n^{-1}\sum_{i=1}^{n}\frac{\hat{R}}{\pi(A_i;X_i)},
$$

and the estimated value function $\hat{R}$ is bounded by τ, we know that $\|\hat{f}\|_k \leq \tau^{1/2}\lambda_n^{-1/2}$. Furthermore, since

$$
\lambda_n\|\tilde{f}_{\lambda_n}\|_k^2 \leq \underset{f\in\mathscr{H}_k}{inf}\ \lambda_n\|f\|_k^2+E\left[\frac{R\phi\{Af(X)\}}{\pi(A;X)}\right] \leq E\left[\frac{R\phi(0)}{\pi(A;X)}\right],
$$

we have $\|\tilde{f}_{\lambda_n}\|_k \leq \tau^{1/2}\lambda_n^{-1/2}$. Combining with Lemma 2, (I) and (II) are bounded by

$C_1\lambda_n^{-1/2}\{2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d}+(b/\{(1-u)n2^{-\lceil log_2 k_n\rceil-1}\})^{1/2}+16\,ln\,n[(1{-}u)n2^{-\lceil log_2 k_n\rceil-1}+2]e^{-b}+e^{-u^2n2^{-\lceil log_2 k_n}}$
for both R_1 and R_2, where C_1 is some constant. Following the results in [26], (III) is bounded by $M_v(n\lambda_n/c_n)^{-2/(v+2)}+M_v\lambda_n^{-1/2}(c_n/n)^{2/(d+2)}+K\rho(n\lambda_n)^{-1}+2K\rho n^{-1}\lambda_n^{-1/2}$
with probability larger than $1-2e^{-\rho}$, where M_v is a constant depending on v and K is a sufficiently large positive constant. Finally, combining (I), (II) and (III), we have

$$\mathrm{pr}(\underset{f \in \mathscr{F}}{sup} V'(f) \le V'(\hat{f}_n) + \varepsilon_1) \ge 1 - 2e^{-\rho}, \quad (15)$$

where

$$\varepsilon_1 = a(\lambda_n) + M_v(n\lambda_n/c_n)^{-2/(v+2)} + M_v\lambda_n^{-1/2}(c_n/n)^{2/(d+2)} + K\rho(n\lambda_n)^{-1} + 2K\rho n^{-1}\lambda_n^{-1/2} + C_1\lambda_n^{-1/2}\{2^{-\{(1-r)\lceil log_2 k_n}$$

.

For the second part in Equation (13),

$$\begin{aligned} &V(f) - V'(f) \\ &= E\left(\frac{RI[A = \text{sign}\{f(X)\}]}{\pi(A;X)}\right) \\ &- E\left(\frac{\hat{R}I[A = \text{sign}\{f(X)\}]}{\pi(A;X)}\right) \\ &= E(\{E(T|X,A) - \hat{E}(T|X,A)\}\frac{I[A = \text{sign}\{f(x)\}]}{\pi(A;X)}) \end{aligned}$$

if $R = R_1$. For $R = R_2$, we have

$$V(f) - V'(f) = E((1-\delta)\{E(T|X,A,T>Y,Y) - \hat{E}(T|X,A,T>Y,Y)\}\frac{I[A = \text{sign}\{f(X)\}]}{\pi(A;X)}).$$

By Lemma 2,

$$\begin{aligned} &\underset{f \in \mathscr{F}}{sup}|V(f) \\ &-V'(f)| \le C_2\lambda_n^{-1/2}\{2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d} + (b/\{(1-u)n2^{-\lceil log_2 k_n\rceil - 1}\})^{1/2} + 16\, ln\, n[(1-u)n2^{-\lceil log_2 k_n\rceil - 1} + 2]e^{-b} + e^{-u^2 n2^{-\lceil l}} \end{aligned} \quad (16)$$

where C_2 is some constant. Now, combining (15) and (16) we have

$$\mathrm{pr}(V(f*) \le V(\hat{f}_n) + \varepsilon) \ge 1 - 2e^{-\rho},$$

where

$$\varepsilon = a(\lambda_n) + M_v(n\lambda_n/c_n)^{-2/(v+2)} + M_v\lambda_n^{-1/2}(c_n/n)^{2/(d+2)} + K\rho(n\lambda_n)^{-1} + 2K\rho n^{-1}\lambda_n^{-1/2} + C\lambda_n^{-1/2}\{2^{-\{(1-r)\lceil log_2 k_n\rceil\}/d} + (b/\{(1-u)n$$

This completes the proof. □

Additional simulation results for different censoring rates

We summarize the additional simulation results in this section. For each simulation scenario considered in Section 4, we alter the first constant term in the censoring distribution to achieve 30% (Table 4 and Figure 4), and 60% (Table 5 and Figure 5) censoring rates.

References

1. Biau G. Analysis of a random forests model. Journal of Machine Learning Research. 2012; 13:1063–1095.
2. Bredensteiner, EJ., Bennett, KP. Computational Optimization. Springer; 1999. Multicategory classification by support vector machines; p. 53-79.
3. Breiman L. Random forests. Machine learning. 2001; 45:5–32.
4. Breiman, L., Friedman, J., Stone, CJ., Olshen, RA. Classification and regression trees. CRC press; 1984.
5. Chang CC, Lin CJ. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST). 2011; 2:27.
6. Geng Y, Zhang HH, Lu W. On optimal treatment regimes selection for mean survival time. Statistics in medicine. 2015; 34:1169–1184. [PubMed: 25515005]
7. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine learning. 2006; 63:3–42.
8. Goldberg Y, Kosorok MR. Q-learning with censored data. Annals of statistics. 2012; 40:529. [PubMed: 22754029]
9. Hothorn T, Lausen B, Benner A, Radespiel-Tröger M. Bagging survival trees. Statistics in medicine. 2004; 23:77–91. [PubMed: 14695641]
10. Ishwaran H, Kogalur UB. Consistency of random survival forests. Statistics & probability letters. 2010; 80:1056–1064. [PubMed: 20582150]
11. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The annals of applied statistics. 2008:841–860.
12. Laber E, Zhao Y. Tree-based methods for individualized treatment regimes. Biometrika. 2015; 102:501–514. [PubMed: 26893526]
13. Laber EB, Lizotte DJ, Qian M, Pelham WE, Murphy SA. Dynamic treatment regimes: Technical challenges and applications. Electronic journal of statistics. 2014; 8:1225. [PubMed: 25356091]
14. LeBlanc M, Crowley J. Relative risk trees for censored survival data. Biometrics. 1992:411–425. [PubMed: 1637970]
15. Lee Y, Lin Y, Wahba G. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. Journal of the American Statistical Association. 2004; 99:67–81.
16. Ma J, Hobbs BP, Stingo FC. Statistical methods for establishing personalized treatment rules in oncology. BioMed research international. 2015:2015.
17. Murphy SA. Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2003; 65:331–355.
18. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. Annals of statistics. 2011; 39:1180. [PubMed: 21666835]
19. Socinski MA, Schell MJ, Peterman A, Bakri K, Yates S, Gitten R, Unger P, Lee J, Lee JH, Tynan M, et al. Phase iii trial comparing a defined duration of therapy versus continuous therapy followed by second-line therapy in advanced-stage iiib/iv non–small-cell lung cancer. Journal of Clinical Oncology. 2002; 20:1335–1343. [PubMed: 11870177]
20. Steinwart I, Scovel C. Fast rates for support vector machines using gaussian kernels. The Annals of Statistics. 2007:575–607.
21. Tian L, Zhao L, Wei L. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. Biostatistics. 2014; 15:222–233. [PubMed: 24292992]

22. Zhang B, Tsiatis AA, Laber EB, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics. 2012; 68:1010–1018. [PubMed: 22550953]

23. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association. 2012; 107:1106–1118. [PubMed: 23630406]

24. Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics. 2011; 67:1422–1433. [PubMed: 21385164]

25. Zhao YQ, Zeng D, Laber EB, Kosorok MR. New statistical learning methods for estimating optimal dynamic treatment regimes. Journal of the American Statistical Association. 2015; 110:583–598. [PubMed: 26236062]

26. Zhao YQ, Zeng D, Laber EB, Song R, Yuan M, Kosorok MR. Doubly robust learning for estimating individualized treatment with censored data. Biometrika. 2015; 102:151–168. [PubMed: 25937641]

27. Zhou X, Mayer-Hamblett N, Khan U, Kosorok MR. Residual weighted learning for estimating individualized treatment rules. Journal of the American Statistical Association. 2015:00–00.

28. Zhu R, Kosorok MR. Recursively imputed survival trees. Journal of the American Statistical Association. 2012; 107:331–340. [PubMed: 23125470]

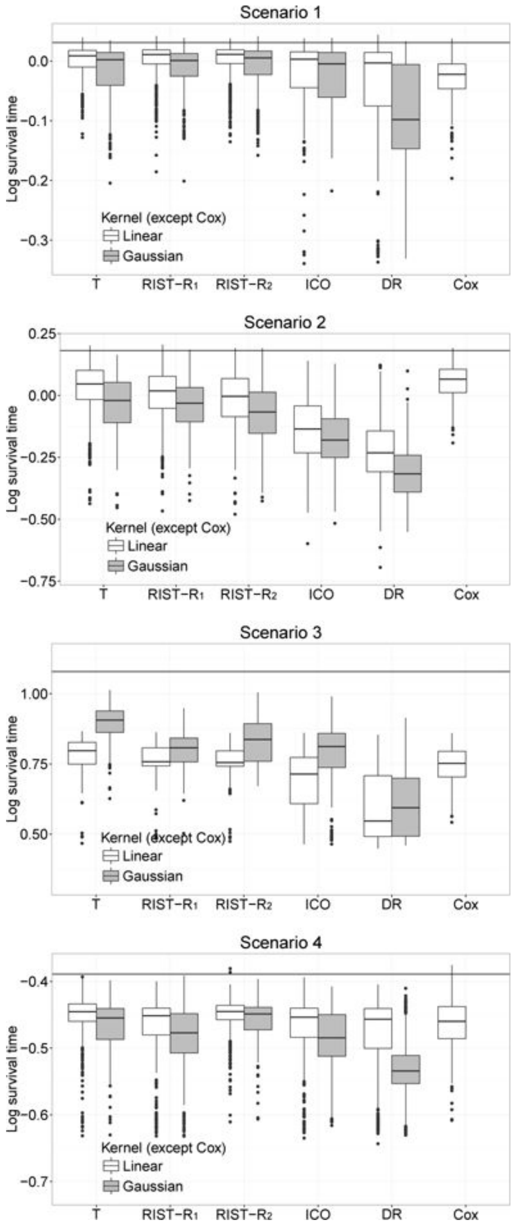

Fig 1.

Boxplots of mean log survival time for different treatment regimes. Censoring rate: 45%. T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning. The black horizontal line is the theoretical optimal value.

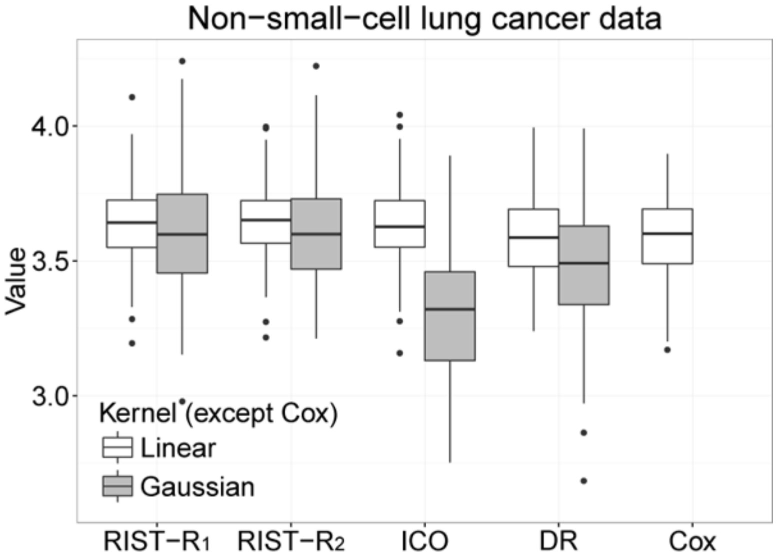

Fig 2.

Boxplots of cross-validated value of survival weeks on the log scale. RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning.

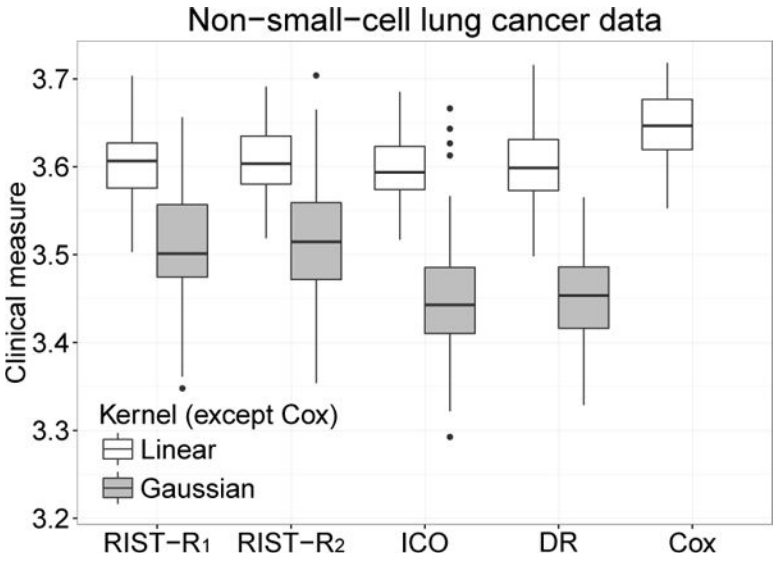

Fig 3.

Boxplots of cross-validated value of survival weeks on the log scale. RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning.

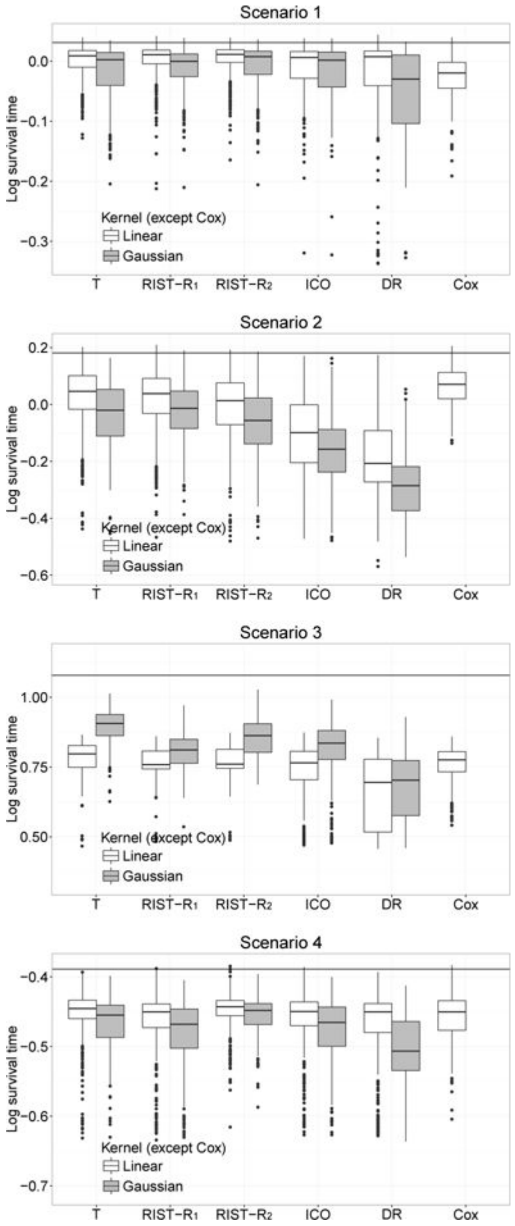

Fig 4.

Boxplots of mean log survival time for different treatment regimes. Censoring rate: 30%. T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning. The black horizontal line is the theoretical optimal value.

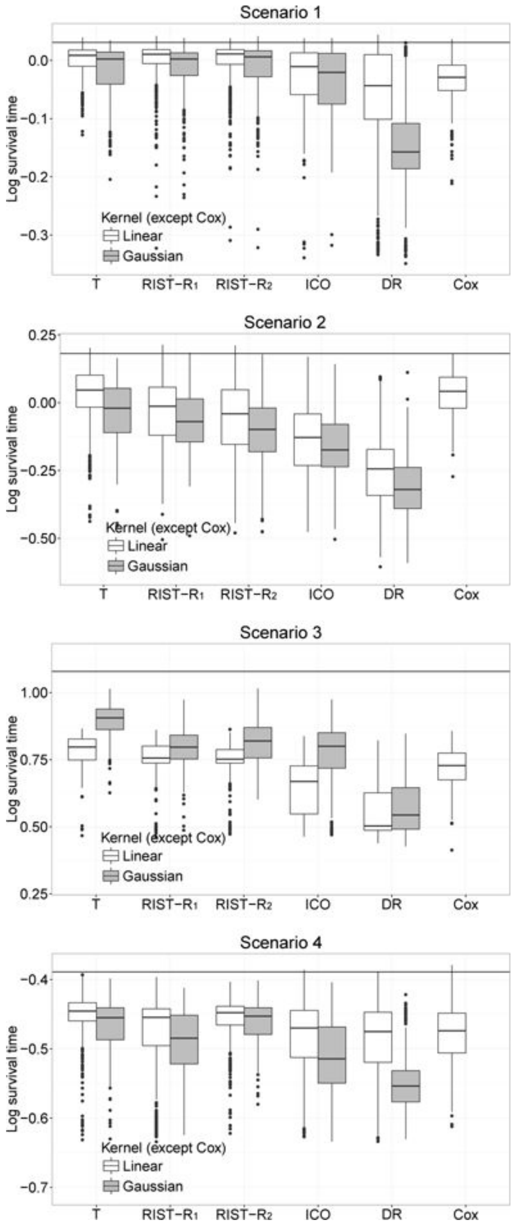

Fig 5.

Boxplots of mean log survival time for different treatment regimes. Censoring rate: 60%. T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning. The black horizontal line is the theoretical optimal value.

Table 1

Simulation results: Mean ($\times 10^3$) and (sd) ($\times 10^3$). Censoring rate: 45%. For each scenario, the theoretical optimal value ($\times 10^3$) is 31, 181, 1079, and −389, respectively.

	kernel	T	RIST-R_1	RIST-R_2	ICO	DR	Cox
1	Linear	0 (26)	0 (31)	1 (30)	−20 (54)	−39 (76)	−29 (33)
	Gaussian	−17 (44)	−11 (35)	−8 (36)	−25 (50)	−88 (79)	
2	Linear	22 (113)	−1 (112)	−24 (125)	−137 (131)	−232 (132)	53 (69)
	Gaussian	−39 (115)	−40 (103)	−72 (114)	−175 (120)	−311 (106)	
3	Linear	785 (52)	766 (59)	763 (51)	683 (113)	598 (120)	745 (64)
	Gaussian	896 (61)	803 (56)	834 (71)	785 (105)	606 (115)	
4	Linear	−453 (37)	−469 (47)	−451 (27)	−469 (48)	−481 (59)	−464 (36)
	Gaussian	−465 (35)	−482 (44)	−457 (28)	−487 (45)	−531 (43)	

T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning; Cox: Cox proportional hazards model using covariate-treatment interactions.

Table 2

Analysis of non-small-cell lung cancer data: Mean (standard deviation) of value function

kernel	RIST-R_1	RIST-R_2	ICO	DR	Cox
Linear	3.641 (0.144)	3.641 (0.138)	3.633 (0.158)	3.590 (0.174)	3.582 (0.158)
Gaussian	3.611 (0.215)	3.615 (0.220)	3.302 (0.221)	3.470 (0.233)	

RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning; Cox: Cox proportional hazards model using covariate-treatment interactions.

Table 3

Analysis of non-small-cell lung cancer data: Mean (standard deviation) of a clinical measure

kernel	RIST-R_1	RIST-R_2	ICO	DR	Cox
Linear	3.603 (0.040)	3.606 (0.037)	3.598 (0.037)	3.601 (0.042)	3.646 (0.039)
Gaussian	3.511 (0.064)	3.514 (0.068)	3.451 (0.062)	3.456 (0.052)	

RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning; Cox: Cox proportional hazards model using covariate-treatment interactions.

Table 4

Simulation results: Mean (×10^3) and (sd) (×10^3). Censoring rate: 30%. For each scenario, the theoretical optimal value (×10^3) is 31, 181, 1079, and −389, respectively.

	kernel	T	RIST-R_1	RIST-R_2	ICO	DR	Cox
1	Linear	0 (26)	1 (31)	2 (28)	−10 (40)	−20 (63)	−26 (33)
	Gaussian	−17 (44)	−10 (34)	−7 (37)	−18 (45)	−48 (65)	
2	Linear	22 (113)	17 (105)	−14 (126)	−110 (136)	−193 (133)	65 (63)
	Gaussian	−39 (115)	−25 (101)	−62 (113)	−164 (119)	−285 (112)	
3	Linear	785 (52)	768 (53)	771 (52)	737 (95)	667 (124)	763 (61)
	Gaussian	896 (61)	810 (54)	854 (69)	817 (124)	679 (123)	
4	Linear	−453 (37)	−465 (46)	−448 (27)	−461 (42)	−471 (54)	−457 (32)
	Gaussian	−465 (35)	−477 (42)	−456 (27)	−474 (41)	−505 (48)	

T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning; Cox: Cox proportional hazards model using covariate-treatment interactions.

Table 5

Simulation results: Mean (×10^3) and (sd) (×10^3). Censoring rate: 60%. For each scenario, the theoretical optimal value (×10^3) is 31, 181, 1079, and −389, respectively.

	kernel	T	RIST-R_1	RIST-R_2	ICO	DR	Cox
1	Linear	0 (26)	−2 (39)	−5 (43)	−29 (57)	−64 (92)	−34 (36)
	Gaussian	−17 (44)	−12 (40)	−12 (45)	−35 (55)	−144 (78)	
2	Linear	22 (113)	−36 (123)	−61 (135)	−138 (133)	−248 (129)	31 (79)
	Gaussian	−39 (115)	−69 (108)	−102 (115)	−165 (117)	−313 (101)	
3	Linear	785 (52)	753 (77)	748 (69)	646 (104)	556 (94)	721 (70)
	Gaussian	896 (61)	796 (63)	819 (67)	775 (106)	573 (93)	
4	Linear	−453 (37)	−478 (55)	−458 (33)	−486 (55)	−492 (59)	−480 (43)
	Gaussian	−465 (35)	−492 (48)	−461 (29)	−513 (53)	−551 (38)	

T: using true survival time as weight; RIST-R_1 and RIST-R_2: using the estimated R_1 and R_2 respectively as weights, while the conditional expectations are estimated using recursively imputed survival trees; ICO: inverse probability of censoring weighted learning; DR: doubly robust outcome weighted learning; Cox: Cox proportional hazards model using covariate-treatment interactions.