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

Abstract

Zhao et al. (2012) and Zhang et al. (2012) proposed outcome weighted learning to estimate individualized treatment rules directly through maximizing the expected clinical response. In this paper, we extend 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 Zhao et al. (2015). To accomplish this, we take advantage of the tree based approach proposed in Zhu & Kosorok (2012) 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 the second method imputes the expected failure time conditional on the observed censoring time. 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.

Key words — Individualized treatment rule; Nonparametric estimation; Right censored data; Consistency.

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 (Murphy, 2003; Qian & Murphy, 2011; Zhang et al., 2012; Zhao et al., 2012). While each of these approaches has strengths and weaknesses, we highlight the approach in Zhao et al. (2012) because of its robustness to model misspecification. This approach is commonly referred to as outcome weighted learning.

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 (Zhu & Kosorok, 2012). Since the true failure times \( T \) are only partially known, they cannot be used directly as weights in the outcome weighted learning (Zhao et al., 2012) framework. However, recursively imputed survival trees (Zhu & Kosorok, 2012) provide an alternative approach to weighting by using the conditional expectations of censored observations without requiring inverse weighting.

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 Ishwaran & Kogalur (2010).

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. 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 \( X \in \mathcal{X} \) be the observed patient-level covariate vector, where \( \mathcal{X} \) 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 \( \tau \), i.e., \( T = \min(\tilde{T}, \tau) \). 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 \( D \), which is a map from the patient-
level covariate space $X$ to the treatment space $\{+1, -1\}$, for each individual which maximizes the expected reward. In the survival outcome setting, we will use $R = T$ or $\log(T)$ as done in Zhao et al. (2015).

To achieve this maximization, we define the value function as

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

where $I\{\cdot\}$ is an indicator function, $\pi(a; X) = \Pr(A = a \mid X) > M'$ a.s. for some $M' > 0$ and each $a \in \{+1, -1\}$. The function $\pi$ 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 $D^*$ which maximizes the value function. The definition of $D^*$ is equivalent to $D^*(x) = \arg \max_a E(R \mid A = a, X = x)$ (Zhao et al., 2012). Instead of maximizing the value function, the outcome weighted learning approach searches for the optimal rule $D^*$ by minimizing the weighted misclassification error, i.e.,

$$D^* = \arg \min_D E[R I\{A \neq D(X)\}/\pi(A; X)].$$

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 $\delta = I(T \leq C)$. 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 $\hat{E}(T \mid X, A)$. Letting $R_1 = E(T \mid X, A)$ and bringing the expectation of $T$ inside,
we have

\[ E[T I\{A = D(X)\}/\pi(A; X)] = E[R_1 I\{A = D(X)\}/\pi(A; X)]. \]  

(2)

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 \( \delta \), can be written as

\[ R_2 := E(T|X,A,Y,\delta) = I(\delta = 1)Y + I(\delta = 0)E(T|X,A,T > Y,Y). \]  

(3)

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 \( T \) accordingly. And it is easy to see that the value function defined using \( R_2 \) as the reward is equivalent to the left side of equation (2) by further taking expectations with respect to \( Y \) and \( \delta \). 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.

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:

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

(4)

and

\[ n^{-1} \sum_{i=1}^{n} \frac{\{\delta_i Y_i + (1 - \delta_i)\hat{E}(T_i | X_i, A_i, T_i > Y_i, Y_i)\}I\{A_i = D(X_i)\}}{\pi(A_i; X_i)}. \]  

(5)

### 2.3 Outcome weighted learning with survival trees

The recursively imputed survival trees method proposed by Zhu & Kosorok (2012) 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 \( \hat{E}(T \mid X,A) \) is obtained by averaging the trees from the last step.

Following (Zhao et al., 2012), we next use support vector machines to solve for the optimal treatment rule. A decision function \( f(x) \) is learned by replacing \( I\{A_i = D(X_i)\} \) in Equations (4) or (5) with \( \phi\{A_i f(X_i)\} \), 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 \( \lambda_n \) 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 (6). Both settings can be efficiently solved by quadratic programming. For further details regarding solving weighted classification problems using support vector machines, we refer to (Zhao et al., 2012, 2015; Chang & Lin, 2011).

**Algorithm 1**: Pseudo algorithm for the proposed method

1. Use \( \{(X_i^T, A_i, A_i X_i^T)\}^{n}_{i=1} \) to fit recursively imputed survival trees. Obtain the estimation \( \hat{E}(T_i \mid A_i, X_i) \) if reward \( R_1 \) is used or \( \hat{E}(T_i \mid X_i, A_i, T_i > Y_i, Y_i) \) if \( R_2 \) is used.

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

   \[
   \hat{f}(x) = \arg \min_f \sum_{i=1}^{n} W_i \frac{\phi\{A_i f(X_i)\}}{\pi(A_i; X_i)} + \lambda_n \|f\|^2. \tag{6}
   \]

3. Output the estimated optimal treatment rule \( \hat{D}(x) = \text{sign}\{\hat{f}(x)\} \).
3 Theoretical results

3.1 Preliminaries

The risk function is defined as $R(f) = E\left[\frac{R}{\pi(A; X)} I\{A \neq \text{sign}(f(X))\}\right]$, where the reward $R = R_1$ for the first approach, or $R = R_2$ 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, respectively, $L_\phi(f) = R\phi\{Af(X)\}/\pi(A; X)$ and $L'_\phi(f) = \hat{R}\phi\{Af(X)\}/\pi(A; X)$, respectively.

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

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

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

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 Zhu & Kosorok (2012). To the best of our knowledge, this is the first consistency result for a tree-based survival model, even under somewhat restricted conditions. Moreover, the implication of this result is interesting in its own right.

For simplicity, in this section, we assume that the training sample is $Q_n = \{(Y_i, \delta_i, X_i, A_i)\}$, 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 fixed $x$, our goal is to estimate the cumulative hazard function of failure time $r(\cdot, x, a) = \Lambda_T(\cdot \mid X = x, A = a)$, hereinafter, we write it as $\Lambda(\cdot \mid X = x, A = a)$.

A random forest is a collection of randomized base regression trees $\{r_n(\cdot, x, a, \Theta_m, Q_n)\}$, $m \geq 1$. The randomizing variable $\Theta$ is used to indicate how the successive cuts are performed when an individual tree is built. Here, we consider a simplified scenario in which the selection of the coordinate
is completely random and independent from the training data (Biau, 2012). We denote our aggregated estimator as \( \hat{r}_n(\cdot, X, A) = E_\Theta[r_n(\cdot, X, A, \Theta)] \).

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 \( \tau \), 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 \( \Lambda(t | X, A) \) is \( L \)-Lipschitz continuous in terms of \( X \), and the hazard function \( \lambda(t | X, A) \) is \( L' \)-Lipschitz continuous in terms of \( X \) with the Euclidean norm.

The following theorem proves consistency of the proposed tree based survival model. 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 hold. Further assume that \( k_n \to \infty \), \( n/k_n \to \infty \) and \( n/k_n \to \infty \) as \( n \to \infty \), where \( k_n \) is a tuning parameter denoting the number of terminal nodes. Then the estimator of the survival tree model is consistent. Moreover, for \( t < \tau \) and any \( b > 288 \), with probability larger than \( 1 - w_n \) we have

\[
\sup_{t < \tau} |\hat{r}_n(t, X, A) - r(t, X, A)| \leq C[d^{1/2}2^{-(1-r)[\log_2 k_n]}/d + b^{1/2}((1-u)n2^{-(\log_2 k_n)-1})^{-1/2}],
\]

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

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

### 3.3 Consistency and Excess Value Bound

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 1 Based on Theorem 1, the estimated conditional expectations converge in probability, i.e., with probability larger than $1 - 2w_n$,

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

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

for some constant $C_1$, $C_2$ (depending on $L, L', \tau, M, d$).

We will use the above lemma 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 $V'(f) = E(\tilde{R}I[A = \text{sign}(f(X))]/\pi(A; X))$, 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 $\epsilon$-ball based on this norm. By Theorem 2.1 in Steinwart & Scovel (2007), we restate the bound for covering numbers:

Lemma 2 (Theorem 2.1 in Steinwart & Scovel (2007)) For any $\beta > 0$, $0 < v < 2$, $\epsilon > 0$ we have

$$\sup_{P_n}\log N(B_{H_k}, \epsilon, L_2(P_n)) \leq c_{v, \beta, d}\sigma_n^{(1-v/2)(1+\beta)d} \epsilon^{-v},$$

where $B_{H_k}$ is the closed unit ball of $H_k$, and $d$ is the dimension of $\mathcal{X}$.

Lastly, for $\tilde{f} = \arg\min_{f \in \mathcal{X}} E\{L_\phi(f)\}$, we define the approximation error function $a(\lambda) = \inf_{f \in H_k}[E\{L_\phi(f)\} + \lambda\|f\|^2] - E\{L_\phi(\tilde{f})\}$. Then we have following theorem, the proof of which is given in Appendix.

Theorem 2 Based on Theorem 1 and assume that the sequence $\lambda_n > 0$ satisfies $\lambda_n \to 0$, $\lambda_n(\log_2 k_n)^2 \to \infty$ and $\lambda_n(n/k_n) \to \infty$. Then we have $\text{pr}(V(f^*) \leq V(\hat{f}_n) + \epsilon) \geq 1 - 2w_n - 2e^{-\rho}$, where $f^*$ maximize the true value function $V$, $\rho > 0$ and $c_n = c_{v, \beta, d}\sigma_n^{(1-v/2)(1+\beta)d}$, $\epsilon = a(\lambda_n) + M_v(n\lambda_n/c_n)^{-2/(v+2)} + M_v\lambda_n^{-1/2}(c_n/n)^{2/(d+2)} + K\rho(n\lambda_n)^{-1} + 2Kpn^{-1}\lambda_n^{-1/2} + C\lambda_n^{-1/2}[2^{-((1-r)[\log_2 k_n]})]/d + (b/((1 - u)n2^{-[\log_2 k_n]-1}))^{1/2}],$ 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 $H_k$. The second part $O_P\{k_n^{-1/d} + (n/k_n)^{-1}\} \lambda_n^{-1/2}$ controls the approximation error due to using the proposed tree-based method to estimate the conditional expectation.
The behavior of the approximation error function $a(\lambda)$ can be characterized in terms of the size of the points which are close to boundary (Zhao et al., 2015; Tsybakov, 2004; Steinwart & Scovel, 2007). There exist constant $c_1$ such that $a(\lambda_n) \leq c_1 \lambda_n^{q/(q+1)}$ when Gaussian kernel has parameter $\sigma_n = \lambda_n^{-1/[(q+1)d]}$ and $c_n = \lambda_n^{-(1-v/2)(1+\beta)/(1+q)}$, where $q$ is the noise exponent. A larger $q$ indicates that the two groups benefiting from the two different treatments are well separated in the data.

As Zhao et al. (2015) point out, an optimal choice of $\lambda_n$ balances the bias and variance. When $k_n = n^{(1+d)^{-1}}$ and $\lambda_n = \max(n^{-2(1+q)/[(4+v)q+2+(2-v)(1+\beta)]}, n^{-2(q+1)(d+1)^{-1}(2q+1)^{-1}})$, the optimal balance happens in our setting, so that the optimal rate we could achieve is $O_p(n^{-\min(2q/[4+v]q+2+(2-v)(1+\beta)], -2q(d+1)^{-1}(2q+1)^{-1})}$. When the data is well separated, the geometric noise exponent $q$ can be arbitrarily large. In this case, the optimal rate is close to $n^{-(d+1)^{-1}}$.

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 (Zhao et al., 2015). 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.

We mostly use the default values for tuning parameters. The number of variables considered at each split is the integer part of $\sqrt{d}$ as suggested by Ishwaran et al. (2008) and Geurts et al. (2006). We set
the total number of trees to be 50 as suggested by Zhu & Kosorok (2012) 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 (Chang & Lin, 2011). 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 \(\lambda_n\) 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, \(\epsilon\) is generated from a standard normal distribution. For all Cox proportional hazards models, the baseline hazard function \(\lambda_0(t) = 2t\).

**Scenario 1.** Both \(\tilde{T}\) and \(C\) are generated from the accelerated failure time model. \(\tau = 2.5, d = 10\) and the censoring rate is about 30%. The optimal decision function is linear. The value of the optimal treatment rule is approximately 0.031.

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

\[
\log(C) = 0.5 - 0.8X_1 + 0.4X_2 + 0.4X_3 + (0.5 - 0.1X_1 - 0.6X_2 + 0.3X_3)A + \epsilon.
\]

**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. \(\tau = 8, d = 10\) and the censoring rate is about 60%. The value of the optimal treatment rule is approximately 1.070.

\[
\lambda_{\tilde{T}}(t \mid A, X) = \lambda_0(t) \exp\{-1 - 1.5X_1^{0.5} + 0.5X_2 + (0.8 - 0.7X_1^{0.5} - 1.2X_2^3)A\},
\]

\[
\log(C) = -0.2 + 0.6X_1 - 0.1X_2^2 + 0.3X_3 + (0.2 + X_1^{0.5} - 2X_2 + 0.5X_3)A + \epsilon.
\]

**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. \(\tau = 8, d = 5\) and the censoring rate is about 40%. The value of the optimal treatment rule is approximately 1.079.

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

\[
I(X_4 < 0.3)I(X_5 < 0.3) + I(X_4 > 0.7)I(X_5 >
\]
\[ \lambda_C(t \mid A, X) = \lambda_0(t) \exp\{-1.5 + X_1 + (1 + 0.6X_2^1)A\}. \]

Scenario 4. \( \tilde{T} \) is generated from an accelerated failure time model. \( C \) is generated from a Cox model. \( \tau = 2, \ d = 10 \) and the censoring rate is about 45%. The value of the optimal treatment rule is approximately \(-0.389\).

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

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

### 4.2 Simulation results

Figure 1 shows the boxplot of values based on the logarithm of \( T \) calculated from the test data. 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.

![Boxplots of mean log survival time for different treatment regimes.](image)

Figure 1: Boxplots of mean log survival time for different treatment regimes. 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.
5 Data Analysis

We apply the proposed method to a non-small-cell lung cancer randomized trial dataset described in Socinski et al. (2002). 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 within range from 90% to 100% and 109 subjects within 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 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 resulting empirical values. This procedure is then repeated 100 times. To calculate the empirical value on the testing data, we use
\[
\sum_{i=1}^{n} R_i I\{A_i = D(X_i)\}/\sum_{i=1}^{n} I\{A_i = D(X_i)\}.
\]
We also compare with the procedures proposed in Zhao et al. (2015), where $R$ is defined as
\[
\frac{\Delta \log(Y)}{\hat{S}_C(Y \mid A, X)} - \int \hat{E}_T\{\log(T) \mid T > t, A, X\} \left\{ \frac{dN_C(t)}{\hat{S}_C(t \mid A, X)} + I(Y_i \geq t) \frac{d\hat{S}_C(t \mid A, X)}{\hat{S}_C(t \mid A, X)^2} \right\},
\]
where $\hat{S}_C(t \mid A, X)$ and $\hat{E}_T(T \mid T > t, A, X)$ are estimated from the Cox model. The boxplot is shown in 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.
Figure 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.

References


