Transformation-Invariant Learning of Optimal Individualized Decision Rules with Time-to-Event Outcomes

Yu Zhou\textsuperscript{a}, Lan Wang\textsuperscript{b}, Rui Song\textsuperscript{c}, and Tuoyi Zhao\textsuperscript{b}

\textsuperscript{a}Roku, San Jose, CA; \textsuperscript{b}Department of Management Science, University of Miami, Coral Gables, FL; \textsuperscript{c}Department of Statistics, North Carolina State University, Raleigh, NC

ABSTRACT

In many important applications of precision medicine, the outcome of interest is time to an event (e.g., death, relapse of disease) and the primary goal is to identify the optimal individualized decision rule (IDR) to prolong survival time. Existing work in this area have been mostly focused on estimating the optimal IDR to maximize the restricted mean survival time in the population. We propose a new robust framework for estimating an optimal static or dynamic IDR with time-to-event outcomes based on an easy-to-interpret quantile criterion. The new method does not need to specify an outcome regression model and is robust for heavy-tailed distribution. The estimation problem corresponds to a nonregular \(M\)-estimation problem with both finite and infinite-dimensional nuisance parameters. Employing advanced empirical process techniques, we establish the statistical theory of the estimated parameter indexing the optimal IDR. Furthermore, we prove a novel result that the proposed approach can consistently estimate the optimal value function under mild conditions even when the optimal IDR is nonunique, which happens in the challenging setting of exceptional laws. We also propose a smoothed resampling procedure for inference. The proposed methods are implemented in the R-package \texttt{QTOCen}. We demonstrate the performance of the proposed new methods via extensive Monte Carlo studies and a real data application. Supplementary materials for this article are available online.

1. Introduction

The problem of estimating the optimal individualized decision rule (IDR) has recently received substantial attention in precision medicine and other domains. A treatment can be a drug, a therapy, or any other actionable choice (or a sequence of such choices) such as a policy or program. The goal of optimal IDR estimation is to determine a decision rule that assigns a subject to one of the treatment options based on individual information available at each decision point such that some functional of the potential outcome distribution is optimized.

For completely observed data, several successful approaches exist for estimating the optimal IDR, including Q-learning (Watkins and Dayan 1992; Murphy 2005b; Chakraborty, Murphy, and Streicher 2010; Song et al. 2015), A-learning (Robins, Hernan, and Brumback 2000; Murphy 2003, 2005a; Moodie and Richardson 2010), model-free or policy search methods (Robins, Orellana, and Rotnitzky 2010; Orellana, Rotnitzky, and Robins 2010; Zhu et al. 2012, 2015b), the interpretation-enhanced tree or list-based methods (Laber and Zhao 2015; Cui, Zhu, and Kosorok 2017; Zhu et al. 2017; Zhang et al. 2018) among others. See also the books of Chakraborty and Moodie (2013) and Kosorok and Moodie (2016) for a general introduction and additional references.

The focus of this article is on estimating the optimal static (one-stage) or dynamic (multi-stage) IDR for time-to-event or survival data, where the outcome is possibly censored. Several new challenges arise when analyzing such data compared with the complete data case. Censoring occurs when an individual drops out from the study or the study ends before the subject experiences the event of interest. The distribution of survival time (e.g., time to death, onset of disease) is often highly skewed. The situation gets even more complicated for estimating the optimal dynamic IDR, where the data are collected longitudinally. Consider the setting where the treatment decisions for a patient are made at \(k\) prespecified decision points. Each decision is allowed to depend on the patient’s characteristics (e.g., gender, age) and treatment history (e.g., disease progression status and how the individual responds to previous treatments) up to that decision point. The patient may be censored at any stage of treatment. Direct application of existing complete-data techniques could result in severe bias, as demonstrated in the Monte Carlo studies in Section 5.

Several authors have recently investigated estimating the optimal IDR with survival data; see Goldberg and Kosorok (2012), Xu et al. (2016), Jiang et al. (2017a), Jiang et al. (2017b), Bai et al. (2017), Hager, Tsiatis, and Davidian (2018), Xu et al. (2016), Diaz, Savenkov, and Ballman (2018), Simonoe
et al. (2020), among others. For time-to-event outcomes, new
criterion is needed to evaluate the effectiveness of an IDR.
Of crucial importance is that such a criterion can be reliably
estimated under censoring. The existing work have been mostly
focused on maximizing the restricted mean survival time in the
population.

We consider time-to-event outcomes and propose a new
robust framework for estimating the optimal IDR using an alter-
native criterion based on the marginal quantile of the poten-
tial outcome distribution. The new optimality criterion is easy
to interpret. Median survival time has already been popularly
used clinically to evaluate the success of cancer treatment. It
can be reliably estimated even under relatively heavy censoring
(e.g., the censoring rate is more than 50% in the real data
example of this article). The resulted optimal IDR is invariant
under monotone-transformation of the outcome. We develop
robust estimation methods for both static and dynamic opti-
mal IDRs. The robust approach circumvents the difficulty of
specifying a reliable outcome regression model, especially for
the dynamic setting which demands a sequence of generative
regression models.

We consider estimating the optimal IDR in a class of candi-
date decision rules indexed by a finite-dimensional parameter.
The estimation problem corresponds to a challenging non-
regular M-estimation problem with both finite and infinite-
dimensional nuisance parameters, due to the unknown censor-
ing distribution. For the optimal static IDR, we rigorously estab-
lish the cube-root convergence rate for the estimated parameter
by employing modern empirical processes techniques. We prove
that its asymptotic distribution corresponds to the maximizer of
a centered Gaussian process with a parabolic drift. It is worth
emphasizing that the nonstandard asymptotics is due to the
intrinsic nature of the decision problem, which relates to a
sharp edge effect in the decision function. Due to the nature of
nonstandard asymptotics, the problem is substantially harder
than regular M-estimation problem with infinite-dimensional
nuisance parameters.

Moreover, we establish a useful novel result that shows the
optimal value can be consistently estimated under weak con-
ditions for the challenging setting of exceptional laws. Under
exceptional laws, there exists a subgroup of patients for whom
the treatment is neither beneficial nor harmful. In this case, the
optimal IDR is nonunique, see, for example, the discussions in
Robins and Rotnitzky (2014) and Luedtke and van der Laan
(2016).

Theoretically, our work complements existing results and
significantly enhances the knowledge about optimal IDR esti-
mation with survival data. The existing results have been mostly
focused on prediction error bounds and have not studied the
properties of the estimated parameter indexing the optimal IDR.
Furthermore, to the best of our knowledge, existing work on
optimal IDR estimation with survival outcomes assume nonex-
ceptional laws and thus avoid the problem of optimal value
estimation under exceptional laws.

The rest of the article is organized as follows. Section 2
introduces the new framework and the robust method for esti-
mating an optimal static IDR with survival data, with theo-
retical properties developed in Section 3. Section 4 presents
the estimation method and the theory for the dynamic IDR
setting. In Section 5, we report results from extensive Monte
Carlo studies. In Section 6, we illustrate the application on the
analysis of a breast cancer dataset. The proposed methods are
implemented in the R-package qtoce. The regularity con-
ditions are given in the Appendix. The technical derivations
and additional numerical results are given in the supplementary
materials.

2. Robust Estimation for Static Optimal IDR

2.1. Preliminaries

We first consider the single-stage setting. Let $A \in \{0, 1\}$ denote
the binary treatment, $X \in \mathbb{R}^p$ denote the vector of covariates
with support $\mathcal{X}$, and $T \in \mathbb{R}^+$ denote the time to the event
of interest (or a transformation thereof). We often refer to $T$
as survival time. Without loss of generality, we assume that
a larger value of $T$ indicates better treatment effect. The outcome
$T$ may not be observed due to censoring. Let $C \in \mathbb{R}^+$ denote
the censoring variable and $\Delta = I(T \leq C)$. If the observation is
censored (i.e., $\Delta = 0$), then we only observe $C$. Let $Y = \min(T, C)$
be the observed outcome. The observed data consist of
$(X_i, A_i, Y_i, \Delta_i), i = 1, \ldots, n$, which are independent copies
of $(X, A, Y, \Delta)$.

To assess the treatment effect, we adopt the potential outcome
framework (Neyman 1923; Rubin 1978) in causal inference. The
ith subject has two potential outcomes: $T^*_i(0)$ and $T^*_i(1)$, where
$T^*_i(0)$ is the survival time had the subject received treatment 0
and $T^*_i(1)$ is defined similarly, $i = 1, \ldots, n$. In practice, a subject
receives one and only one of the two possible treatments. Under
the stable unit treatment value assumption (Rubin 1986), we have
$T_i = A_i T^*_i(1) + (1 - A_i) T^*_i(0)$. That is, the survival time of
the ith subject is the potential survival time corresponding to
the treatment that subject actually received. Furthermore, we
assume the no unmeasured confounders assumption (Rosen-
baum and Rubin 1983) is satisfied, that is, $(T^*_i(0), T^*_i(1))$
are independent of $A_i$ conditional on $X_i$. This is a common assump-
tion in causal inference and is automatically satisfied for a ran-
donized trial. Mathematically, an IDR $d(X)$ is a mapping from
the space of covariates $\mathcal{X}$ to the space of candidate treatments
$\{0, 1\}$. The potential outcome associated with $d(X)$ is denoted by
$T^*(d(X))$. We have $T^*(d(X)) = T^*_i(1)d(X) + T^*_i(0)[1 - d(X)]$.

2.2. New Optimality Criterion for Evaluating IDR with
Time-to-Event Outcome

Different from the completely observed data case, the mean of
the outcome is usually difficult to estimate accurately at the
presence of censoring. The most popular criterion in the existing
literature for time-to-event data is the restricted mean survival
time $E[\min(T^*(d(X)), L)]$, where $L$ is a user-supplied cutoff
time. In this article, we consider an alternative criterion for com-
paring IDRs for time-to-event outcomes based on the marginal
quantile treatment effect. The new criterion enjoys three appeal-
ing properties: easy to interpret, robust to long-tailed survival
distribution, and invariant to the monotone transformation of
survival time. The marginal $r$th quantile ($0 < r < 1$) of the
potential outcome $T^*(d(X))$ is defined as
function estimator for indexing the optimal IDR is also theoretically substantially harder with the presence of an infinite-dimensional nuisance parameter due to censoring. Wahed (2009) studied estimating the survival quantiles in two-stage randomization designs with fixed IDRs but had not investigated the more challenging problem of optimal IDR estimation.

In practice, $D$ is usually chosen to be a class of IDRs indexed by a Euclidean parameter for interpretability. Same as Zhang et al. (2012) and others, we focus on the class of index rules $D = \{d_\beta(X) = 1(\beta^T X > 0) : |\beta_1| = 1, \tilde{\beta} \in \mathbb{B} \}$, where $\beta = (\beta_1, \ldots, \beta_p)^T = (\beta_1, \tilde{\beta})^T$, $\mathbb{B}$ is a compact subset of $\mathbb{R}^{p-1}$ and $1(\cdot)$ denotes the indicator function. For identifiability, we assume there exists a continuous covariate whose coefficient has absolute value one. Without loss of generality, we assume $|\beta_1| = 1$. The population parameter $\beta_0 = (\beta_{01}, \ldots, \beta_{0p})^T$ indexing the optimal IDR is
\[
\beta_0 = \arg \max_{\beta \in \mathbb{B}^p} Q_\tau \{ T^\tau (d(X)) \},
\]
where $\mathbb{B}^p = \{ \beta \in \mathbb{R}^p : |\beta_1| = 1, \tilde{\beta} \in \mathbb{B} \}$.

### 2.3 A Robust Estimation Procedure

Based on the observations $\{X_i, A_i, Y_i, \Delta_i\}, i = 1, \ldots, n$, our goal is to estimate the population parameter $\beta_0$ indexing the optimal IDR. It is known that a misspecified generative regression model can result in severe bias in estimating the optimal treatment (Qian and Murphy 2011; Zhang et al. 2012; Zhao et al. 2012, 2015a). We introduce a robust estimator that accounts for censoring while at the same time circumvents the difficulty of specifying a reliable generative regression model.

Given an IDR $d_\beta(X)$, the treatment it would recommend to subject $i$ may or may not coincide with the treatment the subject actually received. Even if $A_i = d_\beta(X_i)$, we may not observe $T^\tau_i [d_\beta(X_i)]$ if the subject is censored. To obtain a consistent estimator for $Q_\tau \{ T^\tau_i [d_\beta(X_i)] \}$, we adapt the induced missing data framework in Zhang et al. (2012) to time-to-event data. Specifically, we consider an artificial missing data structure with the missing data indicator $R_i(\beta) = \left[ A_i d_\beta(X_i) + (1 - A_i) \right] / \Delta_i$. The observed outcome $Y_i$ is equal to the potential outcome $T^\tau_i [d_\beta(X_i)]$ only if $R_i(\beta) = 1$. In this framework, the “full data” that we may not completely observe consist of $\{ X_i, T^\tau_i [d_\beta(X_i)] \} \Delta_i = \{ X_i, R_i(\beta), (R_i(\beta) T^\tau_i [d_\beta(X_i)] \} = \{ X_i, R_i(\beta), (R_i(\beta) Y_i) \} n_i$.

Let $\pi_A(X_i) = P(A_i = 1 | X_i)$ be the propensity score; and let $G_C(t | X,A) = P(C > t | X,A)$ denote the conditional survival function of $C$ given $\{ X,A \}$. Let $\pi(\beta) = P[R_i(\beta) = 1 | X_i, T^\tau_i (1), T^\tau_i (0)]$ be the probability of missingness conditional on the full data. We observe
\[
\pi_i(\beta) = \pi_A(X_i) \beta d_\beta(X_i) \left[ \Delta_i = 1 | X_i, T^\tau_i (1), T^\tau_i (0), A_i = 1 \right] + (1 - \pi_A(X_i))(1 - \beta d_\beta(X_i)) \left[ \Delta_i = 0 | X_i, T^\tau_i (1), T^\tau_i (0), A_i = 0 \right] = \left\{ \pi_A(X_i) \beta d_\beta(X_i) + (1 - \pi_A(X_i))(1 - \beta d_\beta(X_i)) \right\} G_C \left( T^\tau_i [d_\beta(X_i)] \right),
\]
where $\pi(\beta)$ is the conditional survival function of $C_i$ given $\{ X_i, A_i \}$. Let $\pi_i(\beta) = \pi(\beta) G_C \left( T^\tau_i [d_\beta(X_i)] \right)$. Note that for the complete cases (corresponding to $R_i(\beta) = 1$), we have $Y_i = T^\tau_i [d_\beta(X_i)]$ and the corresponding conditional survival function $\pi_i(\beta) = \pi(\beta) G_C \left( T^\tau_i [d_\beta(X_i)] \right)$.

To estimate $\beta_0$, we propose the following two-step estimator. First, we estimate $Q_\tau \{ T^\tau (d_\beta(X)) \}$ by the following inverse probability weighted estimator
\[
\hat{Q}_\tau \{ T^\tau (d_\beta(X)) \} = \arg \min_{\beta \in \mathbb{B}^p} \sum_{i=1}^n R_i(\beta) \rho_i(Y_i - b),
\]
where $\hat{\pi}_i$ is an estimate of $\pi_i(\beta)$ and $\rho_i(u) = u(\tau - I(u < 0))$ is the quantile loss function (Koenker 2005). By convention, we define $0/0 = 0$. Next, employing the policy-search idea, we estimate $\beta_0$ by
\[
\hat{\beta}_n = \arg \max_{\beta \in \mathbb{B}^p} \hat{Q}_\tau \{ T^\tau (d_\beta(X)) \}.
\]

The above estimator can be computed using the genetic algorithm in R package xgenoud (Mebane Jr and Sekhon 2011). The estimate of the optimal IDR is $d^\tau_\beta(X) = I(\hat{\beta}_n^T X > 0)$.

**Remark 1.** A key quantity in estimating $\pi_i(\beta)$ is the conditional survival function of the censoring variable $G_C(t | X,A) = P(C > t | X,A)$. There are several approaches for estimating $G_C(t | X,A)$. For clarity of presentation, as in Goldberg and Kosorok (2012) and Jiang et al. (2017a), we assume that $C \perp \{ T^\tau_i (0), T^\tau_i (1), A_i \}$ in the theoretical development. This is often satisfied in real applications where administrative censoring occurs. In this case, $G_C(t | X,A)$ can be estimated by $G_C(\cdot)$, the classical Kaplan–Meier estimator applied to $\{ Y_i, 1 - \Delta_i \}, i = 1, 2, \ldots, n$. When necessary, we can relax the independent censoring assumption to the conditionally independent censoring assumption $C \perp \{ T^\tau_i (0), T^\tau_i (1) \} | \{ X,A \}$ and employs the local Kaplan–Meier estimator. Without loss of generality, we assume that the first $n_1$ subjects receive treatment...
A = 0, and the other \((n - n_1)\) subjects receive treatment \(A = 1\). The local Kaplan–Meier estimator Gonzalez-Manteiga and Cadarso-Suarez (1994) for \(G_C(\cdot | X, A = 0)\) is given by

\[ \hat{G}_C(\cdot | X, A = 0) = \prod_{j=1}^{n_1} \left(1 - \frac{B_{n1}(X)}{\sum_{k=1}^{n_1} I(C_j \geq C_k)B_{n1}(X)}\right)^{\eta_j(t)}, \]

where \(\eta_j(t) = I(C_j \leq t, \Delta_j = 0)\), and \(\{B_{n1}(X), k = 1, \ldots, n_1\}\) is a sequence of nonnegative weights adding up to 1. A popular choice is the Nadaraya–Watson’s type weights for univariate covariate: \(B_{n1}(X) = \left[\sum_{i=1}^{n_1} K(h_{-1}(X-X_i))\right]^{-1}K(h_{-1}(X-X_0))\), where \(K(\cdot)\) is a positive kernel function and \(h_n\) is a sequence of bandwidths converging to zero as \(n \to \infty\). We can obtain \(\hat{G}_C(\cdot | X, A = 1)\) similarly. A third approach is to estimate the conditional survival function using a working model, such as the Cox proportional hazards regression model, as investigated in Zhao et al. (2015a).

**Remark 2.** The Kaplan–Meier estimator \(\hat{G}_C\) in (2) is sometimes unstable at the tail of the distribution. Practically, a simple approach to improve the stability is by employing an artificial censoring technique in Zhou (2006), based on the intuition that any alteration of a random variable’s distribution beyond the quantile of interest would have no impact on the quantile. Specifically, assume there exists a large positive constant \(M \in \mathbb{R}\) such that \(\sup_{\beta} Q_t \left(T^*(d_\beta)\right) < M\) and \(\sup_{t: G_C(t) > 0} > M\). The first requirement means the largest achievable \(t\) th quantile using IDRs belonging to \(D\) is smaller than \(M\); and the second one ensures every data point has a positive probability of not being censored. Note that these conditions are weak, especially if we are interested in lower quantiles. Let \(Y^M = Y \wedge M\) and \(\Delta^M = \Delta + (1 - \Delta)I(Y \geq M)\). Then it is straightforward to show that \(\hat{\beta}_n^M\) obtained using the transformed dataset \((X_i, A_i, Y^M_i, \Delta^M_i)\), \(i = 1, \ldots, n\), is the same as \(\hat{\beta}_n\) in (3).

**Remark 3.** As the optimization problem is nonconcave and nonsmooth, multiple local optimal may exist. Popular algorithms based on derivatives do not work for this challenging setting. At the same time, it is impractical to exhaustively enumerate all possible solutions and pick the best one. In our numerical experiments, we use the genetic algorithm in the R package  `rgenoud`, which is useful in such a challenging setting when the objective function is nonconcave and the derivatives do not exist. The genetic algorithm (a type of evolutionary algorithm) is inspired from the biological evolution process. In a genetic algorithm, the problem is encoded in a series of bit strings that are manipulated by the algorithm. It is a stochastic, population-based algorithm that searches randomly by mutation and crossover among population members. It is based on searching for the best solutions using inheritance and strengthening of useful features of multiple objects of a specific application in the process of imitation of their evolution. We refer to Mitchell (1998) and Mebane Jr and Sekhon (2011) for more detailed description of the algorithm and other references. In our numerical experience, the algorithm provides high-quality solutions with desirable statistical properties.

### 3. Asymptotic Theory

In this section, we present two results regarding the asymptotic properties of the estimated optimal IDR with survival data.

- First, we show that the estimated parameter indexing the optimal IDR has nonstandard asymptotics, which is characterized by the cube-root convergence rate and the nonnormal limiting distribution.
- Second, we show that under rather weak conditions, which do not require the optimal IDR to be unique at the population or sample level, the theoretically optimal value can be estimated at a near \(n^{-1/2}\)-rate.

Both results are novel for optimal IDR estimation with time-to-event outcomes. The first result corresponds to a nonstandard estimation problem with both finite-dimensional and infinite-dimensional nuisance parameters. The second result deals with the challenging setting of exceptional law where the optimal IDR is nonunique.

#### 3.1. Asymptotic Distribution of the Estimated Parameter Indexing the Optimal IDR

Write \(X = (X_1, \ldots, X_p)^T \equiv (X_1, \tilde{X}^T)^T\). Let \(G_C(\cdot)\) denote the survival function of \(C\). Let \(F_{T^*(0)}(t|X)\) and \(F_{T^*(1)}(t|X)\) be the cumulative distribution functions of the potential survival times \(T^*(0)\) and \(T^*(1)\), respectively; and let \(f_0(t|X)\) and \(f_1(t|X)\) be the corresponding conditional density functions. Given any \(\beta \in \mathbb{R}^p\), let \(f_{T^*(d_\beta)}(\cdot)\) denote the marginal density function of the distribution of the potential survival time \(T^*(d_\beta(X))\).

To avoid complications irrelevant to the main results of the article, we consider data collected from a randomized study where \(\pi_A(X_i) = 0.5\), but the results can be extended to observational data under mild assumptions. The estimator \(Q_t(T^*(d_\beta))\) in (2) simplifies to

\[ \hat{Q}_t(\hat{\beta}; \hat{G}_C) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \frac{R_i(\beta) \rho_i(Y_i - b)}{0.5 \hat{G}_C(Y_i)}, \]

where \(\hat{G}_C(\cdot)\) is the classical Kaplan–Meier estimator of \(G_C(\cdot)\). We then estimate \(\beta_0\) by \(\hat{\beta}_n = \arg \max_{\beta \in \mathbb{R}^p} \hat{Q}_t(\beta; \hat{G}_C)\). Write \(\tilde{\beta}_n = (\beta_{n1}, \tilde{\beta}_n^T)^T\), where \(\beta_{n1}\) satisfies the identifiability condition \(|\beta_{n1}| = 1\). In the proof of Theorem 1 in the online supplement, it was shown that \(\tilde{\beta}_n\) is consistent for \(\beta_0\). Hence, we have \(\hat{\beta}_{n1} = \beta_{n1}\) with probability approaching one. Theorem 1 below states the nonstandard convergence rate and nonnormal limiting distribution of \(\hat{\beta}_n\).

**Theorem 1.** Suppose conditions (C1)–(C4) are satisfied. Then as \(n \to \infty\),

\[ n^{1/3} (\hat{\beta}_n - \beta_0) \to \arg \max_t \left\{ \Psi(t) + \mathbb{W}(t) \right\} \]

in distribution, where \(\Psi(t)\) is a deterministic function whose form is given in (17) of the supplementary materials and \(\mathbb{W}(t)\) is a mean-zero Gaussian process with covariance function given in (19) in the supplementary materials.
The proof of Theorem 1 is given in the supplementary materials. Theoretical analysis of the asymptotic distribution of \( \hat{\beta}_n \) in (5) is challenging, as it is defined via a bilevel optimization problem. The proof involves reformulating \( \hat{\beta}_n \) as an M-estimator with a nonsmooth objective function that has two nuisance parameters: a finite dimensional nuisance parameter \( m_0 \) and an infinite dimensional nuisance parameter \( G_C(\cdot) \). The nonstandard asymptotics arise from the so-called sharp-edge effect, see Kim and Pollard (1990) for an informative example of the short estimator that illustrates this phenomenon. It is worth noting that the theory in Kim and Pollard (1990) can only handle a finite dimensional nuisance parameter, hence, is not applicable in our setting.

**Remark 4.** Our results are related to recent work on nonstandard estimation problem in Banerjee and McKeague (2007), Sen, Banerjee, and Woodroofe (2010), Matsouaka, Li, and Cai (2014), Wang et al. (2018), Shi, Lu, and Song (2018), Patra, Seijjo, and Sen (2018) and Banerjee, Durot, and Sen (2019). However, none of the above work involves an infinite-dimensional nuisance parameter as we face in the current setting. In fact, our estimation method involves both finite-dimensional and infinite-dimensional nuisance parameters, the role of the latter is for estimating the censoring distribution. Due to the nature of nonstandard asymptotics, the problem is different from and much harder than regular M-estimation problem with infinite-dimensional nuisance parameters. Advanced empirical process techniques from van der Vaart and Wellner (1996), Kosorok (2008), and Delso and Van Keilegom (2020) were adapted here to help deal with the theoretical challenges.

### 3.2. Estimating the Optimal Value with Possibly Nonunique Optimal IDR

Besides the optimal IDR itself, a quantity of interest is the optimal value, defined as

\[
V_{opt} = \sup_{\beta \in \mathbb{R}^p} Q_1 \left\{ T^n (d_\beta (X)) \right\}. \tag{7}
\]

This quantity is the maximally achievable marginal \( r \)th quantile of the potential distribution of all IDRs in the given class of candidate rules. It is an important measure of the performance of the potential distribution of all IDRs in the given class of candidate rules. A natural estimator of this quantity is \( \hat{V}_n = \hat{Q}_1 (\hat{\beta}_n; G_C) \).

**Theorem 2.** Suppose condition (C1) is satisfied. We have \( \hat{V}_n = V_{opt} + o_p(n^{-1/2} + \gamma_0) \), for an arbitrary \( \gamma_0 > 0 \).

**Remark 5.** It is worth emphasizing that Theorem 2 requires much weaker conditions than Theorem 1 does. In particular, it does not require the optimal IDR to be unique at the population or sample level. This corresponds to a well-known challenging situation where there exists a subpopulation who responds similarly to the two treatment options. If one is willing to assume unique optimal IDR, then the above result can be strengthened to parametric convergence rate, that is, \( n^{-1/2} \) rate.

### 3.3. Smoothed Resampling Inference

Statistical inference for \( \hat{\beta}_0 \) is challenging due to the nonstandard asymptotic distribution. A natural idea is to use bootstrap. However, the standard nonparametric bootstrap procedure is generally inconsistent for cube-root \( M \)-estimators (e.g., Abrevaya and Huang 2005; Léger and MacGibbon 2006) even for the relatively simpler setting without nuisance functions. As a remedy, \( m \)-out-of-\( n \) bootstrap (Bickel, Götze, and van Zwer 2012), which draws subsamples of size \( m \) from the original sample of size \( n \) with replacement, has been shown to be consistent for \( M \)-estimators with a cube root convergence rate in some settings (Delgado, Rodriguez-Poo, and Wöl 2001). Theoretically, \( m \) depends on \( n \), tends to infinity with \( n \), and satisfies \( m = o(n) \). Practically, choosing an optimal \( m \) is not a simple task. Several data-driven approaches for selecting \( m \) were investigated but require intensive computation (e.g., Banerjee and McKeague 2007; Bickel and Sakov 2008; Chakraborty, Laber, and Zhao 2013; Qian et al. in press).

In this section, we consider an alternative smoothed resampling-based procedure which is computationally more convenient. This approach is motivated by the alternative expression of \( \beta_0 \) (see the derivation of Lemma 1 in the supplementary materials), given by

\[
\beta_0 = \arg \max_{\beta \in \mathbb{R}^p} P_g (\cdot, \beta, V_{opt}, G_C),
\]

where \( V_{opt} \) is the optimal value and \( g (\cdot, \beta, v, G_C) \) is defined in (10). That is, \( \beta_0 \) is the parameter indexing the IDR that achieves the optimal value \( V_{opt} \). This naturally leads to an alternative representation of \( \beta_n \), given by \( \hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} g (\cdot, \beta, \hat{V}_n, G_C) \).

To implement the smoothed resampling-based inference, we first obtain the estimator \( \hat{\beta}_n \) and then estimate the optimal value function by \( \hat{V}_n = \arg \min_{\beta} \sum_{i=1}^{n} \frac{\hat{\beta}_n (\beta ) \rho (y_i - b)}{0.5 G_C (y_i)} \). Motivated by Wu and Wang (2021) for mean-optimal treatment regime with complete data, we consider the following smoothed estimator.
\( \bar{\beta}_n = \arg \max_{\beta \in \mathbb{R}^\infty} \frac{1}{n} \sum_{i=1}^{n} (2A_i - 1) \frac{\Delta_i I(Y_i > \widehat{V}_n)}{G_c(Y_i)} K \left( \frac{X_i^T \beta}{h_n} \right), \)

where \( K(\cdot) \) is a kernel function and \( h_n \) is a bandwidth. The kernel function \( K(\cdot) \) is only required to satisfy some general conditions, for example, we can take it to be the cumulative distribution function of the standard normal distribution. Replacing the indicator function in the treatment regime by the kernel function helps alleviate the sharp edge effect. Write \( \bar{\beta}_n = (\bar{\beta}_n, 1, \bar{\beta}_n)^T \) Similarly as in Wu and Wang (2021), it is expected that \( \sqrt{n} h_n (\beta - \bar{\beta}_0) \) is asymptotically normal. Note that \( \bar{\beta}_n \) minimizes the loss function \( -n^{-1} \sum_{i=1}^{n} (2A_i - 1) \frac{\Delta_i I(Y_i > \widehat{V}_n)}{G_c(Y_i)} K \left( \frac{X_i^T \beta}{h_n} \right), \) which is a smoothed estimator of a weighted misclassification error. We choose \( h_n \) by 5-fold cross-validation based on this loss function.

The asymptotic covariance matrix is complex and involves unknown counter-factual distributions. For inference, we consider the following perturbed smoothed estimator

\[
\tilde{\beta}_n = \arg \max_{\beta \in \mathbb{R}^\infty} \frac{1}{n} \sum_{i=1}^{n} \tilde{\xi}_i (2A_i - 1) \frac{\Delta_i I(Y_i > \widehat{V}_n)}{G_c(Y_i)} K \left( \frac{X_i^T \beta}{h_n} \right),
\]

where \( \tilde{\xi}_1, \ldots, \tilde{\xi}_n \) are positive random weights independent of the data, with mean one and variance one. To obtain the bootstrap distribution of \( \tilde{\beta}_n \), we repeatedly generate independent random weights and solve for the smoothed estimator. Write \( \tilde{\beta}_n = (\tilde{\beta}_{n1}, \tilde{\beta}_{n2}, \ldots, \tilde{\beta}_{np}) \) and \( \tilde{\beta}_n^* = (\tilde{\beta}_{n1}^*, \tilde{\beta}_{n2}^*, \ldots, \tilde{\beta}_{np}^*) \), where \( |\tilde{\beta}_{n1}| = 1 \). For \( j = 2, \ldots, p \), let \( \eta_{j1-\alpha/2} \) and \( \eta_{j1+\alpha/2} \) be the \((\alpha/2)\)th and \((1-\alpha/2)\)th quantile of the bootstrap distribution of \((nh_n)^{1/2}(\tilde{\beta}_{nj} - \tilde{\beta}_{nj}^*)\), respectively, where \( \alpha \) is a small positive number. We can estimate \( \eta_{j1-\alpha/2} \) and \( \eta_{j1+\alpha/2} \) from a large number of bootstrap samples. An asymptotic \( 100(1-\alpha)\% \) bootstrap confidence interval for \( \beta_{0j} \), \( j = 2, \ldots, p \), is given by \([\tilde{\beta}_{nj} - (nh_n)^{-1/2}\eta_{j1-\alpha/2}, \tilde{\beta}_{nj} - (nh_n)^{-1/2}\eta_{j1+\alpha/2}]\).

### 4. Estimation of Optimal Dynamic IDR with Censored Data

In this section, we consider the extension to the dynamic decision problem which involves multiple decision points. The decision at a later stage can depend on baseline covariates, treatment history, and intermediate variables such as how the subject responds to earlier treatment(s). For survival data, complications arise as the subject may be censored anytime before the end of the study, which results in an incomplete trajectory of treatments.

#### 4.1. Potential Outcome Framework

For clarity, we focus on a two-stage dynamic decision problem with random right censoring. We consider a setup similar as Jiang et al. (2017a) but will define the potential outcome more carefully. At the beginning of a study (time point 0), baseline covariates \( X_1 \) of patients would be collected, and each of them is assigned one of two second-stage treatment options, say \( A_1 \) and \( A_2 \). Then the second stage starts from a prespecified time \( s \) with \( s > 0 \). Additional intermediate covariates \( X_2 \) reflecting the reaction to first stage treatment up to time \( s \) would be collected if applicable, and those subjects who remain at risk at time \( s \) is assigned one of two first-stage treatment options, say \( B_1 \) and \( B_2 \). \( (B_1, B_2) \) may not overlap with \((A_1, A_2)\). For example, in making decisions for cancer patients, \((A_1, A_2)\) could be induction treatments, while \((B_1, B_2)\) may represent maintenance treatment or salvage treatment.

Similarly as in the one-stage setting, the potential survival time is defined when censoring is absent, and we would like to estimate the optimal dynamic IDR with a criterion based on this potential survival time when the real data is complicated by censoring. Let \( D_i \) denote the random treatment at stage \( i \) when the subject is eligible. Note that \( D_2 \) may not exist if the patient is not at risk at time \( s \). Consider a sequential IDR \( d = (d_1, d_2) \), where \( d_1(X_1) \in \{A_1, A_2\} \) and \( d_2(X_1, D_1, X_2) \in \{B_1, B_2\} \). A subject is considered to be consistent with \( d \) if he/she receives a first treatment \( D_1 \) that equals \( d_1(X_1) \) and a second treatment \( D_2 \) which equals \( d_2(X_1, D_1, X_2) \) (full compliance); or receives treatment \( D_1 \) complying with the rule \( d_1 \) at stage one but does not survive long enough to be eligible for stage two treatment.

Let \( R(d_1) = I(\hat{T}(d_1, \theta) > s) \) indicate the subject’s eligibility status for stage-two treatment when complying with rule \( d_1 \), where \( \hat{T}(d_1, \theta) \) is shorthand notation of \( T(d_1(X_1), \theta) \), which represents the potential survival time if the subject receives \( d_1 \) without stage-two action. Let \( \hat{T}(d_1, d_2) \) be the potential survival time if the subject receives the full sequence of treatments \((d_1, d_2)\). Implicitly, \( \hat{T}(d_1, d_2) > s \). Let \( T^*(d) \) be the potential survival time if the subject is consistent with the treatment sequence \( d \). We can write

\[
T^*(d) = \hat{T}(d_1, \theta)(1 - R(d_1)) + \hat{T}(d_1, d_2)R(d_1).
\]

We are interested in estimating the optimal sequential decision \( \hat{d} = (d_1, d_2) \) in some class \( D \), that is, \( \hat{d}_{opt} = \arg \max_{d \in D} T^*(d) \).

Define \( H_1 = \{X_1\} \), and define \( H_2(d_1) = \{X_1, d_1, X_2(d_1)\} \), where \( X_2(d_1) \) denotes the potential intermediate information between decision 1 and decision 2 had the subject started with treatment \( d_1(X_1) \) and given \( R(d_1) = 1 \). Denote the set of potential outcomes as

\[
O^*(d) = \{I(d_1, \theta), \hat{R}(d_1), \hat{R}(d_1)X_2^*(d_1), \hat{R}(d_1)d_2(H_2^*(d_1)), T^*(d) \}.
\]

#### 4.2. Robust Estimation of Optimal IDR

When censoring is absent, for a given subjects, we would observe the first stage treatment \( D_1 \) and the corresponding \( \hat{R}(D_1) \). The consistency assumption for causal inference, similar as in the one-stage setting, ensures that the observed survival time \( T \) would satisfy

\[
T = \begin{cases} 
\sum_{j \in [1,2]} I(D_1 = A_j) \hat{T}(A_j, \theta), & \text{if } \hat{R} = 0 \\
\sum_{j \in [1,2]} \sum_{k \in [1,2]} I(D_1 = A_j, D_2 = B_k) \hat{T}(A_j, B_k) & \text{if } \hat{R} = 1.
\end{cases}
\]

Due to censoring, we may not observe \( T \). Denote the actually observed survival time under possible censoring as \( Y = \min(T, C) \), and let \( \Delta = I(T \leq C) \) be the censoring indicator. Further, let \( \Gamma = I(C > s) \) denote whether censoring occurs
in the first stage. As a result of censoring, only those subjects that survived longer than \( s \) and are not censored before \( s \) are eligible for the second-stage treatment, for whom the trajectory observed up to time \( s \) is \( H_2 = \{X_1, D_1, X_2\} \). When the trial ends, the observed data is

\[
\{X_{1i}, D_{1i}, \tilde{R}_{1i}, \Gamma_{1i}, \tilde{R}_{1i}, \Gamma_{1i}, X_{12i}, \tilde{R}_{1i}, \Gamma_{1i}, D_{2i}, Y_{i}, \Delta_{i}\}, \text{ for } i = 1, \ldots, n. \tag{13}
\]

Based on the observed data, our goal is to estimate the optimal sequential decision rule within a given class \( D \). Extending the one-stage formulation, we consider sequential IDR of the form \( d_t = \{d_{t,1}(H_1), d_{t,2}(H_2)\} \), where \( d_{t,1}(H_1) = d_{t,1}(X_1) = I(X_1^T \beta > 0), d_{t,2}(H_2) = d_{t,2}(X_{1i}, D_{1i}, X_{2i}) = I(H_2^T \xi > 0) \) and \( \xi = (\beta^T, \gamma^T)^T \). Without loss of generality, we assume that if \( d_{t,1}(X_1) = 0 \), then the recommended first-stage treatment is \( A_1 \), otherwise it is \( A_2 \); and if \( d_{t,2}(H_2) = 0 \), then the recommended second-stage treatment is \( B_1 \), otherwise it is \( B_2 \). For identifiability, we assume \( \beta \) and \( \xi \) satisfy \( |\beta|_1 = 1 \) and \( |\xi|_1 = 1 \). Hence, \( D = \{d_{t} : \xi \in \mathbb{C}\} \) where \( \mathbb{C} = \{\{-1, 1\} \times \{-1, 1\} \times \{-1, 1\} \times \mathbb{Z}^+\} \), with \( \mathbb{Z}^+ \) being a compact subset of \( \mathbb{R}^{P_{1-1}} \), \( \mathbb{Z} \) being a compact subset of \( \mathbb{R}^{P_{2-1}} \), \( \mathbb{P}_1 \) is the dimension of \( X_1 \), and \( \mathbb{P}_2 \) is the dimension of \((X_1^T, D_1, X_2^T)^T\). The parameter \( \xi_0 \) indexing the optimal dynamic IDR in \( D \) is defined by

\[
\xi_0 = \arg \max_{\xi \in \mathbb{C}} Q_{\xi}(T^*(d_\xi)),
\]

where \( Q_{\xi}\{\cdot\} \) denotes the marginal \( \tau \)-th quantile \((0 < \tau < 1)\), and \( T^*(d_\xi) \) is obtained by setting \( d = d_\xi \) in (11). To extend the policy-search method to estimate \( \xi_0 \), we define

\[
\bar{R}(d_\xi) = \Delta I(D_1 = d_{t,1}(X_1)) \{I(Y \leq s) + I(Y > s)I(D_2 = d_{t,2}(H_2))\}.
\]

For subjects with \( \bar{R}(d_\xi) = 1 \), we observe the potential survival time of interest, that is, \( Y = T^*(d_\xi) \).

For simplicity in presentation, we assumed that the data are from a sequential multiple assignment randomized trial (SMART, (Lavori and Dawson 2000; Murphy 2008)), where the randomization probabilities at each stage are designed by that. That is, at stage one, \( P(D_1 = A_2) = 1 - P(D_1 = A_1) = \pi_{1,1} \), while at stage two \( P(D_2 = B_2 | Y_1 > s, C_1 > s) = 1 - P(D_2 = B_1 | Y_1 > s, C_1 > s) = \pi_{2,1} \). Given \( d = (d_{1}, d_{2}) \in D \), let \( \pi_{d_{1}}(X_{1i}) = \pi_{1,1}d_{1}(X_{1i}) + (1 - \pi_{1,1})(1 - d_{1}(X_{1i})) \) denote the probability of compliance to \( d_1 \) at stage one; and let \( \pi_{d_{2}}(H_{2i}) = \pi_{2,2}d_{2}(H_{2i}) + (1 - \pi_{2,2})(1 - d_{2}(H_{2i})) \) denote the probability of compliance to \( d_2 \) at stage two, given that stage one’s target potential data is observed and \( Y > s, C > s \). Overall, the probability to observe \( T^*(d) \) is

\[
\tilde{W}_{d,1} = P(\bar{R}_{1i}(d) = 1 | X_{1i}, O_{1i}(d)) = \pi_{d_{1}}(X_{1i})G_{C}(Y_1)\{I(Y_1 \leq s) + \pi_{d_{2}}(H_{2i})I(Y_1 > s)\}. \tag{14}
\]

We estimate \( Q_{\xi}(T^*(d_\xi)) \) by

\[
\hat{Q}_{\xi}(T^*(d_\xi); \hat{G}_{C}) = \arg \min_{\xi} \frac{\sum_{i=1}^{n} \tilde{R}_{1i}(d_{1})(Y_{i} - b)}{\tilde{W}_{d,1}},
\]

where \( \tilde{R}_{d_{1}} \) is obtained by plugging in the Kaplan–Meier estimator for \( G_{C} \) in (14). For brevity, we use the shorthand notation \( \hat{Q}_{\xi}(\xi; \hat{G}_{C}) \) for \( \hat{Q}_{\xi}(T^*(d_\xi); \hat{G}_{C}) \). Let \( L \) denote the end of the study. Assume there exists a constant \( \eta > 0 \) such that \( G_{C}(L) > \eta > 0 \). Furthermore, assume \( C \) has a continuously differentiable density function which is bounded away from infinity on \((0, L)\). Also, \( m_{0} = \sup_{x \in \mathbb{C}} Q_{\xi}(T_{d_{1}}(d_{1}) < L) \). Consider an arbitrary treatment sequence \( d = (d_{1}, d_{2}) \), with \( d_{1}(H_{1}) \in \{A_{1}, A_{2}\} \) and \( d_{2}(H_{2}) \in \{B_{1}, B_{2}\} \). Marginally, \( T(d_{1}, \theta) \) and \( T(d_{1}, d_{2}) \) have continuous distributions with continuously differentiable density functions. \( \forall x \in \mathbb{C} \), let \( f_{T_{d_{1}}(d_{1})}(\cdot) \) denote the marginal density function of the distribution of the potential survival time \( T_{d_{1}}(d_{1}) \). There exist positive constants \( \kappa_{1} \) and \( \delta \), such that \( \inf_{x \in \mathbb{C}} \inf_{m_{0} - m_{0}^{(\eta)}} f_{T_{d_{1}}(d_{1})}(m) > \kappa_{1} \).

The following lemma states the consistency of \( \hat{Q}_{\xi}(\xi; \hat{G}_{C}) \) for the marginal quantile of \( T_{d_{1}}(d_{1}) \).

**Lemma 1.** For all \( d_{1} \in D \), we have \( \hat{Q}_{\xi}(\xi; \hat{G}_{C}) \rightarrow Q_{\xi}(T^*(d_\xi)) \) in probability.

Hence, an estimator of the parameter \( \xi_0 \) is

\[
\hat{\xi}_{n} = \arg \max_{\xi \in \mathbb{C}} \hat{Q}_{\xi}(T^*(d_\xi); \hat{G}_{C}). \tag{15}
\]

The optimal value function, the maximally achievable marginal \( \tau \)-th quantile of the potential outcome distribution considering all IDRs in \( D \), is given by \( V_{\text{opt}} = \sup_{\xi \in \mathbb{C}} Q_{\xi}(T^*(d_\xi)) \). An estimator of this quantity is \( \hat{V}_{\hat{\xi}} = \hat{Q}_{\xi}(\hat{\xi}_{n}, \hat{G}_{C}) \). Similarly as the one-stage setting, we can re-express \( \hat{V}_{\hat{\xi}} \) as

\[
\hat{V}_{n} = \sup_{\xi \in \mathbb{C}} \left\{ m : \sup_{n^{-1} \sum_{i=1}^{n} g_{\xi}(\cdot, \xi, m, \hat{G}_{C}) \geq 0 \right\}, \tag{16}
\]

where \( g(\cdot, \xi, m, \hat{G}_{C}) = \frac{\tilde{R}(d_{1})(Y_{m} > 0)}{m_{0}} \). The following theorem shows that in the dynamic setting, the optimal value function can be estimated with a near parametric rate.

**Theorem 3.** For the estimator \( \hat{V}_{n} \) defined in (16), we have \( \hat{V}_{n} = V_{\text{opt}} + o_{p}(n^{-1/2+\varepsilon}) \) for an arbitrary \( \varepsilon > 0 \).

It is worth noting that the above result does not require the optimal IDR to be unique. Similar nonregular asymptotic distribution for \( \hat{\xi}_{n} \) can also be established using the same idea as in Section 3 but more complex notations.

**Remark 6.** The method we propose for the dynamic setting is different from the Q-learning approach, which searches for optimal treatment regimes starting from the last stage and moving backward. The Q-learning approach was extended to censored data by Goldberg and Kosorok (2012). There exist several distinct differences between these two methods. First, the proposed method is model-free in the sense that it does not require to specify an outcome regression model. The Q-learning approach is model-based and requires a survival time model that incorporates both the covariate effects and treatment–covariate interaction effects. Second, the proposed method considers a quantile-optimal criterion while Goldberg and Kosorok (2012) adopts a restricted mean criterion. Finally, from a theoretical perspective, this work focuses on the statistical properties of the estimated parameters indexing the optimal IDR while Goldberg
and Kosorok (2012) focuses on the finite sample bound of the generalization error of the estimated optimal IDR.

5. Numerical Studies

5.1. Monte Carlo Simulations

We report simulation results for three different settings. In the first example, we estimate one-stage optimal treatment under random censoring; in the second example, we estimate one-stage optimal treatment under covariate-dependent censoring; while in the third example, we consider a two-stage dynamic optimal IDR estimation problem.

**Example 1 (random censoring).** We generate the random sample \( \{X_i, A_i, Y_i, \Delta_i\}, i = 1, 2, \ldots, n \), from the model: \( X_1 \sim U(0, 1), T^*(0)[X_1] \sim \text{Weibull}(\text{shape} = 1, \text{scale} = 1) + 1, \) \( T^*(1)[X_1] \sim \text{Weibull}(\text{shape} = 3, \text{scale} = 0.5 + X_1) + 2X_1, \) \( A[X, T^*(0), T^*(1)] \sim \text{Bernoulli}(0.5) \). The response variable in the absence of censoring is generated by \( T = T^*(0)(1 - A) + T^*(1)A \). The censoring time \( C \) has a constant density function \( 0.22 \) on \([0, 2]\) and a constant density function \( 0.07 \) on \([2, 10]\). The observed response is \( Y = \min(T, C) \) and the censoring indicator is \( \Delta = I[T \leq C] \). This setup achieves an overall censoring rate of 35%.

To illustrate the heterogeneous treatment effects, we split \( X_1 \) into two strata: \([0, 0.5]\) and \((0.5, 1]\). Figure 1 displays the histograms of \( T^*(0) \) and \( T^*(1) \) in each stratum. This plot provides strong evidence that \( X_1 \) has a qualitative interaction with the treatment. Intuitively, the optimal IDR should depend on \( X_1 \). We will apply the proposed method to estimate the quantile-optimal IDR for \( \tau = 0.25 \) and \( \tau = 0.5 \), respectively. We consider the following class of IDRs \( D = \{d_\beta(X) = I(X_1 \beta_1 + \beta_2 > 0) : |\beta_1| = 1, \beta_2 \in \mathbb{R} \} \). Denote the parameter indexing the \( \tau \) th quantile optimal IDR in \( D \) by \( \beta_\tau^{(r)} \).

For each \( \tau \), we use a large Monte Carlo dataset of size \( n = 10^4 \) to estimate \( \beta_0^{(r)} \) and the \( \tau \) th quantile of the potential outcome in the above class of IDRs (denoted by \( Q_\tau \)) and treat the results as population parameter values, see Table 1. Consider, for example, the row corresponding to \( \tau = 0.5 \). We apply the 0.5-quantile optimal IDR to assign treatment in a large independent Monte Carlo sample. Assume everyone in the population follows the recommended treatment and records his/her outcome. The median of the potential outcome distribution is \( 2.258 \), the first quartile of the potential outcome distribution is \( 1.587 \).

We compare the proposed estimator (denoted by New) with the naive estimator (denoted by Naive), which ignores censoring and pretends all observations are complete (Wang et al. 2018). We conduct the simulation experiment with 400 replications for sample size \( n = 300, 500, \) and \( 1000 \). In this experiment, we observed that New always correctly estimates the sign of \( \beta_0 \) for both \( \tau = 0.25 \) and \( \tau = 0.5 \), while Naive has 4%, 2%, and 1% error rate for \( n = 300, 500, \) and \( 1000 \), respectively, in estimating the sign of \( \beta_0 \) for \( \tau = 0.25 \) (0% error rate for \( \tau = 0.5 \)). Considering this phenomenon, we conservatively compare the estimates for \( \beta_0^{(r)} \) for the two methods in cases where \( \beta_0 = 1 \). Table 2 summarizes the bias (with standard deviation in the parenthesis) of New and Naive for estimating \( \beta_0^{(r)} \) and \( Q_\tau \) for different combinations of \( \tau \) and \( n \). To estimate \( Q_\tau \), New plugs \( \hat{\beta}_0 \) into the formula of \( \hat{Q}_\tau \) in (5); and Naive plugs in similarly pretending all observations were complete. We observe that New has satisfactory performance, while Naive exhibits substantial bias for estimating both \( \beta_0^{(r)} \) and \( Q_\tau \).

Finally, we demonstrate the smoothed resampling procedure in Section 3.3 for inference. We consider 90% and 95% confidence intervals for \( \beta_0 \) for \( \tau = 0.25 \) and \( 0.5 \), respectively. The empirical coverage probabilities and average confidence interval lengths are reported in Table 3 for \( n = 1000 \) based on 400 bootstrap samples. The observed empirical coverage probabilities are close to the nominal levels with reasonable lengths.

**Table 1. Example 1: Parameters indexing the quantile-optimal IDRs (\( \tau = 0.25 \) and 0.5) in \( D \) and the \( \tau \) th quantile of the potential outcome distribution (denoted by \( Q_\tau \)), based on a Monte Carlo experiment (\( n = 10^4 \)).

| \( \tau \) | \( \beta_0^{(r)} \) | \( \beta_0^{(r)} \) | \( Q_{0.25} \) | \( Q_{0.5} \) |
|----------|----------------|----------------|-------------|-------------|
| 0.25     | 1              | -0.428         | 1.658       | 2.215       |
| 0.50     | 1              | -0.552         | 1.587       | 2.258       |

**Figure 1.** Histograms of \( T^*(0) \) and \( T^*(1) \) stratified by \( X_1 \).
Table 2. Bias (with standard deviation in the parenthesis) of New and Naive for estimating $\beta_{02}^{(r)}$ and $Q_r$ for Example 1.

| $\tau$ | n   | New          |         | Naive         |         |
|--------|-----|--------------|---------|--------------|---------|
|        |     | $\hat{\beta}_{02}^{(r)}$ | $Q_r$   | $\hat{\beta}_{02}^{(r)}$ | $Q_r$   |
| 0.25   | 300 | 0.005(0.066) | −0.025 | 0.056(0.113) | −0.025 |
|        | 500 | −0.001(0.054) | −0.043 | 0.027(0.082) | −0.043 |
| 0.50   | 300 | 0.001(0.043) | −0.048 | 0.020(0.055) | −0.048 |
|        | 500 | 0.003(0.080) | 0.122  | 0.048(0.124) | 0.122  |

Table 3. Confidence intervals for $\hat{\beta}_{02}^{(r)}$ using smoothed resampling.

| n   | $\tau$ | Coverage | Length | Coverage | Length |
|-----|--------|----------|--------|----------|--------|
| 500 | 0.5    | 0.89     | 0.17   | 0.92     | 0.21   |
|     | 0.25   | 0.91     | 0.29   | 0.95     | 0.34   |
| 1000| 0.25   | 0.88     | 0.13   | 0.93     | 0.16   |

Table 4. Parameters indexing the quantile-optimal IDR ($\tau = 0.1$ and 0.25) and the most achievable $r$th quantile of the potential outcome (denoted by $Q_r$) in $\mathbb{D}$ in Example 1, based on a Monte Carlo experiment ($n = 10^7$).

| $\tau$ | $\hat{\beta}_{01}^{(r)}$ | $\hat{\beta}_{02}^{(r)}$ | $\hat{\beta}_{03}^{(r)}$ | $Q_r(\beta)$ |
|--------|--------------------------|--------------------------|--------------------------|---------------|
| 0.10   | −1                       | 0.89                     | −0.774                   | 1.853         |
| 0.25   | −1                       | 1.140                    | −2.247                   | 2.247         |

Example 2 (covariate-dependent censoring). Let $X_1 = (X_{11}, X_{12})^T$, where $X_{11}, X_{12}$ are independent Uniform (0, 1) random variables. The binary treatment $A_i$ is independent of $X_i$ and satisfies $P(A_i = 1) = 0.5$. The distribution of censoring variable $C_i$ is

$$C_i = \begin{cases} 
4 + (2 - X_{11}) \omega_1, & \text{if } A_i = 0 \\
2 + I(X_{11} < 0.5 \land X_{12} < 0.5) + \omega_1, & \text{if } A_i = 1
\end{cases},$$

where the $\omega_i$s are independent $N(0, 1)$ random variables. The survival time $T_i$ is generated by

$$T_i = 1 + X_{11} + X_{12} + A_i (3 - 3X_{11} - 1.5X_{12}) + [0.5 + A_i (1 + X_{11} + X_{12})] \epsilon_i,$$

where the $\epsilon_i$s are independent normal random variables with mean zero and standard deviation 0.5. The observed response is $Y_i = \min(T_i, C_i)$. This configuration yields a 30% censoring rate. We consider estimating the $r$th quantile of the optimal IDR ($\tau = 0.1$ and 0.25) within the class $\mathbb{D} = \{1(\beta_1 X_{11} + \beta_2 + \beta_3 X_{12} > 0) : |\beta_1| = 1, (\beta_2, \beta_3) \in \mathbb{R}^2\}$. Similarly as for example 1, the parameters indexing the quantile-optimal IDR ($\tau = 0.1$ and 0.25) and the most achievable $r$th quantile of the potential outcome (denoted by $Q_r$) in $\mathbb{D}$ were estimated based on a large Monte Carlo experiment with sample size $n = 10^7$ and treated as population parameter values, see Table 4.

To incorporate covariate-dependent censoring, we adopt the local Kaplan–Meier estimator (with bandwidth $h_n = 0.08, 0.1, 0.12, 0.14$) described in Remark 1 of Section 2 to estimate the propensity score. Table 5 summarizes the simulation results for New and Naive for $n = 500$ based on 300 replications. We observe that New has satisfactory performance and its performance is stable with respect to different choices of the bandwidth $h$. In contrast, Naive exhibits substantial bias for estimating $\beta_{02}$ and $Q_r$.

Example 3 (Two-stage dynamic individualized decision rule). The simulation setup is motivated by the example in Jiang et al. (2017a). The censoring time $C \sim \text{Unif}(0, C_0)$, where $C_0$ is a positive constant. Let $s = 1$. Generate the data up to time $s$, $\{X_1, D_1, Z^C = I(\min(T_1, C) > s)\}$, from the following distributions: $X_1 \sim \text{Unif}(0, 4), D_1 \mid X_1 \sim \text{Bernoulli}(0.5)$ and $T_1 \mid X_1, D_1 \sim \text{Exp}(\lambda_1(X_1, D_1))$ where $\lambda_1(\cdot)$ is a rate function to be specified later, and $Z^C$ is an auxiliary variable that equals the product of $R$ and $\Gamma$ in observed data model (13).

If $Z^C = 1$, then the simulated patient is eligible for stage-two treatment. We generate the intermediate covariate $X_2$, second stage treatment $D_2$ and time $T_2$, representing the survival time after time $s$ according to: $e \sim \text{Unif}(0, 2), X_2 \mid X_1, D_1 \sim \text{Bernoulli}(0.5), T_2 \mid X_1, D_1, X_2 \sim \text{Exp}(\lambda_2(X_1, D_1, X_2, D_2))$, where $\lambda_2(\cdot)$ is a rate function to be specified later. For $Z^C = 0$, the observed survival time and censoring status are $Y = \min(s + T_2, C)$ and $\Delta = I(s + T_2 \leq C)$, respectively; for $Z^C = 0$, the observed survival time and censoring status are $Y = \min(T_1, C)$ and $\Delta = I(T_1 \leq C)$, respectively. Let $H_1 = \{X_1 \} \land H_2 = \{X_1, D_1, X_2\}$. For rate functions for $T_1$ and $T_2$, consider three scenarios:

(a) $\lambda_1(H_1, D_1) = 0.5 \exp(1.75(D_1 - 0.5)(X_1 - 2))$;
(b) $\lambda_2(H_1, D_1) = 0.5 \exp(2.5(D_2 - 0.4)(X_2 - 2) - D_1(X_1 - 2))$;
(c) $\lambda_3(H_1, D_1) = 0.3 \exp(3(D_1 - 0.3)(X_1 - 3))$.

For each scenario, we consider two different choices of $C_0$ to achieve 15% and 40% overall censoring rate, respectively.

In this setup, $T_1$ serves as the underlying $\sum_{i \in [1, 2]} I(D_1 = A_i) \hat{I}(A_i, \theta)$ in equation (12); $T_2$ is the survival time after initiation of the second stage treatment conditional on $Z^C = 1$, hence, $T_2 + s$ serves as $\sum_{i \in [1, 2], k \in [1, 2]} I(D_1 = A_i, D_2 = B_k) \hat{I}(A_i, B_k)$ in (12). By the interaction between $D_2$ and $H_2$ in $\lambda_2(\cdot)$ functions, the three scenarios share the same true optimal second-stage strategy for patients with $Z^C = 1$, which is $d_{\text{opt}}^{\text{opt}}(H_2) = I(\neg X_2 + 2 > 0)$. Suppose the class of IDRs is $\mathcal{D} = \{d_k = (d_1, \beta, d_2, \xi) : d_1, \beta(X_1) = I(\beta_1 X_1 + \beta_2 > 0), d_2, \xi(X_2) = I(\xi_1 X_2 + \xi_2 > 0), |\beta| = 1, |\xi| = 1\}$. Thus, $d_{\text{opt}}^{\text{opt}}(H_2)$ is contained in $\mathcal{D}$, and the parameter indexing $d_{\text{opt}}^{\text{opt}}(H_2)$ in $\mathcal{D}$ is $\zeta_0 = -1, \zeta_0 = 2$. However, for all three cases, the true value for $\beta_0$ indexing the optimal first-stage treatment in $\mathcal{D}$ does not have closed form representations. We used grid-search with sample size $n = 10^7$ for $\beta \in [-1] \times [0, 4] \cup [1] \times [-4, 0]$ to obtain $\beta_0$ in these cases, where the search space is the largest set of identifiable $\beta$ since $X_1$ has support $[0, 4]$.

With the above setup, we generate a random sample $\{X_{1i}, D_{1i}, Z^C_{i1}, X_{12i}, Z^C_{12i}, Y_i, \Delta_i, i = 1, \ldots, n\}$, where $n = 300, 500$, and 1000 are considered. We assume the random-
organization probability $\pi_1$ and $\pi_2$ are known, and applied the proposed method to estimate the optimal IDR with $\tau = 0.3$. Table 6 reports the simulation estimates of bias and standard deviations of the parameter indexing the quantile optimal dynamic regime $\xi_n$. It also reports estimates of bias and standard deviation of the plug-in estimator of maximal achievable 0.3-quantile, $Q_{0.3}(\hat{f}_{\tau,\theta})$. We observe the proposed method reliably estimated the optimal two-stage IDR. The average biases and standard deviations of $\hat{\beta}_{02}$ and $\hat{\beta}_{03}$ decrease as sample size increases. The lower censoring rate corresponds to better performance.

6. Analysis of GBSG2 Study Data

To illustrate the proposed method, we analyze the data from the GBSG2 study conducted by the German Breast Cancer Study Group (Schumacher et al. 1994; Schmoller, Olschewski, and Schumacher 1996). The study investigated the efficacy of four combinations of treatments: three versus six cycles of chemotherapy with or without the adjuvant hormonal therapy with Tamoxifen. The outcome of interest is the recurrence-free survival time in days. The dataset, available in the R package TH.data (Hothorn 2017), contains information on 686 patients of whom 56% had censored outcomes. The covariates include the age at diagnosis, the menopausal status, tumor size, tumor grade, the number of positive lymph nodes, estrogen receptor (ER) and progesterone receptor (PR) expression level in the tumor tissue. Earlier work on this study provided strong evidence that six cycles of chemotherapy is not superior to three cycles with respect to recurrence-free survival. Our analysis therefore focuses on IDRs regarding the assignment of adjuvant Tamoxifen therapy. Figure 2 depicts the estimated Kaplan–Meier curves for the groups of patients with and without Tamoxifen therapy.

First, we estimate the probability of receiving Tamoxifen. In this study, about two thirds of the recruited patients were randomized, and those who were not randomized chose the treatment by personal or professional preference. Because the randomization status is masked in the anonymous version of this data, we consider a working propensity score model by logistic regression. We observe that the randomization status is not significantly associated with the survival (Schmoller, Olschewski, and Schumacher 1996). Let $A$ denote the hormonal therapy...

---

**Table 5.** Bias (with standard deviation in the parenthesis) of New and Naive for estimating $\beta_0(\tau)$ and $\xi_0$ for Example 2.

| Method         | $\beta_0(\tau)$ | $\xi_0(\tau)$ | $Q_{\tau}$ |
|----------------|-----------------|----------------|------------|
| Naive          | -0.107(0.150)   | 0.007(0.295)   | -0.169(0.665) |
| New ($h_n=0.08$) | -0.022(0.094)   | 0.020(0.162)   | 0.025(0.059) |
| New ($h_n=0.10$) | -0.016(0.091)   | 0.029(0.165)   | 0.030(0.058) |
| New ($h_n=0.12$) | -0.009(0.086)   | 0.020(0.157)   | 0.027(0.058) |
| New ($h_n=0.14$) | -0.020(0.094)   | 0.035(0.170)   | 0.038(0.054) |

**Table 6.** Simulation results about the bias (standard deviation of the estimates given in the parentheses) of $\hat{\xi}_n$ relative to $\xi_0$ and of $Q_{0.3}(\hat{f})$ relative to $Q_{0.3}(\xi_0)$.

| Case | n | C% | Bias in $\beta_0$ | Bias in $\xi_0$ | Bias in $Q_{0.3}(\xi_0)$ |
|------|---|----|-------------------|-----------------|------------------------|
| (a)  | 300 | 40 | -0.032(0.499)     | 0.035(0.391)   | 0.525(0.445) |
| 500  | 40 | 0.036(0.416)    | 0.031(0.339)   | 0.353(0.358) |
| 1000 | 40 | -0.005(0.317)   | 0.043(0.282)   | 0.206(0.220) |
| 300  | 15 | 0.016(0.402)    | -0.008(0.356)  | 0.331(0.345) |
| 500  | 15 | 0.010(0.329)    | 0.006(0.303)   | 0.247(0.262) |
| 1000 | 15 | -0.008(0.255)   | 0.001(0.230)   | 0.139(0.187) |
| (b)  | 300 | 40 | -0.030(0.648)     | -0.023(0.371)  | 0.280(0.225) |
| 500  | 40 | -0.031(0.581)   | -0.015(0.344)  | 0.181(0.149) |
| 1000 | 40 | -0.016(0.429)   | -0.013(0.279)  | 0.120(0.105) |
| 300  | 15 | -0.031(0.600)   | 0.005(0.332)   | 0.235(0.178) |
| 500  | 15 | -0.037(0.485)   | 0.012(0.314)   | 0.155(0.137) |
| 1000 | 15 | 0.006(0.414)    | -0.004(0.260)  | 0.101(0.090) |
| (c)  | 300 | 40 | -0.049(0.433)     | -0.009(0.342)  | 0.484(0.396) |
| 500  | 40 | -0.056(0.372)   | 0.001(0.299)   | 0.380(0.347) |
| 1000 | 40 | 0.003(0.277)    | 0.008(0.250)   | 0.198(0.231) |
| 300  | 15 | -0.074(0.404)   | 0.002(0.304)   | 0.379(0.329) |
| 500  | 15 | -0.029(0.336)   | -0.012(0.254)  | 0.261(0.286) |
| 1000 | 15 | -0.015(0.249)   | 0.001(0.229)   | 0.150(0.167) |

**Figure 2.** Plot of the Kaplan–Meier estimator of survival functions of $T$ for $A = 0, 1$, respectively.

**Table 7.** True values of $\xi_0 = (\beta_0^T, \xi_0^T)^T$ indexing the 0.3 quantile-optimal DTR.

| Case  | $\beta_0$ | $\xi_0$ | $Q_{0.3}(\xi_0)$ |
|-------|-----------|---------|-----------------|
| (a)   | $(-1, 2.00)^T$ | $(-1, 2)^T$ | 1.524 |
| (b)   | $(1, -1.95)^T$ | $(1, -1)^T$ | 1.566 |
| (c)   | $(1, -2.94)^T$ | $(1, -2)^T$ | 2.132 |
status \((A = 0: \text{did not receive}; A = 1: \text{received})\). We first fit a logistic regression model using \(A\) as the response and all available covariates. We then perform a best subset selection using the R package bestglm (McLeod and Xu 2010), and obtained the model \(\logit(\pi_A (X; \gamma)) = \gamma_0 + \gamma_1 \text{MNST}\), where MNST is the binary menopausal status of patients. Using this selected model, we obtain the estimated propensity score 0.203 for premenopausal patients, and 0.472 for postmenopausal patients. The dependency of propensity score on MNST is mostly due to a modification on the protocol for randomization starting from the third year of GBSG2 recruitment. We also tried the propensity score model with all covariates and found that it leads to almost the same recommendations as measured by the match ratio (percentage of times two decision rules make the same treatment recommendations), which is above 98% for both of the two classes of regimes under consideration.

Motivated by the extensive work in the medical literature on Tamoxifen’s molecular level mechanism and its clinical long-term effects, we consider IDRs that depend on the following three variables.

1. ER: The role of estrogen receptor expression as a predictive factor guiding the allocation of tamoxifen is well recognized. A large meta-analysis of randomized clinical trials demonstrated that high-ER patients respond better to Tamoxifen compared with low-ER patients (Early Breast Cancer Trialists’ Collaborative Group 1998).

2. PR: Progesterone receptor expression is routinely measured for breast cancer patients as an important prognostic factor. However, its predictive power for the efficacy of Tamoxifen is still not well understood. It was observed that breast cancer patients with both high ER and high PR (“double positive”) have the best chance of surviving (Bardou et al. 2003).

3. Age: Age is an important risk factor in breast cancer. We speculate that it may also contribute to how well patients respond to the adjuvant Tamoxifen therapy.

Because ER and PR are both highly skewed and have the minimal value 0 in this dataset, we adopt the transformation \(\text{LER} = \log_{10}(\text{ER} + 1)\) and \(\text{LPR} = \log_{10}(\text{PR} + 1)\). Age is linearly normalized to be between 0 and 1, and is denoted by NAGE.

We first consider the class of IDRs \(D_1\) that depend on all three variables:

\[
D_1 = \{ \{ [\beta_1 \text{LER} + \beta_2 + \beta_3 \text{LPR} + \beta_4 \text{NAGE} > 0 ] : \beta_1 = 1, \\
\beta_2, \beta_3, \beta_4 \in \mathbb{R} \} \}.
\]

We restrict the sign of ER to be positive based on evidence from the clinical practice (Hammond et al. 2010). We estimate the IDR in the class \(D_1\) that maximizes the first quartile \((\tau = 0.25)\) of the recurrence-free survival time. We examine the dependence of the censoring time \(C\) on \(A\) and all seven prognostic factors by Cox regression and conclude that the independent censoring assumption is plausible. Furthermore, since GBSG2 has a high censoring rate and relatively short follow-up time, hereafter we use the artificial censoring technique (with \(M\) being set as 1550 days) in Remark 2 in Section 2.3 to improve stability of the proposed method.

The estimated parameter indexing the quartile-optimal IDR is \(\widehat{\theta}_{n;D_1} = (1, -1.23, 0.94, -0.14)^T\). This regime leads to an estimated quartile survival time of \(\widehat{Q}_{0.25} (\widehat{\theta}_{n;D_1}) = 1246\) days with approximately 81.6% of patients being recommended to treatment. In contrast, the Kaplan–Meier estimator of the first quartile of the observed survival time is 727 (90% confidence interval = (622, 805)). The first row in Table 8 reported the 90% smoothed bootstrap confidence interval (Section 3.3) for each coefficient except \(\beta_1\) based on 400 bootstrap samples. The coefficient of NAGE, \(\beta_4\), is insignificant at the 0.1 level, which suggests that age may not be an important variable for determining Tamoxifen.

Next, we estimate quartile-optimal IDR in the following simplified class of IDRs to obtain a concise rule,

\[
D_2 = \{ \{ [\beta_1 \text{LER} + \beta_2 + \beta_3 \text{LPR} > 0 ] : \beta_1 = 1, \beta_2, \beta_3 \in \mathbb{R} \} \}.
\]

The estimated parameter indexing the quartile-optimal IDR in \(D_2\) is \(\widehat{\theta}_{n;D_2} = (1, -1.26, 0.97)\), which leads to an estimated quartile survival time of \(\widehat{Q}_{0.25} (\widehat{\theta}_{n;D_2}) = 1246\) days with approximately 82.3% of patients being recommended to treatment. The second row of Table 8 reported the 90% smoothed bootstrap confidence intervals for \(\beta_2\) and \(\beta_3\). The coefficient for LPR is significant. Also, because the estimated \(\beta_2\) for LPR is about 1, we conclude that LPR is as important as the well-established predictive factor LER in developing an IDR that optimizes the first quartile survival time.

**Table 8. Estimated parameters indexing the quartile-optimal IDR and 90% smoothed bootstrap confidence intervals for the GBSG2 study.**

| Regimes | \(\beta_2\) | \(\beta_3\) | \(\beta_4\) |
|---------|------------|------------|------------|
| \(D_1\) | -1.23      | 0.94       | -0.14      |
|         | (-3.39, -0.88) | (0.87, 2.02) | (-1.21, 2.39) |
| \(D_2\) | -1.26      | 0.97       | /          |
|         | (-2.41, -1.07) | (0.43, 2.01) | /          |

**Appendix: Regularity Conditions**

We introduce below a set of regularity conditions needed to establish the statistical theory. The proof of the theory is given in the supplementary materials.

**C1** Let \(L\) denote the end of the study. The censoring variable \(C\) has a continuously differentiable density function which is bounded away from infinity on \((0, L)\). There exists a constant \(\eta > 0\) such that \(G_C (L) > \eta > 0\). The densities \(f_0 (t | X)\) and \(f_1 (t | X)\) are uniformly bounded away from infinity, almost surely in \(X\), and \(\sup_{\theta \in \mathbb{R}^p} \mathbb{E}_T \{ T^d (d \theta) \} < L\). There exist positive constants \(\epsilon_1\) and \(\delta\), such that \(\inf_{\theta \in \mathbb{R}^p} \inf_{m - m_0 \geq \delta} \mathbb{E}_T \{ T^d (d \theta) \} \geq \epsilon_1\), where \(m_0 = \sup_{\theta \in \mathbb{R}^p} \mathbb{E}_T \{ T^d (d \theta) \} \).

**C2** The probability density function of \(X_1\) conditional on \(X\) is continuously differentiable. The angular component of \(X\), considered as a random element of the sphere \(\mathbb{S} \subseteq \mathbb{R}^p\), has a bounded and continuous density.

**C3** The population parameter \(\theta_0 = (\theta_{01}, \ldots, \theta_{0p})^T\) indexing the optimal IDR is unique in \(\mathbb{R}^n\).

**C4** The \((p-1) \times (p-1)\) matrix \(\Lambda (\hat{\theta}, h) |_{\theta = \hat{\theta}, h = h_0}\) defined in the proof of Lemma 1 in the supplementary materials, is negative definite.

**Remark.** Condition (C1) is common in survival analysis, where \(L\) is the maximum follow-up time. The survival time is not observed if it exceeds \(L\). Condition (C2) has to do with population parameter
identifiability, as discussed in Section 2.2. We assume after a possible rearranging of elements in $\mathbf{X}$, the density of $X_1$ conditional on $\mathbf{X}$ is continuously differentiable for every $\mathbf{X}$ almost surely. If all covariates are discrete, the problem of estimating an optimal IDR actually becomes simpler in some sense as there are finite many decision rules. One can directly compare the estimated value functions. Condition (C3) is standard for index models. Condition (C4) is needed for evaluating the Hessian matrix when establishing the limiting distribution of $\hat{\beta}_n$. 

Supplementary Materials

The supplementary materials contain the technical derivations and additional numerical results.

Acknowledgments

We thank Coeditor Professor McKeague, the AE and the anonymous referees for their helpful comments.

Funding

Wang, Zhou, and Zhao's research was supported by NSF FRGMS-1952373. Song's research was supported by NSF DMS-2113637.

References

Abrevaya, J., and Huang, J. (2005), “On the Bootstrap of the Maximum Score Estimator,” Econometrica, 73, 1175–1204. [2636]
Bai, X., Tsiatis, A. A., Lu, W., and Song, R. (2017), “Optimal Treatment Regimes for Survival Endpoints Using a Locally-Efficient Doubly-Robust Estimator from a Classification Perspective,” Lifetime Data Analysis, 23, 585–604. [2632]
Banerjee, M., Durot, C., and Sen, B. (2019), “Divide and Conquer in Non-standard Problems and the Super-Efficiency Phenomenon,” The Annals of Statistics, 47, 720–757. [2636]
Banerjee, M., and McKeague, I. W. (2007), “Confidence Sets for Split Points in Decision Trees,” The Annals of Statistics, 35, 543–574. [2636]
Bardou, V.-I., Arsino, G., Eldredge, R. M., Osborne, C. K., and Clark, G. M. (2003), “Progestosterone Receptor Status Significantly Improves Outcome Prediction Over Estrogen Receptor Status Alone for Adjuvant Endocrine Therapy in Two Large Breast Cancer Databases,” Journal of Clinical Oncology, 21, 1973–1979. [2642]
Bickel, P. J., Götze, F., and van Zwart, W. R. (2012), “Resampling Fewer than $n$ Observations: Gains, Losses, and Remedies for Losses,” in Selected Works of Willem van Zwet, eds. S. van de Geer, and M. Wegkamp, pp. 267–297, New York: Springer. [2636]
Bickel, P. J., and Sákov, A. (2008), “On the Choice of m in the m out of n Bootstrap and Confidence Bounds for Extrema,” Statistica Sinica, 18, 967–985. [2636]
Chakraborty, B., Laber, E. B., and Zhao, Y. (2013), “Inference for Optimal Dynamic Treatment Regimes Using an Adaptive m-out-of-n Bootstrap Scheme,” Biometrics, 69, 714–723. [2636]
Chakraborty, B., and Moodie, E. E. (2013), Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine, New York: Springer. [2632]
Chakraborty, B., Murphy, S., and Strecher, V. (2010), “Inference for Non-regular Parameters in Optimal Dynamic Treatment Regimes,” Statistical Methods in Medical Research, 19, 317–343. [2632]
Cui, Y., Zhu, R., and Kosorok, M. (2017), “Tree Based Weighted Learning for Estimating Individualized Treatment Rules with Censored Data,” Electronic Journal of Statistics, 11, 3927–3953. [2632]
Delgado, M. A., Rodriguez-Poo, J. M., and Wolf, M. (2001), “Subsampling Inference in Cube Root Asymptotics with an Application to Manski’s Maximum Score Estimator,” Economics Letters, 73, 241–250. [2636]
Delsol, L., and Van Kelegem, I. (2020), “Semiparametric m-estimation with Non-smooth Criterion Functions,” Annals of the Institute of Statistical Mathematics, 72, 577–605. [2636]
Diaz, I., Savenkov, O., and Ballman, K. (2018), “Targeted Learning Ensembles for Optimal Individualized Treatment Rules with Time-to-Event Outcomes,” Biometrika, 105, 723–738. [2632]
Early Breast Cancer Trials’ Collaborative Group. (1998), “Tamoxifen for Early Breast Cancer: An Overview of the Randomised Trials,” The Lancet, 351, 1451–1467. [2642]
Goldberg, Y., and Kosorok, M. R. (2012), “Q-Learning with Censored Data,” Annals of Statistics, 40, 529–568. [2632,2634,2638,2639]
Gonzalez-Manteiga, W., and Cadarso-Suárez, C. (1994), “Asymptotic Properties of a Generalized Kaplan–Meier Estimator with Some Applications,” Communications in Statistics-Theory and Methods, 4, 65–78. [2635]
Hager, R., Tsiatis, A. A., and Davidian, M. (2018), “Optimal Two-Stage Dynamic Treatment Regimes from a Classification Perspective with Censored Survival Data,” Biometrics, 74, 1180–1192. [2632]
Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., and Hicks, D. G. (2010), “American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer (Unbridged Version),” Archives of Pathology & Laboratory Medicine, 134, e48–e72. [2642]
Hothorn, T. (2017), “TH.data: TH’s Data Archive,” Available at https://cran.r-project.org/web/packages/TH.data/. [2641]
Jiang, R., Lu, W., Song, R., and Davidian, M. (2017a), “On Estimation of Optimal Treatment Regimes for Maximizing t-year Survival Probability,” Journal of the Royal Statistical Society, Series B, 79, 1165–1185. [2632,2634,2637,2640]
Jiang, R., Lu, W., Song, R., Hudgens, M. G., and Napravnik, S. (2017b), “Doubly Robust Estimation of Optimal Treatment Regimes for Survival Data with Application to an HIV/AIDS Study,” The Annals of Applied Statistics, 11, 1763–1786. [2632]
Kim, J. K., and Pollard, D. (1990), “Cube Root Asymptotics,” The Annals of Statistics, 18, 191–219. [2636]
Koenker, R. (2005), Quantile Regression, New York: Cambridge University Press. [2634]
Kosorok, M. R. (2008), Introduction to Empirical Processes and Semiparametric Inference, New York: Springer. [2636]
Kosorok, M. R., and Moodie, E. E. (2016), Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA. [2632]
Laber, E., and Zhao, Y. (2015), “Tree-Based Methods for Individualized Treatment Regimes,” Biometrika, 102, 501–514. [2632]
Lavori, P. W., and Dawson, R. (2000), “A Design for Testing Clinical Strategies: Biased Adaptive Within-Subject Randomization,” Journal of the Royal Statistical Society, Series A, 163, 29–38. [2638]
Léger, C., and MacGibbon, B. (2006), “On the Bootstrap in Cube Root Asymptotics,” Canadian Journal of Statistics, 34, 29–44. [2636]
Luedtke, A., and van der Laan, M. J. (2016), “Statistical Inference for the Mean Outcome Under a Possibly Non-unique Optimal Treatment Strategy,” The Annals of Statistics, 44, 713–742. [2633]
Matosuaka, R. A., Li, J., and Cai, T. (2014), “Evaluating Marker-Guided Treatment Selection Strategies,” Biometrics, 70, 489–499. [2636]
McLeod, A., and Xu, C. (2010), “bestglm: Best subset glm,” Available at https://cran.r-project.org/web/packages/bestglm/. [2642]
Mebane, W. R., Jr., and Sekhon, J. S. (2011), “Genetic Optimization Using Derivatives: The rgenoud package for r,” Journal of Statistical Software, 42, 1–26. [2634,2635]
Mitchell, M. (1998), An Introduction to Genetic Algorithms, Cambridge, MA: MIT press. [2635]
Moodie, E. E., and Richardson, T. S. (2010), “Estimating Optimal Dynamic Regimes: Correcting Bias Under the Null,” Scandinavian Journal of Statistics, 37, 126–146. [2632]
Murphy, S. A. (2003), “Optimal Dynamic Treatment Regimes,” Journal of the Royal Statistical Society, Series B, 65, 331–366. [2632]
