Estimating Quantile-optimal Treatment Regime with Survival Data

Lan Wang
(with Yu Zhou and Rui Song)

School of Statistics
University of Minnesota

IMA workshop on Precision Medicine, 2018
Motivating example: German GBSG2 breast cancer study

- The outcome of interest is the recurrence-free survival time in days.
- 686 patients: 56% have censored outcomes.
- **Treatments**: chemotherapy with or without the adjuvant hormonal therapy with Tamoxifen. Let $A$ denote the hormonal therapy status ($A = 0$: received; $A = 1$: did not receive).
Main objective: to introduce a robust approach for estimating an optimal treatment regimes for censored data using a quantile criterion.

Challenges for censored data:
- Mean criterion often does not serve as a suitable measure of performance.
- Mean cannot be estimated reliably when the censoring rate is high.

Other challenges:
- Estimated optimal treatment regime is sensitive to the specification of the outcome regression model.
- Theory is considerably harder.
Survival data

- $A_i$: the **treatment** (0 or 1) subject $i$ receives
- Outcome of interest $Y_i$: time to an event, e.g., death, relapse of a certain disease, re-employment or divorce.
- Vector of baseline **covariates**: $X_i \in \mathbb{X}$
- $C_i$: censoring variable, $\Delta_i = I(Y_i \leq C_i)$ is the censoring indicator
- **Observed data**: $(T_i, X_i, A_i, \Delta_i), i = 1, \ldots, n$, where
  \[T_i = \min(Y_i, C_i).\]
Potential outcomes: $Y^*_i(0)$ is the value of the outcome had subject $i$ received treatment 0, $Y^*_i(1)$ is the value of the outcome had subject $i$ received treatment 1.

Consistency assumption:

$$Y_i = Y^*_i(1)A_i + Y^*_i(0)(1 - A_i).$$

No unmeasured confounder assumption:

$$\{Y^*_i(1), Y^*_i(1)\} \perp A_i | X_i.$$

Conditional independence of censoring:

$$C_i \perp \{Y^*_i(1), Y^*_i(1), A_i\} | X_i.$$
A treatment regime is a decision rule that assigns a treatment to an individual based on his/her observed characteristics.

\[ d(X) : \mathbb{X} \rightarrow \{0, 1\}. \]

\[ Y^*(d) = Y^*(1)d + Y^*(0)(1 - d) \]

is the potential outcome that we would obtain if we use treatment regime \(d\) to assign treatment.

**Question**: Assume larger values of the outcome are preferred. Given a class of treatment regimes \(\mathcal{D}\), how to estimate the optimal treatment regime for survival outcome?
There exist a large literature for the complete outcome scenario.

- **Q-learning**: Watkins and Dayan (1992), Murphy (2005b), Chakraborty et al. (2010), Moodie and Richardson (2010), Qian and Murphy (2011), Goldberg and Kosorok (2012), Song et al. (2015), among others.

- **A-learning**: Robins et al. (2000), Murphy (2003, 2005a), Moodie and Richardson (2010); Shi et al. (2018), among others.

- **Model-free methods**: Robins and Rotnitzky (2008), Orellana and Robins (2010), Zhang et al. (2012a), Zhao et al. (2012, 2015a), Athey and Wager (2017), Linn et al. (2017); Zhou et al. (2017); Zhu et al. (2017); Wang et al. (2018), among others.

For survival outcomes,

- Goldberg and Kosorok (2012) extended Q-learning to right-censored data with a flexible number of stages using a **restricted-mean criterion**.

- Zhao et al. (2015a) extended outcome weighted learning (O-learning) to censored data using a restricted-mean criterion.

- Jiang et al. (2017a) investigated a counting-process based approach that maximizes the **t-year survival probability**; and Jiang et al. (2017b) extended this framework to estimate optimal treatment regimes that optimize a user-specified function of the survival curve.

- Comparison of a finite set of DTRs for survival outcomes (e.g. Lunceford et al., 2002; Wahed, 2009; Zhang and Schaubel, 2012).
Quantile criterion for optimal treatment regime

- **New quantile criterion:**

\[
d^{\text{opt}}(X) = \arg\max_{d \in \mathcal{D}} Q_\tau(Y^*(d)),
\]

(1)

where \(Q_\tau(Y^*) = \inf\{y : F_{Y^*}(y) \geq \tau\}, 0 < \tau < 1\)

- \(\mathcal{D} = \{d(X, \eta), \eta \in \mathcal{B}\}\) is a class of treatment regimes indexed by \(\eta\):
  - Feasibility, costs, side effects, interpretability
  - Examples: \(I(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0), I(X_1 < \eta_1, X_2 < \eta_2)\)

- We will consider

\[
\mathcal{D} = \{d(X, \eta) = I(\eta^\top X > 0) : |\eta_1| = 1, \tilde{\eta} \in \tilde{\mathcal{B}}\},
\]

where \(\tilde{\mathcal{B}}\) is a compact subset of \(\mathbb{R}^{p-1}\).
The distribution of survival time is often skewed.

Median survival time is a widely used and easily interpretable measure of success of treatment when the outcome of interest is time to event. Marginal median and lower quantiles can usually still be accurately estimated even if censoring is heavy.

In some applications, the lower tail of the survival time is of direct importance.

Robust estimation: a slightly misspecified outcome model can result in biased estimation of the optimal treatment regime.

Challenge of statistical theory: Little existing work on the statistical theory for the estimated optimal treatment regime with survival outcome. Non-standard asymptotics.
Robust estimation of quantile-optimal treatment regime with survival outcome

- $C_i$: censoring variable, $\Delta_i = I(Y_i \leq C_i)$ is the censoring indicator

- **Observed data**: $(T_i, X_i, A_i, \Delta_i), i = 1, \ldots, n$, where $T_i = \min(Y_i, C_i)$.

- **Induced missing data framework**: for fixed $\eta$, let

$$R_i(\eta) = [A_i d(X_i, \eta) + (1 - A_i) (1 - d(X_i, \eta))] \Delta_i.$$ 

- If $R_i(\eta) = 1$, then $Y_i = Y_i^*(d_\eta)$, then $Y_i^*(d_\eta)$ is observed.

- **Full data**: $\{Y^*(d_\eta), X\}$;  
  **Observed data**: $\{R_\eta, R_\eta Y^*(d_\eta), X\}$. 

Robust estimation of quantile-optimal treatment regime with survival outcome (cont’d)

- Let $\pi_i(\eta) = P[R_i(\eta) = 1 \mid X_i, T_i^*(1), T_i^*(0)]$
  $\pi_A(X_i) = P(A_i = 1 \mid X_i)$. We have

  $\pi_i(\eta) = (\pi_A(X_i)d(x_i, \eta) + (1 - \pi_A(X_i))(1 - d(x_i, \eta)))$
  $\times G_C(T_i^*(d(x_i, \eta)) \mid X, A_i = d(x_i, \eta)).$

- For the complete cases (corresponding to $R_i(\eta) = 1$),

  $\pi_i(\eta) = (\pi_A(X_i)d(x_i, \eta) + (1 - \pi_A(X_i))(1 - d(x_i, \eta)))G_C(Y_i \mid X, A_i)$.

- We can estimate $Q_\tau(T^*(d_\eta))$ by

  $$\hat{Q}_\tau(T^*(d_\eta)) = \arg \min_b \sum_{i=1}^n \frac{R_i(\eta)}{\hat{\pi}_i} \rho_\tau(Y_i - b).$$
Quantile-optimal treatment regime with survival outcome (cont’d)

- **Independent censoring**: Let $\hat{G}(\cdot)$ be the estimator of the survival function of the censoring time, obtained by applying the classical Kaplan-Meier estimator.

- **Covariate-dependent censoring**: The local Kaplan-Meier estimator of $G_C(\cdot | \mathbf{X}, A = 0)$ is given by

$$
\hat{G}_C(\cdot | \mathbf{X}, A = 0) = \prod_{j=1}^{n_1} \left\{ 1 - \frac{B_{n_1j}(\mathbf{X})}{\sum_{k=1}^{n_1} I(C_k \geq C_j) B_{n_1k}(\mathbf{X})} \right\}^{\eta_j(t)},
$$

where $\eta_j(t) = I(C_j \leq t, \Delta_j = 0)$, and $\{ B_{n_1k}(\mathbf{X}), k = 1, ..., n_1 \}$ is a sequence of non-negative weights adding up to 1.
Quantile-optimal treatment regime with survival outcome (cont’d)

Lemma

Under some regularity conditions, for all $d_η \in D$, 

$$\hat{Q}_T(η, \hat{G}_C) \xrightarrow{P} Q_T\{ T^*(d_η) \}.$$ 

- We estimate $η_0$, the parameter indexing the quantile-optimal treatment regime in $D$, by 
  $$\hat{η}_n = \arg \max_{\beta: |β_1| = 1, \tilde{η} \in \tilde{B}} \hat{Q}_T(η; \hat{G}_C).$$

Thus, the estimate of quantile-optimal treatment regime is 

$$d_{\hat{η}_n}(X) = I(\hat{η}_n^T X > 0).$$

Reformulation of the estimator:

\[
g(\cdot, \eta, m, G) = \frac{R(\eta)}{0.5 G_C(Y)} I(Y - m > 0),
\]

\[
m_0 = \sup \{ m : \sup_{\eta : |\eta_1| = 1, \tilde{\eta} \in \tilde{B}} Pg(\cdot, \eta, m, G_C) \geq 1 - \tau \},
\]

\[
\eta_0 = \arg \max_{\eta : |\eta_1| = 1, \tilde{\eta} \in \tilde{B}} Pg(\cdot, \eta, m_0, G_C).
\]

Intuition: For a randomized trial, for any given \( m \),
\[ P(g(\cdot, \eta, m)) = 0.5 P(Y^*(d_\eta) > m), \] which is equal to
\[ (1 - \tau)/2 \] if \( m = Q_\tau(Y^*(d_\eta)) \). For any given \( \eta \), \( g(\cdot, \eta, m) \) is monotonically decreasing in \( m \). As the result, the estimator of \( Q_\tau(Y^*(d_\eta)) \) is the largest value of \( m \) that \( P_n g(\cdot, \eta, m) \) is greater than or equal to \( (1 - \tau)/2 \).
An alternative representation of $\hat{\eta}_n$ is given by

$$\hat{\eta}_n = \arg \max_{\eta:|\eta_1|=1, \tilde{\eta} \in \tilde{B}} n^{-1} \sum_{i=1}^{n} g(\cdot, \eta, \hat{m}_n, \hat{G}_C),$$

where

$$\hat{m}_n = \sup \left\{ m : \sup_{\eta:|\eta_1|=1, \tilde{\eta} \in \tilde{B}} n^{-1} \sum_{i=1}^{n} g(\cdot, \eta, m, \hat{G}_C) \geq 1 - \tau \right\}.$$

Theoretical challenge: Nonstandard M-estimation problem with both a finite dimensional nuisance parameter $m_0$ and an infinite dimensional nuisance parameter $G_C(\cdot)$.

We can still prove cube-root asymptotics using advanced empirical process theory.
Theorem

Under some regularity conditions, \( n^{1/3} (\tilde{\eta}_n - \tilde{\eta}_0) \) converges to the maximizer of the process \( t \mapsto \Psi(t) + \mathbb{W}(t) \), where \( \Psi(t) \) is a deterministic function and \( \mathbb{W}(t) \) is a mean-zero Gaussian process.
Inference using $m$-out-of-$n$ bootstrap

- The standard nonparametric bootstrap procedure is generally inconsistent for cuberoot M-estimators (e.g., Abrevaya and Huang (2005)).
- In a different setting, Chakraborty et al. (2013) investigated $m$-out-of-$n$ bootstrap for Q-learning inference.
- Let the original data be $W_n = \{(A_i, X, Y_i, \Delta_i) : i = 1, 2, \ldots, n\}$. Denote the subsample of size $m$ (with replacement, $m < n$) by $W_m^o = \{(A_i^o, X_i^o, Y_i^o, \Delta_i^o) : i = 1, 2, \ldots, m\}$. Let $\tilde{\eta}_m^{(b)}$ be the estimators calculated from independent subsamples, $b = 1, \ldots, M$.
- The $(1 - \alpha)\%$ confidence interval for $c^T \tilde{\eta}_0$ is:
  \[
  \left( c^T \tilde{\eta}_n - m^{-1/3} \hat{u}, \ c^T \tilde{\eta}_n - m^{-1/3} \hat{l} \right),
  \]
  where $\hat{l}$ and $\hat{u}$ are the $\alpha/2$ and $1 - \alpha/2$ quantiles of $m^{1/3} \left( c^T \tilde{\eta}_m^{(b)} - c^T \tilde{\eta}_n \right)$ from the $B$ subsamples, respectively.
A Monte Carlo example

We generate the random sample \( \{X_i, A_i, Y_i, \Delta_i\}, \ i = 1, 2, \ldots, n \), from the model:

\[ X_1 \sim U(0, 1), \]
\[ T^*(0) \mid X_1 \sim \text{Weibull}(\text{shape} = 1, \text{scale} = 1) + 1, \]
\[ T^*(1) \mid X_1 \sim \text{Weibull}(\text{shape} = 3, \text{scale} = 0.5 + X_1) + 2X_1, \]
\[ A \mid \{X, T^*(0), T^*(1)\} \sim \text{Bernoulli}(0.5). \]

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\%. 
Figure: Histograms of $T^*(0)$ and $T^*(1)$ stratified by $X_1$
A Monte Carlo example (cont’d)

We consider estimating the quantile-optimal treatment regime in the class of treatment regimes
\[ \mathcal{D} = \{ I(X_1 \beta_1 + \beta_2 > 0) : |\beta_1| = 1, \beta_2 \in \mathbb{R} \} . \]

Table: Parameters indexing the quantile-optimal treatment regimes \((\tau = 0.25 \text{ and } 0.5)\) and the maximally achievable \(\tau\)th quantile of the potential outcome (denoted by \(Q_\tau\)) in \(\mathcal{D}\), based on a Monte Carlo experiment \((n = 10^7)\).

<table>
<thead>
<tr>
<th>(\tau)</th>
<th>(\eta_{01}^{(\tau)})</th>
<th>(\eta_{02}^{(\tau)})</th>
<th>(Q_{0.25})</th>
<th>(Q_{0.5})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>1</td>
<td>-0.428</td>
<td>1.658</td>
<td>2.215</td>
</tr>
<tr>
<td>0.50</td>
<td>1</td>
<td>-0.552</td>
<td>1.587</td>
<td>2.258</td>
</tr>
</tbody>
</table>

The overall censoring rate is 35%.
Table: Bias (with standard deviation in the parenthesis) of New and Naive for estimating $\eta_0^{(\tau)}$ and $Q_\tau$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$n$</th>
<th>New $\eta_0^{(\tau)}$</th>
<th>New $Q_\tau$</th>
<th>Naive $\eta_0^{(\tau)}$</th>
<th>Naive $Q_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>300</td>
<td>0.005(0.066)</td>
<td>0.056(0.113)</td>
<td>-0.025(0.204)</td>
<td>-0.555(0.057)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.001(0.054)</td>
<td>0.027(0.082)</td>
<td>-0.043(0.213)</td>
<td>-0.568(0.050)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.001(0.043)</td>
<td>0.020(0.055)</td>
<td>-0.048(0.202)</td>
<td>-0.585(0.031)</td>
</tr>
<tr>
<td>0.50</td>
<td>300</td>
<td>0.002(0.098)</td>
<td>0.048(0.124)</td>
<td>0.129(0.082)</td>
<td>-0.618(0.078)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.003(0.080)</td>
<td>0.023(0.101)</td>
<td>0.122(0.064)</td>
<td>-0.617(0.065)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.002(0.051)</td>
<td>0.022(0.061)</td>
<td>0.134(0.051)</td>
<td>-0.649(0.049)</td>
</tr>
</tbody>
</table>
CI from $m$-out-of-$n$ bootstrap

- We obtain $B = 400$ bootstrap samples with subsample size $m = 150$ to construct 90% confidence intervals for $\beta_{02}^{(\tau)}$. We did 200 replications.

- For $\tau = 0.25$, the confidence intervals based on New have an associated average coverage probability of 88%, with an average length 0.175.

- For $\tau = 0.5$, the associated empirical confidence intervals have an empirical coverage probability of 91%, with an average length 0.221.
Analysis of GBSG2 study data

- GBSG2 study conducted by the German Breast Cancer Study Group.
- The outcome of interest is the recurrence-free survival time in days.
- 686 patients: 56% have censored outcomes.
- **Treatments**: chemotherapy with or without the adjuvant hormonal therapy with Tamoxifen. Let $A$ denote the hormonal therapy status ($A = 0$: received; $A = 1$: did not receive).
Analysis of GBSG2 study data

We consider treatment regimes 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.

2. **PR**: Progesterone receptor expression is routinely measured for the management of breast cancer as a positive prognostic factor. However, its predictive power for the efficacy of Tamoxifen is still not well understood.

3. **Age**: Age is an important risk factor in breast cancer. It is not clear whether it has qualitative interaction with the hormonal therapy.
\( \mathcal{D}_1 = \{ I(\eta_1\text{LER} + \eta_2 + \eta_3\text{LPR} + \eta_4\text{NAGE} > 0) : \eta_1 = 1 \} \).

\( \mathcal{D}_2 = \{ I(\eta_1\text{LER} + \eta_2 + \eta_3\text{LPR} > 0) : \eta_1 = 1 \} \).

**Table:** Estimated parameters indexing the quartile-optimal treatment regime and 90% \( m \)-out-of-\( n \) bootstrap confidence intervals for the GBSG2 study

<table>
<thead>
<tr>
<th>Regimes</th>
<th>( \eta_2 )</th>
<th>( \eta_3 )</th>
<th>( \eta_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{D}_1 )</td>
<td>-1.597</td>
<td>1.102</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>(-3.361, -0.757)</td>
<td>(0.512, 2.498)</td>
<td>(-1.256, 1.664)</td>
</tr>
<tr>
<td>( \mathcal{D}_2 )</td>
<td>-1.549</td>
<td>1.100</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>(-2.516, -0.569)</td>
<td>(0.493, 1.907)</td>
<td>/</td>
</tr>
</tbody>
</table>

The estimated optimal rule leads to an estimated quartile survival time of \textbf{1246} days with approximately 81.0% of patients being recommended to treatment. The Kaplan-Meier estimator of the first quartile of the observed survival time is \textbf{727} days.
**Dynamic (sequential) treatment regimes**

- **A dynamic treatment regime** is a list of sequential decision rules to assign treatments based on an individual’s characteristics and history.
- **Example**: Adaptive pharmacological and behavioral treatments for children with ADHD Trial (W. Pelham (PI); Nahum-Shani et al., 2012a; Lei et al., 2012; Laber et al., 2014).
Two-stage dynamic decision problem: complete data

- Observe \((X_{i1}, A_{i1}, X_{i2}, A_{i2}, T_i), i = 1, \ldots, n\)
  - \(X_{i1}\) is baseline vector of covariates for subject \(i\)
  - \(A_{i1}\) is the treatment subject \(i\) receives at stage 1
  - \(X_{i2}\) is intermediate information observed between the two stages
  - \(A_{i2}\) is the treatment subject \(i\) receives at stage 2
  - \(T_i\) is the observed outcome for subject \(i\) (assume larger is better)

- **Two-stage sequence of treatment regimes:** \(\bar{d} = (d_1, d_2) \in D = D_1 \times D_2\).

- Potential outcomes: \(X_2^*(d_1), T^*(\bar{d})\).
  - \(X_2^*(d_1)\) is the intermediate covariate information before decisions 2 had the subject received treatment \(d_1\);
  - \(T^*(\bar{d})\) is the outcome had the subject received treatment \(\bar{d}\).

- **Quantile criterion:**
  \[
  \bar{d}^{\text{opt}}(X) = \arg\max_{\bar{d} \in D} Q_\tau(T^*(\bar{d})).
  \]
We focus on a two-stages study in which a subject receives a treatment $A_1 \in \{0, 1\}$ at stage one and receives a treatment $A_2 \in \{0, 1\}$ at stage two.

We assume that treatment $A_1$ is assigned at baseline (time point 0), and treatment $A_2$ is assigned at an intermediate time point $s$.

Let $X_1$ denote the vector of baseline covariates, and let $X_2$ denote the intermediate covariates collected at time point $s$ just before $A_2$ starts.
Let $\hat{T}^*(d_1, \emptyset)$ is the potential survival time if the subject only follows treatment $d_1$.

$\hat{T}^*(d_1, d_2)$ is the potential survival time if the subject receives the full sequence of treatments $(d_1, d_2)$. Implicitly, $T^*(d_1, d_2) > s$.

Define $Z^C = \mathbb{I}(\min(A_1 \hat{T}^*(1, \emptyset) + (1 - A_1) \hat{T}^*(0, \emptyset), C) > s)$.

The observed data consist of

$\{X_1, A_1, Z^C, Z^C X_2, Z^C A_2, Y, \Delta\}$ where $Y = \min\{ T, C \}$, and

$$T = \begin{cases} \sum_{a_1 \in \{0,1\}} \mathbb{I}(A_1 = a_1) T^*(a_1, \emptyset) & \text{if } Z^C = 0 \\ \sum_{(a_1, a_2) : a_i \in \{0,1\}, i=1,2} \mathbb{I}(A_1 = a_1, A_2 = a_2) T^*(a_1, a_2) & \text{if } Z^C = 1 \end{cases}$$
Two-stage dynamic decision problem: censored data

- Given a two-stage treatment regime $d = (d_1, d_2)$, the potential survival time of a patient following this sequence of treatment regimes is

$$T^*(d) = \dot{T}^*(d_1, \emptyset)I(\dot{T}^*(d_1, \emptyset) \leq s) + \dot{T}^*(d_1, d_2)I(\dot{T}^*(d_1, \emptyset) > s),$$

where $\dot{T}^*(d_1, \emptyset) = d_1(X_1)\dot{T}^*(1, \emptyset) + (1 - d_1(X_1))\dot{T}^*(0, \emptyset)$.

- Define

$$R_i^{(2)}(d) = \Delta_i I(A_{i1} = d_1(X_{i1}))\left[ I(Y_i \leq s) + I(Y_i > s)I\{A_{i2} = d_2(X_{i1}, A_{i1}, X_{i2})\} \right].$$

For those subjects with $R_i^{(2)}(d) = 1$, the observed response is the corresponding potential survival time $T^*(d)$. 
We consider a sequential multiple assignment randomized trial (SMART). Assume at stage one,
\[ P(A_1 = 1) = 1 - P(A_1 = 0) = \pi_1, \]
while at stage two
\[ P(A_2 = 1 | Y_i > s) = 1 - P(A_2 = 0 | Y_i > s) = \pi_2. \]

Let \( w_{d,i}^{(2)} = P(R_i^{(2)}(d) = 1 | O_i). \) Then it can be shown that for those subjects with \( R_i^{(2)}(d) = 1, \)
\[ w_{d,i}^{(2)} = \pi_{A_1}(X_{i1}) G_C(Y_i) \left\{ I(Y_i \leq s) + \pi_{A_2}(X_{i1}, A_{i1}, X_{i2}) I(Y_i > s) \right\}. \]

We estimate the \( Q_T \{ T^*(d_\eta) \} \) by
\[
\hat{Q}_T \{ T^*(d) \} = \arg \min_b \sum_{i=1}^n \frac{R_i^{(2)}(d) \rho_T(Y_i - b)}{\hat{w}_{d,i}^{(2)}}.
\]

The estimator of the parameter indexing the quantile-optimal DTR in the class \( \mathcal{D} \) is
\[
\hat{\eta} = \arg \max_{\eta: d_\eta \in \mathcal{D}} \hat{Q}_T(T^*(d_\eta)).
\]
Conclusions and acknowledgments

- Quantile criterion for optimal treatment regime estimation is useful for survival data.
- Robust estimation of quantile-optimal treatment regimes for both one-stage and dynamic decision problems with censored responses.
- Nonstandard asymptotic statistical theory.
- R package: on the way.