Quantile-optimal Treatment Regimes

Lan Wang

School of Statistics
University of Minnesota
(Joint work with Yu Zhou, Rui Song and Ben Sherwood)

IMA Workshop on Innovative Statistics and Machine Learning for Precision Medicine (Sep 14-16, 2017)
Motivations of the proposed work

- In a variety of applications, criteria other than the mean (or the average) may be more sensible.

**Examples:**
- When the outcome has a skewed distribution (e.g., survival time of patients), it may be desirable to consider the treatment regime that maximizes the median of the distribution of the potential outcome.
- Sometimes, the tail of the potential outcome distribution is of direct importance.
- Optimizing expectations is especially difficult for various financial outcomes including equities, commodities, and electricity markets; for example, daily volatilities of 20-30% are common in electricity markets.
A toy example

- \( Y_i = 1 + 3A_i + X_i - 5A_iX_i + (1 + A_i + 2A_iX_i)\epsilon_i, \ X_i \sim U[0, 1], \ \epsilon_i \sim N(0, 1) \). The treatment is randomly assigned with probability 0.5.

- The optimal mean treatment regime is to assign the treatment when \( X \leq 3/5 \).

- We consider the following 6 treatment regimes:
  1. Control assigned to everyone, \( A_i = 0, \ \forall \ i \),
  2. \( A_i = I(X_i \leq 3/5) \),
  3. \( A_i = I(X_i \leq 1/2) \),
  4. \( A_i = I(X_i \leq 1/5) \),
  5. \( A_i = I(X_i \leq 1/10) \),
  6. Treatment assigned to everyone, \( A_i = 1, \ \forall \ i \).
A toy example (cont’d)

Table: An Monte Carlo experiment with $10^6$ observations

<table>
<thead>
<tr>
<th>$d$</th>
<th>mean</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.50</td>
<td>0.80</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>2.40</td>
<td>1.10</td>
<td>-0.03</td>
</tr>
<tr>
<td>3</td>
<td>2.37</td>
<td>1.14</td>
<td>0.20</td>
</tr>
<tr>
<td>4</td>
<td>2.00</td>
<td>1.01</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>1.78</td>
<td>0.91</td>
<td>0.26</td>
</tr>
<tr>
<td>6</td>
<td>2.00</td>
<td>-0.02</td>
<td>-2.29</td>
</tr>
</tbody>
</table>
Motivations of the proposed work (cont’d)

- **Difficulty of specifying a reliable outcome regression model**: a slightly misspecified outcome model can result in biased estimation of the optimal treatment regime.

- **Challenge of statistical theory**: little existing work on asymptotic theory for the estimated optimal treatment regime.

- **We propose**: a new quantile criterion; a robust estimation method that does not rely on the outcome regression model; asymptotic theory on the estimated optimal treatment regime (cubic-root asymptotics, non-normal limiting distribution); extensions to alternative criteria.
Existing work have been focused on the **mean criterion**:

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

- **A-learning**: Robins et al. (2000), Murphy (2003, 2005a), among others.


Potentially, the recent work on **discrete Q-learning** in Moodie et al. (2014) can be applied to first estimate the probabilities and then invert them to estimate quantiles, but this application has not been systematically studied.

Linn, Laber and Stefanski (2016) independently considered estimating quantile-optimal treatment regime. However, their approach depends on applying threshold **interactive model-based Q-learning** at a sequence of thresholding values and then performing inversion.

In the causal inference context, several authors have considered estimating quantile treatment effects for **comparing several pre-determined treatment regimes**: Rubin (1974), Rosenbaum and Rubin (1983), Abadie, Angrist and Imbens (2002), Bitler et al. (2002), Chernozhukov and Hansen (2005), Firpo (2006), Chen et al. (2012), among others. But they have not investigated the fundamental problem of estimating the optimal treatment regimes.
One-stage (static) decision problem

- $A_i$: the **treatment** (0 or 1) subject $i$ receives

**Potential outcomes (counterfactual outcomes)**
- $Y_i^*(0)$: the value of the outcome had subject $i$ received treatment 0
- $Y_i^*(1)$: the value of the outcome had subject $i$ received treatment 1

**Observed outcome**: $Y_i = Y_i^*(1)A_i + Y_i^*(0)(1 - A_i)$.

- **Vector of baseline covariates**: $X_i \in \mathbb{X}$

**Observed data**: $(Y_i, X_i, A_i)$, $i = 1, \ldots, n$.

**Unconfoundedness assumption**: $\{Y_i^*(1), Y_i^*(1)\} \perp A_i | X_i$. 
A **treatment regime** is a decision rule: $d(X) : \mathbb{X} \rightarrow \{0, 1\}$

Given a class of decision rules $\mathbb{D}$, we wish to find the optimal one. Given $d \in \mathbb{D}$,

$$Y^*(d) = Y^*(1)d + Y^*(0)(1 - d)$$

is the **potential outcome** that we would obtain if we assign treatment according to $d$.

Assume a larger value of the outcome is preferred.

**Quantile criterion**:  

$$d^{\text{opt}}(X) = \arg\max_{d \in \mathbb{D}} Q_\tau(Y^*(d)),$$  \hspace{1cm} (1)

where $Q_\tau(Y^*) = \inf\{y : F_{Y^*}(y) \geq \tau\}, \ 0 < \tau < 1$
Estimation of the optimal treatment regime

- We consider a class of decision rules index by $\eta$: $D = \{d(X, \eta), \eta \in \mathcal{B}\}$.
  - Feasibility, costs, side effects, interpretability
  - Examples: $I(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0)$, $I(X_1 < \eta_1, X_2 < \eta_2)$

- Motivated by Zhang et al. (2012), we consider an **induced missingness structure**: For a fixed $\eta$, let

  $$C_{\eta} = Ad(X, \eta) + (1 - A)(1 - d(X, \eta)).$$

  If $C_{\eta} = 1$, then $Y^*(d_{\eta})$ is observed; if $C_{\eta} = 0$ then $Y^*(d_{\eta})$ is “missing”. The “full data” are $\{Y^*(d_{\eta}), X\}$ and the observed data are $\{C_{\eta}, C_{\eta} Y^*(d_{\eta}), X\} = \{C_{\eta}, C_{\eta} Y, X\}$.

- We can show **MAR** assumption is satisfied: $Y^*(d_{\eta}) \perp C_{\eta} | X$. 
Let $\pi(X) = P(A = 1|X)$, the propensity score is:

$$P(C_\eta = 1 | X) = \pi_c(X, \eta) = \pi(X)d(X, \eta) + (1-\pi(X))(1-d(X, \eta)).$$

**How to estimate $\pi(X)$?**

- Randomized clinic trial: constant
- Popular approach: logistic regression,
  $$\pi(X, \gamma) = \frac{\exp(X^T\gamma)}{1 + \exp(X^T\gamma)}$$
- Semiparametric and nonparametric approaches: index models, additive models, etc
The estimate of the $\tau$th quantile of $Y^*(d_\eta)$ is:

$$\hat{Q}_\tau(\eta) = \arg\min_a n^{-1} \sum_{i=1}^n \frac{C_{\eta,i}}{\hat{\pi}_c(X_i, \eta)} \rho_\tau(Y_i - a),$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function.

The estimator of $\eta$ corresponding to the optimal treatment regime from the class $\mathbb{D}$ is defined as

$$\hat{\eta}_n = \arg\max_{\eta \in \mathbb{B}} \hat{Q}_\tau(\eta). \quad (2)$$

The estimated optimal treatment regime is $d_{\hat{\eta}_n} = d(X, \hat{\eta}_n)$.

**Computation:** genetic algorithm (R package *rgenoud*)
Consider the class of candidate treatment regimes, $d(X, \eta) = I\{X^T \eta > 0\}$, $\eta \in \mathbb{B}$.

**Estimating the value function of the treatment regime:**

$$\sup_{\eta \in \mathbb{B}} |\hat{Q}_\tau(\eta) - Q_\tau(Y^*(d_\eta))| \to 0.$$
Recall: \( \hat{\eta}_n = \operatorname{argmax}_{\eta \in B} \hat{Q}_\tau(\eta) \), where
\[
\hat{Q}_\tau(\eta) = \operatorname{argmin}_a n^{-1} \sum_{i=1}^{n} \frac{C_{\eta,i}}{\hat{\pi}_c(X_i,\eta)} \rho_\tau(Y_i - a).
\]

**A new alternative formulation:**

\[
g(\cdot, \eta, m) = C(\eta) I\{Y - m > 0\},
\]
\[
m_0 = \sup_{\eta \in B} \{ m : \sup_{\eta \in B} P g(\cdot, \eta, m) \geq (1 - \tau)/2 \}
\]
\[
\eta_0 = \operatorname{argmax}_{\eta \in B} P g(\cdot, \eta, m_0).
\]

**Intuition:** For a randomized trial, for any given \( m \),
\[
P(g(\cdot, \eta, m)) = \frac{1}{2} P(Y^*(d_\eta) > m),
\]
which is equal to \( \frac{1-\tau}{2} \) 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 \( P_{ng}(\cdot, \eta, m) \) is greater than or equal to \( \frac{1-\tau}{2} \).
Asymptotic theory (cont’d)

- $m_0$ is the largest achievable $\tau$th quantile of $Y^*(d_\eta)$ over $\eta \in B$; and $\eta_0$ is the population value of the parameter that indexes the optimal treatment regime.
- $\hat{m}_n = \sup\{ m : \sup_{\eta \in B} P_n g(\cdot, \eta, m) \geq (1 - \tau)/2 \}$, estimator of the maximally achievable value function.
- $\hat{\eta}_n = \arg\max_{\eta \in B} P_n g(\cdot, \eta, \hat{m}_n)$.

Lemma

**Under some regularity conditions,**

(1) $\hat{m}_n = m_0 + O_p(n^{-1/2})$.
(2) $P_n g(\cdot, \hat{\eta}_n, m_0) \geq \sup_{\eta \in B} P_n g(\cdot, \eta, m_0) - o_p(n^{-2/3})$. 
Asymptotic theory (cont’d)

- **Asymptotic normality**

  \[ n^{1/3}(\hat{\eta}_n - \eta_0) \to \arg \max_t Z(t), \]

  where the process \( Z(t) = -\frac{1}{2} t^T V t + W(t) \), \( V \) is an \( l \times l \) positive definite matrix and \( W(t) \) is a centered Gaussian process with continuous sample paths.
Remark: If the mean-optimal criterion is of interest, then we let

\[ g^*(\cdot, \eta, \mu) = C(\eta)(Y - \mu), \]
\[ \hat{\mu}_n = \sup_{\eta \in B} \{ \mu : \sup_{\eta \in B} P_n g^*(\cdot, \eta, \mu) > 0 \}. \]

The estimated parameter indexing the mean-optimal treatment regime has the representation

\[ \hat{\eta}_n^{\text{mean}} = \arg\max_{\eta \in B} P_n g^*(\cdot, \eta, \hat{\mu}_n). \]
Extension to optimal treatment regimes with respect to alternative criteria

- Given a treatment regime \( d_\beta \), let
  \[
  G(Y^*(d_\beta)) = -E|Y_1^*(d_\beta) - Y_2^*(d_\beta)|,
  \]
  where \( Y_1^*(d_\beta) \) and \( Y_2^*(d_\beta) \) are independent copies of the potential outcome \( Y^*(d_\beta) \). The optimal treatment regime that minimizes Gini’s mean difference is defined as \( \arg \max_{d \in \mathbb{D}} G(Y^*(d_\beta)) \).

- For a randomized study, we can consistently estimate \( G(Y^*(d_\beta)) \) using the following second-order U-statistic
  \[
  \hat{G}(\beta) = -\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} 4C_i(\beta)C_j(\beta)|Y_i - Y_j|.
  \]

  The estimated parameter indexing the optimal treatment regime is \( \hat{\beta}_n = \arg \max_{\beta \in \mathbb{B}} \hat{G}(\beta) \).
Extension to optimal treatment regimes with respect to alternative criteria

Let \( g(Z_i, Z_j, \xi, m) = 4C_i(\beta)C_j(\beta)(-|Y_i - Y_j| - m), \)
\( \hat{m}_n = \sup\{ m : \sup_{\beta \in B} U_ng(\cdot, \cdot, \xi, m) \geq 0 \}, \)
\( m_0 = \sup\{ m : \sup_{\beta \in B} P_g(\cdot, \cdot, \xi, m) \geq 0 \}, \)
\( \beta_0 = \arg\max_{\beta \in B} P_g(\cdot, \cdot, \xi, m_0), \) where \( Z_i = \{X_i, Y_i\}, \)
\( U_n g(\cdot, \cdot, \beta, m) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} g(Z_i, Z_j, \xi, m). \)

We have

\( \hat{\beta}_n = \arg\max_{\beta \in B} U_n g(\cdot, \cdot, \beta, \hat{m}_n). \)
A numerical example

- Random data are generated from

\[
Y = 1 + X_1 - X_2 + X_3^3 + \exp(X_4) \\
+ A(3 - 5X_1 + 2X_2 - 3X_3 + X_4) \\
+(1 + A(1 + X_1 + X_2 + X_3 + X_4))\epsilon,
\]

where \( X_i \sim U[0, 1] \), \( \epsilon \sim N(0, 1) \).

- \[
\log \left( \frac{P(A = 1|X)}{P(A_i = 0|X)} \right) = -0.5 - 0.5(X_1 + X_2 + X_3 + X_4),
\]
  where \( X = (X_1, \ldots, X_4)' \).

- The optimal treatment regime according to the mean criterion is \( d(X) = I(3 - 5X_1 + 2X_2 - 3X_3 + X_4 > 0) \).

- We consider the class of treatment regimes \( I(\eta_0 + \eta^T X > 0) \), where \((\eta_0, \eta_1, \ldots, \eta_4)^T\) has \( L_2 \)-norm 1.
For the model-based approach we impose models for \( \mu(a, X) \) and then estimate the mean-optimal treatment regime by \( I(\hat{\mu}(1, X) > \hat{\mu}(0, X)) \), where \( \hat{\mu} \) is the estimated value of \( \mu \). We consider two posited models for \( \mu(a, X) \): correctly specified regression function and misspecified regression function.

For the model-free approach, we consider the estimator in Zhang et al. (2012a).

We denote these three estimators by mean\(_{RG\mu_t}\), mean\(_{RG\mu_m}\) and mean\(_{ZTLD}\), respectively.
Table: Population parameters and summary values for optimal treatment regimes under different criteria for Example 1 based on a Monte Carlo experiment with \( n = 10^5 \).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>( \eta_0 )</th>
<th>( \eta_1 )</th>
<th>( \eta_2 )</th>
<th>( \eta_3 )</th>
<th>( \eta_4 )</th>
<th>( Q_{mean} )</th>
<th>( Q_{0.25} )</th>
<th>( Q_{0.1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean criterion</td>
<td>0.43</td>
<td>-0.72</td>
<td>0.29</td>
<td>-0.43</td>
<td>0.14</td>
<td>3.99</td>
<td>2.28</td>
<td>0.55</td>
</tr>
<tr>
<td>0.25qt criterion</td>
<td>0.42</td>
<td>-0.60</td>
<td>0.41</td>
<td>-0.43</td>
<td>-0.34</td>
<td>3.79</td>
<td>2.46</td>
<td>1.18</td>
</tr>
<tr>
<td>0.1qt criterion</td>
<td>0.27</td>
<td>-0.68</td>
<td>0.38</td>
<td>-0.43</td>
<td>-0.37</td>
<td>3.44</td>
<td>2.36</td>
<td>1.55</td>
</tr>
</tbody>
</table>
### A numerical example (cont’d)

<table>
<thead>
<tr>
<th>Method</th>
<th>n</th>
<th>$\hat{\eta}_0$</th>
<th>$\hat{\eta}_1$</th>
<th>$\hat{\eta}_2$</th>
<th>$\hat{\eta}_3$</th>
<th>$\hat{\eta}_4$</th>
<th>$\bar{Q}_{\text{mean}}$</th>
<th>$\bar{Q}_{0.25}$</th>
<th>$\bar{Q}_{0.1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean$<em>{RG</em>{\mu_t}}$</td>
<td>500</td>
<td>0.42</td>
<td>-0.71</td>
<td>0.28</td>
<td>-0.41</td>
<td>0.14</td>
<td>3.99</td>
<td>2.29</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.07)</td>
<td>(0.13)</td>
<td>(0.11)</td>
<td>(0.12)</td>
<td>(0.21)</td>
<td>(0.19)</td>
<td>(0.40)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.43</td>
<td>-0.71</td>
<td>0.29</td>
<td>-0.43</td>
<td>0.14</td>
<td>3.99</td>
<td>2.28</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.09)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td>(0.14)</td>
<td>(0.13)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>mean$<em>{RG</em>{\mu_m}}$</td>
<td>500</td>
<td>0.26</td>
<td>-0.71</td>
<td>0.30</td>
<td>-0.38</td>
<td>0.37</td>
<td>3.96</td>
<td>2.23</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.11)</td>
<td>(0.08)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.21)</td>
<td>(0.19)</td>
<td>(0.38)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.27</td>
<td>-0.71</td>
<td>0.31</td>
<td>-0.39</td>
<td>0.37</td>
<td>3.97</td>
<td>2.22</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.09)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td>(0.15)</td>
<td>(0.13)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>mean$_{ZTLD}$</td>
<td>500</td>
<td>0.36</td>
<td>-0.63</td>
<td>0.31</td>
<td>-0.38</td>
<td>0.12</td>
<td>4.31</td>
<td>2.31</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2)</td>
<td>(0.14)</td>
<td>(0.24)</td>
<td>(0.2)</td>
<td>(0.27)</td>
<td>(0.21)</td>
<td>(0.21)</td>
<td>(0.53)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.38</td>
<td>-0.67</td>
<td>0.29</td>
<td>-0.4</td>
<td>0.17</td>
<td>4.18</td>
<td>2.29</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.15)</td>
<td>(0.11)</td>
<td>(0.19)</td>
<td>(0.15)</td>
<td>(0.19)</td>
<td>(0.13)</td>
<td>(0.16)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>0.25qt criterion</td>
<td>500</td>
<td>0.38</td>
<td>-0.57</td>
<td>0.37</td>
<td>-0.37</td>
<td>-0.31</td>
<td>3.85</td>
<td>2.65</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.19)</td>
<td>(0.18)</td>
<td>(0.2)</td>
<td>(0.26)</td>
<td>(0.16)</td>
<td>(0.39)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.4</td>
<td>-0.59</td>
<td>0.35</td>
<td>-0.43</td>
<td>-0.28</td>
<td>3.81</td>
<td>2.57</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.17)</td>
<td>(0.12)</td>
<td>(0.15)</td>
<td>(0.18)</td>
<td>(0.11)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>0.10qt criterion</td>
<td>500</td>
<td>0.24</td>
<td>-0.56</td>
<td>0.3</td>
<td>-0.4</td>
<td>-0.33</td>
<td>3.5</td>
<td>2.45</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.23)</td>
<td>(0.2)</td>
<td>(0.25)</td>
<td>(0.22)</td>
<td>(0.25)</td>
<td>(0.26)</td>
<td>(0.16)</td>
<td>(0.15)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.27</td>
<td>-0.61</td>
<td>0.32</td>
<td>-0.44</td>
<td>-0.33</td>
<td>3.47</td>
<td>2.42</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.18)</td>
<td>(0.14)</td>
<td>(0.22)</td>
<td>(0.15)</td>
<td>(0.19)</td>
<td>(0.18)</td>
<td>(0.11)</td>
<td>(0.11)</td>
</tr>
</tbody>
</table>
### Theoretically optimal regime

<table>
<thead>
<tr>
<th>Estimated optimal regime</th>
<th>n</th>
<th>mean criterion</th>
<th>0.25qt criterion</th>
<th>0.10qt criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean criterion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.84(0.06)</td>
<td>0.73(0.08)</td>
<td>0.57(0.09)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.86(0.04)</td>
<td>0.74(0.07)</td>
<td>0.57(0.08)</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.92(0.02)</td>
<td>0.76(0.04)</td>
<td>0.57(0.05)</td>
</tr>
<tr>
<td>0.25qt criterion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.74(0.07)</td>
<td>0.85(0.05)</td>
<td>0.73(0.08)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.74(0.06)</td>
<td>0.88(0.04)</td>
<td>0.76(0.07)</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.75(0.03)</td>
<td>0.92(0.02)</td>
<td>0.77(0.04)</td>
</tr>
<tr>
<td>0.10qt criterion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.58(0.08)</td>
<td>0.76(0.07)</td>
<td>0.88(0.04)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.57(0.05)</td>
<td>0.75(0.05)</td>
<td>0.91(0.03)</td>
</tr>
<tr>
<td></td>
<td>5000</td>
<td>0.57(0.03)</td>
<td>0.75(0.03)</td>
<td>0.94(0.01)</td>
</tr>
</tbody>
</table>
Double robust estimation

Assume that $Q_\tau(Y^*(1)|X) = X^T \eta_1(\tau)$ and $Q_\tau(Y^*(0)|X) = X^T \eta_0(\tau)$, for all $\tau \in (0, 1)$. Let 
\{$u_{11}, \ldots, u_{1n}\}$ and \{$u_{01}, \ldots, u_{0n}\}$ be independent random samples from the uniform$(0, 1)$ distribution. We generate random responses by 
$Y_i^{**} = X_i^T \eta_1(u_{1i})d(X_i, \eta) + X_i^T \eta_0(u_{0i})(1 - d(X_i, \eta)),$ 
$i = 1, \ldots, n$. This gives a random sample from $Y^*(d_\beta)$.

In practice, we replace $\eta_1(\tau)$ and $\eta_0(\tau)$ by their estimators and denote the corresponding generated outcomes by $\hat{Y}_i^{**}$, $i = 1, \ldots, n$.

We then estimate the $\tau$th quantile of $Y^*(d_\eta)$ by

$$\arg\min_{\eta} n^{-1} \sum_{i=1}^{n} \left[ \frac{C_i(\eta)}{\hat{\pi}_c(X_i, \eta)} \rho_\tau(Y_i - a) + (1 - \frac{C_i(\beta)}{\hat{\pi}_c(X_i, \eta)}) \rho_\tau(\hat{Y}_i^{**} - a) \right].$$

This estimator can be seen to enjoy the double robustness properties.
A numerical example

- We generate a random sample of observations \((Y_i, A_i, X_i)\), \(i = 1, \ldots, 1000\), with \(X_i = (X_{i1}, X_{i2})'\), such that

\[
Y_i = 2 + 0.2(X_{i1} + X_{i2}) - 1.5X_{i1}^2 - 2X_{i2}^2 + 3X_{i1}X_{i2} + A_i(-0.1 - X_{i1} + X_{i2}) + 0.2\epsilon_i,
\]

where \(X_{i1}, X_{i2}\) are independent and uniformly distributed on \((-1.5, 1.5)\); \(A_i\mid X_i\) is Bernoulli with success probability satisfying \(\text{logit}\{\Pr (A_i = 1 \mid X_i)\} = -1.2 + 0.8X_{i1}^2 + 0.8X_{i2}^2\), and the random error \(\epsilon_i\) follows a Gamma distribution with shape parameter 2 and rate parameter 0.5.

- Propensity score model: One is the correctly specified model \(\text{logit}\{\pi (X_i; \gamma)\} = \gamma_0 + \gamma_1 X_{i1}^2 + \gamma_2 X_{i2}^2\); the other is a misspecified model \(\text{logit}\{\pi (X_i; \gamma)\} = \gamma_0 + \gamma_1 X_{i1} + \gamma_2 X_{i2}\).
A numerical example (cont’d)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\eta}_0$</th>
<th>$\hat{\eta}_1$</th>
<th>$\hat{\eta}_2$</th>
<th>$\hat{Q}_{0.5}$</th>
<th>$\hat{Q}_{0.2}$</th>
<th>$\hat{Q}_{\text{mean}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS correct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPWE</td>
<td>-0.05(0.16)</td>
<td>-0.68(0.08)</td>
<td>0.69(0.09)</td>
<td>1.60(0.09)</td>
<td>-1.00(0.31)</td>
<td>0.67(0.14)</td>
</tr>
<tr>
<td>DR</td>
<td>-0.05(0.15)</td>
<td>-0.70(0.07)</td>
<td>0.68(0.08)</td>
<td>1.57(0.08)</td>
<td>-1.08(0.31)</td>
<td>0.61(0.14)</td>
</tr>
<tr>
<td>PS incorrect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPWE</td>
<td>-0.42(0.22)</td>
<td>-0.34(0.39)</td>
<td>0.68(0.14)</td>
<td>1.64(0.07)</td>
<td>-0.74(0.28)</td>
<td>0.85(0.11)</td>
</tr>
<tr>
<td>DR</td>
<td>-0.04(0.14)</td>
<td>-0.68(0.07)</td>
<td>0.70(0.08)</td>
<td>1.56(0.08)</td>
<td>-1.00(0.26)</td>
<td>0.68(0.11)</td>
</tr>
</tbody>
</table>

Table: The median optimal treatment regime is $d^{\text{opt}}(X_i) = I(\eta_0 + \eta_1 X_{i1} + \eta_2 X_{i2} > 0)$, where $\eta_0 = -0.071$, $\eta_1 = -0.705$ and $\eta_2 = 0.705$. The maximally achievable median is $\max_\eta Q_{0.5}(Y^*(d_\eta)) = 1.54$. 
Dynamic treatment regime

- **$K$-stage sequence of treatments:**
  \[
  \overline{d}_k = (d_1, \ldots, d_k) \in D = D_1 \times \ldots \times D_1
  \]

- Baseline vector of covariates $X_1$, potential outcomes:
  $X_2^*(\overline{d}_1), \ldots, X_k^*(\overline{d}_{k-1}), Y^*(\overline{d}_k)$.
  - $X_k^*(\overline{d}_{k-1})$ is the covariate information between decisions $(k-1)$ and $k$ had the subject received treatment $\overline{d}_{k-1}$
  - $Y^*(\overline{d}_k)$ is the outcome had the subject received treatment $\overline{d}_k$

- **Quantile criterion:**
  \[
  \overline{d}_k^{opt}(X) = \arg\max_{\overline{d}_k \in D} Q_\tau(Y^*(\overline{d}_k)).
  \]
Two-stage decision problem

- Observe \((X_{i1}, A_{i1}, X_{i2}, A_{i2}, Y_i), i = 1, \ldots, n\)
  - \(X_{i1}\) 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
  - \(Y_i\) is the observed outcome for subject \(i\) (assume larger is better)

- **No unmeasured confounder assumption:**
  \[
  A_{i1} \perp \{X^*_2, Y^*\}|X_1; \ A_{i2} \perp Y^*|(X^*_1, X^*_2, A_{i1})
  \]
  Treatment \(A_j\) received in the \(j\)th stage is independent of any future (potential) covariate or outcome conditional on the history
Robust quantile-optimal dynamic treatment regime estimator

- For a given decision $\bar{d}_\eta = (d_{\eta_1}, d_{\eta_2})$, “full data”: $(X_1, X_2^*(d_{\eta_1}), Y^*(\bar{d}_\eta))$.

- **Induced monotone dropout**: Let
  - $C_\eta = \infty$ if $A_1 = d_{\eta_1}$ and $A_2 = d_{\eta_2}$. In this case $(X_1, X_2, Y) = (X_1, X_2^*(d_{\eta_1}), Y^*(\bar{d}_\eta))$.
  - $C_\eta = 2$ if $A_1 = d_{\eta_1}$ but $A_2 \neq d_{\eta_2}$ (dropout before decision 2)
  - $C_\eta = 1$ if $A_1 \neq d_{\eta_1}$ and $A_2 \neq d_{\eta_2}$ (dropout before decision 1)

- We can show MAR assumption is satisfied: missingness may be related to the observed information but is conditionally independent of what’s missing.
Let $\pi_1(H_1) = P(A_1 = 1 \mid H_1)$ and $\pi_2(H_2) = P(A_2 = 1 \mid H_2)$, where $H_1 = \{X_1\}$ and $H_2 = \{X_1, A_1, X_2\}$. Then

$$
\pi(\eta) = P(C_\eta = \infty \mid H_2) = [\pi_1(H_1)d_1\eta + (1 - \pi_1(H_1))(1 - d_1\eta)]
\times [\pi_2(H_2)d_1\eta + (1 - \pi_2(H_2))(1 - d_2\eta)].
$$

We can posit models $\pi_1(X_1, \gamma_1)$ and $\pi_2(X_1, A_1, X_2, \gamma_2)$.

The estimate of the $\tau$th quantile of $Y^*(\bar{d}_\eta)$ is:

$$
\hat{Q}_\tau(\eta) = \arg\min_{\eta} \frac{1}{n} \sum_{i=1}^{n} I(C_{\eta,i} = \infty) \frac{\rho_{\tau}(Y_i - a)}{\hat{\pi}(\eta)} 
$$

The estimator corresponding to the optimal rule from the class $\mathcal{D}$ is defined as

$$
\hat{\eta} = \arg\max_{\eta \in \mathcal{B}} \hat{Q}_\tau(\eta).
$$

The estimated optimal quantile regime is $\bar{d}_{\hat{\eta}} = \bar{d}(X, \hat{\eta})$. 
Asymptotic theory

- Sequential, Multiple, Assignment Randomized Trials (SMART, Murphy, 2008): a useful design to help find the optimal dynamic treatment regime

- For the SMART design, $\pi_1(H_1)$ and $\pi_1(H_2)$, hence $\pi(\eta)$ is known.

- For the smart design, if we consider the class of dynamic treatment regimes $\{I(H_1^T \eta_1 \geq 0), I(H_2^T \eta_2 \geq 0)\}$, with $\eta = (\eta_1^T, \eta_2^T)^T \in \mathbb{B}$, then the limiting distribution of $n^{1/3}(\hat{\eta}_n - \eta_0)$ converges in distribution to $\arg \max_t Z^*(t)$, where the process $Z^*(t) = -\frac{1}{2} t^T V^* t + W^*(t)$, $V^*$ is an $l \times l$ positive definite matrix and $W^*(t)$ is a centered Gaussian process with continuous sample paths and covariance kernel function $K^*(C_1, C_2)$. 

A numerical example

- The data are generated from
  \[ Y = 1 + X_1 + A_1 \left[ 1 - 3 (X_1 - 0.2)^2 \right] + X_2 + A_2 \left[ 1 - 5 (X_2 - 0.4)^2 \right] + (1 + 0.5A_1 - A_1 X_1 + 0.5A_2 - A_2 X_2)\epsilon, \]
  where \( \epsilon \sim N(0, 0.4) \), \( X_1 \sim \text{Uniform}(0, 1) \), \( X_2|\{X_1, A_1\} \sim \text{Uniform}(0.5X_1, 0.5X_1 + 0.5) \), \( A_1|X_1 \sim \text{Bernoulli}(\text{expit}(-0.5 + X_1)) \), and \( A_2|\{X_1, A_1, X_2\} \sim \text{Bernoulli}(\text{expit}(-1 + X_2)) \) with \( \text{expit}(t) = e^t/(1 + e^t) \).

- We consider sequential treatment regimes of the form \((A_1, A_2)\), where \( A_1 = I\{X_1 < \eta_1\} \), and \( A_2 = I\{X_2 < \eta_2\} \).

- We note that this class contains the mean-optimal sequential treatment regimes which are given by \( A_1 = I(X_1 < 0.777) \) and \( A_2 = I(X_2 < 0.847) \).
A numerical example

Table: Population parameters and summary values for optimal treatment regimes under different criteria for Example 2 based on a Monte Carlo experiment with \( n = 10^5 \).

<table>
<thead>
<tr>
<th>Method</th>
<th>( \eta_1 )</th>
<th>( \eta_2 )</th>
<th>( Q_{\text{mean}} )</th>
<th>( Q_{0.50} )</th>
<th>( Q_{0.75} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean criterion</td>
<td>0.777</td>
<td>0.847</td>
<td>3.331</td>
<td>3.323</td>
<td>3.821</td>
</tr>
<tr>
<td>0.50qt criterion</td>
<td>0.753</td>
<td>0.808</td>
<td>3.327</td>
<td>3.327</td>
<td>3.827</td>
</tr>
<tr>
<td>0.75qt criterion</td>
<td>0.729</td>
<td>0.795</td>
<td>3.322</td>
<td>3.325</td>
<td>3.828</td>
</tr>
</tbody>
</table>
Table: Estimated optimal treatment regimes and their corresponding estimated value functions under different criteria for Example 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>n</th>
<th>$\eta_1$</th>
<th>$\eta_2$</th>
<th>$\hat{Q}_{\text{mean}}$</th>
<th>$\hat{Q}_{0.50}$</th>
<th>$\hat{Q}_{0.75}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean_Qlearning</td>
<td>500</td>
<td>0.755(0.041)</td>
<td>1.176(0.294)</td>
<td>3.319(0.090)</td>
<td>3.309(0.102)</td>
<td>3.815(0.122)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.752(0.027)</td>
<td>1.131(0.144)</td>
<td>3.321(0.065)</td>
<td>3.305(0.07)</td>
<td>3.819(0.079)</td>
</tr>
<tr>
<td>mean_ZTLD</td>
<td>500</td>
<td>0.773(0.073)</td>
<td>0.846(0.067)</td>
<td>3.370(0.095)</td>
<td>3.376(0.097)</td>
<td>3.862(0.118)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.768(0.055)</td>
<td>0.852(0.059)</td>
<td>3.356(0.065)</td>
<td>3.354(0.068)</td>
<td>3.848(0.081)</td>
</tr>
<tr>
<td>0.50qt criterion</td>
<td>500</td>
<td>0.751(0.08)</td>
<td>0.815(0.079)</td>
<td>3.357(0.090)</td>
<td>3.391(0.102)</td>
<td>3.858(0.119)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.750(0.062)</td>
<td>0.813(0.069)</td>
<td>3.343(0.063)</td>
<td>3.366(0.068)</td>
<td>3.849(0.081)</td>
</tr>
<tr>
<td>0.75qt criterion</td>
<td>500</td>
<td>0.734(0.108)</td>
<td>0.785(0.103)</td>
<td>3.328(0.095)</td>
<td>3.331(0.109)</td>
<td>3.892(0.123)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.723(0.084)</td>
<td>0.795(0.095)</td>
<td>3.322(0.067)</td>
<td>3.326(0.075)</td>
<td>3.865(0.077)</td>
</tr>
</tbody>
</table>
The outcome of interest is the CD4 count at $96 \pm 5$ weeks from baseline (denoted as cd496).

We consider the problem of how to assign treatment to the HIV-infected patients who had taken AZT before, either to the AZT+ddI combination therapy ($A_i = 1$) or to the ddI monotherapy ($A_i = 1$).

Two covariates are considered for estimating the optimal treatment regimes: $X_1$ (baseline weight of patient, measured in kg) and $X_2$ (baseline CD4 T cell count, denoted by cd40). It has been observed that body weight has a significant role on AZT pharmacokinetic profile.

We consider the class of candidate regimes of the form
\[ I \{ \eta_0 + \eta_1 X_1 + \eta_2 X_2 < 0 \} , \text{ where } \| \eta \| = 1. \]
Histograms of CD4 counts at 96 weeks for AZT+ddI arm and ddI arm, grouped by CD4 count at week 0.
Figure:

Histograms of CD4 counts at 96 weeks for AZT+ddI arm and ddI arm, grouped by baseline weight.
Table: Estimated optimal treatment regimes and summary values for ACTG175 data analysis.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{\eta}_0$</th>
<th>$\hat{\eta}_1$</th>
<th>$\hat{\eta}_2$</th>
<th>$\hat{Q}_{0.50}$</th>
<th>$\hat{Q}_{0.25}$</th>
<th>$\hat{Q}_{\text{mean}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50qt criterion</td>
<td>-0.571</td>
<td>0.691</td>
<td>0.444</td>
<td>360</td>
<td>220</td>
<td>375.4</td>
</tr>
<tr>
<td>0.25qt criterion</td>
<td>-0.210</td>
<td>0.958</td>
<td>-0.194</td>
<td>333</td>
<td>263</td>
<td>346.5</td>
</tr>
<tr>
<td>Mean criterion</td>
<td>-0.526</td>
<td>0.799</td>
<td>0.292</td>
<td>331</td>
<td>219</td>
<td>403.9</td>
</tr>
</tbody>
</table>
Figure: Graphical representation of the estimated optimal treatment regimes for ACTG175 data analysis.
Conclusions

- Quantile criterion for optimal treatment regime estimation is an interesting and useful alternative to the mean criterion.
- Robust estimation of quantile-optimal treatment regime for both one-stage and dynamic decision problems.
- Asymptotic theory:
  - The estimated parameter indexing the quantile-optimal treatment regime converges at a cube-root rate to a nonnormal limiting distribution that is characterized by the maximizer of a centered Gaussian process with a parabolic drift;
  - The value function corresponding to the quantile optimal treatment regime can be estimated at an $O_p(n^{-1/2})$ rate.