Inference for Nonregular Parameters in Optimal Dynamic Treatment Regimes

Bibhas Chakraborty
Department of Statistics, University of Michigan

Joint work with Susan Murphy

ENAR Spring Meetings
San Antonio, Texas
March 17, 2009

Dynamic Treatment Regimes

- Diseases like depression, schizophrenia, substance abuse, HIV infection and cancer are often treated in multiple stages
  - At each stage, treatment type and dosage are adapted to the individual patient's ongoing response, adherence, burden, side effects, preference, and past treatments
- Dynamic treatment regimes are individualized treatment rules that offer a way to operationalize this adaptive clinical practice
  - thus provide an opportunity to improve the clinical practice
- Formally, a dynamic treatment regime (DTR) is a sequence of decision rules or functions, one per stage
  - Each decision rule takes a patient's treatment and covariate history as input, and outputs a recommended treatment

Data Structure

- Two stages on a single patient:

$$O_1, A_1, O_2, A_2, O_3$$

O_j	:	Observation (pre-treatment) at the j-th stage
A_j	:	Treatment (action) at the j-th stage, $A_j \in \{-1, 1\}$
H_j	:	History at the j-th stage, $H_j = \{O_1, A_1, \ldots, O_{j-1}, A_{j-1}, O_j\}$
Y_j	:	Outcome at the j-th stage, $Y_j = f_j(H_{j+1})$, with known f_j

- Assume that A_j's are sequentially randomized in the data, e.g.,

$$P[A_j = 1|H_j] = P[A_j = -1|H_j] = 0.5$$

- A DTR is a sequence of decision rules:

$$d \equiv (d_1, d_2) \text{ with } d_j(H_j) \in \{-1, 1\}$$

Regression-based Methods for Constructing Optimal DTRs

- Q-learning (*Watkins, 1989*)
  - A popular method from Reinforcement (Machine) Learning
  - A generalization of least squares regression to multistage decision problems (*Murphy, 2005*)
- A-learning (*Murphy, 2003*) or Optimal Structural Nested Mean Models (SNMM) (*Robins, 2004*)
  - Under simple conditions, Q-learning is an inefficient version of Robins' SNMM (*Chakraborty et al., 2009*)
- The problem of non-regularity arises in both
  - For simplicity, we will focus on Q-learning here

Motivation for Q-learning

- The intuition comes from Dynamic Programming (*Bellman, 1957*) in case the multivariate distribution of the data is known
- Move backward in time to take care of the delayed effects
- Define the so-called Q-functions:

$$
\begin{aligned}
Q_2(h_2, a_2) &= \mathbb{E}\Big[Y_2 \Big| H_2 = h_2, A_2 = a_2\Big] \\
Q_1(h_1, a_1) &= \mathbb{E}\Big[Y_1 + \underbrace{\max_{a_2} Q_2(H_2, a_2)}_{\text{delayed effect}} \Big| H_1 = h_1, A_1 = a_1\Big]
\end{aligned}
$$

- Optimal DTR: $d_j(h_j) = \arg\max_{a_j} Q_j(h_j, a_j), \; j = 1, 2$
- In real life, we do not know the true distribution of the data
- All we have is a data set: $\{O_{1i}, A_{1i}, O_{2i}, A_{2i}, O_{3i}\}, \; i = 1, \ldots, n$

Q-learning with Linear Regression

- Model for Q-functions (where S_j' and S_j are summaries of H_j):

$$Q_j(H_j, A_j; \beta_j, \psi_j) = \beta_j^T S_j' + (\psi_j^T S_j) A_j, \ j = 1, 2$$

- Stage-2 Regression:

$$(\hat{\beta}_2, \hat{\psi}_2) = \arg\min_{\beta_2, \psi_2} \frac{1}{n} \sum_{i=1}^{n} \Big(Y_{2i} - Q_2(H_{2i}, A_{2i}; \beta_2, \psi_2) \Big)^2$$

- Stage-1 Pseudo-outcome: $\hat{Y}_{1i} = Y_{1i} + \max_a Q_2(H_{2i}, a; \hat{\beta}_2, \hat{\psi}_2)$
- Stage-1 Regression:

$$(\hat{\beta}_1, \hat{\psi}_1) = \arg\min_{\beta_1, \psi_1} \frac{1}{n} \sum_{i=1}^{n} \Big(\hat{Y}_{1i} - Q_1(H_{1i}, A_{1i}; \beta_1, \psi_1) \Big)^2$$

- Optimal DTR: $\hat{d}_j(h_j) = \arg\max_a \ Q_j(h_j, a; \hat{\beta}_j, \hat{\psi}_j), \ j = 1, 2$

Bias in $\hat{\psi}_1$

$$\hat{Y}_{1i} = Y_{1i} + \max_a Q_2(H_{2i}, a; \hat{\beta}_2, \hat{\psi}_2) = Y_{1i} + \hat{\beta}_2^T S'_{2i} + |\hat{\psi}_2^T S_{2i}|$$

- Maximization is a non-smooth, non-linear operation. Even when $Q_2(H_{2i}, a; \hat{\beta}_2, \hat{\psi}_2)$ is unbiased for $Q_2(H_{2i}, a; \beta_2, \psi_2)$, we have

$$E\left[\max_a Q_2(H_{2i}, a; \hat{\beta}_2, \hat{\psi}_2)\right] \geq \max_a Q_2(H_{2i}, a; \beta_2, \psi_2)$$

- Thus $\hat{Y}_{1i}$ is a biased predictor of $Y_{1i} + \max_a Q_2(H_{2i}, a; \beta_2, \psi_2)$
- Bias in $\hat{Y}_{1i},\ i = 1, \ldots, n$, can induce bias in $\hat{\psi}_1$. Possibly sub-optimal DTR!

Non-regularity in Inference

- We want to construct CIs for ψ_1 or test $H_0 : \psi_1 = 0$. Why?
  - reduce the variety of information to be collected for future implementations of the DTR
  - know when there is insufficient evidence in the data to recommend one treatment over another – allow expert opinion
- Because of the lack of smoothness (non-differentiability) of $\hat{Y}_{1i}$:

$$\hat{Y}_{1i} = Y_{1i} + \max_a Q_2(H_{2i}, a; \hat{\beta}_2, \hat{\psi}_2) = Y_{1i} + \hat{\beta}_2^T S'_{2i} + |\hat{\psi}_2^T S_{2i}|$$

  - The asymptotic distribution of $\hat{\psi}_1$ is normal if $P[\psi_2^T S_2 = 0] = 0$ and non-normal if $P[\psi_2^T S_2 = 0] > 0$
  - The change between the two asymptotic distributions is abrupt
  - Details on non-regularity in this problem: *Robins (2004)*

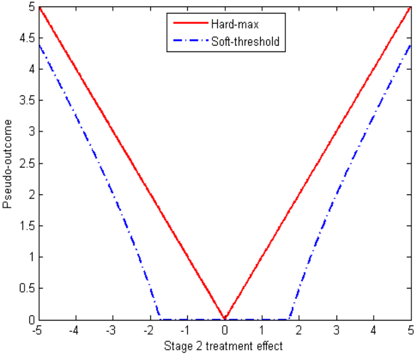
Confidence Intervals and Non-regularity

- CIs can be wrongly centered due to the bias in $\hat{\psi}_1$
- Non-regularity can cause additional problems in tail behavior
- Whenever true $\psi_2^T S_2 \approx 0$, Wald type CIs show poor frequentist properties (e.g., *Robins, 2004; Moodie and Richardson, 2008*)
- Usual bootstrap CIs also show poor frequentist properties
  - Bootstrap is inconsistent due to non-differentiability (e.g., *Shao, 1994; Andrews, 2000*)

So what could be a remedy?

Soft-threshold Estimator: A Solution

$$\hat{Y}_{1i}^{ST} = Y_{1i} + \hat{\beta}_2^T S'_{2i} + |\hat{\psi}_2^T S_{2i}| \cdot \left(1 - \frac{\lambda_i}{|\hat{\psi}_2^T S_{2i}|^2}\right)^+, \ \lambda_i > 0$$



The corresponding estimator $\hat{\psi}_1^{ST}$ from the stage-1 regression is the soft-threshold estimator

Soft-threshold Estimator

- The soft-threshold estimator is akin to the shrinkage estimators in machine learning, e.g., adaptive lasso (Zou, 2006)
- $\hat{Y}_{1i}^{ST}$ is still a non-smooth function, but the problematic term $|\hat{\psi}_2^T H_{21,i}|$ is shrunk (thresholded) towards zero
- For $\lambda_i = 3H_{21,i}^T \hat{\Sigma}_2 H_{21,i}/n$, the soft-threshold pseudo-outcome $\hat{Y}_{1i}^{ST}$ is an empirical Bayes "estimator" of the true quantity of interest ($\hat{\Sigma}_2/n$ is the estimated covariance matrix of $\hat{\psi}_2$)
  - It follows from a Bayesian formulation originally used for image processing by *Figueiredo and Nowak (2001)*

Simulation Design

- 1000 simulated data sets, each of size $n = 300$
- Generative Model:

$$\begin{aligned} O_1, A_1, A_2 &\in \{-1, 1\} \quad \text{with probability } 0.5 \\ O_2 &\in \{-1, 1\} \quad \text{with } P[O_2 = 1 | O_1, A_1] \text{ varied in the examples} \end{aligned}$$

- Analysis Model:

$$\begin{aligned} Q_2 &= \beta_{20} + \beta_{21} O_1 + \beta_{22} A_1 + \beta_{23} O_1 A_1 + \underbrace{(\psi_{20} + \psi_{21} O_2 + \psi_{22} A_1)}_{\psi_2^T S_2} A_2 \\ Q_1 &= \beta_{10} + \beta_{11} O_1 + (\psi_{10} + \psi_{11} O_1) A_1 \end{aligned}$$

- The stage-2 treatment effect $\psi_2^T S_2$ governs the results
- 1000 bootstrap replications to construct CIs

Example (Non-regular)

Bias is the main issue!

$$P[O_2 = 1|O_1, A_1] = \frac{\exp(0.5(O_1 + A_1))}{1 + \exp(0.5(O_1 + A_1))}$$
$$Y_1 = 0; \quad Y_2 = -0.5A_1 + 0.5A_2 + 0.5A_1A_2 + \epsilon, \ \epsilon \sim N(0, 1)$$

Estimation of ψ_{10}

Estimator	Bias	MSE	Coverage	Length
hard-max	-0.0401	0.0075	88.4*	0.2987
soft-threshold	-0.0185	0.0058	93.4·	0.2923

Coverage = coverage of nominal 95% percentile bootstrap CI

Example (Non-regular)

Tail behavior is the main issue!

$$P[O_2 = 1|O_1, A_1] = \frac{\exp(0.5(O_1 + A_1))}{1 + \exp(0.5(O_1 + A_1))}$$
$$Y_1 = 0; \quad Y_2 = \epsilon, \ \epsilon \sim N(0,1)$$

Estimation of ψ_{10}

Estimator	Bias	MSE	Coverage	Length
hard-max	0.0003	0.0045	96.8*	0.2687
soft-threshold	0.0009	0.0036	95.3	0.2498

Coverage = coverage of nominal 95% percentile bootstrap CI

Example (Regular)

Price of shrinkage!

$$P[O_2 = 1|O_1, A_1] = \frac{\exp(0.1(O_1 + A_1))}{1 + \exp(0.1(O_1 + A_1))}$$

$$Y_1 = 0;\ Y_2 = -0.5A_1 + 0.25A_2 + 0.5O_2A_2 + 0.5A_1A_2 + \epsilon,\ \epsilon \sim N(0,1)$$

$\psi_2^T S_2$ is far from 0

Estimation of ψ_{10}

Estimator	Bias	MSE	Coverage	Length
hard-max	0.0009	0.0067	95.0	0.3094
soft-threshold	0.0052	0.0074	94.8	0.3191

Coverage = coverage of nominal 95% percentile bootstrap CI

Discussion

- Soft-threshold estimator addresses the issues of bias and coverage of CIs under non-regular settings
- The original hard-max estimator would be preferred over the soft-threshold estimator when the stage-2 treatments are "too different" (so-called regular settings)
  - However regular settings are less likely to occur in clinical trial data due to ethical considerations, e.g., "the principle of clinical equipoise" (*Freedman, NEJM, 1987*)
- The developed methodology can be used for both randomized and observational data
  - Here we focused on randomized data since we wanted to separate causal inference issues from the issue of non-regularity

- Main Reference:

B. Chakraborty, V. Strecher, and S. Murphy (2009). Inference for Nonregular Parameters in Optimal Dynamic Treatment Regimes. *To appear in Statistical Methods in Medical Research.*

- Slides can be found at: http://www.stat.lsa.umich.edu/∼bibhas/presentations.html

- Email me with questions: bibhas@umich.edu

Thank you!