1. Abstract

In recent years, many Sequential Multiple Assignment Randomized Trials (SMART) have been conducted. The data collected from this type of trial is extremely useful to answer questions such as “what treatment policy (or, dynamic treatment regime) will yield highest mean outcome if followed by patients from the same population”. In this research project, we propose novel approach to estimating the value of treatment policies and policy search method based on the policy value estimators. The proposed assisted estimator for policy value is based on Structured Nested Mean Model (SNMM) that was developed by Robins (1994) in causal inference research. Moreover, we propose some computationally more efficient variation of the original policy search problem.

2. Framework and Data Structure

Treatment policies (dynamic treatment regimes, or adaptive treatment strategies): a sequence of decision rules:

- Take measurements of patients’ time-varying covariates as inputs, output recommended treatments. Characteristics: Dynamic, Personalized.
- A policy class: A class of policies parameterized by a finite number of parameters.

Goal of this project: Identify the optimal treatment policy within a given policy class, using data from two-stage Multiple Assignment Randomized Trials.

Data structure: five components for each individual as in the following flowchart:

- Stage one: measured covariates.
- Stage one: measured covariates.
- Stage one: measured covariates.
- Stage one: measured covariates.
- Stage one: measured covariates.

X1, X2, X3, X4, X5: five components for each individual as in the following flowchart:

Example: In the alcohol dependence study, X1, X2 can include time-varying measurements of alcohol craving scores, number of heavy drinking days per week, number of drinking days per week. X3 typically includes baseline demographics (gender, race, etc), X4, X5 can be medication or behavioral therapy, interesting Y can be (1) end of study percentage of heavy drinking days; (2) end of study percentage of heavy drinking days; (3) end of study craving scores, etc.

A two-stage treatment policy consists of two mappings \( \pi \equiv (\pi_1, \pi_2) \):

- Stage one: \( \pi_1: H_1 \rightarrow A_1 \)
- Stage two: \( \pi_2: H_2 \rightarrow A_2 \)

\( H_1 \) is space of \( H_1 \equiv X_1; H_2 \) is space of \( H_2 \equiv (X_1, X_2, X_3) \).

The value of \( \pi \) is \( V_2 = E(Y|\pi) \). \( E(Y|\pi) \) is the expectation w.r.t. distribution of \( (X_1, A_1, X_2, A_2, Y) \) with \( A_1 \) and \( A_2 \) assigned according to rules in \( \pi \) (the other components of the joint distribution are the same as population in the trial). The regret of \( \pi \) is \( \rho_{\pi} = E(Y|\pi) - E(Y|\pi^*) \).

3. Assisted Estimator for Policy Value / Regret

Define the treatment effect functions \( \mu_1, \mu_2 \):

- \( \mu_1(H_1, A_1) = E(Y|H_1, A_1) - E(Y|H_1, A_2) = 0 \)
- \( \mu_2(H_1, A_1) = E(Y|H_1, A_1) - E(Y|H_2, A_2) = 0 \)

\( \mu_1, \mu_2 \) represent the differences in the expected outcome of \( Y \) by inducing following \( A_1 \) other than treatment coded by zero at stage one.

\( \mu_2 \) represents the difference in the expected outcome of \( Y \) by inducing following \( A_1 \) other than treatment coded by zero at stage two.

We model treatment effects parametrically (first proposed by Robins (1994) as structural nested mean model):

\[ \mu_1(H_1, A_1) = \beta_1(h_1(H_1)T_1b_1 + h_1(H_1)\pi_1T_1b_2), \]

\[ \mu_2(H_1, A_1) = \beta_2(h_2(H_2)\pi_1T_2b_1 + h_2(H_2)\pi_2T_2b_2), \]

where \( h_1, h_2 \) are known functions with constraints \( h_1(H_1) = 0 \) and \( h_2(H_2) = 0 \), respectively.

Example: In alcohol dependence study, one can model \( h_1(H_1) = A_1(1; C_S 2, P_H)^T \) and \( h_2(H_2) = A_2(1; C_S 2, P_H)^T \) if it is believed that each stage of treatment has interactive effect with the recent alcohol craving score (C_S) and the recent percentage of heavy drinking days (P_H).

Suppose we already have consistent and asymptotically normal estimator \( \hat{\beta}_1, \hat{\beta}_2 \) for \( \beta_1, \beta_2 \). Then the following assisted estimator for policy value is consistent and asymptotically normal for \( V_2 \) under regularity conditions:

\[ \hat{\rho}_{\pi_1, \pi_2} = \frac{\hat{V}_2 - \hat{V}_2(\hat{\beta}_1, \hat{\beta}_2)}{\text{var}(\hat{\beta}_1, \hat{\beta}_2)}, \]

where \( \hat{f}_{\beta_1}(H_1) \) is the known treatment assignment probability in the randomized trial.

Intuition: Construct a pseudo-outcome by subtracting the effects of observed treatment sequence, and adding in the treatment effects of the sequence specified by \( \pi \).

Similarly, we propose the assisted estimator for policy regret:

\[ \hat{\rho}_{\pi_1, \pi_2} = \frac{\hat{V}_2 - \hat{V}_2(\hat{\beta}_1, \hat{\beta}_2)}{\text{var}(\hat{\beta}_1, \hat{\beta}_2)}, \]

where \( \hat{\rho}_{\pi_1, \pi_2} \) is consistent and asymptotically normal for \( \rho_{\pi_1, \pi_2} \) as long as treatment effect model is correctly specified, and does not rely on correct modeling of \( m(H_1, A_1) \).

4. Efficiency Correction on Assisted Estimators

Motivated by augmented estimator from missing data theory, consider the following augmented-assisted estimator for policy regret:

\[ \hat{\rho}_{\pi_1, \pi_2} = \frac{\hat{V}_2 - \hat{V}_2(\hat{\beta}_1, \hat{\beta}_2)}{\text{var}(\hat{\beta}_1, \hat{\beta}_2)}, \]

where \( m(H_1, A_1) \) is a model for \( E(\beta_2|H_2, \pi_2)|H_1, A_1 \).

\[ \hat{\rho}_{\pi_1, \pi_2} \] is consistent and asymptotically normal for \( \rho_{\pi_1, \pi_2} \) as long as treatment effect model is correctly specified, and does not rely on correct modeling of \( m(H_1, A_1) \).

5. Policy Search

For simplicity, consider assisted policy search using non-augmented assisted estimator for policy regret.

Consider motivating example: two-stage setting with binary treatments. Consider policy class as follows (this type of policy also called “linear decision boundary”):

\[ \pi_1(H_1, b_1) \equiv I(\hat{X}_1 > \theta), \pi_2(H_2, b_2) \equiv I(\hat{X}_2 > \theta). \]

With binary treatments, models for treatment effects can only take the form:

\[ \rho_{\pi_1}(H_1, A_1) = A_1\hat{f}_{\beta_1} \]
\[ \rho_{\pi_2}(H_2, A_2) = A_2\hat{f}_{\beta_2} \]

(\( S_1, S_2 \) are certain features of \( H_2, H_2 \) could be some functions of covariates). Optimize \( \rho_{\beta_1, \beta_2} \) over \( b_1, b_2 \):

\[ \min_{b_1, b_2} \frac{1}{b_1, b_2} \max_{b_1, b_2} P_{\pi_1}(\hat{X}_1 > \theta) \hat{f}_{\beta_1} + \frac{(2\theta - 1)\hat{f}_{\beta_2}(\hat{X}_2 > \theta)}{\pi_2(H_2, A_2)} \]

\[ \pi_1(H_1, b_1) \equiv I(\hat{X}_1 > \theta) \]
\[ \pi_2(H_2, b_2) \equiv I(\hat{X}_2 > \theta) \]

Natural idea: iteratively optimize (4) over \( b_1 \) and \( b_2 \).

Fixing \( b_2 \), equivalent to

\[ \min_{b_1} P_{\pi_1}(\hat{X}_1 > \theta) \]

(5)

where \( W_1 = \hat{f}_{\beta_1} + \frac{(2\theta - 1)\hat{f}_{\beta_2}(\hat{X}_2 > \theta)}{\pi_2(H_2, A_2)} \). Other ways to decompose \( W_1 \) may have better property. Furthermore equivalent to

\[ \min_{b_1} P_{\pi_1}(\hat{X}_1 > \theta) \]

(6)

where \( W_1 = \max_{b_1} -W_1, W_2 = \max_{b_2} -W_2 \).

One possible convex relaxation of (6):

\[ \min_{b_1, b_2} P_{\pi_1}(\hat{X}_1 > \theta) \]

We are currently investigating properties of different ways of relaxing the optimization. The step of optimizing over \( b_2 \) when fixing \( b_1 \) can proceed similarly.

6. Simulation

Figure 1: Comparison between estimated and true treatment policy and iteratively optimizing the related problem. In the left panel, \( \tilde{\theta} \) is rescaled so that in each experiment \( \tilde{\theta} \) corresponds to the mean value of \( \theta \) under randomized treatment. 1 corresponds to the policy value of true optimal policy.