Stabilized Dynamic Treatment Regimes

Yingqi Zhao

Fred Hutchinson Cancer Research Center

IMA Innovative Statistics and Machine Learning for Precision Medicine, Minneapolis,
Sep 14, 2017

Outline

1. Motivation
2. Methodology: stabilized dynamic treatment regimes
3. Simulation
4. Discussion

Active Surveillance in Prostate cancer

- Active surveillance represents a strategy to address the overtreatment of prostate cancer.
- Delay intervention in patients whose tumors initially have features consistent with a low risk cancer and treat only when a more clinically significant malignancy is identified.
- Patients treated with active surveillance undergo serial monitoring with serum prostate specific antigen (PSA) measurements, clinical examinations and repeat biopsies.
- When to recommend intervention (surgery)?

Canary Prostate cancer Active Surveillance Study (PASS)

- A multi-institutional active surveillance cohort established in 2008.
- Broad eligibility criteria were used to sample the full spectrum of men using active surveillance; 905 participants enrolled.
- Participants were followed with serum PSA measurements every 3 months, clinical and digital rectal examination every 6 months, and repeat prostate biopsy 6 to 12, 24, 48 and 72 months after diagnosis.
- Outcome: disease reclassification
- Can we identify a *fixed* rule over time to recommend intervention as needed?

Diabetes in Complex Patients

- Current diabetes guidelines: tight control of glycosylated hemoglobin (A1c) (< 7 %)
  - Healthy patients.
  - Based on trials of younger patients without severe diabetes complications or other comorbidities.
  - Relatively low risk of tight control; significant benefits in reducing incidence of vascular events
  - Tight control of BP ($<$130/80 mm Hg) and LDL cholesterol ($<$100 mg/dl) for patients with diabetes

Diabetes in Complex Patients

- Inappropriate for complex diabetes patients, i.e., older patients (age $>$ 65 years) and/or those with comorbid conditions.
  - Evidence for these guidelines was mainly obtained from the results of randomized clinical trials (RCTs)
  - Complex patients usually meet the exclusion criteria of clinical trials
  - Increased risk of drug-related morbidity, e.g., hypoglycemia, hypotension
- How can we strengthen the current guidelines for complex patients?

A1c Control Observational Study

- Linked claims and EHR data for Medicare beneficiaries in the University of Wisconsin Medical Foundation (UWMF) system.
- 8,304 diabetes patients active during 2003-2011, recorded each 90-day quarter in which they were alive at the start of the quarter
- Outcome: adverse outcomes occurring during these quarters (e.g., emergency department use or hospitalizations, death), documented from claims information.
- Covariates: sociodemographics, and indicators for comorbidities; time varying patient complexity
- Can we identify a *fixed* rule to target tight control of A1c?

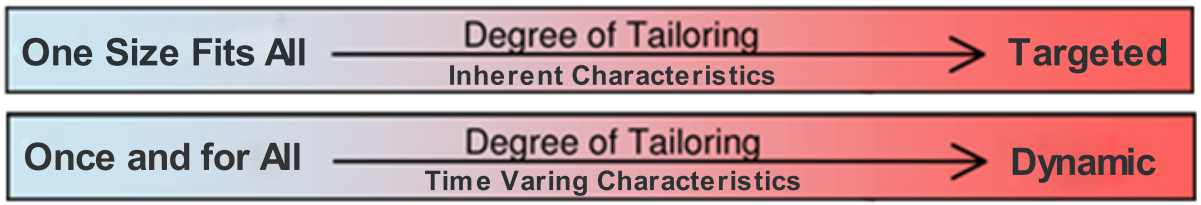
Dynamic Treatment Regime

- At any decision point
  - Input: available historical information on the patient to that point.
  - Output: next treatment.

- Dynamic treatment regimes (DTRs) are sequential decision rules for individual patients that can adapt over time to an evolving illness.
  - One decision rule for each time point.
  - Each rule: recommends the treatment at that point as a function of accrued historical information.
  - An algorithm for treating any patient.
  - Aim to optimize some cumulative clinical outcome.

DTR Goals

Learn adaptive treatment strategies: tailor (sequences of) treatments based on patient characteristics.



Maximize the benefit of dynamic treatment regimes:

- Well chosen tailoring variables.
- Well devised decision rules.

Dynamic Treatment Regimes (DTR)

Observe data on n individuals, T stages for each individual,

$$X_1, A_1, R_1, X_2, A_2, \ldots, X_T, A_T, R_T, X_{T+1}$$

X_j: Patient covariates available at stage j.
A_j: Treatment at stage j, $A_j \in \{-1, 1\}$.
R_j: Outcome following stage j.
H_j: History available at stage j, $H_j = \{X_1, A_1, R_1, \ldots, A_{j-1}, R_{j-1}, X_j\}$.

A DTR is a sequence of decision rules:

$$d = (d_1(H_1), \ldots, d_T(H_T)), d_j(H_j) \in \{-1, 1\}.$$

Goal

Maximize the expected sum of outcomes if the DTR is implemented in the future.

Value Function and Optimal DTR for Multiple Stages

- The value function: $\mathcal{V}(d) = E^d(R_1 + \ldots + R_T)$.
- Optimal DTR: $d^* = \text{argmax}_d \, \mathcal{V}(d)$.
- Two main challenges in developing optimal DTRs:
  - Taking individual information into account in decision making.
  - Incorporating long-term benefits and risks of treatment due to delayed effects.

Q-learning

- Backwards and recursively estimates the following Q-function:

$$Q_j(h_j, a_j) = E(R_j + \max_{a_{j+1} \in \{-1,1\}} Q_{j+1}(H_{j+1}, a_{j+1}) | H_j = h_j, A_j = a_j),$$

  where $Q_{T+1} = 0$, and $h_j \in \mathcal{O}_j, a_j \in \mathcal{A}_j, j = 1, \ldots, T$.

- The estimated optimal sequence of decision rules

$$\hat{d}_j(h_j) = \operatorname*{argmax}_{a_j \in \{-1,1\}} \hat{Q}_j(h_j, a_j).$$

- Q learning with regression: estimate the Q-functions from data using regression and then find the optimal DTR.

- Decision not shared.

Current methods

- Simultaneous g-estimation: the empirical performance is largely unknown
- Q learning with shared parameters: simultaneous estimation procedure for the shared parameters based on Q-learning
- Rely on the assumption that models are correct

Stabilized Dynamic Treatment Regimes

- Learn the optimal regimes at all stages simultaneously.
- Any $d_j(h_j)$ can be written as $d_j(h_j) = \text{sign}\{f_j(h_j)\}$; $A_j = d_j(h_j)$ is equivalent to $A_j f_j(h_j) > 0$.
- Decompose H_j into $H_j^{(s)}$ and $H_j^{(u)}$, which represent variables that are shared or not shared.
- Let $\beta_j = (\beta_j^{(u)\mathsf{T}}, \beta^{(s)\mathsf{T}})^{\mathsf{T}}$, where $\beta_j^{(u)}$ are the unshared parameters, and $\beta^{(s)}$ are the shared parameters across stages.

Stabilized Dynamic Treatment Regimes

- An inverse probability of treatment estimator (IPWE) of $\mathcal{V}(d)$

$$\begin{aligned}\hat{V}^{IPWE}(f) &= \mathbb{P}_n\left(\frac{\sum_j R_j I\{A_1 f_1(H_1) > 0, \ldots, A_T f_T(H_T) > 0\}}{\prod_{j=1}^{T} p_j(A_j|H_j)}\right) \\ &= \mathbb{P}_n\left(\frac{\sum_j R_j I[\min_{j=1,\ldots,T}\{A_j f_j(H_j)\} > 0]}{\prod_{j=1}^{T} p_j(A_j|H_j)}\right),\end{aligned}$$

where $f_j(h_j) = h_j^{(s)}\beta^{(s)} + h_j^{(u)}\beta_j^{(u)}$, and $\mathbb{P}_n$ denotes empirical averages.

Stabilized Dynamic Treatment Regimes

- Replace a concave surrogate for the indicator to alleviate the computation difficulties.

$$\max_{\beta_j^{(u)},\beta^{(s)}} \mathbb{P}_n \left(\frac{R\psi[\min_{j=1,\ldots,T}\{A_j(H_j^{(u)\mathsf{T}}\beta_j^{(u)} + H_j^{(s)\mathsf{T}}\beta^{(s)})\}]}{\prod_{j=1}^{T} p_j(A_j|H_j)} \right),$$

  denoted by $\Phi_n(\beta_j^{(u)}, \beta^{(s)})$ and ψ is a concave function.

- Use $\psi(t) = -\log(1 + e^{-t})$
- Use soft minimum to approximate the min function.
- $p_j(A_j|H_j)$ can be estimated using e.g., logistic regression
- Apply the LASSO penalty for sparsity, i.e., maximize

$$\max \Phi_n(\beta_j^{(u)}, \beta^{(s)}) - \lambda_n \Big(\sum_j |\beta_j^{(u)}| + \sum |\beta^{(s)}| \Big),$$

Stabilized Dynamic Treatment Regimes

- The IPWE $\hat{V}^{IPWE}(f)$ has potentially high variance.
- Consider a doubly robust method – construct augmented IPWE for $\mathcal{V}(d)$.

$$\hat{V}^{AIPWE}(f) = \mathbb{P}_n \left[\sum_{a_1,\ldots,a_T} I[\min_{j=1,\ldots,T}\{a_j f_j(H_j)\} \geq 0] W_a(Y, H_T; \lambda, L) \right]$$

  - Incorporate contributions from the subjects who did not receive the specified treatment assignments across all stages via the weighting function $W_a(Y, H_T; \lambda, L)$.
  - Weighting functions involves propensity scores, λ, and regression functions, L, at each stage.
  - Robust against misspecification.

Properties and Computation

- Fisher consistency
- Dervatives can be derived and we implement Orthant-Wise Limited-memory Quasi- Newton algorithm to find the solution.
- Posit parametric models to approximate propensity scores and regression functions.

Simulation Studies: Generative Model

- Baseline covariates $X_{1,1}, \ldots, X_{1,p} \sim N(0,1)$
- 4 stages.
- $X_{j,1} = X_{j-1,1} + N(0,1), 2 \leq j \leq 4.$
- $A_j \in \{-1,1\}$ w.p. 0.5 in Scenarios 1, 2
- $A_j \in \{-1,1\}$ and $logit\{P(A_j = 1)\} = X_{j,1} + X_{j,2} - 0.5$ in Scenarios 3, 4

Simulation Studies: Generative Model

- Scenario 1/3: covariates shared throughout

$$Y \sim 10 + X_{1,1}X_{1,2} + X_{1,3} - \sum_{j=1}^{4} |X_{j,1} - 1|\{I(A_j > 0) - I(X_{j,1} - 1 > 0)\}^2 + N(0,1).$$

- Scenario 2/4: 25 covariates in stage 1 unshared, covariates shared from stages 2 to 4

$$Y \sim 10 + X_{1,1}X_{1,2} + X_{1,3} - |X_{1,1} + X_{1,2} - 2|\{I(A_1 > 0) - I(X_{1,1} + X_{1,2} - 2 > 0)\}^2 - \sum_{j=2}^{4} |X_{j,1} - 1|\{I(A_j > 0) - I(X_{j,1} - 1 > 0)\}^2 + N(0,1).$$

Simulation Studies

- Optimal decisions:

  Scenario 1/3: $d_j^*(h_j) = \text{sign}(X_j - 1)$
  Scenario 2/4:
  $d_1^*(h_1) = \text{sign}(X_1 + X_2 - 2), d_j^*(h_j) = \text{sign}(X_j - 1), j = 2, 3, 4$

- Training data sample size $n = 300$.
- Testing data sample size 10000.
- $p = 5, 10$.
- 500 replications; tuning parameters selected via cross validation.
- Evaluate using the values of the estimated DTRs.

Simulation Studies

Table 1: Results

	Scenario	1	2	3	4
p=5	SDTR-IPW	9.92	9.40	9.38	8.86
	SDTR-AIPW	9.89	9.48	9.68	9.15
	Q learning	8.12	8.15	8.65	8.82
	Shared Q learning	9.74	9.25	9.40	9.19
p=10	SDTR-IPW	9.87	9.38	9.21	8.71
	SDTR-AIPW	9.85	9.56	9.63	9.04
	Q learning	8.14	8.12	8.53	8.65
	Shared Q learning	9.58	9.07	9.30	9.02

Simulation Studies

- Mimic PASS study
- Baseline variables: $(PSA_0, Coreratio_0) \sim$ a multivariate normal distribution with mean (5.2,13) and the covariance matrix diag(12.5,49); 5 other variables $\sim N(0,1)$
- If $PSA_{j-1} < 8$, do not intervene at stage j; otherwise, intervene, i.e., $A_j = 1$ with probability

$$\frac{\exp(4 - 0.2PSA_j - 0.01Coreratio_{j-1}}{1 + \ exp(4 - 0.2PSA_j - 0.01Coreratio_{j-1}))}$$

- The probability of reclassification at stage j is generated from

$$\frac{\exp(-2 + 0.08(PSA_j \geq 10)PSA_j - 0.5A_j)}{1 + \exp(-2 + 0.08(PSA_j \geq 10)PSA_j - 0.5A_j)}$$

Simulation Studies

Table 2: Results

SDTR-IPW	SDTR-AIPW	Q learning	Shared Q learning
35%	35%	42%	40%

Decision rule by SDTR-IPW:
$\hat{d}(x_j) = 26.7 - 3.7 PSA_j - 0.12 Coreratio_j$

Discussion

- Survival outcomes.
- Multiple treaments.
- Incorporating Cost.
- Landmark analysis.

Acknowledgement

IMA workshop organizers
Ruoqing Zhu
Yingye Zheng
Guanhua Chen