Interpretable treatment regimes via decision lists

Eric B. Laber

Department of Statistics, North Carolina State University

September 14, 2017

Acknowledgments

Thanks to

- IMA organizing committee
- NCSU/UNC AI and Precision Medicine Lab
- PR Department of Natural Resources
- National Science Foundation
- National Institutes of Health

Joint work with

- Students in Laber Labs
- Yichi Zhang
- John Drake
- Krishna Pacifici
- Brian Reich
- Marie Davidian
- Butch Tsiatis

Outline: talk in two parts

- Interpretable treatment regimes
- Open problems

Outline: talk in two parts

- **Interpretable treatment regimes**
- Open problems

Precision medicine

- "The right treatment for the right patient at the right time." –Mantra of precision medicine advocates
  - Widely recognized that best clinical care requires treatment decisions tailored to individual patient characteristics
  - Improve patient outcomes, reduce cost and patient burden
- Treatment regimes
  - Formalize clinical decision making as sequence of decision rules
    - One rule per stage of clinical intervention
    - Maps current patient info to recommended treatment
  - Optimal regime maximizes the mean of some cumulative clinical outcome if applied to population of interest

Ex. Treatment regime: mHealth for PTSD in cancer patients (PI S. Smith)

First stage decision rule

If distress ≥ 3 **then:** Cancer Distress Coach (CDC)

Else if PTSD symptom score ≥ 20 **then:** CDC

Else: usual care

Second stage decision rule

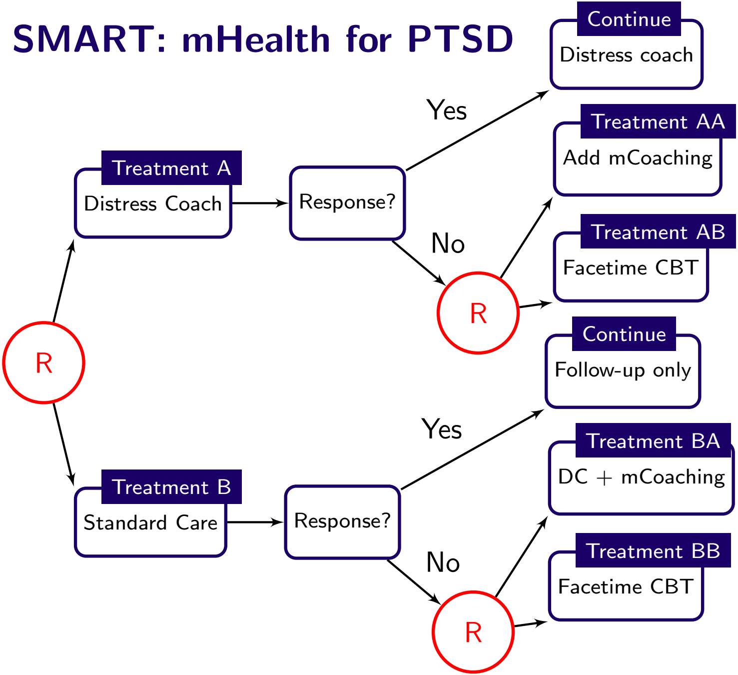
If responder **then:** continue first stage treatment

Else if using CDC and PSTD change ≥ 3 **then:** add mCoaching

Else if using CDC and distress ≥ 4 **then:** add FaceTime CBT

Else FaceTime CBT only

Ex. SMART: mHealth for PTSD

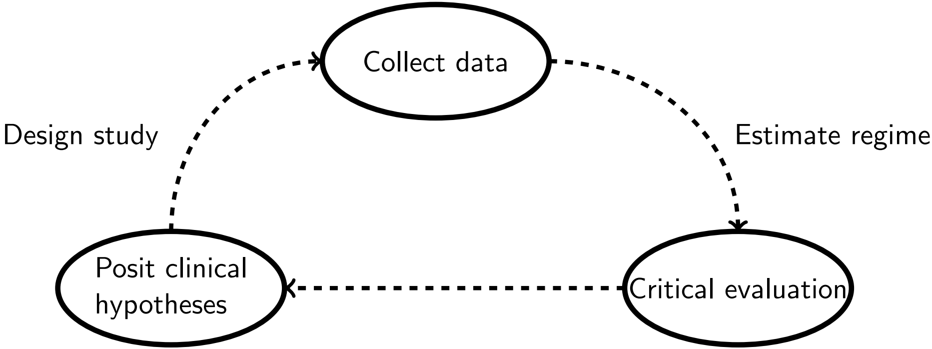


Estimation of treatment regimes: recent trend

- Current paradigm: assume estimated regime used to dictate treatment choice $\Rightarrow$ emphasize flexibility over interpretability
- Surge in machine-learning/semi-parametric methods
  - Direct-search with large-margin classifiers (Zhao et al. 2012, 2013ab, 2014; Kang et al., 2014; Zhao and L., 2015)
  - Generalized additive models (Moodie et al. 2014)
  - Nearest-neighbor methods (Zhou and Kosorok, 2016)
  - JSM JASA T&M (2016) and A&CS (2016, 2017) discussed papers about non-parametric estimation of treatment regimes (Xu et al. 2016; Chen et al., 2016, Rashid et al., 2017)

Treatment regimes and clinical research

- Reality: estimated regimes part of secondary, exploratory analyses ⇒ emphasize interpretability in domain context
- Treatment regimes for application in clinical practice requires longterm, joint-effort of statisticians and clinical scientists



List-based regimes

- Goal: estimation of an interpretable yet expressive regime
  - List-based rules are a sequence of if-then statements mapping logical clauses to treatment recommendations
  - Regime comprising list-based decision rules are immediately interpretable in a domain context

If c_1 **then:** a_1
Else if c_2 **then:** a_2
$\cdots$
Else if c_{L-1} **then:** a_{L-1}
Else: a_L

If distress ≥ 3 **then:** CDC
Else if PTSD ≥ 20 **then:** CDC
Else: usual care

List-based regimes cont'd

- Simple but powerful
  - Txt a if and only if all $\{c_1, \ldots, c_k\}$ hold
  - Txt a if any of $\{c_1, \ldots, c_k\}$ hold
  - Different treatment for each level of categorical variable
  - $\ldots$
- Restrict attention to clauses dictated by union or intersection of thresholded regions
  - $X_j > \tau_j$
  - $X_j \leq \tau_j$
  - $X_j > \tau_j$ and $X_k > \tau_k$
  - $X_j > \tau_j$ and $X_k \leq \tau_k$
  - $X_j > \tau_j$ or $X_k > \tau_k$
  - $X_j \geq \tau_j$ or $X_k \leq \tau_k$
  - $X_j \leq \tau_j$ and $X_k \leq \tau_k$
  - $X_j \leq \tau_j$ and $X_k > \tau_k$
  - $X_j \leq \tau_j$ or $X_k \leq \tau_k$
  - $X_j \leq \tau_j$ or $X_k > \tau_k$

List-based regimes and short-circuiting

- List-based regimes need not be unique

π
If c_1 **then:** a_1
Else if c_2 **then:** a_2
Else: a_3

π'
If $\neg c_1$ and $\neg c_2$ **then:** a_3
Else if $\neg c_1$ **then:** a_2
Else: a_1

- Choose regime that requires least patient burden/cost among equivalence class dictated by marginal mean outcome
  - Ex. π need not evaluate c_2 among patients satisfying c_1
  - Focus on estimating a single member of equivalence class

Setup and notation

- Observe $\{(\boldsymbol{X}_{1,i}, A_{1,i}, \ldots, \boldsymbol{X}_{T,i}, A_{T,i}, Y_i)\}_{i=1}^{n}$ *i.i.d.* from P
  - $\boldsymbol{X}_t \in \mathbb{R}^{p_t}$ subject info during stage t
  - $A_t \in \mathcal{A}_t$ assigned txt at stage t
  - $Y \in \mathbb{R}$ outcome coded so that higher is better
- Define history $\boldsymbol{H}_1 = \boldsymbol{X}_1$, $\boldsymbol{H}_t = (\boldsymbol{H}_{t-1}^{\intercal}, A_{t-1}, \boldsymbol{X}_t^{\intercal})^{\intercal}$, $t \geq 2$
- Treatment regime $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)$ where

$$\pi_t : \operatorname{supp} \boldsymbol{H}_t \to \operatorname{supp} A_t,$$

patient presenting with $\boldsymbol{H}_t = \boldsymbol{h}_t$ assigned txt $\pi_t(\boldsymbol{h}_t)$

Optimal regime via potential outcomes

- Define $\overline{a}_t = (a_1, \ldots, a_t)$ and set of potential outcomes

$$W^* = \left\{ \boldsymbol{H}_1, \boldsymbol{H}_2^*(a_1), \ldots, \boldsymbol{H}_T^*(\overline{a}_{T-1}), Y^*(\overline{a}_T) \,:\, \overline{a}_T \in \otimes_{t=1}^{T} \mathcal{A}_t \right\}$$

- Potential outcome under a regime $\boldsymbol{\pi}$

$$Y^*(\boldsymbol{\pi}) = \sum_{\overline{a}_T} Y^*(\overline{a}_T) \prod_{t=1}^{T} 1\left[\pi_t \left\{ \boldsymbol{H}_t^*(\overline{a}_{t-1}) \right\} = a_t\right]$$

- Optimal regime in class Π satisfies $\mathbb{E}Y^*(\boldsymbol{\pi}^{\mathrm{opt}}) \geq \mathbb{E}Y^*(\boldsymbol{\pi})$, where $\boldsymbol{\pi}^{\mathrm{opt}}$, $\boldsymbol{\pi} \in \Pi$

Characterizing the optimal regime

- For each $t = 1, \ldots, T$ assume

  (C1) Sequential ignorability: $A_t \perp W^* \Big| \boldsymbol{H}_t$

  (C2) Positivity: $P(A_t = a_t \Big| \boldsymbol{H}_t) \geq \epsilon$ wp1 for some $\epsilon > 0$

  (C3) Consistency: $\boldsymbol{H}_t = \boldsymbol{H}_t^*(\overline{A}_t)$, $Y = Y^*(\overline{A}_T)$

  (C4) Stable unit treatment value assumption (SUTVA)

- (C1)-(C3) are satisfied by construction in a sequential multiple assignment randomized trial (SMART; Murphy, 2005)

Characterizing the optimal regime cont'd

- Let Π_t denote space of list-based rules on $\operatorname{supp} \boldsymbol{H}_t$ for $t \geq 1$
- Assume (C1)-(C4) then
  - Define

$$Q_T(\boldsymbol{h}_T, a_T) = \mathbb{E}\left(Y \middle| \boldsymbol{H}_T = \boldsymbol{h}_T, A_T = a_T\right)$$

    then $\pi_T^{\text{opt}} = \arg\max_{\pi_T \in \Pi_T} \mathbb{E} Q_T \left\{\boldsymbol{H}_T, \pi_T(\boldsymbol{H}_T)\right\}$

  - Recursively, define

$$Q_t(\boldsymbol{h}_t, a_t) = \mathbb{E}\left[Q_{t+1}\left\{\boldsymbol{H}_{t+1}, \pi_{t+1}^{\text{opt}}(\boldsymbol{H}_{t+1})\right\} \middle| \boldsymbol{H}_t = \boldsymbol{h}_t, A_t = a_t\right]$$

    then $\pi_t^{\text{opt}} = \arg\max_{\pi_t \in \Pi_t} \mathbb{E} Q_t \left\{\boldsymbol{H}_t, \pi_t(\boldsymbol{H}_t)\right\}$

Q-learning with policy search

- Let $\mathcal{Q}_t$ denote postulated class of models for $Q_t(\boldsymbol{h}_t, a_t)$
- Estimation algorithm
  - Compute
$$\widehat{Q}_T = \arg\min_{Q_T \in \mathcal{Q}_T} \left[\mathbb{P}_n \left\{Y_T - Q_T(\boldsymbol{H}_T, A_T)\right\}^2 + \mathcal{P}_T(Q_T)\right]$$
and subsequently $\widehat{\pi}_T = \arg\max_{\pi_T \in \Pi_T} \mathbb{P}_n \widehat{Q}_T \left\{\boldsymbol{H}_T, \pi_T(\boldsymbol{H}_T)\right\}$
  - Recursively, compute
$$\widehat{Q}_t = \arg\min_{Q_t \in \mathcal{Q}_t} \left\{\mathbb{P}_n \left[\widehat{Q}_{t+1}\left\{\boldsymbol{H}_{t+1}, \widehat{\pi}_{t+1}(\boldsymbol{H}_{t+1})\right\} - Q_t(\boldsymbol{H}_t, A_t)\right]^2 + \mathcal{P}_t(Q_t)\right\}$$
and subsequently $\widehat{\pi}_t = \arg\max_{\pi \in \Pi_t} \mathbb{P}_n \widehat{Q}_t \left\{\boldsymbol{H}_t, \pi_t(\boldsymbol{H}_t)\right\}$

Computation

- Π_t class of list-based rules $\Rightarrow$ computation of $\arg\max_{\pi_t \in \Pi_t} \mathbb{P}_n \widehat{Q}_t \left\{ \boldsymbol{H}_t, \pi_t\left(\boldsymbol{H}_t\right) \right\}$ discrete opt. problem

- Basic idea: stepwise splitting procedure in spirit of CART
  - Optimize marginal mean outcome plus complexity penalty
  - Stopping criteria based on significance test

Lemma

Let $m_t = \#\mathcal{A}_t$, $d_t = dim(\boldsymbol{h}_t)$, and L_t is the maximum list depth. Then, the time complexity for computing $\widehat{\boldsymbol{\pi}} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_T)$ is

$$O\left\{ n \log(n) \sum_{t=1}^{T} L_t m_t d_t^2 \right\}.$$

Convergence of Q-learning with policy search

- Overview of technical assumptions
  - (A1) $\boldsymbol{H}_t$ and Y are bounded with probability one
  - (A2) Q-functions are sufficiently smooth (weakly differentiable in sense of Eberts and Steinwart, 2013, Defn. 2.1 with L_P norm)
  - (A3) Margin condition on the treatment effects
  - (A4) Estimation via kernel ridge regression with Gaussian kernel
  - (A5) Rate conditions on complexity penalty in splitting criteria

Convergence of Q-learning with policy search cont'd

Lemma

Assume (C1)-(C4) and (A1)-(A5). Then, for each $t = 1, \ldots, T$ $P\left\{\widehat{\pi}_t(\boldsymbol{H}_t) \neq \pi_t^{\mathrm{opt}}(\boldsymbol{H}_t)\middle|\mathbb{P}_n\right\}$ *converges to zero in probability.*

Corollary

Assume (C1)-(C4) and (A1)-(A5). Then $\mathbb{E}Y^*(\boldsymbol{\pi}^{\mathrm{opt}}) - \mathbb{E}Y^*(\widehat{\boldsymbol{\pi}})$ *converges to zero.*

Q-learning with policy search discussion

- Divorce form of Q-function from class of regimes
- Decision lists offer interpretable yet expressive class of regimes
  - Diagnostics via comparison of $\arg\max_{a_t} \widehat{Q}_t(\boldsymbol{h}_t, a_t)$ with $\widehat{\pi}_t(\boldsymbol{h}_t)$
  - Short-circuiting can save cost/burden
- Potential drawbacks and open problems
  - Sub-parametric rates of convergence (not shown)
  - Methods for inference and uncertainty quantification needed

Simulation experiments: overview

- Settings taken from Zhao et al. (2015) and Murphy (2003)
- Basic framework
  - Normal covariates with AR-1 updates
  - Txt assign. mimics both randomized and observational study
  - Mean outcome nonlinear in history

Model	Stages	Covariates	Propensity	Source
1	2	50	Constant	Zhao et al.
2	2	50	History-dependent	Zhao et al.
3	3	3	Constant	Zhao et al.
4	10	10	History-dependent	Murphy

Simulation experiments: overview cont'd

- Performance measure $\mathbb{E}\{\mathbb{E}Y^*(\widehat{\boldsymbol{\pi}})\}$
  - Q-learning + Policy search with decision lists
  - Non-parametric Q-learning with random forests
  - Linear Q-learning with lasso penalty
  - Backward outcome weighted learning (BOWL)
- Additional numerical details
  - Use 1K Monte Carlo replications
  - Performance estimated with test set of size 10K
  - Consider training sets of size $n = 100$ and $n = 400$
  - Maximum list-depth set to $L_t = 5$ for all t

Simulation experiments: results

Model	n	Decision List	Q-RF	Q-Lasso	BOWL
1	100	6.63	**6.70**	6.50	**6.70**
1	400	**6.94**	6.70	6.66	6.70
2	100	3.66	3.41	**3.68**	2.68
2	400	3.73	3.71	**3.75**	3.19
3	100	**14.49**	12.94	5.74	8.12
3	400	**18.60**	18.02	8.42	13.62
4	100	**23.68**	17.83	12.39	NA
4	400	**26.80**	24.73	16.58	NA

Simulation experiments: summary

- Q-learning + policy search with decision lists performs favorably to parametric and non-parametric competitors
  - Parsimony of list provides automatic variable selection
  - Did not consider dense but weak signals
  - Significantly more interpretable than alternatives
  - R package **decisionList** freely available on CRAN

Discussion of list-based optimal regimes

- Estimation of optimal treatment regimes often part of secondary, hypothesis-generating analyses
- Q-learning + policy search with decision lists is a powerful tool for estimating high-quality interpretable regimes
  - Computationally efficient
  - Consistent under weak conditions
  - Because of underlying regression framework extends to high-dimensional data, censored data, continuous txts etc.
  - Applies to infinite horizon problems modeled as Markov Decision Processes through Bellman optimality est. equations

Outline: talk in two parts

- Interpretable treatment regimes
- Open problems

Outline: talk in two parts

- Interpretable treatment regimes
- **Open problems**

Research-practice gap

- Active methods work
  - Machine learning
  - Infinite horizon problems
  - Clustered designs and regimes
- Most active methods work under framework considered here
  - Single outcome
  - Single decision maker
  - Stationary preferences
  - ...

Research-practice gap cont'd

- Need decision-support systems that faithfully reflect the complexities and realities of clinical decision making
  - Heterogeneous patient preferences
  - Multiple stakeholders with different objectives
  - Implementation costs
  - Non-stationarity
- Need to expand current mathematical framework

Individual patient preference

- Clinical decision-making requires balancing multiple, possibly competing outcomes
  - Ex., side-effects and efficacy
  - Ex., cost and local availability
  - ...
- Preferences across outcomes can vary across patients and within patients over time
  - Ex., become averse to side-effect after experiencing it
  - Ex., lethargy and lifestyle
  - Ex., hypertension and weight gain

Composite outcomes

- Idea: create a single composite summary across all outcomes
  - Assumes homogeneous preference across all patients
  - Thall et al. have applied this approach successfully in cancer
- Potential problems
  - Preferences vary across patients
  - Preferences change over time

Preference elicitation

- Idea! ask patients about their preferences
  - Administer questionnaire, link answers to utility function through latent preference model, e.g., item response model
  - Questionnaire becomes part of treatment package
  - Leads to regime based on individual patient preference
- Potential problems
  - Need high-quality instruments
  - Patients unable/unwilling to communicate preferences

Set-valued regimes

- Idea! screen-out treatment choices that are dominated across all outcomes under consideration
  - Recommend set of treatment options
  - Includes optimal regime for large class of preferences
  - Avoid elicitation or composite outcomes
- Potential problems
  - Little guidance on how to choose among set
  - Individual preference only incorporated indirectly

Informing, not dictating

- Supplement domain expertise with data
    - Distill data into information
    - Communicate strength of evidence provided by data
- Health communication is notoriously difficult
    - Health literacy varies widely
    - Shared decision model can be time and resource intensive

Informing, not dictating cont'd

- Methodology must be informed by clinical need
  - Not all statistical/methodological challenges
  - Information communication tools needed
- Many important open problems
  - Extensions to more exotic data types/gen models
  - Research-translation

Thank you.

laber@stat.ncsu.edu