Section II: developing a marker-based treatment rule

- Treatment decision rule
- Optimal Treatment Regime
- Estimating optimal treatment regime
  - Q-learning (Regression modeling)
  - A-learning (Advantage learning)
  - Direct optimization

Treatment Decision Rule

Treatment Decision Rule

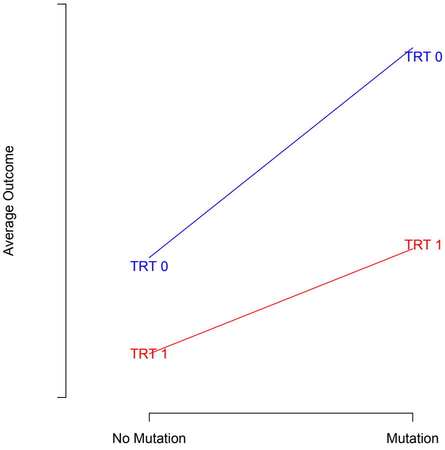
Assume: There is a clinical outcome by which treatment benefit can be assessed

- Survival time, CD4 count, indicator of no myocardial infarction within 30 days, ...
- Lower outcomes are better

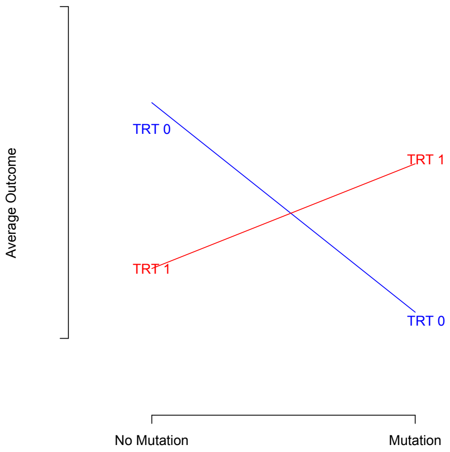
Intuitively: Rules should depend on characteristics (variables , covariates) that exhibit a qualitative interaction with treatment

- Tailoring variables/ treatment selection biomarker

Tailoring Variables



Tailoring Variables



Statistical Framework

Simplest setting: A single decision with two treatment options

Observed data: (D_i, X_i, A_i), $i = 1, \ldots, n$, independently and identically distributed (iid)

- D_i outcome, X_i baseline covariates, $A_i = 0, 1$ treatment received

Treatment regime: An individualized treatment rule

- A function $d : X \rightarrow \{0,\ 1\}$

Simple example

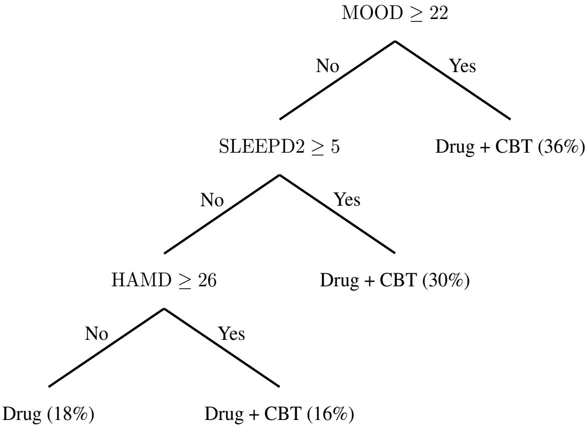
Which treatment to give patients who present with *nonpsychotic Chronic Major Depressive Disorder*?

- *Options:* Nefazodone (Drug) or Drug + Cognitive Behavioral Therapy (CBT)
- *Data:* 681 patients in the Nefazodone-CBASP clinical trial (Keller et al., 2000)
- *Available information:* 50 prognostic variables, e.g., age, baseline depression score
- *Outcome:* Hamilton Rating Scale for Depression

Keller et al. (*NEJM* 2000)

Simple example

A regime example:



Simple example

- Even simpler example: If MOOD $\geq 22 \Rightarrow$ Drug + CBT; otherwise $\Rightarrow$ Drug
- *Mathematically:* The formal rule is

$$d(\text{MOOD}) = I(\text{MOOD} > 22)$$

- Optimal regime : If followed by *all patients* in the population, would lead to *smallest expected outcome* among all regimes in $\mathcal{D}$

Simple example

- Even simpler example: If MOOD $\geq 22 \Rightarrow$ Drug + CBT; otherwise $\Rightarrow$ Drug

- *Mathematically:* The formal rule is

$$d(\text{MOOD}) = I(\text{MOOD} > 22)$$

Optimal Treatment Regime

Considerations

- Identify the subset that are good tailoring variables
- Regime $d(X)$: a function of X
- There are many possible regimes d:

  $\mathcal{D}$: class of all possible treatment regimes

- Can we find the optimal treatment regime in $\mathcal{D}$?
- Optimal regime : If followed by *all patients* in the population, would lead to *smallest expected outcome* among all regimes in $\mathcal{D}$

Potential Outcomes

Single decision: Possible treatment options $a \in \{0, 1\}$

- For a *randomly chosen patient* from the population, define the *random variable* $D(a)$ = the outcome the patient *would experience* if s/he *were to receive* treatment option a
- "*Potential outcome*"
- E.g., $D(1)$= the outcome a patient would have if s/he were given treatment 1, and similarly for $D(0)$

Optimal regime

- Potential outcome for a regime: $D(d)$ = the outcome a patient would have if s/he received treatment *according to a regime* $d \in \mathcal{D}$

- E.g., if the patient has information X

$$D(d) = D(1)I\{d(X) = 1\} + D(0)I\{d(X) = 0\}$$

- $E\{D(d)|X = x\}$ is the expected outcome for a patient with information x if s/he were to receive treatment according to regime $d \in \mathcal{D}$.

- $E\{D(d)\} = E[E\{D(d)|X = x\}]$ is the expected outcome for the population if all patients were to receive treatment according to regime $d \in \mathcal{D}$.

Optimal regime

- The optimal treatment regime $d^* \in \mathcal{D}$ minimizes the expected outcome

$$d^* = \underset{d \in \mathcal{D}}{\operatorname{argmin}}\, E\{D(d)\}$$

- That is, $E\{D(d^*)\} \leq E\{D(d)\}$ for all $d \in \mathcal{D}$
- Also, $E\{D(d^*)|X = x\} \leq E\{D(d)|X = x\}$ for all $d \in \mathcal{D}$ and for all patient subgroups defined by x.

Identifying the optimal treatment regime

- Discover optimal regimes based on data.
- The optimal regime is defined in terms of potential outcomes, not the observed data
- Possible under certain assumptions

Potential Outcomes

Positivity: $P(A = a|X = x)$ strictly positive for all x, i.e., $0 < P(A = 1|X) < 1$ almost surely, usually satisfied in a randomized trial

Consistency: $D(a) = D$ whenever treatment a is actually received, usually satisfied in a randomized trial

No unmeasured confounders: Assume that

$$D(0), D(1) \perp\!\!\!\perp A|X$$

- X contains all information used to assign treatments in the data
- Automatically satisfied for data from a randomized trial

Potential Outcomes

- Implies that

$$\begin{aligned} E\{D(1)\} &= E[E\{D(1)|X\}] \\ &= E[E\{D(1)|A=1,X\}] \\ &= E\{\,E(D|A=1,X)\,\} \end{aligned}$$

and similarly for $E\{D(0)\}$

Optimal Regime

- Under certain assumptions (positivity, consistency and no unmeasured confounders)

$$
\begin{aligned}
E\{D(d)\} &= E[E\{D(d)|X\}] \\
&= E[E(D(1)|A=1,X)I\{d(X)=1\} \\
&\quad +E(D(0)|A=0,X)I\{d(X)=0\}] \\
&= E[E(D|A=1,X)I\{d(X)=1\} \\
&\quad +E(D|A=0,X)I\{d(X)=0\}].
\end{aligned}
$$

- Optimal treatment regime can be derived in terms of observed outcomes.

Optimal Regime

- Optimal Individualized Treatment Rule:

$$E(D|X, A=1) \leq E(D|X, A=0) \Rightarrow d^*(X) = 1$$
$$E(D|X, A=1) > E(D|X, A=0) \Rightarrow d^*(X) = 0$$

  If $E(D|X, A)$ were known, we could find d^*.

- Problem: $E(D|X, A)$ is unknown.

Estimating optimal treatment regime

Q-learning (Regression modeling)

- If we had a sample of data $(X_i, A_i, D_i), i = 1, \ldots, n$, we can posit a regression model

$$E(D|A, X) = \mu(A, X; \beta)$$

and estimate $\hat{\beta}$ using e.g. least squares/logistic regression/cox regression.

- The estimator for the optimal treatment regime

$$\hat{d}_n(x) = I\{\mu(1, x; \hat{\beta}_n) \geq \mu(0, x; \hat{\beta}_n)\},$$

- $\hat{E}(D(d)) = \hat{E}[\hat{E}(D|A = 1, X, D(d) = 1)I(D(d) = 1) + \hat{E}(D|A = 0, X, D(d) = 0)I(D(d) = 0)]$

Q-learning (Regression modeling)

- *Estimate* $E\{D(d)\}$ by

$$n^{-1}\sum_{i=1}^{n}[\mu(1, X_i, \hat{\beta}_n)d(X_i) + \mu(0, X_i, \hat{\beta}_n)\{1 - d(X_i)\}]$$

- Estimate for $E\{D(d^*)\}$

$$n^{-1}\sum_{i=1}^{n}[\mu(1, X_i, \hat{\beta}_n)\hat{d}_n(X_i) + \mu(0, X_i, \hat{\beta}_n)\{1 - \hat{d}_n(X_i)\}],$$

  where d^* is estimated by $\hat{d}_n$.

- $\mu(A, X; \beta)$ may be misspecified.

A-learning (Advantage learning)

- Advantage learning (A-learning) is a more robust method for estimating the optimal treatment regime
- One does not need to know the entire function $E(D|A, X)$.
- It suffices to only consider

$$\Delta(X) = E(D|A = 0, X) - E(D|A = 1, X)$$

- $d^*(x) = I\{\Delta(x) \geq 0\}$.

Advantage function

- A function of A, X
- Advantage: the increase in the expected response had the individual received the optimal treatment
- $\Delta(X) < 0$, $d^*(X) = 0$
  - $A = 0$, no advantage
  - $A = 1$, the advantage in the mean response would be $\Delta(X)$
- $\Delta(X) \geq 0$, $d^*(X) = 1$
  - $A = 1$, no advantage
  - $A = 0$, the advantage in the mean response would be $-\Delta(X)$
- Advantage function: $-\Delta(X)[I\{\Delta(X) \geq 0\} - A]$

Murphy (*JRSSB*, 2003)

Advantage learning

- $E\{D(d^*)\} = E\{E(D - \Delta(X)[I\{\Delta(X) \geq 0\} - A]|X)\}$.
- Consider models for $\Delta(x), \Delta(x;\psi)$
- Estimate $E\{Y(d^*)\}$ by

$$n^{-1}\sum_{i=1}^{n}(D_i - \Delta(X_i;\hat{\psi}_n)[I\{\Delta(X_i;\hat{\psi}_n) \geq 0\} - A_i]),$$

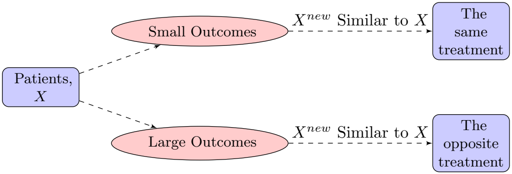
$\hat{\psi}_n$ can be estimated using semiparametric techniques.

Direct Optimization: Classification Perspective

Intuition: Classification

Given a new observation X^{new}, predict the class label $d^{*,\text{new}}$.

- No direct information on the true class labels, d^*.
- Can we assign the right treatment based on the observed information?



Directly estimating the Optimal Regime

Thoughts: Minimize a "*good*" estimator for $E\{D(d)\}$

- $\pi(X) = P(A = 1|X)$ is the propensity score for treatment
- $\pi(X)$ known for randomized study; Can also be estimated using the data $(A_i, X_i), i = 1, \ldots, n$, e.g., logistic regression $\pi(X; \gamma)$ and estimate γ by $\hat{\gamma}$.
- The propensity of receiving treatment consistent with $d(X)$

$$
\begin{aligned}
P\{d(X)|X\} &= P(A = d(X)|X) \\
&= E[Ad(X) + (1 - A)\{1 - d(X)\}|X] \\
&= \pi(X)d(X) + \{1 - \pi(X)\}\{1 - d(X)\}.
\end{aligned}
$$

Directly estimating the Optimal Regime

Identify estimators for $E\{D(d)\}$:

- Inverse probability weighted estimator

$$IPWE(d) = n^{-1} \sum_{i=1}^{n} \frac{I\{A_i = d(X_i)\} D_i}{P\{d(X_i) | X, \hat{\gamma}\}}. \tag{1}$$

- Consistent for $E\{D(d)\}$ if $\pi(X;\gamma)$, and hence $P\{d(X_i)|X,\hat{\gamma}\}$ is correctly specified

Outcome Weighted Learning (OWL)

- Minimize $IPWE(d)$ (1)
- For any rule d, $2d(X) - 1 = \text{sign}\{f(X)\}$ for some function f.
- Hence, minimize:

$$n^{-1} \sum_{i=1}^{n} \frac{-D_i}{P\{d(X_i)|X, \hat{\gamma}\}} I\{(2A_i - 1) \neq \text{sign}(f(X_i)\}.$$

- Can be treated as recoding $\mathcal{A} = \{-1, 1\}$

Zhao et al. (*JASA* 2012)

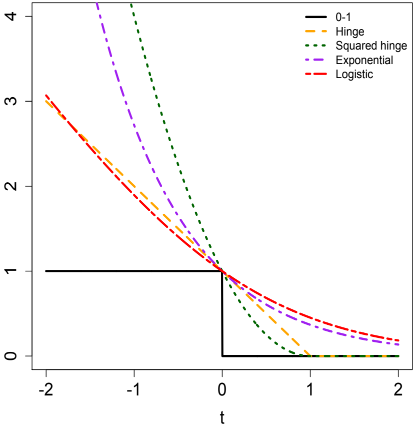
Convex Surrogate Losses for Computation Relaxation

Computation challenges: non-convexity and discontinuity of 0-1 loss.

Replace 0-1 loss by convex surrogate loss

- Hinge loss, $\phi(t) = \max(1 - t, 0)$.
- Exponential loss, $\phi(t) = e^{-t}$.
- Logistic loss, $\phi(t) = \log(1 + e^{-t})$.
- Squared hinge loss, $\phi(t) = \{\max(1 - t, 0)\}^2$.

Convex Surrogate Losses



Outcome Weighted Learning

Objective Function: Regularization Framework

$$\min_f \frac{1}{n}\sum_{i=1}^{n} \frac{-D_i}{P\{d(X_i)|X,\hat{\gamma}\}}\phi\{(2A_i-1)f(X_i))\} + \lambda_n\|f\|^2. \qquad (2)$$

- $\|f\|$ is some norm for f, and λ_n controls the severity of the penalty on the functions.
- A linear decision rule: $f(X) = X^T\beta + \beta_0$, with $\|f\|$ as the Euclidean norm of β.

Outcome Weighted Learning

- Estimated individualized treatment rule:

$$\hat{d}_n(X) = \text{sign}(\hat{f}_n(X)),$$

  where $\hat{f}_n$ is the solution to (2).

- Variable selection is possible, e.g., change $\|f\|^2$ to $\|f\|$.

Efficient Augmentation and Relaxation Learning

- Doubly robust augmented inverse probability weighted estimator

$$
\begin{aligned}
AIPWE(d) &= n^{-1}\sum_{i=1}^{n}\left\{\frac{I\{A_i = d(X_i)\}D_i}{P\{d(X_i)|X,\hat{\gamma}\}}\right. \qquad (3)\\
&- \left.\frac{I\{A_i = d(X_i)\} - P\{d(X_i|X,\hat{\gamma}\}}{P\{d(X_i)|X,\hat{\gamma}\}} m(X_i;\hat{\beta})\right\},
\end{aligned}
$$

where

$$m(X;\beta) = \mu(1,X;\beta)d(X) + \mu(0,X;\beta)\{1 - d(X)\}$$

is a model for $E\{D(d)|X\}$ and $\mu(A,X;\beta)$ is a model for $E(D|A,X)$

- Consistent if either $\pi(X;\gamma)$ or $\mu(A,X;\beta)$ is correct

Efficient Augmentation and Relaxation Learning

- Outcome weighted learning is a special case with $m(X_i; \hat{\beta}) \equiv 0$
- $AIPWE(d)$ is more efficient than $IPWE(d)$ for estimating $E\{D(d)\}$
- A similar solution: replacing 0-1 loss with a convex loss function

Direct optimization: Optimal Restricted Regime

A class of regimes

$$d(X, \beta) = I\{\mu(1, X; \beta) > \mu(0, X; \beta)\},$$

indexed by β,

- E.g.,

$$E(D|A, X) = \exp\{1 + X_1 + 2X_2 + 3X_1X_2 + A(1 - 2X_1 + X_2)\}$$

$$\Rightarrow d^*(X) = I(X_2 \geq 2X_1 - 1)$$

Zhang et al. (*Biometrics* 2012)

Direct optimization: Optimal Restricted Regime

- *Posit*

$$\mu(A, X; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$$

- The regimes $I\{\mu(1, X; \beta) > \mu(0, X; \beta)\}$ define a *class* $\mathcal{D}_\eta$ with elements

$$I(X_2 \geq \eta_1 X_1 + \eta_0) \text{ or } I(X_2 \leq \eta_1 X_1 + \eta_0), \quad \eta_0 = -\beta_3/\beta_5,\, \eta_1 = -\beta_4/\beta_5$$

depending on the sign of β_5

- The optimal regime in this case is contained in $\mathcal{D}_\eta$

Optimal Restricted Regime

Consider directly regimes of the form $\mathcal{D}_\eta = \{d(X, \eta)\}$ indexed by η

- Write $d_\eta(X) = d(X, \eta)$, e.g., $d(X, \eta) = I(X_2 \leq \eta_1 X_1 + \eta_0)$

- Defined based on clinical practice, cost, and interpretability, without reference to a regression model.

- d^* may or may not be in $\mathcal{D}_\eta$ but still of interest

- Optimal restricted regime $d^*_\eta(X) = d(X, \eta^*)$,

$$\eta^* = \underset{\eta}{\text{argmin}}\; E\{D(d_\eta)\}$$

- Estimate the optimal restricted regime by estimating η^*

Estimating the Optimal Restricted Regime

- Minimize a "*good*" estimator for $E\{D(d_\eta)\}$ in η:
- Estimators $\hat{\eta}$ for η^* obtained by minimizing $IPWE(d_\eta)$ or $AIPWE(d_\eta)$ in η
- Non-smooth functions of η; must use suitable optimization techniques (RGENOUD package in R)
- Estimators for $E\{D(d_\eta)\}$

$$IPWE(d_{\hat{\eta}_{ipwe}}) \text{ or } AIPWE(d_{\hat{\eta}_{aipwe}})$$

Can calculate standard errors

Depression Data

- Compare drug therapy ($A = 0$) with drug + behavioral therapy ($A = 1$)
- Five covariates: Age, Gender, HAMABase (pre-treatment total Hamilton Anxiety Rating Scale score), Sleep (sleep disturbance score), Mood (mood cognition score)
- Response: 24-item Hamilton Rating Scale for Depression
- Number of patients: 436

Analyzing Depression Data

- Q-learning: $D \sim 1 + X + A + XA$
- Efficient Augmentation and Relaxation Learning:
  - Logistic loss: $\phi(t) = \log(1 + e^{-t})$
  - Outcome model: $D \sim 1 + X + A + XA$
  - Propensity model: $A \sim X$

Results

- Q-learning: $\hat{d}(X) = I(-0.83 + 0.01 Age - 0.55 Gender + 0.06 HAMABase + 0.01 Sleep - 0.04 Mood < 0)$.
- Efficient Augmentation and Relaxation Learning: $\hat{d}(X) = I(-0.94 + 0.00 Age - 0.33 Gender + 0.05 HAMABase + 0.02 Sleep - 0.01 Mood < 0)$.