Statistical Learning
Optimal Dynamic Treatment Regimes

Outline

Point exposure case:

- Static vs dynamic regime
- Defining optimal point dynamic regime
- Estimating optimal dynamic regime: Outcome Regression vs IPW estimation
- Optimal dynamic regime as a classification problem

Introduction

- In health sciences, there is growing interest in leveraging heterogeneity in treatment effects to tailor treatment assignment to maximize mean outcome for each patient and therefore for average outcome for overall population.
- Heterogeneity may be due to:
    - Genetic/genomic profile
    - Demographic characteristics
    - Physiological characteristics
    - Environment, lifestyle factors
    - Medical history, adverse reactions, adherence to prior treatment
    - etc...

Introduction

- Basic premise of precision medicine: A patient's characteristics are implicated in which treatment option(s) he/she should receive
- Clinical practice: Clinicians make a series of treatment decisions over the course of a patient's disease or disorder
  - Key decision points in the disease/disorder process
  - Multiple treatment options at each

Introduction

How are treatment decisions made?

- Clinical judgment, practice guidelines
- Synthesis of all information on a patient up to the point of a decision to determine next treatment action from among the possible options
- Goal: Make the "best" decisions so as to achieve the most beneficial outcome for this patient

Introduction

- Precision medicine: Formalize clinical decision-making and make it evidence-based
- At any decision point: Would like a formal rule that takes as input all available information on the patient to that point and outputs a recommended treatment action from among the possible, feasible options.

Cancer example

- Two decision points:
  - Decision 1: Induction chemotherapy (2 options: C_1, C_2)
  - Decision 2: Maintenance treatment for patients who respond (2 options: M_1, M_2)
    Salvage chemotherapy for those who don't respond (2 options: S_1, S_2)
- Examples of rules: Acute leukemia
  - Decision 1: If age$<$ 50 and WBC$<$10.0, give chemotherapy C_2, otherwise, give C_1.
  - Decision 2: If patient responded and baseline WBC$<$11.2, current WBC$<$10.5, no grade 3+ hematologic adverse event, current ECOG Performance Status 2, give maintenance M_1, otherwise, give M_2; otherwise if patient did not respond and age$>$60, current WBC$<$11.0, ECOG 2 give salvage S_1, otherwise, give S_2.

Treatment regime

- Treatment regime: A set of decision rules, each corresponding to a decision point.
- It defines an algorithm for treating an individual patient
    - Static rule: Recommended treatment action does not depend on accrued information
    - Dynamic rule: Recommended treatment action varies depending on accrued information
    - Dynamic treatment regime, adaptive treatment strategy, adaptive intervention

Static vs Dynamic regime

- Static regime: Give C_1 until end of study
- Dynamic treatment regimes:
  - Simplest:
    Decision 1: Give C_1
    Decision 2: If response, give M_2, if nonresponse, give S_1
  - More individualized and complex as rules incorporate available accrued patient information ("tailoring variables")

Two stage example

- At baseline: Information L_1, accrued information $H_1 = L_1 \in \mathcal{H}_1$
- Decision 1: Set of options $\mathcal{A}_1 = \{C_1, C_2\}$; rule 1: $d_1(h_1) : \mathcal{H}_1 \to \mathcal{A}_1$
- Between Decisions 1 and 2: Collect additional information L_2, including responder status
- Accrued information $H_2 = (L_1\,,\text{ chemotherapy at decision } 1, L_2) \in \mathcal{H}_2$
- Decision 2: Set of options $\mathcal{A}_2 = \{M_1, M_2, S_1, S_2\}$; rule 2: $d_2(h_2) : \mathcal{H}_2 \to \{M_1, M_2\}$ (responder), $d_2(h_2) : \mathcal{H}_2 \to \{S_1, S_2\}$ (nonresponder)
- Treatment regime: $d = (d_1, d_2)$
- Generalizes to K stages.

K stages

- Baseline information $L_1 \in \mathcal{L}_1$, intermediate information/time varying factors $L_k \in \mathcal{L}_k$ between Decisions $k-1$ and $k, k = 2, \ldots, K$
- Set of treatment options $\mathcal{A}_k$ at decision k, elements $a_k \in \mathcal{A}_k$
- Accrued information or history $H_1 = L_1 \in \mathcal{H}_1, \ldots, H_k = (L_1, A_1, \ldots, L_{k-1}, A_{k-1}, L_k) \in \mathcal{H}_k, k = 2; \ldots, K;$
- Decision rules $d_1\left(h_1\right); d_2\left(h_2\right), \ldots, d_K\left(h_K\right), d_k : \mathcal{H}_k \rightarrow \mathcal{A}_k$
- Treatment regime: $d = \left(d_1, d_2, \ldots, d_K\right)$

Challenges

- Clearly there is an infinitude of possible regimes d
- $\mathcal{D} =$ class of all possible treatment regimes
- Can we find the "best" set of rules, i.e., the "best" treatment regime in $\mathcal{D}$?
- How do we define "best"?

Optimal treatment regime

Intuitively: An optimal treatment regime $d_{opt} = (d_{opt,1}, \ldots, d_{opt,K})$ should satisfy:

- If a patient with history h_1 at baseline were to receive treatment options at all K decisions according to the rules in d_{opt}, his/her expected outcome would be as large as possible.
- Larger than the expected outcome for a patient with h_1 if he/she were to receive treatment options at all K decisions according to any other strategy for selecting treatments
- If all patients in the population were to receive treatment options at all K decisions according to d_{opt}, the population average outcome would be as large as possible.

Individualized treatment regime

Single occasion $K = 1$:

- Baseline information L_1
- Set of treatment options $\mathcal{A}_1$, elements $a_1 \in \mathcal{A}_1$
- Accrued information or history $H_1 = L_1 \in \mathcal{H}_1$
- Decision rule $d_1\left(h_1\right), d_1 : \mathcal{H}_1 \to \mathcal{A}_1$

Potential outcomes

- Recall that $Y(a)$ is the potential outcome a person would experience had he/she taken treatment $a \in \mathcal{A}_1$.
- We then defined the population ATE as $E(Y(1)) - E(Y(0))$ for binary a.
- We will now also consider the potential outcome $Y(d)$ a person would experience had he/she followed the treatment regime d. If $\mathcal{A}_1 = \{0, 1\}$, then for regime $d(h_1) = d_1(h_1)$

$$d_1(h_1) \text{ is either 0 or 1}$$

- Furthermore,

$$Y(d) = Y(1)d(H_1) + Y(0)(1 - d(H_1)) \text{ a.s.}$$

and

$$\mathcal{V}(d) = E[Y(d)]$$

is the population average outcome if all patients in the population were to receive treatment according to d.

Optimal dynamic regime

- Formally: For any $d \in \mathcal{D}$
  - $E\left[Y(d) \mid H_1 = h_1\right]$ is the expected outcome for a patient with history h_1 if he/she were to receive treatment according to the rule d in $\mathcal{D}$
  - $E\left[E\left[Y(d) \mid H_1\right]\right]$ is the population average outcome if all patients in the population were to receive treatment according to the rule d in $\mathcal{D}$
- Optimal regime:

$$d_{opt} = \arg\max_{d \in \mathcal{D}} \mathcal{V}(d)$$

is a regime in $\mathcal{D}$ such that

$$E\left[Y\left(d_{opt}\right) \mid H_1 = h_1\right] \geq E\left[Y(d) \mid H_1 = h_1\right]$$

for all $d \in \mathcal{D}$, $h_1 \in H_1$ and

$$E\left[E\left[Y\left(d_{opt}\right) \mid H_1\right]\right] \geq E\left[E\left[Y(d) \mid H_1\right]\right]$$

for all $d \in \mathcal{D}$

- Can be extended to K decisions/stages

Evaluating the value function

- Using no unmeasured confounding (e.g., in randomized trial), consistency, and positivity assumption

$$
\begin{aligned}
\mathcal{V}(d) &= E[Y(d)] \\
&= E[E[Y(d) \mid H_1]] \\
&= E\left\{E\left[Y(1) \mid H_1\right] d\left(H_1\right) + E\left[Y(0) \mid H_1\right]\left(1 - d\left(H_1\right)\right)\right\} \\
&= E\left\{E\left[Y \mid A_1 = 1, H_1\right] d\left(H_1\right) + E\left[Y \mid A_1 = 0, H_1\right]\left(1 - d\left(H_1\right)\right)\right\} \\
&= E\left\{b\left(1, H_1\right) d\left(H_1\right) + b\left(0, H_1\right)\left(1 - d\left(H_1\right)\right)\right\}
\end{aligned}
$$

Estimating the value function

In practice as $b\left(a_1, h_1\right)$ is not known and can be estimated by a parametric working model, say

$$b\left(A_1, H_1; \eta\right) = \left(A_1, H_1', A_1 H_1'\right)\eta$$

estimated using OLS. Then

$$\widehat{\mathcal{V}}(d) = \mathbb{P}_n\left\{b\left(1, H_1; \widehat{\eta}\right) d\left(H_1\right) + b\left(0, H_1; \widehat{\eta}\right)\left(1 - d\left(H_1\right)\right)\right\}$$

Characterization of optimal dynamic regime

- Recall that the optimal dynamic regime

$$\begin{aligned} d_{opt}(H_1) &= \arg\max_{d \in \mathcal{D}} E\left(Y(d) \mid H_1\right) \\ &= 1\left\{E\left(Y(1) \mid H_1\right) \geq E\left(Y(0) \mid H_1\right)\right\} \\ &= 1\left\{b\left(1, H_1\right) \geq b\left(0, H_1\right)\right\} \\ &= 1\left\{b\left(1, H_1\right) - b\left(0, H_1\right) \geq 0\right\} \\ &= 1\left\{\gamma\left(H_1\right) \geq 0\right\} \end{aligned}$$

  where

$$\begin{aligned} \gamma\left(H_1\right) &= b\left(1, H_1\right) - b\left(0, H_1\right) \\ &= E\left(Y(1) - Y(0) \mid H_1\right) \end{aligned}$$

  is the average causal effect given H_1.

- We see that H_1 determines the optimal dynamic regime if and only if it interacts with treatment A on the additive scale, in fact, if no such interaction exists, then $d_{opt} = 1\left\{\gamma\left(H_1\right) \geq 0\right\} = 1\{E(Y(1) - Y(0)) \geq 0\}$.

Characterization of optimal dynamic regime

- Outcome regression can be used to estimate optimal dynamic regime

$$\begin{aligned}\widehat{d}_{opt}^{OR}(H_1) &= \arg\max_{a_1 \in \{0,1\}} b\left(a_1, H_1; \widehat{\eta}\right) \\ &= 1\left\{b\left(1, H_1; \widehat{\eta}\right) - b\left(0, H_1; \widehat{\eta}\right) \geq 0\right\}\end{aligned}$$

  with corresponding estimator of optimal value

$$\begin{aligned}&\widehat{\mathcal{V}}\left(\widehat{d}_{opt}^{OR}\right) \\ =&\mathbb{P}_n\left\{b\left(1, H_1; \widehat{\eta}\right)\widehat{d}_{opt}^{OR}\left(H_1\right) + b\left(0, H_1; \widehat{\eta}\right)\left(1 - \widehat{d}_{opt}^{OR}\left(H_1\right)\right)\right\}\end{aligned}$$

- There is a limitation because $b\left(a_1, H_1; \widehat{\eta}\right)$ could be poorly specified and therefore $\widehat{d}_{opt}^{OR}\left(H_1\right)$ could be far from $d_{opt}\left(H_1\right)$.

Robust estimation of optimal dynamic regime

- Define a class of regimes

$$d_w(\gamma) = 1\left\{\gamma_w\left(H_1;\eta\right) \geq 0\right\}$$

  where $\gamma_w\left(H_1;\eta\right)$ is a known function indexed by parameter η, e.g., $\gamma_w\left(H_1;\eta\right) = H_1'\eta$
- Note that the induced class of regimes $\{d_w(\gamma)\}$ need not contain the true optimal regime d_{opt}. However, $d_w(\gamma)$ can generate a fairly large (albeit restricted) class of regime, and we can aim to find the optimal regime within this restricted class.

$$\begin{aligned} d_w\left(\gamma\left(\eta_{opt}\right)\right) &= \arg\max_{\eta} \mathcal{V}\left(d_w(\gamma(\eta))\right) \\ &= \arg\max_{\eta} E\left[Y\left(d_w(\gamma(\eta))\right)\right] \end{aligned}$$

Robust estimation of optimal dynamic regime

- NCSU group proposes the following semiparametric approach:
  - Obtain an estimator $\widehat{\mathcal{V}}\left(d_w(\eta)\right)$ for the value $\mathcal{V}\left(d_w(\eta)\right)$ for any fixed η.
  - Treat $\widehat{\mathcal{V}}\left(d_w(\eta)\right)$ as a function of η and maximize in η
  - That is, estimate η_{opt} by

    $$\widehat{\eta}_{opt} = \arg\max_{\eta} \widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)$$

    which gives

    $$\widehat{d}_{w,opt} = d_w\left(\gamma\left(\widehat{\eta}_{opt}\right)\right)$$

- Referred to as value search or direct search estimation.

A missing data analogy

- The approach requires a good estimator for $\mathcal{V}\left(d_w(\gamma(\eta))\right)$.
- They propose to take a missing data analogy and define

$$R(\eta) = d_w(\gamma(\eta))A_1 + \left(1 - d_w(\gamma(\eta))\right)(1 - A_1)$$

- Therefore $R(\eta) = 1$ if the person follows regime $d_w(\gamma(\eta))$ and zero otherwise, i.e., person's who do not follow the regime are missing. That is

$$Y\left(d_w(\gamma(\eta))\right) = \begin{cases} \quad Y & \text{if } R(\eta) = 1 \\ \text{missing} & \text{if } R(\eta) = 0 \end{cases}$$

A missing data analogy

- It follows that under no unmeasured confounding, consistency and positivity,

$$\begin{aligned}\mathcal{V}\left(d_w(\gamma(\eta))\right) &= E\left[Y\left(d_w(\gamma(\eta))\right)\right] \\ &= E\left[\frac{R}{\Pr\left(R=1 \mid H_1\right)} Y\right]\end{aligned}$$

where

$$\begin{aligned}&\Pr\left(R=1 \mid H_1\right) \\ =&d_w(\gamma(\eta)) \Pr\left(A_1=1 \mid H_1\right) \\ &+\left(1-d_w(\gamma(\eta))\right)\left(1-\Pr\left(A_1=1 \mid H_1\right)\right)\end{aligned}$$

A missing data analogy

- In a randomized trial $\Pr(R = 1 \mid H_1)$ is known because $\Pr(A_1 = 1 \mid H_1)$ is known, but in an observational study we can estimate the propensity score using say a logistic regression model, and therefore $\Pr(R = 1 \mid H_1)$ can be estimated.
- Let

$$\begin{aligned}&\Pr(R(\eta) = 1 \mid H_1, \widehat{\alpha})\\ =&d_w(\gamma(\eta)) \Pr(A_1 = 1 \mid H_1, \widehat{\alpha})\\ &+ (1 - d_w(\gamma(\eta)))(1 - \Pr(A_1 = 1 \mid H_1, \widehat{\alpha}))\end{aligned}$$

  where $\Pr(A_1 = 1 \mid H_1, \widehat{\alpha})$ is a consistent estimator of $\Pr(A_1 = 1 \mid H_1)$.

A missing data analogy

- Then

$$\begin{aligned}&\widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)\\=&\mathbb{P}_n\left[\frac{R(\eta)}{\Pr\left(R(\eta)=1 \mid H_1,\widehat{\alpha}\right)}Y\right]\end{aligned}$$

- Might be more stable to use

$$\begin{aligned}&\widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)\\=&\mathbb{P}_n\left[\frac{R(\eta)}{\Pr\left(R(\eta)=1 \mid H_1,\widehat{\alpha}\right)}\right]^{-1}\mathbb{P}_n\left[\frac{R(\eta)}{\Pr\left(R(\eta)=1 \mid H_1,\widehat{\alpha}\right)}Y\right]\end{aligned}$$

Doubly robust estimation

- The approach is guaranteed to be consistent in a randomized trial where $\Pr\left(A = 1 \mid H_1\right)$ is known, but not necessarily so in an observational study where it must be estimated and therefore may be misspecified.
- A more robust and possibly more efficient approach is given by the doubly robust estimator $\widehat{\mathcal{V}}_{dr}\left(d_w(\gamma(\eta))\right) =$

$$\mathbb{P}_n\left[\frac{R(\eta)}{\Pr\left(R(\eta) = 1 \mid H_1, \widehat{\alpha}\right)} Y - \frac{R(\eta) - \Pr\left(R(\eta) = 1 \mid H_1, \widehat{\alpha}\right)}{\Pr\left(R(\eta) = 1 \mid H_1, \widehat{\alpha}\right)} Q\left(H_1, \widehat{\beta}\right)\right]$$

where

$$\begin{aligned} Q\left(H_1, \widehat{\beta}\right) &= b\left(1, H_1; \widehat{\beta}\right) d_w(\gamma(\eta)) \\ &\quad + b\left(0, H_1; \widehat{\beta}\right)\left(1 - d_w(\gamma(\eta))\right) \end{aligned}$$

is a consistent estimator of

$$\begin{aligned} Q\left(H_1\right) &= E\left[Y \mid R(\eta) = 1, H_1\right] \\ &= b\left(1, H_1\right) d_w(\gamma(\eta)) + b\left(0, H_1\right)\left(1 - d_w(\gamma(\eta))\right) \end{aligned}$$

Robust estimation of optimal dynamic regime

- This estimator is doubly robust in the sense that it remains consistent and asymptotically normal if either $\Pr\left(R(\eta)=1 \mid H_1, \widehat{\alpha}\right)$ is consistent for $\Pr\left(R(\eta)=1 \mid H_1, \alpha\right)$ or $b\left(a_1, H_1; \widehat{\beta}\right)$ is consistent for $b\left(a_1, H_1\right)$ but both do not necessarily hold
- Estimation of the optimal regime within the class

$$\left\{d_w(\gamma)=1\left\{\gamma_w\left(H_1;\eta\right)\geq 0\right\}\right\}$$

entails finding the parameter η that maximizes the estimated value $\widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)$, that is

$$\begin{aligned}\widehat{\eta}^{opt} &= \arg\max \widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)\\ \widehat{\eta}^{opt}_{dr} &= \arg\max \widehat{\mathcal{V}}_{dr}\left(d_w(\gamma(\eta))\right)\end{aligned}$$

such that

$$\begin{aligned}\widehat{d}^{opt}_w(\gamma) &= 1\left\{\gamma_w\left(H_1;\widehat{\eta}^{opt}\right)\geq 0\right\}\\ \widehat{d}^{opt,dr}_w(\gamma) &= 1\left\{\gamma_w\left(H_1;\widehat{\eta}^{opt}_{dr}\right)\geq 0\right\}\end{aligned}$$

Some theory

- Note that solving for $\widehat{\eta}^{opt}$ and $\widehat{\eta}^{opt}_{dr}$ is not a standard optimization problem because the objective function is nonsmooth in η. They have used grid search algorithms to find solution.
- They also argue that under regularity conditions,

$$\begin{aligned}&n^{1/2}\left(\widehat{\mathcal{V}}\left(d_w\left(\gamma\left(\widehat{\eta}^{opt}\right)\right)\right)-\mathcal{V}\left(d_w\left(\gamma\left(\eta^{opt}\right)\right)\right)\right)\\=&n^{1/2}\left(\widehat{\mathcal{V}}\left(d_w\left(\gamma\left(\eta^{opt}\right)\right)\right)-\mathcal{V}\left(d_w\left(\gamma\left(\eta^{opt}\right)\right)\right)\right)+o_p(1)\end{aligned}$$

  where $\widehat{\eta}^{opt}\xrightarrow{P}\eta^{opt}$.
- Therefore the variance of $\widehat{\mathcal{V}}\left(d_w\left(\gamma\left(\widehat{\eta}^{opt}\right)\right)\right)$ can be determined in large samples from the variance of $\widehat{\mathcal{V}}\left(d_w\left(\gamma\left(\eta^{opt}\right)\right)\right)$ which can be obtained by a standard sandwich variance formula.
- The same results hold for $\widehat{\mathcal{V}}_{dr}\left(d_w\left(\gamma\left(\widehat{\eta}^{opt}_{dr}\right)\right)\right)$.

Breast cancer example

- Clinical trial in patients with primary operable breast cancer: Conducted by National Surgical Adjuvant Breast and Bowel Project, reported by Gail and Simon (1985).
- $n = 1276$
- Treatment options : $A_1 = 1$ if L-phenylalanine mustard and 5-fluorouracil (PF); $A_1 = 0$ if PF +tamoxifen (PFT)
- Baseline information: H_1={age (years), progesterone receptor (PR) level (fmol), estrogen receptor (ER) level (fmol), tumor size (cm), number of positive nodes}
- Outcome: $Y = 1$ if subject survived disease-free to 3 years, $Y = 0$ otherwise.

Breast cancer example

- Gail and Simon evaluated rules based on age and PR interaction and estimated using subgroup analysis that optimal regime within this class is given by $I\left(\text{age} < 50, PR < 10\right)$ with value function given by $\widehat{\mathcal{V}} \approx 0.68$.
- NCSU group considered a class of regimes of form

$$
\begin{aligned}
\left\{d_w(\gamma) = 1\left\{\gamma_w\left(H_1;\eta\right) \geq 0\right\}\right\} \\
\gamma_w\left(H_1;\eta\right) = & \eta_0 + \eta_1 \text{ age } + \eta_2 \log(1 + PR)
\end{aligned}
$$

- Applying their robust approach, they estimated a similar $\widehat{\mathcal{V}}$
- Both approaches agree and suggest giving PF to younger patients with low PR and lead to about the same estimated values.

Estimation of optimal regimes as a classification problem

- Classification perspective:
  - The rule characterizing a regime d is analogous to a classifier
  - This allows work on classification and machine learning to be exploited
- Generic classification problem
  - Z = outcome or class label; here $Z = \{0, 1\}$ (binary); X = vector of covariates or features taking values in a feature space $\mathcal{X}$
  - d is a classifier: $d : \mathcal{X} \rightarrow \{0, 1\}$
  - $\mathcal{D}$ denotes a family of classifiers, e.g., with $X = (X_1, X_2)$: Hyperplanes of form $d(X) = I\left(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0\right)$; Rectangular region of the form $d(X) = I\left(X_1 < \eta_0\right) + I\left(X_1 \geq \eta_0, X_2 < \eta_0\right)$

Generic classification problem

- Training set $(X_i, Z_i)\,, i = 1, \ldots, n$
- Find classifier $d \in \mathcal{D}$ that minimizes
    - Classification error

$$\mathbb{P}_n \left\{Z_i - d\left(X_i\right)\right\}^2 = \mathbb{P}_n I\left\{Z_i \neq d\left(X_i\right)\right\}$$

    - Weighted classification error

$$\mathbb{P}_n w_i \left\{Z_i - d\left(X_i\right)\right\}^2 = \mathbb{P}_n w_i I\left\{Z_i \neq d\left(X_i\right)\right\}$$

      for w_i, fixed, known weights, $i = 1, \ldots n$.
    - This problem has been studied extensively by statisticians and computer scientists, also known in machine learning as supervised learning
    - Multiple methods and software widely available including CART which yields rectangular regions and support vector machines which yields hyperplanes.

Classification analogy

- Recall that for a restricted class $\mathcal{D}_\eta = \left\{d_\eta\right\}$, the optimal restricted regime $d_{\widehat{\eta}^{opt}}$ satisfies

$$\widehat{\eta}^{opt} = \arg\max_{\eta} \widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right)$$

where for fixed η

$$\widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right) = \mathbb{P}_n\left[d_w(\gamma(\eta))\widehat{\Delta}\left(H_1\right) + \widehat{\Delta}_0\left(H_1\right)\right]$$

$\widehat{\Delta}\left(H_1\right)$ is an estimator of $b(1, H_1) - b(0, H_1)$, $\widehat{\Delta}_0\left(H_1\right)$ is an estimator of $b(0, H_1)$

Classification analogy

- Therefore, maximizing $\widehat{\mathcal{V}}_{dr}\left(d_w(\gamma(\eta))\right)$ in η is equivalent to maximizing

$$\mathbb{P}_n\left[d_w(\gamma(\eta))\widehat{\Delta}\left(H_1\right)\right]$$

- One can further show that

$$\begin{aligned}&d_w(\gamma(\eta))\widehat{\Delta}\left(H_1\right)\\ =&1\left\{\widehat{\Delta}\left(H_1\right)\geq 0\right\}\left|\widehat{\Delta}\left(H_1\right)\right|\\ &-\left|\widehat{\Delta}\left(H_1\right)\right|\left\{1\left\{\widehat{\Delta}\left(H_1\right)\geq 0\right\}-d_w(\gamma(\eta))\right\}^2\end{aligned}$$

Classification analogy

So that

$$\begin{aligned}
\widehat{\eta}^{opt} &= \arg\max_{\eta} \widehat{\mathcal{V}}\left(d_w(\gamma(\eta))\right) \\
&= \arg\min_{\eta} \mathbb{P}_n \left|\widehat{\Delta}\left(H_1\right)\right| \left\{1\left\{\widehat{\Delta}\left(H_1\right) \geq 0\right\} - d_w(\gamma(\eta))\right\}^2 \\
&= \arg\min_{\eta} \mathbb{P}_n \left|\widehat{\Delta}\left(H_1\right)\right| 1\left\{1\left\{\widehat{\Delta}\left(H_1\right) \geq 0\right\} \neq d_w(\gamma(\eta))\right\}
\end{aligned}$$

is minimizing a weighted classification error with

- Label: $Z_i = 1\left\{\widehat{\Delta}\left(H_1\right) \geq 0\right\}$
- Weight: $w_i = \left|\widehat{\Delta}\left(H_1\right)\right|$
- Classifier: $d = d_w(\gamma(\eta))$

Classification analogy

- Note that this framework is quite general as one can choose various estimators to generate the labels $Z_i = 1\left\{\widehat{\Delta}\left(H_1\right) \geq 0\right\}$ with $\widehat{\Delta}\left(H_1\right) =$ an individual level estimator of $\Delta\left(H_1\right) = E\left\{Y(1) - Y(0) \mid H_1\right\}$, e.g.,
    - $\widehat{\Delta}\left(H_1\right) = \widehat{b}\left(1, H_1\right) - \widehat{b}\left(0, H_1\right)$ outcome regression estimator
    - or $\widehat{\Delta}\left(H_1\right) = \frac{(A_1 - \widehat{\pi}(H_1))Y}{(1-\widehat{\pi}(H_1))\widehat{\pi}(H_1)}$ IPW
    - or DR estimator given above

Classification analogy

- In summary, value search estimation of an optimal restricted regime d_w by maximizing an estimator of value function is equivalent to minimizing a weighted classification error
- The choice of classification approach determines the restricted class
- This can be implemented using off-the-shelf software and algorithms for classification problems, e.g., CART and SVM
- Important to note that this analogy does not circumvent the need to optimize a nonsmooth function of η

Classification analogy

- The source of difficulty is that the 0-1 loss function

$$l_{0-1}(x) = 1(x < 0)$$

  is nonconvex
- Optimization involving nonconvex loss functions is challenging as standard techniques cannot be used
- This problem is well studied in classification literature
- In case of SVM replace 0-1 loss function with a convex surrogate such as the hinge loss function

$$l_{\text{hinge}}(x) = (1 - x)^{+}, \text{where } x^{+} = \max(x, 0)$$

Outcome weighted learning

- Recall that for IPW, assuming the Y is bounded and nonnegative,

$$\begin{aligned}
\widehat{\Delta}\left(H_1\right) &= \frac{\left(A_1 - \widehat{\pi}\left(H_1\right)\right) Y}{\left(1 - \widehat{\pi}\left(H_1\right)\right) \widehat{\pi}\left(H_1\right)} \\
1\left\{\widehat{\Delta}\left(H_1\right) \geq 0\right\} &= A_1 \text{ (label)} \\
\left|\widehat{\Delta}\left(H_1\right)\right| &= \frac{Y}{\left(1 - A_1\right)\left(1 - \widehat{\pi}\left(H_1\right)\right) + A_1 \widehat{\pi}\left(H_1\right)} \text{ (weight)}
\end{aligned}$$

- Assume randomized trial with known $\pi\left(H_1\right) = \pi$
- Recode $\mathcal{A} = \{-1, 1\}$ now the classifier $d_w(\gamma(\eta)) = \operatorname{sign}\left(\gamma\left(H_1; \eta\right)\right)$ for decision function $\gamma\left(H_1; \eta\right)$

Outcome weighted learning

- Weighted classification error can be re-written as

$$
\begin{aligned}
&\mathbb{P}_n \frac{Y}{(1-A_1)/2 + A_1\pi} 1\left\{A_1 \neq \text{sign}\left(\gamma\left(H_1;\eta\right)\right)\right\} \\
=&\mathbb{P}_n \frac{Y}{(1-A_1)/2 + A_1\pi} 1\left\{A_1\gamma\left(H_1;\eta\right) < 0\right\} \\
=&\mathbb{P}_n \frac{Y}{(1-A_1)/2 + A_1\pi} l_{0-1}\left\{A_1\gamma\left(H_1;\eta\right)\right\}
\end{aligned}
$$

  which involves non-convex 0-1 loss

- Zhao et al (2012) introduced outcome weighted learning (OWL) which minimizes the following convex relaxation

$$
\mathbb{P}_n \frac{Y}{(1-A_1)/2 + A_1\pi} \left\{1 - A_1\gamma\left(H_1;\eta\right)\right\}^+ + \lambda_n \|\gamma\|^2
$$

Breast cancer example revisited

- Clinical trial in patients with primary operable breast cancer: Conducted by National Surgical Adjuvant Breast and Bowel Project, reported by Gail and Simon (1985).
- $n = 1276$
- Treatment options: $A_1 = 1$ if L-phenylalanine mustard and 5-fluorouracil (PF); $A_1 = 0$ if PF + tamoxifen (PFT)
- Baseline information: H_1={age (years), progesterone receptor (PR) level (fmol), estrogen receptor (ER) level (fmol), tumor size (cm), number of positive nodes}
- Outcome: $Y = 1$ if subject survived disease-free to 3 years, $Y = 0$ otherwise.

Breast cancer example revisited

- Gail and Simon evaluated rules based on age and PR interaction and estimated using subgroup analysis that optimal regime within this class is given by $I\left(\text{age} < 50, PR < 10\right)$ with value function given by $\widehat{\mathcal{V}} \approx 0.68$.
- Using DR and CART, they get a similar $\widehat{\mathcal{V}}$