Precision Medicine: Lecture 01
Overview of Precision Medicine

Michael R. Kosorok,
Nikki L. B. Freeman and Owen E. Leete

Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Fall, 2021

Outline

Precision Medicine

- Precision medicine
  - Data-driven decision support for treating patients in the presence of heterogeneity (dynamic treatment regimes or DTRs)
  - Treatment can include drug choice, administrative actions, dosing, timing, potentially modifiable risk factors, and/or other potentially beneficial actions to the patients
  - Must be reproducible and generalizable (empirically and inferentially valid)

Precision Health

- Precision Public Health
  - Data-driven decision support for families, communities, clinics, social networks, and other entities
  - Potential treatments can also be policies
- Precision Health
  - Union of precision medicine and precision public health
  - We always consider the consequences of our actions on the populations involved not just narrow subgroups or individuals

Operating Principles

- Observable Constituents:
  - Tailoring variables (X)
  - Choice of treatments and/or potentially modifiable risk factors (A)
  - Vector of outcomes or utilities (Y)
  - Could be multiple (X, A, Y) triples over time for each patient
- Dynamic Treatment Regime (DTR):
  - Single decision: make a single recommendation for treatment
  - Multiple decision: make a series of interdependent recommendations
  - Continual monitoring: for diabetes, mHealth

Heterogeneity

- Role of Heterogeneity in the data:
  - Heterogeneity of patients is beneficial (essential) for good precision medicine analysis so that our treatment rules are broadly applicable
  - Need heterogeneity of treatment assignment (either naturally or by design) in the data so we can determine best treatment under a variety of situations

Outline of Overall Pipeline

- Dynamic Treatment Regime:
  - $d(X)$ gives recommended A to maximize Y in future patients
  - Regression: model Y as a function of X and A ($Q(X, A) = E[Y|X, A]$ is the "value"), with interaction between X and A being most important
  - Policy estimation: directly estimate $d(X)$ without necessarily needing $Q(X, A)$ (e.g., outcome weighted learning)
  - Prediction versus prescriptive decision support:
    - Suppose $Y = f(X) + Ag(X) + e$, where bigger Y is better and $A = \{0 \text{ or } 1\}$
    - We only care about $g(x)$, since rule $d(X) = \{1 \text{ if } g(X) > 0, 0 \text{ otherwise}\}$ yields optimal
    - A focus on prediction spends too much energy and focus on $f(X)$ instead of $g(X)$

Pipeline, cont.

- Propensity Score:
  - $P(A|X)$ is propensity score
  - We estimate $P(A|X)$ from data and make sure it is positive for all X
- Causal Methods:
  - Potential outcome validity ($Y(0)$ and $Y(1)$ are well behaved)
  - No unmeasured confounders: $Y(0)$ and $Y(1)$ are independent given X
  - Positivity assumption ($P(A|X) > 0$ for all X)

Relationship to Machine Learning and Artificial Intelligence

- Prediction: This is a common goal of machine learning and AI but it does not necessarily provide meaningful decision support.
- Causal modeling (what if analysis): This requires incorporating causal inference into the prediction and is more difficult than prediction since it requires treatment interaction estimation.
- Dynamic treatment regimes (decision support): This requires optimization over causal models to obtain treatment algorithms which optimize a specified utility as a function of individual-level characteristics.

Outline

The Single-Decision Setting

- The basic ingredients revisited:
  - The data: $\{(X_i, A_i, Y_i)\}_{i=1}^n$, comprising n i.i.d. triples (X, A, Y) where:
    - $X \in \mathcal{X}$ denotes baseline patient characteristics
    - $A \in \mathcal{A}$ denotes the assigned treatment
    - $Y \in \mathbb{R}$ denotes the outcome coded so higher values are better.
  - For each $x \in \mathcal{X}$ define $\psi(x) \subseteq \mathcal{A}$ to be the set of allowable treatments for a patient with $X = x$.
  - Thus a dynamic treatment regime (DTR) is a map $d : \mathcal{X} \to \mathcal{A}$ which satisfies $d(x) \in \psi(x)$ for all $x \in \mathcal{X}$.
- Under d, patients with $X = x$ are assigned treatment $d(x)$.

Single-Decision Setting, cont.

- An optimal DTR yields the maximal mean outcome (or median, etc.) if applied to the population of interest.
- Let $Y^*(a)$ denote the potential outcome under treatment $a \in \mathcal{A}$.
- Accordingly, the potential outcome under DTR d is

$$Y^*(d) = \sum_{a \in \mathcal{A}} Y^*(a) 1_{\{d(X)=a\}}.$$

- The optimal regime, denoted d^{opt}, satisfies:
  - (i) $d^{\text{opt}}(x) \in \psi(x)$ for all $x \in \mathcal{X}$
  - (ii) $EY^*(d^{\text{opt}}) \geq EY^*(d)$ for all d such that $d(x) \in \psi(x)$ and all $x \in \mathcal{X}$.

Some Terminology

- Dynamic treatment regime (or "regimen" or "rule") (DTR)
- Individualized treatment rule (ITR)
- Adaptive treatment strategy
- Treatment policy (or just "policy")
- "Dynamic" often refers to the variation over x not necessarily over time.
- "Action" is often used instead of "treatment"

The Multi-Decision Setting

- The multi-decision setting:
  - Two or more opportunities for treatment decisions (i.e., cancer treatment involving multiple lines of chemotherapy, other chronic diseases, etc.).
  - Interventions can affect patients in multiple ways
    - Immediate affects (proximal)
    - Delayed affects (distal): sometimes the best treatment is initially harmful but sets the patient up for a better response to certain future treatments

Multi-Decision Setting, cont.

- The basic ingredients:
  - The data: $\{(X_{1,i}, A_{1,i}, Y_{1,i}, \ldots, X_{T,i}, A_{T,i}, Y_{T,i})\}_{i=1}^{n}$, comprising n i.i.d. replicates of $(X_1, A_1, Y_1, \ldots, X_T, A_T, Y_T)$, where
    - $X_1 \in \mathcal{X}_1$ denotes baseline information
    - $X_t \in \mathcal{X}_t$ denotes interim information collected during treatment stages $t = 2, \ldots, T$
    - $A_t \in \mathcal{A}_t$ denotes treatment and
    - Y_t denotes proximal outcome measured after treatment at stage t,
    - for $t = 1, \ldots, T$.
  - Define $H_1 = X_1$ and $H_t = (H_{t-1}, A_{t-1}, Y_{t-1}, X_t)$ so that H_t is the available patient history at time t.

Multi-Decision Setting, cont.

- More ingredients:
  - Let $\mathcal{H}_t = \operatorname{dom} H_t$
  - Let $\psi_t(h_t) \subseteq \mathcal{A}_t$ denote the set of allowable treatments for a patient presenting with $H_t = h_t$ at time t.
- A DTR in this setting is a sequence of functions $d = (d_1, \ldots, d_T)$ such that $d_t : \mathcal{H}_t \to \mathcal{A}_t$ satisfies $d_t(h_t) \in \psi_t(h_t)$ for all h_t for $t = 1, \ldots, T$.
- An optimal treatment regime maximizes the expectation (or median, etc.) of some "cumulative" outcome measure $Y = y(Y_1, \ldots, Y_T)$, e.g.,
  - $y(v_1, \ldots, v_T) = \sum_{t=1}^{T} v_t$ or
  - $y(v_1, \ldots, v_T) = \max_t v_t$, or
  - $y(v_1, \ldots, v_T) = v_T$, etc.

Potential Outcome Framework

Interventions at time t affect both Y_t and X_{t+1}, complicating the potential outcomes framework:

- We use an overline to denote history so that $\overline{a}_t = (a_1, \ldots, a_t)$.
- Let $X_t^*(\overline{a}_{t-1})$ be the potential interim measurements at time t under treatment sequence $\overline{a}_{t-1} \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_{t-1}$
- Define $Y_t^*(\overline{a}_t)$ similarly.
- The potential outcome under $\overline{a}_T \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_T$ is thus $Y^*(\overline{a}_T) = y\{Y_1^*(a_1), Y_2^*(\overline{a}_2), \ldots, Y_T^*(\overline{a}_T)\}$.

Potential Outcome Framework, cont.

- For any regime, d, the potential outcome is

$$Y^*(d) = \sum_{\overline{a}_T} Y^*(\overline{a}_T) \prod_{t=1}^{T} 1_{[d_t\{H_t^*(\overline{a}_{t-1})\}=a_t]},$$

  where we have defined $H_1^*(a_0) \equiv H_1$.
- An optimal regime, d^{opt}, satisfies:
  - (i) $d_t^{\text{opt}}(h_t) \in \psi_t(h_t)$ for all $h_t \in \mathcal{H}_t$ and $t = 1, \ldots, T$
  - (ii) $EY^*(d^{\text{opt}}) \geq EY^*(d)$ for all d satisfying $d_t(h_t) \in \psi_t(h_t)$ for all $h_t \in \mathcal{H}_t$ and $t = 1, \ldots, T$.

Decision Support Framework

- In the preceeding, we assume future patients will be treated under the same general time horizon as the patients in the data sample.
- This is not always true, and we need to sometimes allow for an indefinite time horizon:
  - Example 1: glucose control in type 1 diabetes
  - Example 2: infection control in patients with cystic fibrosis
- Extrapolation beyond the time horizon of the sample requires structures to be imposed onto the generative model, including, for example a Markov decision process framework.
- There are other more complex settings such as decision support on a social network for controlling an infectious disease.

Conditional Average Treatment Effect (CATE)

The following are connected concepts which we now explore for the single decision setting:

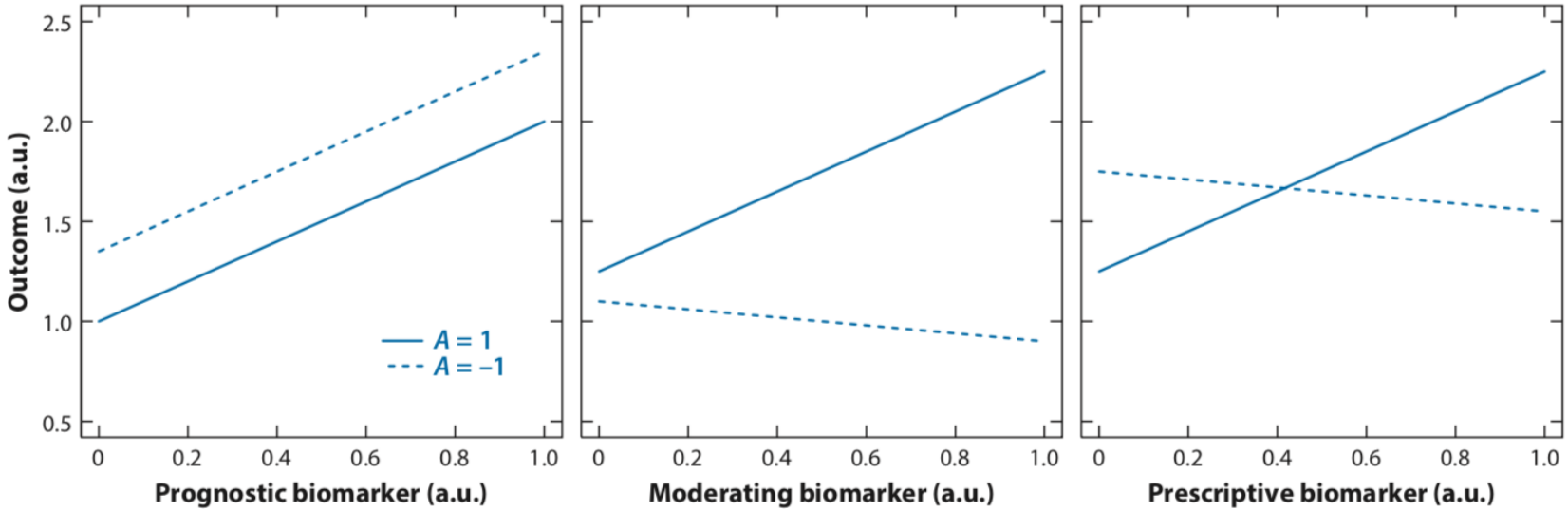
- DTR
- Subgroup identification
  - An optimal DTR $d^{opt}(x)$ would identify subgroups for which each treatment is optimal, i.e.,
    - The optimal subgroup for treatment 1 is $\{x : d^{opt}(x) = 1\}$
    - The optimal subgroup for treatment -1 is $\{x : d^{opt}(x) = -1\}$
- Condtional average treatment effect (CATE)
  - CATE is defined as: $\Delta(x) = E\{Y^*(1) - Y^*(-1)|X = x\}$.
  - It can be shown that $d^{opt}(x) = \text{sign}\{\Delta(x)\}$ yields the optimal DTR (Bayes rule).
  - Strange inconsistencies can happen if we compare the decision rule obtained by using a linear regression model for predicting outcome Y versus using a linear regression model for $\Delta(x)$.

Outline

Identifying biomarkers is a common clinical goal

- In precision medicine research, we often want to identify patient biomarkers that are important for choosing an optimal treatment.
- We use the term **biomarker** to represent a scalar feature constructed from current patient information. Biomarkers include
  - Single component of the available history, and
  - Composite measure constructed from multiple components.
- A biomarker may provide clinical information by being
  - Prognostic,
  - Moderating, or
  - Prescriptive.

Biomarkers



from Kosorok and Laber (2019)

Prognostic biomarkers

- Prognostic biomarkers are useful in predicting the mean (median, etc.) outcome of a patient
- Example: Single-decision binary treatment setting with $A \in \mathcal{A} = \{-1, 1\}$. It follows that

$$E\{Y^*(a)|X = x\} = \mu(x) + a\Delta(x)/2$$

where $\mu(x) = E\{Y^*(1) + Y^*(-1)|X = x\}$

  - Assume $\mu(x; \beta_0^*) = x^\intercal \beta_0^*$ and $\Delta(x; \beta_1^*) = x^\intercal \beta_1^*$
  - $X = (X_1, \ldots, X_p)^\intercal$ are the biomarkers under consideration
  - X_j is a prognostic biomarker if either $\beta_{0,j}^* \neq 0$ or $\beta_{1,j}^* \neq 0$.

Moderating biomarkers

- Moderating biomarkers are useful in predicting contrasts of the mean outcome across different candidate treatments.
- In the linear model example where $\mu(x; \beta_0^*) = x^\intercal \beta_0^*$ and $\Delta(x; \beta_1^*) = x^\intercal \beta_1^*$, we say that X_j is a moderating biomarker if $\beta_{1,j}^* \neq 0$.
- A moderating biomarker is also prognostic.

Prescriptive biomarkers

- Prescriptive biomarkers are useful in selecting the treatment that maximizes the mean outcome.
- Same example: Single-decision binary treatment setting with $A \in \mathcal{A} = \{-1, 1\}$.

$$E\{Y^*(a)|X = x\} = \mu(x) + a\Delta(x)/2$$

where $\mu(x) = E\{Y^*(1) + Y^*(-1)|X = x\}$

  - Assume as before $\mu(x; \beta_0^*) = x^\intercal \beta_0^*$
  - Different assumption $\Delta(x; \beta_1^*, \beta_2^*) = \exp(x^\intercal \beta_2^*) x^\intercal \beta_1^*$
  - Support of $X = (X_1, \ldots, X_p)^\intercal$ is $\mathbb{R}^p$
  - X_j is a prescriptive biomarker if $\beta_{1,j}^* \neq 0$.
  - If $\beta_{1,j}^*$ is zero but $\beta_{2,j}^*$ is nonzero, then X_j is moderating but not prescriptive.
- More general definitions for prognostic, moderating, and prescriptive biomarkers can be found in Kosorok and Laber (2019).

Outline

The Single-Decision Setting

- The goal is to estimate a regime, d^{opt}, that satisfies $EY^*(d^{\text{opt}}) \geq EY^*(d)$ for any other regime d
- To construct an estimator, we need to express d^{opt} in terms of the data-generating model.
- We make the following standard assumptions:
  - Positivity: $P(A = a|X = x) > 0 \; \forall \; a \in \psi(x), x \in X$
  - Consistency: $Y = Y^*(A)$
  - Strong Ignorability: $\{Y^*(a) : a \in A\} \perp A|X$

The Single-Decision Setting, cont.

- Define $Q(x, a) = E(Y|X = x, A = a)$
- Under the preceding assumptions
  $d^{\text{opt}}(x) = \text{argmax}_{a \in \psi(x)} Q(x, a)$ is an optimal regime
- Regression-based estimation:
  - Construct an estimator $\widehat{Q}_n(x, a)$ of $Q(x, a)$
  - Use the plug-in estimator $\widehat{d}_n(x) = \text{argmax}_{a \in \psi(x)} \widehat{Q}_n(x, a)$
  - Example – Linear Models:
    - Define $Q(x, a; \beta) = \sum_{a' \in A} x_{a'} \beta_{a'} 1\{a = a'\}$
    - For the empirical measure $\mathbb{P}_n$, let
      $\widehat{\beta}_n = \text{argmin}_\beta \, \mathbb{P}_n\{Y - Q(X, A; \beta)\}^2$
    - Then $\widehat{d}_n = \text{argmax}_{a \in \psi(x)} Q(x, a; \widehat{\beta})$

Regression-Based Estimators

- Linear models are common because they are easy to interpret
  - The coefficients of $\widehat{\beta}_{a,n}$ can be used to identify important tailoring variables
  - Interpretability may come at the cost of misspecification;
  - Model diagnostics and interactive model-building techniques should be used to ensure a high-quality model
- Flexible model classes (e.g. nonparametric) can be used to estimate $Q(x, a)$ to mitigate misspecification issues
  - Increased flexibility for $\widehat{Q}_n(x, a)$ can render the estimated rule $x \mapsto \text{argmax}_{a \in \psi(x)} \widehat{Q}_n(x, a)$ unintelligible
  - Can obscure the scientific content of the estimated regime and limit its value as a decision support tool

Regression-Based Estimators

- Regression-based policy-search methods
  - The class of regimes is decoupled from the class of estimators used for $Q(x, a)$.
  - For a prespecified class of regimes, $\mathcal{D}$, the optimal within this class is $d_{\mathcal{D}}^{\text{opt}} = \text{argmax}_{d \in \mathcal{D}} \, EQ\{X, d(X)\}$
  - Because the class $\mathcal{D}$ is chosen independently of the class of models for $Q(x, a)$, one can use nonparametric regression estimators while maintaining control of the form of the estimated optimal regime.
- Because these approaches are built upon regression, it is easily extensible to settings with complex data structures, censored data, or measurement error, or to other settings for which regression models have been developed.

Classification-Based Estimators

- A representation based on importance sampling leads to another class of estimators termed direct-search or classification-based estimators
- Assume that treatment is binary and coded so that $A \in \mathcal{A} = \{-1, 1\}$
- Under the causal conditions stated above, the marginal mean outcome under a regime d is

$$V(d) \triangleq EY^*(d) = E\left\{\frac{Y1_{d(X)=A}}{\pi(A;X)}\right\}$$

  where $\pi(a;x) = P(A = a|X = x)$ is the propensity score
- Thus, the optimal regime satisfies $d^{\text{opt}} = \text{argmax}_d\, V(d)$

Classification-Based Estimators

- Given an estimator $\widehat{\pi}_n(a;x)$ of $\pi(a;x)$ the inverse probability weighted estimator of V (d) is given by $\widehat{V}_n(d) = \mathbb{P}_n\{Y 1_{d(X)=A}/\widehat{\pi}_n(A;X)\}$.
- Given a class of regimes, $\mathcal{D}$, one could construct an estimator of d^{opt} by direct maximization, i.e., $\widehat{d}_{\mathcal{D},n} = \text{argmax}_{d\in\mathcal{D}}\, \widehat{V}_n(d)$
- Because of the discontinuous indicator function, direct optimization is not computationally feasible save for settings where $\mathcal{D}$ is small.

Classification-Based Estimators

- However, it can be seen that

$$V(d) = -E\left\{\frac{|Y|1_{A\text{sign}(Y)d(X)<0}}{\pi(A;X)}\right\} + E\left\{\frac{(Y)_+}{\pi(A;X)}\right\}$$

  where $\text{sign}(u) = 1_{u>0} - 1_{u<0}$ is the sign function and $(u)+ = \max(0,u)$ is the positive part function.
- Thus, given a class of regimes, D, the optimal regime within this class satisfies

$$d_{\mathcal{D}}^{\text{opt}} = \underset{d\in\mathcal{D}}{\text{argmin}}\, E\left\{\frac{|Y|1_{A\text{sign}(Y)d(X)<0}}{\pi(A;X)}\right\}$$

- Hence, it can be seen that d^{opt} minimizes a cost-sensitive classification problem

The Multi-Decision Setting

- The multi-decision setting is complicated by the need to account for delayed treatment effects and prognostic effects
- We consider two multi-decision settings
  - A finite time horizon wherein the number of decision points is small and finite
  - An indefinite time horizon wherein the number of decision points is large or indeterminate.
- There are many intermediate settings, but these two encompass many commonly encountered settings in precision medicine.
- The methods discussed apply to both observational and randomized studies.

Finite time horizon

- We can estimate the optimal dynamic treatment regimes through a variety of reinforcement learning techniques (g-estimation, Q- and A-learning, modeling entire longitudinal process, extensions of OWL, etc.)
- In precision medicine, we are primarily interested in estimating the decision rule
  - In many settings, it is much more robust and feasible to not model the entire process if possible
  - Many of the learning methods listed above were motivated, at least in part, to obtain the dynamic treatment regime without needing to model the full process

Finite time horizon

- We derive regression-based and inverse-weighting or classification-based representations of the optimal treatment regime in terms of the data-generating model
- These representations can be used to construct estimators of the optimal regime.
- We make the following assumptions:
  - Positivity: $P(A_t = a_t | H_t = h_t) > 0$ for all $a_t \in \psi_t(h_t)$ and $h_t \in \mathcal{H}_t$
  - Consistency: $H_t = H_t^*(\overline{A}_{t-1})$ for $t = 2, \ldots, T$ and $Y = Y^*(\overline{A}_T)$
  - Sequential ignorability:

$$\{Y^*(\overline{a}_T), H_T^*(\overline{a}_T), \ldots, H_2^*(a_1), H_1 : \overline{a}_T \in \otimes_{t=1}^{T} \mathcal{A}_t\} \perp A_t | H_t$$

for $t = 1, \ldots, T$, where $\otimes$ denotes the Cartesian product taken over the specified range of indices.

Finite time horizon

- Define $Q_T(h_T, a_T) = E(Y|H_T = h_T, A_T = a_T)$, and for $t = T-1, ..., 1$ define $Q_t(h_t, a_t) = E\{\max_{a_{t+1} \in \psi_t(h_t)} Q_{t+1}(H_{t+1}, a_{t+1})|H_t = h_t, A_t = a_t\}$
- It follows that $d^{\text{opt}}(h) = \underset{a_t \in \psi_t(h_t)}{\text{argmax}}\, Q_t(h_t, a_t)$
- Q-learning is an approximate dynamic programming algorithm based on the above equation
  - Construct an estimator $\widehat{Q}_{T,n}(h_T, a_T)$ of $Q_T(h_T, a_T)$ obtained by regressing Y on H_T and A_T
  - Subsequently, for $t = T-1, \ldots, 1$, let $\widehat{Q}_{t,n}(h_t, a_t)$ be an estimator of $Q_t(h_t, a_t)$ obtained by regressing $max_{a_{t+1} \in \psi_{t+1}(H_t+1)} \widehat{Q}_{t+1,n}(H_{t+1}, a_{t+1})$ on H_t and A_t
  - The Q-learning estimator of d^{opt} is thus $\widehat{d}_{t,n}(h_t) = \text{argmax}_{A_t \in \psi_t(h_t)} \widehat{Q}_{t,n}(h_t, a_T)$ for $t = 1, \ldots, T$

Finite time horizon

- Q-learning relies on a series of regression models
  - Easily extensible to a variety of models and data structures
  - Allows the user to interactively construct and critique the models used for the Q-functions
- In the above formulation the estimated optimal decision rule is tied to the class of models used for the Q-functions
  - Trade-off between severe model misspecification and an unintelligible black box
  - Q-learning with policy-search wherein the class of regimes is divorced from the class of models for the Q-functions

Finite time horizon

- An alternative representation of the optimal decision rule is based on inverse probability weighting. For any regime d, it follows that

$$V(d) = E\left(Y \prod_{t=1}^{T} \frac{1_{d_t(H_t)=A_t}}{\pi_t(A_t|H_t)}\right)$$

- Given estimators $\widehat{\pi}_{t,n}(a_t; h_t)$ of $\pi_t(a_t; h_t)$ for $t = 1, \ldots, T$, the plug-in estimator of $V(d)$ is

$$\widehat{V}_n(d) = \mathbb{P}_n\left(Y \prod_{t=1}^{T} \frac{1_{d_t(H_t)=A_t}}{\widehat{\pi}_{t,n}(A_t|H_t)}\right)$$

- For a class of regimes $\mathcal{D}$, an estimator of $d^{\text{opt}}_{\mathcal{D}}$ is $\widehat{d}_{\mathcal{D},n} = \text{argmax}_{d\in\mathcal{D}}\, \widehat{V}_n(d)$

Finite time horizon

- Directly computing $\widehat{d}_{\mathcal{D},n}$ is difficult except in small problems, as the indicators make this into a discontinuous optimization problem.
- This can be avoided by using a surrogate optimization function for $\widehat{V}_n(d)$ which is smooth and has a global optimum.
- Optimizing $\widehat{V}_n(d)$, or a surrogate, can be difficult when the number of time points T is large
  - The product of indicators can rapidly become zero for the majority of subjects
  - The product of the propensity scores can grow quite small, leading to high variance.
- For these reasons, direct search estimators based on $\widehat{V}_n(d)$ work best for settings where T is small, e.g., $T = 2$

Infinite time horizon

- The infinite horizon setting applies when a sequence of similar decisions need to be made over an extended time (e.g. diabetes or other chronic diseases)
- Decision making is typically modeled as a Markov decision process
- We assume that the observed data are n i.i.d. replicates of $(S_1, A_1, S_2, \ldots, S_{T-1}, A_{T-1}, S_T)$,
  - $S_t \in \mathcal{S}$ denotes a summary of the patient's health status at time t
  - $A_t \in \mathcal{A}$ denotes the treatment applied at time $t = 1, \ldots, T$
  - T is the observed time horizon.

Infinite time horizon

- Let $y : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ be the momentary reward function
  - $y(s, a, s')$ captures the momentary goodness for a patient with health status s who receives treatment a and subsequently transitions to a health status s'
  - We write $Y_t = y(S_t, A_t, S_{t+1})$ to denote the observed momentary outcome for time t
- The goal may be to estimate a treatment regime that can be applied well beyond T decision points
- We assume that the observed data are Markov and homogeneous so that for any $\mathcal{Z} \subseteq \mathcal{S}$ and $t \geq 1$,

$$\begin{aligned} &P(S_{t+1} \in \mathcal{Z} | \overline{S}_t = \overline{s}_t, \overline{A}_t = \overline{a}_t) \\ &= P(S_{t+1} \in \mathcal{Z} | S_t = s_t, A_t = a_t) = \mu_{s_t, a_t}(\mathcal{Z}) \end{aligned}$$

- The measure μ_{s_t, a_t} does not depend on time

Infinite time horizon

- For each $s_t \in \mathcal{S}$, let $\psi(s_t) \subseteq A$ denote the set of allowable treatments for a patient with status $S_t = s_t$.
- A treatment regime in this context is a map $d : \mathcal{S} \to \mathcal{A}$ that satisfies $d(s) \in \psi(s)$ for all $s \in \mathcal{S}$
- Let $S_t^*(\overline{a}_{t-1})$ denote the potential patient status at time t under treatment sequence $\overline{a}_{t-1} \in \otimes_{v-1}^{t-1} \mathcal{A}$ so that the potential status under a regime d is

$$S_t^*(d) = \sum_{\overline{a}_{t-1}} S_t^*(\overline{a}_{t-1}) \prod_{v=1}^{t-1} 1_{[d\{S_v^*(\overline{a}_{v-1})\}=a_v]}$$

- The potential momentary outcome for regime d is $Y_t^*(d) = y[S_t^*(d), d\{S_t^*(d)\}, S_{t+1}^*(d)]$

Infinite time horizon

- Define the conditional discounted marginal mean outcome under d to be

$$V(S;d) = E\left\{ \sum_{k \geq 0} \gamma^k Y^*_{t+k}(d) \middle| S_t = s \right\}$$

  where $\gamma \in [0,1)$ is a discount factor that balances the trade-off between immediate and long-term outcomes
- An optimal regime, d^{opt}, satisfies $V(s;d^{\text{opt}}) \geq V(s;d)$ for all $s \in \mathcal{S}$ and all regimes d
- In the context of policy-search methods over a prespecified class of regimes, $\mathcal{D}$, we define the optimal regime within $\mathcal{D}$ with respect to reference population distribution R as $d^{\text{opt}}_{\mathcal{D},R} = \text{argmax}_{d \in \mathcal{D}} V_R(d)$

Mobile health

One important motivation for infinite horizon reinforcement learning is mHealth, which collects information and provides interventions in real time

- Such data can be collected retrospectively or by using sequential multiple assignment randomized trial (SMART) designs
- A somewhat different approach to precision medicine in mHealth involves using data obtained from a microrandomized clinical trial
  - Well suited for developing interventions involving prompting people to take healthy actions to improve health behavior
  - Often designed to improve proximal outcomes, not necessarily longer-term outcomes

Data Sources and Study Design

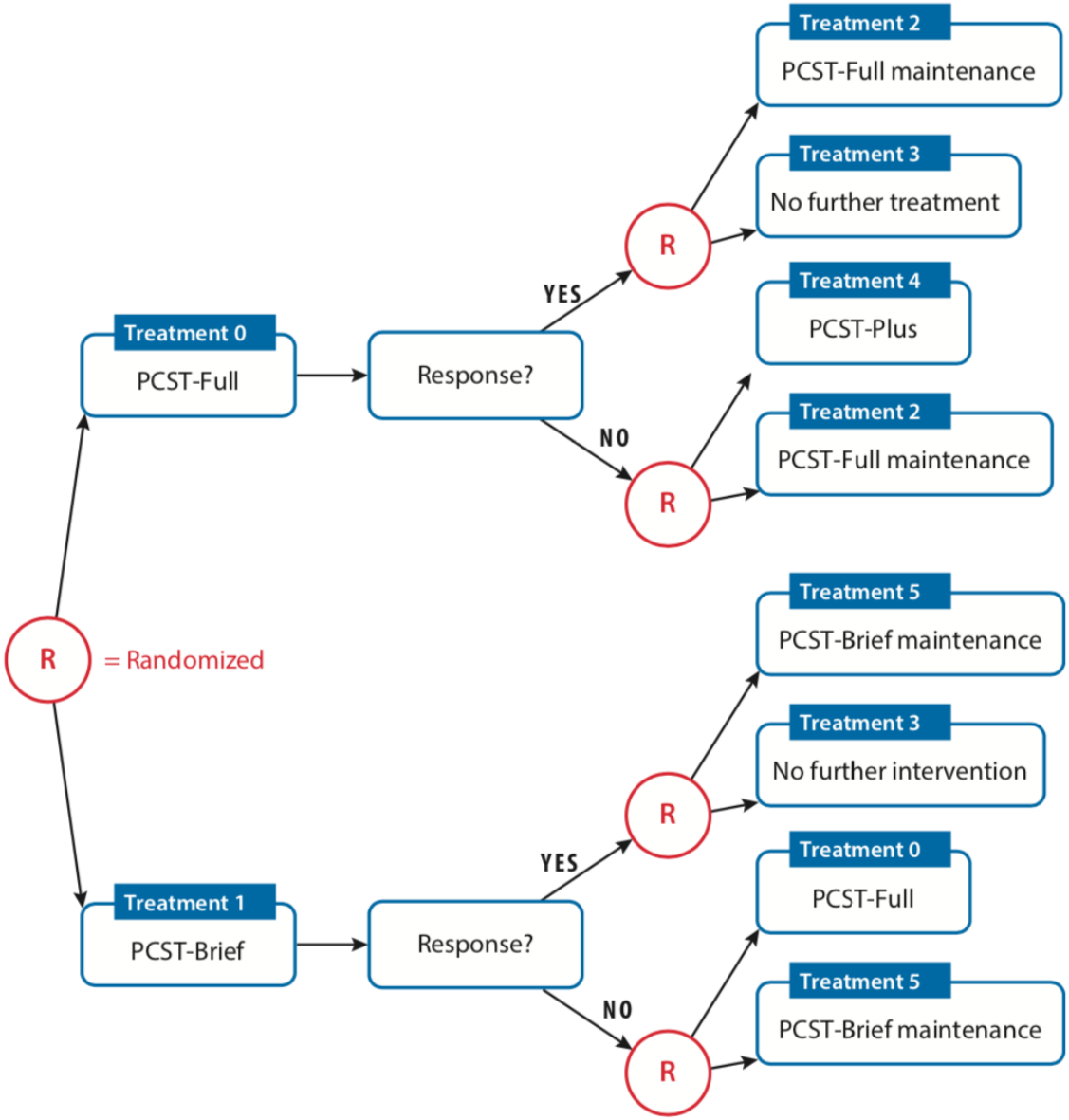
Data for estimating dynamic treatment regimes can come from a range of sources

- In many of these sources, the estimation of an optimal treatment regime is a strictly exploratory-analysis
- Such data can still be a rich resource for estimation and inference for optimal treatment regimes
  - Clinical trials
  - Electronic health records
  - Planned observational studies
- Observational designs are frequently the inspiration for causal inference research
- Can also include careful selection of convenience samples to improve quality of causal inference

Data Sources and Study Design

- Randomized clinical trials are a gold standard for data collection as they protect against unmeasured confounding and can ensure efficient estimation
- For single-stage decision problems, a k-arm randomized clinical trial with equal randomization provides maximal information about ATEs for pairs of treatments
- For multi-stage decision problems, SMARTs allow for the efficient comparison of treatment sequences and fixed treatment regimes
- In a SMART, a patient is randomized at each point in the treatment process where there is clinical equipoise, and thus, each patient may be randomized multiple times throughout the trial

SMART example: pain coping skills training (PCST)



from Kosorok and Laber (2019)

Data Sources and Study Design

- Randomization at each decision point is not always possible or optimal, and various hybrid designs can be considered
- Hybrid designs have both randomization and observational components and often have a pragmatic motivation
  - E.g. enrichment designs allow the first treatment assignment to be nonrandomized but the second treatment assignment is randomized
- Two important points
  - Heterogeneity is good for precision medicine as it is needed to estimate an optimal treatment regime
  - The design of studies used for discovering precision medicine is crucially important

Outline

Multiple outcomes are common in clinical settings

- Often multiple outcomes must be considered when managing treatment decisions.
- Examples
  - In treating schizophrenia, there is a trade-off between side-effects and efficacy (Butler et al 2018)
  - In treating bipolar depression, there is a trade-off between depressive symptoms and risk of mania (Luckett et al 2021+)
- Recent work on this area include
  - Set-valued treatment regimes which recommend a set of 'acceptable' treatments given a patient's history
  - Methods that maximize a primary outcome while constraining a secondary outcome to be within an acceptable region

Examples of multiple outcomes in precision medicine research

Sometimes there is a trade-off between two or more endpoints which depend on patient preferences.

- Butler et al (2018) develop a method for the single-decision setting which uses item response theory to elicit patient preferences
- Preferences are combined with Q-learning to optimize the patient-preferred composite utility which is a convex combination of the two outcomes

Example of multiple outcomes in precision medicine research

Sometimes the trade-offs between two outcomes depend on complex individual-level factors about which clinicians have imperfect information

- Luckett et al (2021+) consider observational data on clinicians prescribing anti-depressants to patients with bipolar disorder
- Outcomes include measures of depression and mania
- It is assumed that clinicians try to act optimally and sometimes, but not always, succeed
- Estimated Q-functions for each outcome are used to estimate the weight in the combined utility of the convex combination of the two outcomes as a function of patient-level covariates as well as the probability of correct treatment assignment as a function of patient-level covariates

Outline

Statistical inference in precision medicine

What are some of the common inferential properties and quantities that statisticians are interested in?

- Consistency (zero-order inference)
- Confidence intervals, hypothesis tests, sample size calculations (first-order inference)

Inference for precision medicine is an open and active area of research

- For many of the methods we've discussed, consistency of the estimators has been proven
- First-order inference is not known for many of the machine learning tools used in precision medicine. Some advances include:
  - SVM setting, random forests
  - Error bounds
  - Sample size formulas for the single-decision setting based on the value function
- Primary emphasis is on inference for performance of a treatment regime

Outline

Overall Conclusions and Future Work

- This is an exciting time for precision medicine and precision health at the confluence of machine learning and statistics.
- There are numerous open questions.
- Inference can be challenging and nonstandard.
- Consistency, or zero order inference, is often an important first step.
- This work is part of the emergence of a new (or renewed) discipline focused on data driven decision making, precision operations research, precision health, and precision medicine, with many connections in many quantitative and nonquantitative disciplines.